

UNIVERSIDADE DE LISBOA
Faculdade de Medicina



Uncertainty and urgency

Tackling ME/CFS research and
assessing the risk of SARS-CoV-2 Omicron subvariants
in vaccinated populations

João Torrado Malato

Orientadores: Prof. Doutor Nuno Henriques dos Santos de Sepúlveda
Prof. Doutor Luís Ricardo Simões da Silva Graça

Tese especialmente elaborada para obtenção do grau de Doutor em

Ciências Biomédicas

2024

UNIVERSIDADE DE LISBOA
Faculdade de Medicina

Uncertainty and urgency
Tackling ME/CFS research and
assessing the risk of SARS-CoV-2 Omicron subvariants
in vaccinated populations

João Torrado Malato

Orientadores: Prof. Doutor Nuno Henriques dos Santos de Sepúlveda
Prof. Doutor Luís Ricardo Simões da Silva Graça

Tese especialmente elaborada para obtenção do grau de Doutor em **Ciências Biomédicas**

Júri: PhD Karl Morten, Investigator da Oxford University (Reino Unido)
Doutor Pedro Nuno Ferreira Pinto de Oliveira, Professor Associado com Agregação do
Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto
Presidente: Doutor Mário Nuno Ramos ed Almeida Ramirez, Professor Associado com Agregação
e Vice-Presidente do Conselho Científico da Faculdade de Medicina da Universidade
de Lisboa
Vogais: Doutor Paulo Jorge da Silva Nogueira, Professor Associado da Faculdade ed Medicina
da Universidade de Lisboa
Doutor Rui Miguel Prudêncio da Cunha Pignateli, Professor Associado Convidado com
Agregação da Faculdade de Medicina da Universidade de Lisboa

Fundação para a Ciência e a Tecnologia. Ref. SFRH/BD/149758/2019

2024

A impressão desta tese foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 9 de março de 2024.

As opiniões expressas nesta publicação são da exclusiva responsabilidade do seu autor.

Todo hombre necesita ser lo que es para hacer lo que hace. Y viceversa.

– Antonio Machado

Somos cientistas [...] da maneira que somos. E ainda bem. Cada um de nós é diferente, pelo menos pelo seu percurso individual e como aquilo que fazemos reflecte aquilo que somos, o que fazemos é também diversificado. E somos o que fomos sendo e o que fomos fazendo da vida que nos foi fazendo.

– António Coutinho

Acknowledgments

This thesis was financially supported by Fundação para a Ciência e a Tecnologia (grant reference SFRH/BD/149758/2019) and was hosted by the Instituto de Medicina Molecular João Lobo Antunes at the Faculty of Medicine, University of Lisbon.

This work was produced despite setbacks and delays in data transfer agreements due to Brexit, isolation and salary payment delays due to the Covid-19 pandemic, and having to shorten academic visits abroad due to the ongoing Russo-Ukrainian War. Over the course of the last four years I inhabited four different homes, suffered from hearing loss, and even did daily therapy on a hyperbaric chamber for a while. During this time, I also experienced the joys of marriage, welcomed my first nephew, and made numerous friends who have enriched my life.

As such, I would like to express my gratitude to all the people who directly and indirectly contributed to and provided invaluable support for this thesis.

Nuno Sepúlveda, my supervisor and mentor. You showed me the magnificent world that lies at the intersection between life sciences and statistics. Six years ago, I started working alongside you as an unproven Biology student (with highly doubtful programming skills). Your expertise, dedication, and genuine curiosity towards science, together with the stubbornness to do things “just because” and think actively instead, have become an inspiring and positive reference. Words cannot express my gratitude to you and everything you have taught me. I can only say that it has been a pleasure to discuss everything, anything, and all the things in between. Muito, muito obrigado.

Luís Graça, my co-supervisor and self-proclaimed “desorientador” (Portuguese for disorienting-supervisor). Thank you for all the insightful discussions and comments during our lab meetings, and for giving me the opportunity to work and analyse official data and develop interesting and impactful work on the topic of SARS-CoV-2 (presented in this thesis in Chapter 9 and Chapter 10). I am hopeful of what is to come.

André Fonseca, who started this journey with me. Your dedication towards work-life balance is remarkable, and your patience and wise comments to my insecurities are only matched by the love I received from your family during my time in Warsaw. I am sure you will accomplish

great things.

The people at the LGraca group. Rodrigo Pedroso, for cutting straight to the point every single time, and being a great teammate. Saumya Kumar, who during the first years of my project took upon herself to teach me the ways of knowing my audience, frenetically studying to connect answers from completely different topics, and also Indian cuisine. I still have a .txt file titled “things-saumya-says” that I open from time to time. JC, who keeps making the extra effort to find the holes in my research ideas and then pesters me with counter-arguments. Miguel Santos, for the endless stream of motivational speeches during the submission of this work. Also, for promptly volunteering to read the introduction of my thesis—without knowing it is almost 40 pages long; and for reading it anyway, in time to give his comments! To the rest of ‘The Bunker’ people, Tomás Gomes, Diana Matias, Rui Vieira, and Diogo Fonseca, for including me in their den and making me feel welcome coming to work at iMM, during the final and more stressful stages of the thesis. Filipa Ribeiro, Catarina Mendes, Pedro Ribeiro, Pedro Gaspar, Vladimir Ghilas, Maria Miguel, Deepanwita Ghosh, and all the talented past and present members of the lab, for creating such a fun and stimulating environment to do research at.

My CAML PhD colleagues at iMM, Ana Cachucho, Sara Pinto, and Carol Pacini, for providing enough sangria to spark creativity whenever I felt stuck. Also, for letting me win at every single virtual board game (suck it Sara!) and ensuring the stability of my mental health during early pandemic isolation.

MI².AI research group and Przemysław Biecek, for hosting me at Politechnika Warszawska and providing the fantastic work conditions where I was able to produce most of my written work. With particular emphasis, thank you to Kasia Woźnica, Anna Kozak, and Hubert Baniecki, for the conversations and sporadic hang-outs with lots of tea and cake. You have my contact, come for a visit!

Eliana Lacerda and Luís Nacul, representing the CureME research group and the UK ME/CFS Biobank. Thank you for introducing me to the research done in ME/CFS and contextualising my data-driven questions with clinical insights.

Francisco Westermeier, for directly keeping me updated on parallel research in ME/CFS.

And also, for doing it indirectly, by providing Nuno with a continuous stream of ideas to keep tabs on.

Agora em Português. Este trabalho é fruto de muito e continuado apoio. Assim sendo, quero agradecer a todos aqueles que, directa ou indirectamente, contribuíram para o seu progresso e conclusão.

Marília Antunes, Ruy Ribeiro, e Manuel Carmo Gomes, pelos ensinamentos e disponibilidade em atenderem simpaticamente todas as chamadas e responderem a todos os emails, mesmo quando fora de horas.

Clara Cordeiro, por nunca permitir que incongruências estatísticas passassem por si. E por todo o apoio desde o primeiro dia deste projecto.

Orquestra Académica da Universidade de Lisboa e todos os seus membros.

Grupo de almoços da Cantina Velha, discutir convosco é sempre um prazer.

Constantino Caetano, Miguel Matos, Pedro Moura, Olivia Schroeder, Margarida Camoesas, Rita Bizarro, e Carolina Mendeiro, pela vossa continuada amizade.

Paweł e Adam, por num encontro aleatório terem conseguido que Varsóvia se tornasse uma segunda casa.

Belinha, Paupério, e respectivas cantoras do rés-do-chão. Maria Frederica, e Raúl, porque não custa nada adicionar mais uma linha à tese. Manu, porque sim.

Aos membros da minha família: Mãe e Pité, por incansavelmente zelarem por mim e garantirem que não me disperso em demasia, entre todas as ideias e objectivos que vão surgindo. Devo-vos muito, e um dos grandes prazeres que tenho na vida é poder discutir saúde e ciência convosco. Pai, por toda a companhia e passeios à distância, e por sempre me puxar para ver o ‘panorama geral’ onde o meu trabalho se insere. Meus irmãos, Ana, Madalena, e Martim, por toda a companhia, desde sempre; vejo continuamente a vossa influência em tudo o que faço. Todos os meus tios, tias, primos, e primas, que distribuem tanto carinho quanto desconversa. Minhas Avós Maria, Karin, e Marília. Aos Avós que gostava que celebrassem comigo, Catucha, João José, e Nuno.

Por fim, e mais que todos, à Dida-Borboleta. Completas-me e sou feliz contigo todos os dias.

Abstract

This thesis describes research developed in two topics: research strategies in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), and the risk of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron BA.5 subvariant in the context of highly vaccinated populations.

ME/CFS is a complex disease of unknown cause. The lack of biomarkers for the unequivocal identification of ME/CFS leaves its diagnosis to be made under a degree of uncertainty, relying solely on the assessment of symptoms and the exclusion of other possible diseases that could explain fatigue. Some symptoms, such as long-lasting and profound fatigue that is not alleviated by rest, post-exertional malaise, unrefreshing sleep, or multi-joint pain, are generally required for the diagnosis. However, the spectrum of required symptoms can vary depending on the case definition used, and there are currently more than 20 ME/CFS case definitions proposed.

The first aim of this thesis is to investigate the diagnostic agreement among four commonly used ME/CFS case definitions and study the overall reproducibility of case-control association studies under the assumption that misdiagnosed individuals may be included in the cohort of patients. I simulated data from different scenarios as a function of sample sizes and the strength of the association between candidate risk factor and the disease. The influence of sensitivity and specificity in the outcome of usual serological tests was also taken into account, with possible misclassification events for false negative or false positive tests. In Chapter 2, results showed that although most patients are diagnosed by all, there is no complete agreement between case definitions, thus reinforcing the hypothesis of other misdiagnosed illnesses being included in research studies. In Chapter 3 and Chapter 4, the simulated studies under the assumption of misdiagnosis suggested that the current studies on ME/CFS research have suboptimal statistical power to detect potential true associations with candidate causal factors consistently. Some improvements are suggested, such as increasing sample sizes, reporting the case definitions and exclusion criteria implemented, and focusing on subgroups of more similar ME/CFS patients with a specific pathological mechanism.

There is a growing body of epidemiological studies where patients report an acute infection

as the origin of the disease. In accordance, studies have found evidence of associations with certain infections, suggesting ME/CFS as a disease with immune dysregulation, similar to autoimmune conditions such as multiple sclerosis (MS), rheumatoid arthritis or, more recently, Long Covid.

The second aim of this thesis is to explore the relationship between ME/CFS and past viral infections by herpesviruses, stratifying patients by infection trigger or symptom severity. In Chapter 5, I analysed previously published microarray data from IgG antibody responses to Epstein-Barr virus (EBV) antigens in ME/CFS patients and healthy controls and identified two candidate antigens (EBV proteins EBNA4_0529 and EBNA6_0070) inducing increased antibody production in patients with an infection trigger. Logistic regression with an interaction between the specific antibodies and age/gender displayed the ability to classify the subgroup of individuals reporting an infection trigger, with high sensitivity and specificity. This indicates that the two EBV antigens could potentially be used as biomarkers for the diagnosis of ME/CFS patients with a putative infection trigger. This finding corroborated the proposal of different subtypes of ME/CFS, reinforcing the strategy of stratifying patients adequately. Moreover, since EBV-derived antigen EBNA6_0070 had high sequence homology with human proteins, there is the potential involvement of molecular mimicry in the pathogenesis of the disease.

In Chapter 6, analysis of IgG antibody responses against six different herpesviruses in ME/CFS and MS patients revealed distinct antibody-symptom associations between the two conditions. Notably, symptoms from the immunological domain (sore throat, tender glands, and flu-like symptoms) were the main symptoms differentiating the ME/CFS infection trigger subgroups and MS. However, antibody-symptom associations were more heterogeneous when studying the stratified ME/CFS, being generally clearer in the MS control group. Analysing the population of ME/CFS patients showed a link between exposure to herpes simplex virus-1 (HSV1) and experiencing more exacerbated symptoms from the neurocognitive domain. Interestingly, this relation between a neurotropic virus and the severity of symptoms from the neurocognitive domain was also found in Chapter 7, where ME/CFS patients were grouped by the combined severity of symptoms related to seven specific domains (PEM, immunological, neurocognitive, neuroendocrine, autonomic, neurophysiological, and pain). Once again, these

results suggest the possibility of ME/CFS being an umbrella term, encompassing different specific subgroups with similar symptomatology. In this line, the analysis show that stratification of ME/CFS patients enables for a better understanding between different viral infections and the (chronic) activation or dysregulation of particular mechanisms.

The project of this thesis was developed during the coronavirus disease 2019 (Covid-19) pandemic. During this period of public health emergency, adjustments were made to accommodate research related to the increased risk of ME/CFS patients towards a SARS-CoV-2 infection, relative to healthy controls. A parallel aim of the thesis is then to study whether the expression of the human angiotensin-converting enzyme 2 (ACE2), the major cell entry receptor for SARS-CoV-2, is altered in patients. In Chapter 8, I performed a meta-analysis of public data on CpG DNA methylation and gene expression of this enzyme and its homologous ACE protein in peripheral blood mononuclear cells. The results revealed decreased methylation levels of four CpG probes in the *ACE* locus and one single CpG probe in the promoter region of the *ACE2* gene, suggesting increased expression of the respective genes. Conversely, the meta-analysis revealed decreased expression of *ACE2* but not *ACE* in patients when compared to healthy controls. The results were not particularly clear to provide a definitive answer. However, the finding of increased *ACE:ACE2* ratio in patients was concerning, as it can promote vasoconstriction that could lead to increased production of reactive oxygen species and inflammation.

Nearing the end of 2021, the rise in cases from the SARS-CoV-2 Omicron variant led to concerns about the protection conferred by vaccines and boosters being used, as they were adapted from early lineages of the virus. Omicron displayed an enhanced ability to evade immunity from previous variants. After a period of dominance from Omicron BA.1 and BA.2 subvariants, during the second half of 2022 Portugal became one of the first countries with Omicron BA.5 as the dominant variant. At this time, new adapted vaccines under development were based on BA.1, meaning that there was a need to study if vaccines and previous Omicron infections would grant effective protection against infections (and reinfections) from the new Omicron subvariant. The last aim of this thesis is to study the protection effectiveness and stability over time from infections with past variants and subvariants, towards the Omicron BA.5, using the

Portuguese population as a case study of a highly vaccinated population—with a vaccination cover over 98% of individuals 12 years and older by the end of 2021. In Chapter 9, results showed that hybrid immunity (vaccination + single past infection) reduced the risk of BA.5 overall and previous infection with subvariants BA.1/BA.2 conferred the highest protection efficacy overall (protection effectiveness by different single infection variants in relation to the uninfected group: Wuhan-Hu-1 51.6%, Alpha 54.8%, Delta 61.3%, BA.1/BA.2 75.3%). In addition, in Chapter 10, I estimated that this additional immunity wanes over time, with the relative risk towards a BA.5 infection rapidly increasing from approximately 0.06 to 0.35 between three to eight months following a BA.1/BA.2 infection, only to stabilise around similar values (relative risk approximately 0.37) after that period and to eight months post-infection. These findings suggested that in a population with very high vaccine coverage, BA.1 adapted vaccine boosts would be successful in reducing the risk of breakthrough BA.5 infections.

A large number of patients infected with SARS-CoV-2 has developed a post-acute infection syndrome, Long Covid. Notably, the symptoms that characterise this disease overlap with those of ME/CFS, and it has been discussed whether there is an etiological link between the two. However, there is a clear association between an infection and Long Covid. Still, these similarities reinforce the need to understand better the relationship and impact of external agents, such as viruses, on the possible causes of homeostatic imbalances that could lead to ME/CFS and similar diseases.

Keywords: Myalgic encephalomyelitis/Chronic fatigue syndrome; Misdiagnosis; Patient stratification; SARS-CoV-2 Omicron; Hybrid immunity

Resumo

A presente tese descreve um trabalho de investigação dividido em dois tópicos. No primeiro tópico estudaram-se estratégias de investigação e estratificação de doentes com encefalomielite miálgica/síndrome de fadiga crónica (EM/SFC). O segundo tópico estuda o risco da subvariante Ómicron BA.5 do coronavírus da síndrome respiratória aguda grave 2 (SARS-CoV-2) em contexto de populações com elevada cobertura vacinal.

A EM/SFC é uma doença complexa e de etiologia e patofisiologia desconhecidas. A falta de marcadores biológicos para a identificação inequívoca da EM/SFC torna o seu diagnóstico incerto, baseando-se unicamente na avaliação de sintomas e na exclusão de outras doenças que possam justificar o cansaço crónico. Alguns sintomas são recorrentemente observados e necessários para o diagnóstico da doença. Exemplos incluem a fadiga profunda prolongada e não aliviada pelo repouso, o mal-estar pós-esforço (do inglês *post-exertional malaise*, sigla PEM), o sono não reparador, ou as dores musculares ou poliarticulares. No entanto, o espectro de sintomas avaliados pode variar dependendo dos critérios de diagnóstico implementados. E atualmente existem mais de 20 definições propostas para a EM/SFC.

O primeiro objetivo desta tese passa por investigar a concordância entre quatro dos critérios de diagnóstico mais utilizados em EM/SFC e estudar a reprodutibilidade de estudos de associação do tipo caso-controlo, sob a suposição de que indivíduos com falso diagnóstico (do inglês *misdiagnosis*) podem ser incluídos na coorte de doentes. Para tal, simulei dados relativos a vários cenários em função de diferentes tamanhos amostrais ou diferentes forças de associação entre possíveis fatores de risco e a doença. Também considerei a influência de parâmetros como a sensibilidade e especificidade dos testes serológicos usuais, uma vez que estudos serológicos também podem gerar falsos negativos ou falsos positivos. No Capítulo 2, os resultados mostraram que embora a maioria da população suspeita de EM/SFC seja diagnosticada pelos critérios de diagnóstico considerados, não existe uma concordância total entre eles, o que reforça a hipótese de influência por falso diagnóstico de doentes. No Capítulo 3 e Capítulo 4, os estudos de simulação realizados assumindo o pressuposto de falso diagnóstico sugerem que os estudos atuais em EM/SFC têm um poder estatístico insuficiente para detetar,

de forma consistente, potenciais associações a fatores de risco de interesse. A fim de melhorar a reproduutibilidade dos estudos nesta doença, algumas recomendações são sugeridas, tal como o aumento do tamanho das amostras de ambos os coortes, a comunicação e reporte dos critérios de diagnóstico e de exclusão aplicados, e o foco em subgrupos de doentes com EM/SFC mais semelhantes entre si e com um eventual mecanismo patológico específico.

Existe um número crescente de estudos epidemiológicos em que os pacientes reportam uma infecção aguda como origem da doença. Em concordância, investigações têm encontrado possíveis relações que sugerem uma associação entre a origem da doença e infecções. Isto sugere que a EM/SFC poderá ser uma doença com desregulação do sistema imune, aproximando-a de doenças autoimunes como a esclerose múltipla (do inglês *multiple sclerosis*, sigla MS) ou a artrite reumatoide.

O segundo objetivo desta tese é explorar a relação entre EM/SFC e infecções virais anteriores por herpesvírus, estratificando os doentes de acordo com o reporte de infecções como causa da doença ou pela severidade dos sintomas apresentados. No Capítulo 5 analisei dados de *microarray* previamente publicados relativos a respostas de anticorpos IgG aos抗énios do vírus Epstein-Barr (VEB) em doentes com EM/SFC, comparando-os com controlos saudáveis. Identifiquei dois抗énios candidatos (proteínas do VEB EBNA4_0529 e EBNA6_0070) que induziam maior produção de anticorpos em doentes EM/SFC com infecção na origem da doença. O uso de uma regressão logística com uma interação entre estes anticorpos e idade/sexo demonstrou boa capacidade de classificação para o subgrupo de doentes com origem numa infecção, com sensibilidade e especificidade elevadas. Estes resultados sugerem que os dois抗énios do VEB podem ser utilizados como biomarcadores para o diagnóstico da EM/SFC em indivíduos com uma infecção anterior como origem da doença. Os resultados corroboram a noção de que a EM/SFC é composta por diferentes subtipos, reforçando a proposta de uma estratificação adequada de doentes. Adicionalmente, uma vez que o抗énio viral EBNA6_0070 possui elevada homologia com proteínas humanas, pode existir o envolvimento de mimetismo molecular na patogénesis da doença.

No Capítulo 6, a análise a respostas de anticorpos IgG contra seis herpesvírus diferentes em doentes EM/SFC e MS revelou associações distintas entre anticorpos e sintomas nas duas

doenças. Nomeadamente, os sintomas do domínio imunológico (e.g., odinofagia, adenopatia, e síndrome gripal) foram os principais sintomas que diferenciaram os subgrupos EM/SFC com infecção prévia e MS. No entanto, associações anticorpos-sintomas foram mais heterogéneas quando se estudou a EM/SFC estratificada, sendo geralmente mais claras nos grupos de controlo da MS. A análise da população de doentes com EM/SFC mostrou uma ligação entre a exposição ao vírus herpes simplex-1 (HSV1) e a ocorrência de sintomas mais exacerbados do domínio neurocognitivo. É interessante notar que esta relação entre um vírus neurotrópico e a severidade dos sintomas do domínio neurocognitivo foi também encontrada no Capítulo 7, onde estratifiquei os doentes com EM/SFC de acordo com o perfil de severidade de sintomas, agrupados por domínios específicos (domínios relativos a PEM, imunológico, neurocognitivo, neuroendócrino, autonómico, neurofisiológico e dor). Mais uma vez, estes resultados demonstram a possibilidade de EM/SFC ser uma terminologia que engloba diferentes subgrupos específicos mas com sintomatologia semelhante. Desta forma, esta análise demonstra que a estratificação de doentes EM/SFC possibilita uma melhor compreensão da ligação entre diferentes infecções virais e a ativação (crónica) ou desregulação de mecanismos de particulares.

Esta tese foi desenvolvida durante a pandemia da doença por coronavírus 2019 (Covid-19). Durante este período de emergência de saúde pública, foram realizados alguns ajustes para acomodar a investigação relacionada com o risco acrescido de infecção por SARS-CoV-2 em doentes com EM/SFC relativamente a indivíduos saudáveis. Assim, um objetivo paralelo da tese é estudar se a expressão da enzima conversora de angiotensina humana 2 (ACE2)—o principal recetor de entrada celular do SARS-CoV-2—está alterada nos doentes. No Capítulo 8 efetuei uma meta-análise de dados públicos de metilação de sítios CpG no ADN e de expressão genética desta enzima e da sua proteína homóloga, a ACE, em células mononucleares do sangue periférico. Os resultados mostraram a diminuição dos níveis de metilação em quatro sondas CpG no locus ACE e numa única CpG na região promotora do gene ACE2, sugerindo um aumento da expressão dos respetivos genes. No entanto, a mesma análise demonstrou também que há uma diminuição da expressão da ACE2 mas não da ACE nos doentes, quando comparados com controlos saudáveis. Estes resultados não foram particularmente claros para fornecer uma resposta definitiva. Contudo, a descoberta do aumento na razão ACE:ACE2 em doentes

levantou algumas preocupações, uma vez que esta relação pode promover a vasoconstrição e pode levar ao aumento da produção de espécies reativas de oxigénio e inflamação.

Perto do final de 2021, o aumento dos casos da variante Ómicron do SARS-CoV-2 suscitou preocupações quanto à proteção conferida pelas vacinas e reforços utilizados, uma vez que eram adaptadas às primeiras linhagens do vírus. A Ómicron demonstrou uma maior capacidade de se evadir e escapar à imunidade conferida pelas variantes anteriores. Após um período de dominância das subvariantes da Ómicron, BA.1 e BA.2, durante a segunda metade de 2022 Portugal tornou-se num dos primeiros países sob dominância da subestirpe BA.5. Nesta altura, as novas vacinas em desenvolvimento baseavam-se na BA.1, sendo necessário estudar se as vacinas administradas, juntamente com infeções anteriores por Ómicron, confeririam uma proteção eficaz contra as infeções (e reinfeções) pela nova subvariante da Ómicron.

O último objetivo desta tese é, assim, estudar a eficácia e a estabilidade da proteção ao longo do tempo por infeções anteriores, contra a Ómicron BA.5. Utilizou-se a população portuguesa como caso de estudo de uma população fortemente vacinada—mais de 98% dos indivíduos a partir dos 12 anos receberam pelo menos uma primeira vacina até ao final de 2021. No Capítulo 9, os resultados mostraram que a imunidade híbrida (vacinação + infeção única) reduziu o risco da BA.5 e a infeção por BA.1/BA.2 conferiu a maior eficácia de proteção (eficácia de proteção por diferentes variantes de infeção única em relação ao grupo não infetado: Wuhan-Hu-1 51,6%, Alpha 54,8%, Delta 61,3%, BA.1/BA.2 75,3%). Posteriormente, no Capítulo 10 estimei a perda de imunidade ao longo do tempo, com o aumento rápido do risco relativo de uma infeção por BA.5 de aproximadamente 0,06 para 0,35 entre três a oito meses após uma infeção por BA.1/BA.2, tendendo a estabilizar após esse período. Estes resultados sugerem que numa população com uma cobertura vacinal muito elevada, os reforços de vacinas adaptadas à BA.1 seriam bem-sucedidos na redução do risco de infeções por BA.5, com uma duração de até oito meses.

Um número elevado de doentes infetados pelo SARS-CoV-2 desenvolve um síndrome de infeção pós-aguda denominado Covid Longo. Curiosamente, os sintomas que caracterizam esta doença sobrepõe-se aos da EM/SFC e discute-se até se não será apenas um novo fator

etiológico para a doença, sendo que para o Covid Longo existe uma clara ligação a uma infecção. Estas semelhanças reforçam a necessidade de compreender melhor a relação e impacto de agentes externos, nomeadamente vírus, nas possíveis causas de desequilíbrios homeostáticos que podem conduzir a EM/SFC ou condições semelhantes.

Palavras-chave: Encefalomielite miálgica/Síndrome de fadiga crónica; Falso diagnóstico; Estratificação de pacientes; Ómicron SARS-CoV-2; Imunidade híbrida

List of publications included in the thesis

This thesis is based on the following papers:

Chapter 2

J. Malato, L. Graça, L. Nacul, E. M. Lacerda, and N. Sepúlveda. Statistical challenges of investigating a disease with a complex diagnosis. In P. Milheiro, A. Pacheco, B. de Sousa, I. F. Alves, I. Pereira, M. J. Polidoro, and S. Ramos, editors, *Estatística: Desafios transversais às ciências com dados*, Eds. Sociedade Portuguesa de Estatística, pages 153–167. Sociedade Portuguesa de Estatística, 2021. ISBN 978-972-8890-47-6. URL <https://www.spestatistica.pt/storage/app/uploads/public/609/28f/6d0/60928f6d08a0c016386627.pdf>.

Chapter 3

J. Malato, L. Graça, and N. Sepúlveda. Impact of misclassification and imperfect serological tests in association analyses of ME/CFS applied to COVID-19 data. In R. Bispo, L. Henriques-Rodrigues, R. Alpizar-Jara, and M. de Carvalho, editors, *Recent Developments in Statistics and Data Science*, Springer Proceedings in Mathematics & Statistics, pages 215–225, Cham, 2022. Springer International Publishing. ISBN 978-3-031-12766-3. doi: 10.1007/978-3-031-12766-315.

Chapter 4

J. Malato, L. Graça, and N. Sepúlveda. Impact of misdiagnosis in case-control studies of myalgic encephalomyelitis/chronic fatigue syndrome. *Diagnostics*. 2023; 13(3):531. doi: <https://doi.org/10.3390/diagnostics13030531>.

Chapter 5

N. Sepúlveda, J. Malato, F. Sotzny, A.D. Grabowska, A. Fonseca, C. Cordeiro, L. Graça, P. Biecek, U. Behrends, J. Mautner, F. Westermeier, E.M. Lacerda, and C. Scheibenbogen. Revisiting IgG Antibody Reactivity to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue

Syndrome and Its Potential Application to Disease Diagnosis. *Frontiers in Medicine*. 2022; 9:921101. doi: <https://doi.org/10.3389/fmed.2022.921101>.

Chapter 6

T.D. Domingues, J. Malato, A.D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, P. Biecek, L. Graça, H. Mourão, C. Scheibenbogen, F. Westermeier, L. Nacul, J.M. Cliff, E. Lacerda, and N. Sepúlveda. Association analysis between symptomatology and herpesvirus IgG antibody concentrations in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and multiple sclerosis. *Heliyon*. 2023; 9(7):E18250. doi: <https://doi.org/10.1016/j.heliyon.2023.e18250>.

Chapter 8

J. Malato, F. Sotzny, S. Bauer, H. Freitag, A. Fonseca, A.D. Grabowska, L. Graça, C. Cordeiro, L. Nacul, E.M. Lacerda, J. Castro-Marrero, C. Scheibenbogen, F. Westermeier, and N. Sepúlveda. The SARS-CoV-2 receptor angiotensin-converting enzyme 2 (ACE2) in myalgic encephalomyelitis/chronic fatigue syndrome: A meta-analysis of public DNA methylation and gene expression data. *Heliyon*. 2021; 7(8). doi: <https://doi.org/10.1016/j.heliyon.2021.e07665>.

Chapter 9

J. Malato, R.M. Ribeiro, P.P. Leite, P. Casaca, E. Fernandes, C. Antunes, V.R. Fonseca, M.C. Gomes, and L. Graca. Risk of BA.5 infection among persons exposed to previous SARS-CoV-2 variants. *New England Journal of Medicine*. 2022; 387(10):953–954. doi: <https://doi.org/10.1056/NEJMc2209479>.

Chapter 10

J. Malato, R.M. Ribeiro, E. Fernandes, P.P. Leite, P. Casaca, C. Antunes, V.R. Fonseca, M.C. Gomes, and L. Graca. Stability of hybrid versus vaccine immunity against BA.5 infection over 8 months. *The Lancet Infectious Diseases*. 2023; 23(2):148–150. doi: [https://doi.org/10.1016/S1473-3099\(22\)00833-7](https://doi.org/10.1016/S1473-3099(22)00833-7).

List of publications not included in the thesis

I.C. Machado, T. Nunes, M. Maximino, J. Malato, L. Tavares, V. Almeida, N. Sepúlveda, and S. Gil. Epidemiologic factors supporting triage of infected dog patients admitted to a veterinary hospital biological isolation and containment unit. *Veterinary Sciences*. 2023; 10(3):186. doi: <https://doi.org/10.3390/vetsci10030186>.

List of abbreviations

ACE Human angiotensin-converting enzyme

ACE2 Human angiotensin-converting enzyme 2

ADAM17 A disintegrin and metallopeptidase domain 17 protein

AIC Akaike's information criterion

AUC Area under the curve

BIC Bayesian information criterion

BMI Body mass index

CCC-2003 Canadian Consensus Criteria

CDC United States Center for Disease Control and Prevention

CDC-1994 US Centre for Disease Control and Prevention Criteria

CFS Chronic fatigue syndrome

CI Confidence interval

CMV Human cytomegalovirus

Covid-19 Coronavirus disease

CTLA4 Cytotoxic T-lymphocyte-associated protein 4

DNA Deoxyribonucleic acid

DPP4 Dipeptidyl peptidase-4

EBV Epstein-Barr virus

ED Endothelial dysfunction

EUROMENE European Network on ME/CFS

FDR False discovery rate

GEO database Gene Expression Omnibus database

- HHV6** Human herpesvirus 6
- HLA** Human leukocyte antigen
- HSV-1** Herpes simplex virus 1
- HSV-2** Herpes simplex virus 2
- ICC-2011** International Consensus Criteria
- IFN- γ** Interferon- γ
- Ig** Immunoglobulin
- IOM-2015** Institute of Medicine Criteria
- IQR** Interquartile range
- IRF5** Interferon regulatory factor 5
- LCA** Latent class analysis
- LDA** Linear discriminant analysis
- MAIT** Mucosal-associated invariant T cells
- MDS** Multidimensional scaling
- ME** Myalgic encephalomyelitis
- ME/CFS** Myalgic encephalomyelitis/Chronic fatigue syndrome
- MERS** Middle East respiratory syndrome
- MHC** Major histocompatibility complex
- MIAME** Minimum Information about a Microarray Experiment
- mRNA** Messenger ribonucleic acid
- MS** Multiple sclerosis
- NICE** National Institute for Health and Care Excellence
- NIOF** Neuroinflammatory and Oxidative Fatigue
- NK** Natural killer (cells)
- OR** Odds ratio
- PBMC** Peripheral blood mononuclear cell

- PCA** Principal component analysis
- PEM** Post-exertional malaise
- PTPN22** Tyrosine phosphatase non-receptor type 22
- RA** Rheumatoid Arthritis
- RAAS** Renin-angiotensin-aldosterone system
- RNA** Ribonucleic acid
- ROC** Receiver operating characteristic (curve)
- ROS** Reactive oxygen species
- RR** Relative risk
- SARS-CoV-2** Severe acute respiratory syndrome coronavirus 2
- SEID** Systemic Exertion Intolerance Disease
- SINAVE** National Covid-19 registry
- SL** SuperLearner (algorithm)
- SLE** Systemic Lupus Erythematosus
- SNP** Single nucleotide polymorphism
- Teffs** Effector T cells
- TGF- β** Transforming growth factor- β
- TMPRSS2** Human transmembrane protease serine 2
- TNF** Tumor necrosis factor
- TNF- α** Tumour necrosis factor- α
- Tregs** Regulatory T cells
- TSS** Transcription starting sites
- UKMEB** United Kingdom ME/CFS Biobank
- US** United States
- VOC** Variants of concern
- VZV** Varicella-zoster virus

Contents

Acknowledgments	ix
Abstract	xiii
Resumo	xvii
List of publications included in the thesis	xxiii
List of publications not included in the thesis	xxv
List of abbreviations	xxix
List of tables	xxxix
List of figures	xliii

I General introduction	1
1 General introduction on the diseases of interest	3
1.1 Myalgic encephalomyelitis/Chronic fatigue syndrome	4
1.1.1 Brief history of the disease	5
1.1.2 Epidemiology	6
1.1.3 Proposed pathogenesis and mechanisms	7
1.1.4 Clinical manifestations and diagnosis	16
1.1.5 Connection to other diseases	21
1.1.6 Treatment and prognosis	23
1.1.7 Research challenges	24
1.2 SARS-CoV-2 and Covid-19	28
1.2.1 Brief overview of the pandemic	29
1.2.2 Vaccines and vaccination	30

1.2.3 SARS-CoV-2 variants	32
1.3 Aims of the thesis	35
1.4 Outline of the thesis	35
Bibliography	38
 II Uncertainty and stratification in ME/CFS	 67
 2 Statistical challenges of investigating a disease with a complex diagnosis	 69
2.1 Introduction	69
2.2 The UKMEB	70
2.3 Diagnostic agreement analysis	71
2.4 Symptoms' similarity analysis	73
2.5 Impact of misclassification on an association analysis	75
2.6 Concluding remarks	78
Bibliography	79
 3 Impact of misclassification and imperfect serological tests in association analyses of ME/CFS applied to Covid-19 data	 81
3.1 Introduction	82
3.2 Simulation study	84
3.2.1 Mathematical formulation of the problem	84
3.2.2 Parameterisation using real-word data	85
3.2.3 Simulation structure	87
3.3 Simulation results	87
3.4 Discussion	89
Bibliography	91
 4 Impact of misdiagnosis in case-control studies of ME/CFS	 95
4.1 Introduction	96
4.2 Statistical methodology	97

4.2.1	Formulation of the problem	97
4.2.2	Simulation study	101
4.2.3	Application to two ME/CFS studies	102
4.3	Results	102
4.3.1	Simulation study: impact of ME/CFS misdiagnosis	102
4.3.2	Simulation study: impact of ME/CFS misdiagnosis and misclassification on the candidate causal factor	103
4.3.3	Application to data from two ME/CFS studies	107
4.4	Discussion	111
	Bibliography	113
5	Revisiting IgG antibody reactivity to EBV in ME/CFS and its potential application to disease diagnosis	121
5.1	Introduction	122
5.2	Materials and methods	124
5.2.1	Study participants	124
5.2.2	Peptide array	125
5.2.3	Statistical analysis	125
5.2.4	Statistical software	127
5.3	Results	127
5.3.1	Principal component and linear discriminant analyses	127
5.3.2	Antibody-wide association analysis	129
5.3.3	Analysis of candidate antigens for classifying ME/CFS patients with infectious trigger	132
5.4	Discussion	136
	Bibliography	140
6	Association analysis between symptomology and herpesvirus IgG antibody concentrations in ME/CFS and MS	147

6.1	Introduction	148
6.2	Materials and methods	150
6.2.1	Study participants	150
6.2.2	Symptomology assessment	151
6.2.3	Herpesvirus IgG antibodies	151
6.2.4	Statistical analysis	152
6.2.5	Ethical approval	154
6.3	Results	154
6.3.1	Basic characteristics of the study participants	154
6.3.2	Analysis of symptomology	156
6.3.3	Univariate analysis of herpesvirus IgG antibody data	159
6.3.4	Combined analysis of IgG antibody data using an SL algorithm	161
6.3.5	Association analysis between symptomology and herpesvirus IgG antibodies	164
6.4	Discussion	166
	Bibliography	172
7	Relation between domain-specific severity profiles and IgG antibody responses in ME/CFS	183
7.1	Introduction	183
7.2	Materials and methods	185
7.2.1	Study participants	185
7.2.2	Symptomatology assessment	186
7.2.3	Herpesviruses serological data	186
7.2.4	Symptoms intra- and inter-rater agreement	188
7.2.5	Construction of patient subgroups from symptomatological data	189
7.2.6	Analysis of serological data	190
7.2.7	Statistical software	191
7.3	Results	191

7.3.1 Characterisation of study participants	191
7.3.2 Symptom description and similarity profiles	192
7.3.3 Latent class analysis across symptomatological domains	195
7.3.4 Herpesvirus IgG antibody data	201
7.3.5 Herpesvirus seropositivity	205
7.4 Discussion	207
7.5 Conclusions	212
Bibliography	212

8 The SARS-CoV-2 receptor ACE2 in ME/CFS: A meta-analysis of public DNA methylation and gene expression data	217
8.1 Introduction	218
8.2 Materials and methods	219
8.2.1 Eligible diagnostic criteria of ME/CFS	219
8.2.2 Analysis of published DNA methylation association studies	219
8.2.3 Analysis of gene expression studies	223
8.2.4 Analysis of new RNA data on the ACE/ACE2 gene expression in ME/CFS	226
8.2.5 Statistical software	227
8.3 Results	228
8.3.1 Meta-analysis of ACE/ACE2 DNA methylation in ME/CFS patients .	228
8.3.2 Meta-analysis of ACE/ACE2 gene expression in ME/CFS patients .	230
8.3.3 Analysis of ACE/ACE2 gene expression from a new female cohort .	231
8.4 Discussion	232
8.5 Conclusions	237
Bibliography	237

III Urgency and the case of SARS-CoV-2 Omicron BA.5 in Portugal	249
9 Risk of BA.5 infection among persons exposed to previous SARS-CoV-2 variants	251
9.1 Introduction	251
9.2 Methods	254
9.2.1 Participant selection	254
9.2.2 Vaccination coverage	257
9.2.3 Statistical analysis	257
9.3 Supplementary discussion	258
Bibliography	263
10 Stability of hybrid vs. vaccine immunity against BA.5 infection over 8 months	267
10.1 Introduction	267
10.2 Methods	270
10.2.1 Participant selection	270
10.2.2 Vaccination coverage	274
10.2.3 Statistical analysis	275
Bibliography	277
IV General discussion	279
11 General discussion	281
11.1 Misdiagnosis and stratification of ME/CFS	281
11.2 Results on Covid-19	284
11.3 ME/CFS and Long Covid	285
11.4 Concluding remarks	287
Bibliography	288

V Appendices	293
A Additional information to Chapter 4	A.295
A.1 Supplementary tables	A.295
A.2 Supplementary equations	A.296
B Additional information to Chapter 5	B.297
B.1 Supplementary tables	B.297
B.2 Supplementary figures	B.300
C Additional information to Chapter 6	C.301
C.1 Supplementary tables	C.301
C.2 Supplementary figures	C.305
D Additional information to Chapter 7	D.309
D.1 Supplementary tables	D.309
D.2 Supplementary figures	D.312
E Additional information to Chapter 8	E.315
E.1 Supplementary tables	E.315
F Publications discussed in this thesis	F.319
F.1 Chapter 2	F.319
F.2 Chapter 3	F.335
F.3 Chapter 4	F.348
F.4 Chapter 5	F.365
F.5 Chapter 6	F.379
F.6 Chapter 8	F.396
F.7 Chapter 9	F.408
F.8 Chapter 10	F.411

List of Tables

1.1	Common domains and list of associated symptoms for ME/CFS	17
1.2	List of suggested proposals for patient stratification.	26
2.1	Frequency of suspected cases of ME/CFS according to their diagnostic outcomes using different case definitions	72
2.2	Estimates of the Jaccard's similarity index for the four case definitions of ME/CFS	73
2.3	Augmented version of the observable 2×2 frequency table and probabilities for healthy controls and suspected cases	77
3.1	Augmented version of the observable 2×2 frequency table in the case-control association study scenario with possible misclassification of suspected ME/CFS cases and existence of false positive and false negative serological outcomes observed from serology tests done to assess exposure	86
3.2	Parameter values used in the study	86
3.3	Maximum values of misclassification rate that maintain power if at least 80% to reject the null hypothesis of lack of association, for different values of true odds ratio country of serological survey, and sample sizes	90
4.1	Two-way contingency table of a typical case-control study	98
4.2	Maximum values of misdiagnosis probability that ensure the minimum power of 80% to detect a genuine association as a function of θ_0 and sample size per group	106
4.3	Maximum values of misdiagnosis probability that still ensures a power of rejecting the null hypotheses of at least 80% for $\Delta_T = 3$ and $\theta_0 = 0.25$	108

4.4	Reported associations of a candidate gene association study (Steiner et al. 2020)	109
4.5	Summary of serological findings from Cliff et al. (2019)	109
5.1	Basic characteristics of ME/CFS patients and healthy controls	124
5.2	Estimates of the final complementary log-log model to discriminate ME/CFS patients with an infectious disease trigger from healthy controls	134
6.1	Basic characteristics of the study participants	155
6.2	Seroprevalence, median concentration and respective IQR per study group and herpesvirus IgG antibody	160
6.3	Area under the Receiver Operating Characteristic curve and its 95% confidence interval, optimal cutoff and associated sensitivity and specificity to discriminate ME/CFS_S0, ME/CFS_S1, ME/CFS_S2, ME/CFS_S3 subgroups from patients with multiple sclerosis used as controls	162
7.1	Samples sizes of the 241 ME/CFS patients in each AIC-based latent class estimated on each domain	198
7.2	Selection of significant p-values, both unadjusted and BH-adjusted, from Kruskal-Wallis sum rank test on antibody concentration values across each herpesvirus, comparing different healthy controls and ME/CFS patients under stratification based on symptomatological domains	201
7.3	Selection of significant p-values, both unadjusted and BH-adjusted, from Pearson's χ^2 test on seropositivity values across each herpesvirus, comparing different healthy controls and ME/CFS patients under stratification based on symptomatological domains	206
8.1	Summary of the six DNA methylation studies under analysis	220
8.2	Summary of the 8 microarray-based gene expression studies under analysis . .	224
8.3	Summary statistics for the gene expression of ACE and ACE2 from the German female study participants	232

8.4 Analysis of the linear regression models for the Box-Cox-transformed <i>ACE</i> and <i>ACE2</i> mRNA levels	233
9.1 Risk of omicron BA.5 infection according to previous infection history	253
9.2 Periods of dominance of the different SARS-CoV-2 variants and omicron sub-variants in Portugal	255
9.3 Risk of omicron BA.5 infection according to previous infection history, considering an estimate of unreported cases of infection	258
10.1 Subintervals of BA.1/BA.2 and BA.5 dominance used in the study	272
10.2 Risk of omicron BA.5 infection at different intervals for individuals infected with BA.1/BA.2 in specific periods	273
A.1 Augmented version of the observed 2×2 contingency table in the presence of the misdiagnosis of ME/CFS cases for a classical case-control association study .	A.295
A.2 Augmented version of the observable 2×2 contingency table in the case-control association study with possible misdiagnosis of ME/CFS cases and misclassification of the true serological status	A.295
B.1 The overall and per-EBV-strain number of 15-mer peptides whose antibody responses analysed	B.297
B.2 Comparison among different null models using the Akaike's information criterion .	B.298
B.3 The top 5 most significant antibodies for each association analysis	B.299
C.1 Description of the 59 (yes–no) symptoms available in the data and domain .	C.302
C.2 Percentage of individuals with presence of each symptom, across multiple sclerosis, ME/CFS as a single cohort, and ME/CFS separated into four distinct subgroups .	C.303
C.3 AUC and its 95% confidence interval, optimal cutoff and associated sensitivity and specificity to discriminate patients with multiple sclerosis from healthy controls	C.304
D.1 Basic descriptive characteristics of healthy controls and ME/CFS patients .	D.309

D.2	List of 57 ordinal symptoms available in the data with respective description, and domain	D.310
D.3	Samples sizes of the 241 ME/CFS patients in each BIC-based latent class estimated on each domain	D.311
E.1	Nineteen and eight CpG probes located in ACE and ACE2 and shared between Infinium HumanMethylation450K and Infinium HumanMethylationEPIC arrays by Illumina	E.316
E.2	Summary data of the CpG probes including SNP or coincided with a polymorphic SNP	E.317
E.3	Estimates of the best linear regression models for 5 significant CpG probes shown in Figure 8.1C	E.317

List of Figures

1.1	Example of a diagnostic flowchart for ME/CFS	19
1.2	Total percentage of vaccine coverage among Portuguese residents by age groups	31
1.3	SARS-CoV-2 accumulation of amino acid mutations in the spike protein's S1 subunit as a function of their relative mutational fitness	32
1.4	Weekly relative frequency of most dominant SARS-CoV-2 variants in Portugal between May 2021 and December 2023	34
2.1	Symptom's similarity analysis based on the Cohen's κ coefficient	74
2.2	Estimated probability of rejecting H_0 as function of the misclassification rate . .	78
3.1	Probabilities of rejecting the null hypothesis as function of the misclassification rate	88
4.1	Probabilities of detecting an association as a function of the misdiagnosis rate .	104
4.2	Probabilities of detecting an association as a function of the misdiagnosis rate considering different combinations of sensitivity and specificity	105
4.3	The relationship between the misdiagnosis probability and the probability of detecting an association estimated from simulated data from two previously published studies	110
5.1	Preliminary multivariate analysis of the data	128
5.2	Antibody-wide association analyses when comparing ME/CFS to healthy controls	131
5.3	Statistical analysis of the antibody levels related to EBNA4_0529, EBNA6_0066, and EBNA6_0070	133

5.4	Analysis of the final classification model for predicting ME/CFS patients with an infectious trigger when compared to healthy controls	135
6.1	Age-adjusted OR for the presence of each of 48 symptoms when comparing the whole ME/CFS group to the MS group	157
6.2	Age-adjusted OR for the presence of each of 48 symptoms when comparing subgroups of ME/CFS to the MS group	158
6.3	ROC curves for the predictions based on an SL algorithm trained with 4 different classifiers and 10-fold cross-validation using antibody data and patients with MS as the controls	163
6.4	Smooth-line approximations of the relationship between \log_2 (antibody concentrations) and SL-estimated probability of ME/CFS_S1 patient when compared to patients with MS	164
6.5	Association analysis between symptoms and IgG antibody data for the MS group and ME/CFS overall and stratified into subgroups	165
7.1	Relative frequency of ordinal degrees of severity on each one of the original 57 symptoms available across the population of ME/CFS patients	187
7.2	Estimation of intra- and inter-rater agreement to assess heterogeneity on the combination of severity levels across the study population	194
7.3	Selection of the optimal number of subgroups on each symptomatological domain based on the more parsimonious model	196
7.4	Latent class analysis estimated class membership probabilities of symptom severities on different AIC-based subgroups, when profiling ME/CFS patients with the totality of available symptoms	197
7.5	Latent class analysis estimated class membership probabilities of symptom severities on different AIC-based subgroups, across the different domains	199
7.6	Plasma antibody concentrations against herpesviruses and relative seroprevalence in healthy controls and ME/CFS patients	202

7.7	Plasma antibody concentrations against herpesviruses and relative seroprevalence in healthy controls and subgroups of ME/CFS patients based on distinct domains	203
7.8	Plasma antibody concentrations against herpesviruses and relative seroprevalence in healthy controls and homogenised ME/CFS subgroups of patients based on distinct domains	205
8.1	DNA methylation analysis of 19 and 8 CpG probes located in the <i>ACE</i> and <i>ACE2</i> genes, respectively	222
8.2	Boxplots per study, group and gender of the M-values referring to probes identified in Figure 8.1C and Figure 8.1D	229
8.3	Analysis of <i>ACE</i> / <i>ACE2</i> -related data from eligible microarray-based gene expression studies	231
8.4	Analysis of <i>ACE</i> and <i>ACE2</i> expression levels from the German study	234
9.1	Protective effect of previous SARS-CoV-2 infection on infection with the Omicron BA.5 subvariant	252
9.2	Flowchart describing the population selection	256
9.3	Estimates of the impact of unreported cases of infection among the population absent from the national Covid-19 registry	262
10.1	Stability of hybrid immunity protection against BA.5 infection following infection with BA.1 or BA.2 subvariants	269
10.2	Flowchart describing the population selection	271
10.3	Variation of RR of protection against BA.5 infection over time since BA.1/BA.2 infection	276
11.1	Alternative misdiagnosis assumptions to include patient stratification	283
B.1	Distributions of the Spearman's correlation coefficient between all pairs of EBV-derived antibodies on studied populations	B.300

C.1	Quantitative serology data per study group and herpesvirus/antigen	C.305
C.2	Smooth-line approximations of the relationship between log(antibody values) and SL-estimated probability of ME/CFS_S0 patient when compared to patients with MS.	C.306
C.3	Smooth-line approximations of the relationship between log(antibody values) and SL-estimated probability of ME/CFS_S2 patient when compared to patients with MS.	C.306
C.4	Smooth-line approximations of the relationship between log(antibody values) and SL-estimated probability of ME/CFS_S3 patient when compared to patients with MS.	C.307
D.1	Latent class analysis estimated class membership probabilities of symptom severities on different BIC-based subgroups, when profiling ME/CFS patients with the totality of available symptoms	D.313
D.2	Latent class analysis estimated class membership probabilities of symptom severities on different BIC-based subgroups, across the different domains	D.314

Part I

GENERAL INTRODUCTION

Chapter 1

General introduction on the diseases of interest

The advances made in the medical field over the last century have brought significant benefits to human society. Progress, such as the eradication of diseases through the development and deployment of effective vaccines, reduction of bottleneck-like diseases, such as malaria, and the ability to manage once life-threatening diseases into chronic states (e.g., human immunodeficiency viruses, diabetes, and some forms of cancer) has significantly improved the overall life expectancy and quality of life. This progress is particularly noticeable in Western countries. However, this modern and long-lasting society has also given rise to cultural changes that are beginning to manifest. Lifestyle changes such as promoting increased protection by increasingly reducing contact with external pathogens and gradually extending working hours under more stressful conditions, coupled with continuous exposure to potential new environmental risk factors, have led to a prevalence increase in autoimmunity and complex new diseases.

This thesis aims to enhance our understanding of Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) as a disease with immune dysregulation, akin to other autoimmune conditions. Additionally, it discusses alternative stratification strategies based on distinct mechanisms, such as symptom severity or infection triggers. Our focus includes analysing patterns of antibody reactivity towards specific herpesviruses and examining symptom profiles by subgrouping patients based on disease onsets. Alternatively, we categorise patients' symptoms into domains and explore different hypothesised aetiologies across the sampled population. In the context of ME/CFS, we also explore methods to improve the consistency of case-control studies—widely used in studies on biomarkers and risk factors—by simulating scenarios under the assumption of imperfect diagnosis of ME/CFS patients (i.e., misdiagnosis). All analyses utilise available or published datasets.

Part of the present thesis was produced during the latter half of the coronavirus disease 2019 (Covid-19) pandemic. This was an unprecedented global event with a worldwide impact, which also influenced the focus of research produced at the time. Here, we assessed the protection effectiveness and stability of vaccines, coupled with prior infections (i.e., hybrid immunity) against the Omicron BA.5 subvariant over time, a worrying lineage in that period.

Before delving into the discussion of this thesis, it is necessary to provide proper introductions to both topics of ME/CFS as a heterogeneous disease and the research interest of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus in the context of the pandemic, as well as delineating the objectives and outline of the thesis. Initially, background information on the importance of stratifying ME/CFS patients to identify more consistent patient groups and implement precise and personalised treatments is provided (Section 1.1). Then, a brief introduction is given to the SARS-CoV-2 virus, the Covid-19 pandemic, and the challenges posed by rising variants of concern over time (Section 1.2). Lastly, the research aims are presented (Section 1.3) followed by the thesis structure (Section 1.4).

1.1 Myalgic encephalomyelitis/Chronic fatigue syndrome

Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) is a chronic and debilitating clinical condition with unknown aetiology and pathophysiology. The disease has no biological marker and its diagnosis is essentially one of exclusion, made on the assessment of specific symptoms together with the exclusion of other fatigue-inducing illnesses. This allows for a large combination of symptoms with varying degrees of severity and results in a heterogeneous group of diagnosed patients. Patients affected by ME/CFS usually report unexplained and persistent fatigue that is not alleviated by rest and post-exertional malaise (PEM) as key symptoms. They also experience a plethora of other symptoms related to the dysfunction of immunological, autonomic, cognitive, neuroendocrine, or neurological systems. Generally, these symptoms worsen following what would be considered normal levels of physical, mental, or emotional effort.

1.1.1 Brief history of the disease

The uncertainty around ME/CFS as a disease extends even to its nomenclature, mirroring the long-lasting debate surrounding the definition of the disease. The terms myalgic encephalomyelitis (ME) and chronic fatigue syndrome (CFS), which nowadays collectively characterise the condition, have distinct origins.

ME generally denotes the involvement of the central nervous system and brain or spinal cord inflammation. The term evolved after 1934 when a series of outbreaks of an illness that resembled poliomyelitis had afflicted individuals reporting malaise, severe headaches, sore throat, myalgia, and swollen lymph nodes (Institute of Medicine 2015, chap. 2). Over time, the diagnosis changed, and alternative names and definitions have emerged in continued efforts to describe the heterogeneous clinical syndrome. Examples of names used in the past include chronic infectious mononucleosis (Isaacs 1948), Royal Free disease (Ramsay and O'sullivan 1956; The Medical Staff Of The Royal Free Hospital 1957), benign myalgic encephalomyelitis (195 1956; Galpine and Brady 1957), epidemic neuromyasthenia (Henderson and Shelokov 1959), post-viral syndrome (Behan et al. 1985), idiopathic chronic fatigue and myalgia syndrome (Byrne 1988), and fibrositis-fibromyalgia (Pritchard 1988; Yunus 1989). The interpretation of ME also changed with time. It was used to describe a post-infectious condition due to the presence of flu-like symptoms in large groups of patients. Controversially, it was also linked to cases of mass hysteria and suggested as a psychosocial phenomenon (McEvedy and Beard 1970). Eventually, the term became more broadly used to describe an inflammatory disease with effects on the brain, muscles, and nerves (Evengård and Klimas 2002; Institute of Medicine 2015).

On the other hand, the term CFS emphasises the overwhelming fatigue experienced by patients (Wojcik et al. 2011). Introduced in 1988 by the United States (US) Center for Disease Control and Prevention (CDC), CFS was intended to replace a disease termed chronic Epstein-Barr virus (EBV) syndrome¹ and be a more inclusive name, free from unproven aetiological

¹The name prior to the introduction of the term CFS reflects the link between herpesvirus and the origin of the disease. At the time, a latent chronic infection was considered a key point for the disease. However, only a part of patients diagnosed—albeit a majority—showed high titres of IgG anti-EBV antibodies (Horwitz et al. 1985),

implications (Holmes et al. 1988). However, this definition was eventually deemed too broad and unsatisfactory in practice since it focused predominantly on fatigue, becoming an oversimplification of the true effects and making the extensions of the disease more difficult to understand (Noor et al. 2021). Additionally, the term also received complaints from patients, justifying that it “stigmatises and trivialises” the disease, leading to “social disregard” and disengagement (Institute of Medicine 2015).

Over the past decades, distinct case definitions and diagnostic criteria have been developed for the two terms (e.g., Hyde 2007 and Carruthers et al. 2011 for ME, and Holmes et al. 1988, Sharpe et al. 1991, and Fukuda et al. 1994 for CFS) and the debate for the implementation of a unifying term remained, with additional proposals being suggested as different case definitions and diagnosis criteria were published. The more recent proposes have been neuroinflammatory and oxidative fatigue (NIOF) (Maes 2015) and systemic exertion intolerance disease (SEID) (Institute of Medicine 2015).

Despite the general discussion that a more appropriate name could be adopted (Nacul et al. 2021), the composite name ME/CFS (or CFS/ME in some cases) was eventually popularised and remains the more conventional term used to describe the disease (Carruthers et al. 2003). Thus, ME/CFS can be regarded as an umbrella term that encapsulates all conditions and adjustments considered since the initial description of the disease, even if neither ME nor CFS fully capture the complexity of the disease.

1.1.2 Epidemiology

The prevalence estimates of ME/CFS range between 0.1% and 2.2% (Estévez-López et al. 2020), with an overall pooled estimate of 0.89% (95% CI = [0.60%, 1.33%]) (Lim et al. 2020). This variability is largely attributed to differences in the assumed disease definitions and diagnostic methods, as well as the study region (Vincent et al. 2012; Nacul et al. 2011; Valdez et al. 2018; Słomko et al. 2019). Despite this uncertainty, considering the current world population of 8.1 billion humans, at least 72 million individuals worldwide are affected by

making the term less inclusive.

this disorder, present across all ethnic groups and socioeconomic strata.

ME/CFS affects two to four times more women than men (Castro-Marrero et al. 2017; Lim and Son 2020). While the disease onset can occur at any age, it is most commonly reported between the ages of 20 and 40 years (Valdez et al. 2018), with proposed bimodal incidence peaks in adolescence (10 to 19 years old) and adulthood (30 to 39 years old) (Bakken et al. 2014). The fact that more women are affected and that the onset periods are related to puberty and years post-puberty—relating to pregnancy—suggests a link between sex hormones and the development of ME/CFS. This is in line with the hypothesis of hormonal differences seen in some autoimmune diseases such as systemic lupus erythematosus (SLE), multiple sclerosis (MS), or rheumatoid arthritis (RA), that are also more prevalent in women than in men and with similar ratios (Whitacre 2001; Quintero et al. 2012).

Most cases of ME/CFS are sporadic, but as mentioned in the previous section, there are also reports of cluster outbreaks (The Medical Staff Of The Royal Free Hospital 1957). The disease can manifest gradually, suddenly, or with an abrupt increase in intensity on mild chronic symptoms already present. It is characterised by multiple and heterogeneous symptoms that vary widely in severity among patients. Some patients experience fluctuating stages of remission and relapse, while others sustain mild to moderate severities and are able to work or attend school, albeit with certain physical limitations and expected regular absences. Conversely, the most severe patients are usually house- or even bed-bound, under isolation and with reduced social contact.

ME/CFS has a significant impact on the quality of life of those who suffer from it. The prolonged effects of the disease, coupled with the reduced awareness of it, contribute to feelings of self-doubt, distrust, and social stigma, and are linked with increased isolation and depression (Institute of Medicine 2015; Lacerda et al. 2019).

1.1.3 Proposed pathogenesis and mechanisms

The absence of objective biomarkers for the identification of ME/CFS limits a direct association with the disease's aetiology. As such, multiple mechanisms have been hypothesised. Some of

the proposed triggering factors include previous or persistent viral or bacterial infections (Rasa et al. 2018), genetic predisposition (Schlauch et al. 2016), presence of deleterious autoantibodies against nuclear, membrane, and neurotransmitter receptor structures (Loebel et al. 2016; Sotzny et al. 2018; Wirth and Scheibenbogen 2020), oxidative stress (Wood et al. 2021), environmental factors, severe and longstanding stress, or trauma (Rivera et al. 2019). However, given the disease heterogeneity, there is likely a coexistence of multiple pathological mechanisms occurring within the same individuals. Numerous studies have shown alterations in the immunological, genetic, and metabolic systems, and there is growing evidence for an autoimmune mechanism at the origin of the disease, at least in a subset of patients (Sotzny et al. 2018).

Viral infection prior to disease development

Although no specific pathogen has been directly associated with ME/CFS, there is a general consensus regarding the hypothesis of a triggering infectious agent. There is a history of non-specific chronic post-viral fatigue stages caused by infections such as Ross River virus (a mosquito-borne tropical disease) (Harley et al. 2002; Hickie et al. 2006), Ebola (Hickie et al. 2006; Prevail III Study Group 2019), influenza (Magnus et al. 2015), EBV (infectious mononucleosis) (Pedersen et al. 2019), bacteria *Coxiella burnetti* (Q fever) (Ayres et al. 1998), *Borrelia burgdorferi* (Lyme disease) (Shadick et al. 1994), and more recently SARS-CoV-2 (Covid-19) (Havervall et al. 2021; Choutka et al. 2022).

In fact, a majority of ME/CFS patients report an acute infection-like episode before the disease onset (Chu et al. 2019). The more commonly mentioned viral triggers are those from the *Herpesviridae* family (Blomberg et al. 2019; Ariza 2021), particularly EBV (Ruiz-Pablos et al. 2021), human herpesvirus 6 (HHV6) (Lee et al. 2021; Kasimir et al. 2022), human cytomegalovirus (CMV) (Lerner et al. 2002), herpes simplex 1 and 2 (HSV1 and HSV2) (Bond and Dinan 2006), and varicella-zoster virus (VZV) (Tsai et al. 2014; Halpin et al. 2017). These viruses are globally distributed and highly prevalent in the population, with more than 90% of the adult population being positive for at least one of them and with multiple examples of

co-infections in the population (Rasa et al. 2018). A common feature of the herpesviruses is the ability to persist inside host cells in a lifelong latent state after primary infection, enabling them to remain dormant and undetected for periods, only to reactivate and initiate a lytic replicative cycle after exposure to impactful or stressful situations (Lan and Luo 2017; Rooney et al. 2019).

Under the viral hypothesis, specific antibodies against herpesviruses proteins could potentially be used as candidate biomarkers for the diagnosis of ME/CFS. In this sense, serological studies have investigated the concentrations of these antibodies in patients, which are then related to the presence or triggering mechanism for the disease onset (Loebel et al. 2017; Blomberg et al. 2019), the severity of symptoms experienced (Domingues et al. 2023), or to propose infection-related subsets of ME/CFS (Eriksen 2018). Unfortunately, despite relative differences indicative of the immune system's involvement, there is still no established serological relationship between herpesviruses and ME/CFS (Ariza 2020).

Immunopathological mechanisms

ME/CFS individuals often manifest flu-like and immune-related symptoms such as low-grade fever, sore throat, tender lymph nodes, myalgias, and hypersensitivity to certain stimuli (e.g., intolerance to alcohol, different drugs or chemicals, light, or noise), which may be explained by alterations and abnormalities in the immune system (Underhill 2015). Several studies have proposed mechanisms for both a state of chronic inflammation or a state of immune dysregulation of the immune system (Natelson et al. 2002; Brenu et al. 2011), focusing on altered cytokine levels and immune activation (Patarca 2001; Tomoda et al. 2005; Lorusso et al. 2009), and changes in function of various types of lymphocytes, particularly natural killer (NK) cells (Klimas et al. 1990), and abnormal responses of T and B cells to specific antigens (Lorusso et al. 2009; Brenu et al. 2014b; Rivas et al. 2018). However, the results are not consensual.

Cytokine exacerbation and inflammation

Chronic inflammation and the origin and severity of symptoms experienced by ME/CFS pa-

tients could be explained by cytokines. Case-control studies have shown elevated levels of pro-inflammatory cytokines in peripheral blood and cerebral spinal fluid in patients, namely the transforming growth factor- β (TGF- β) (Montoya et al. 2017) and tumour necrosis factor- α (TNF- α) (Brenu et al. 2011), which are linked with NK cell cytotoxic activity. Other studies looking at the broad interaction of cytokines have proposed distinct cytokine profiles in patients, relating symptoms such as decreased motor activity, sleep and cognition disruption with interleukin-1 β (IL-1 β), IL-2, IL-6, IL-4, or interferon- γ (IFN- γ) (Broderick et al. 2010). Additionally, distinct pro- and anti-inflammatory cytokine signatures have been found at different stages of the disease, suggesting that the immunopathology of ME/CFS may not be static (Hornig et al. 2015).

Alterations in NK cells

NK cells play an essential role in the early response to viral infections and secretion of major inflammatory cytokines before cells from the adaptive immune system become functionally activated (Murphy and Weaver 2017). The cytotoxic activity of this population of cells and its relation to autoimmunity induced by viral reactivation of herpesviruses has been an important topic of discussion in ME/CFS research. Decrease NK cell function in patients has been reported (Caligiuri et al. 1987; Klimas et al. 1990; Aoki et al. 1993; Fletcher et al. 2010; Hardcastle et al. 2014), which led to hypotheses implicating NK cell dysfunction in ME/CFS pathogenesis. However, this interpretation on NK cell function in ME/CFS has been contested, with case-control studies failing to reproduce these findings (Theorell et al. 2017; Cliff et al. 2019) or even suggesting increased activation instead (Rivas et al. 2018).

Implications from the adaptive immune system

Concerning the involvement of the adaptive immune system in ME/CFS, both T and B cells have been suggested as potential contributors to pathogenesis.

Dysfunctional T cell phenotypes have been observed in ME/CFS, and different pathways implicated in the loss of self-tolerance, anergy, exhaustion, and senescence have been proposed for different cell states, typically associated with chronic viral infections (Maya 2023). Activation or differentiation of these cells and respective subsets towards acute viral infections

leads to substantial changes in the host's cellular metabolism, and their dysregulation may contribute to the immunological abnormalities observed (Bantug et al. 2018).

Among T cell subsets, regulatory CD4⁺ T (Treg) cells have an immunoregulatory function and maintain self-tolerance by suppressing autoreactive lymphocytes (Murphy and Weaver 2017). Studies have shown an increased number of this subset in patients with ME/CFS, suggesting a potential regulatory imbalance (Curriu et al. 2013; Brenu et al. 2014b; Ramos et al. 2016). Interestingly, theoretical models by Sepúlveda et al. (2019) looked at the cross-regulation dynamic between Treg and effector T (Teff) cells in the disease. The models propose that the high number of Treg cells might be due to molecular mimicry between viral and self-antigens, which disrupts the homeostatic state between the two cell populations, with Treg cells promoting a chronically activated state with permanent suppression of the immune response towards the virus. Conversely, reduced populations of Treg cells have also been reported in ME/CFS patients (Rivas et al. 2018), which is in accordance with studies on autoimmune diseases such as RA (Yan et al. 2022) and SLE (Barreto et al. 2009).

Implications of other T cell subsets in ME/CFS have been discussed. A case-control study showed lower activation and frequency of effector memory T cells (Curriu et al. 2013). This is consistent with a preliminary study which compared subsets of CD8⁺ T cells in ME/CFS patients with MS and healthy control individuals and proposed an exhausted profile of these cells, especially effector memory CD8⁺ T cells and their migratory potential (i.e., the ability of these cells to migrate to sites of inflammation) (Brenu et al. 2016).

Additional examples include an increased proportion of mucosal-associated invariant CD8⁺ T (MAIT) cells in severe patients (Cliff et al. 2019), heightened values of activated CD8⁺ T cells (Landay et al. 1991), and augmented Th2 response in patients (Ruiz-Pablos et al. 2021). All supporting the hypothesis of an altered immunological state.

Alterations in the frequency of B cell subsets have been studied in ME/CFS, as their involvement could potentially contribute to inflammation and immune abnormalities. Some case-control studies did not report differences (Curriu et al. 2013; Bradley et al. 2013; Mensah et al. 2016). But others have shown increased frequencies of naive, and memory B cells (Brenu et al. 2014b; Ono et al. 2017). Additionally, Klimas et al. (1990) reported elevated numbers of

CD21⁺ and CD20⁺ B cell subsets, proposing an immune profile compatible with chronic viral reactivation.

B cell dysregulation and the production of autoantibodies in ME/CFS have also been researched. In this regard, a double-blind, placebo-controlled phase II study reported early success with the monoclonal anti-CD20 antibody rituximab—used to treat B cell malignancies—improving self-reported fatigue scores in a majority of ME/CFS patients 2 to 7 months after treatment when compared to the placebo group (Fluge et al. 2011). However, a trial with larger sample sizes for both treatment and placebo groups failed to corroborate the early results (Fluge et al. 2019; Rowe 2019).

Molecular mimicry

Multiple studies reporting immunological, genetic, and metabolic alterations in patients have linked ME/CFS with autoimmune mechanisms (Sotzny et al. 2018). As previously noted, the viral hypothesis includes the possibility of molecular mimicry. This mechanism occurs when sequence similarities between foreign- and self-peptides induce cross-activation of autoreactive cellular populations from the adaptive immune system against specific herpesvirus IgG with human autoantigens (Fonseca et al. 2024). This can lead to chronically activated immune responses attempting to control latent infections, posing a high deleterious autoimmune potential (Blomberg et al. 2018; Sepúlveda et al. 2019; Sundaresan et al. 2023). Interestingly, specific EBV antigens have been reported to share sequence homology with certain human peptides derived from myelin basic protein (Lünemann et al. 2008), lactoperoxidase (Loebel et al. 2017), and anoctamin-2 (Tengvall et al. 2019; Sepúlveda 2021). Moreover, predicting models have found that certain subgroups of patients show increased IgG antibody levels against two EBV-related antigens, particularly those with a reported infectious onset of disease (Sepúlveda et al. 2022).

Molecular mimicry has also been identified in autoimmune diseases, such as RA or SLE (Rojas et al. 2018), and sequence homologies specific to EBV antigens and epitopes from human proteins have been identified in MS (Sospedra et al. 2005; Gabibov et al. 2011).

Alternatively, the latent chronicity of the condition could also be explained by the danger

theory (Pradeu and Cooper 2012), which suggests that infections can result in chronically activated immune responses, leading to persistent inflammations and the observed phenotype of ME/CFS (Sepúlveda et al. 2019; Fonseca et al. 2024).

Endothelial dysfunction

Endothelial cells have an active involvement in the regulation of inflammatory responses and the immune system (Pober and Sessa 2007). They produce cytokines and cell-adhesion molecules that promote binding and movement of circulating immune cells towards infected tissues (Murphy and Weaver 2017). This has led to the proposal of endothelial dysfunction (ED) as a possible pathogenesis of ME/CFS, linking the dysfunction with chronic inflammation and even the severity of symptoms (Scherbakov et al. 2020). Blauensteiner et al. (2021) reported altered expression of ED-related microRNAs in plasma from affected individuals, and Bertinat et al. (2022) used an *in vitro* model to show that cultured endothelial cells exposed to plasma from ME/CFS patients have lower ability to produce nitric oxide. The decrease in nitric oxide, together with results for elevated concentrations of the vasoconstrictor Edothelin-1 (Haffke et al. 2022), support the idea that ED could be at the origin of oxidative stress promotion and increased vasoconstriction activity (Kostov 2021). Furthermore, there is evidence of the involvement of ED in other autoimmune disorders linked with chronic inflammation (Steyers and Miller 2014).

Genetic predisposition

As a multifactorial disorder, ME/CFS has been hypothesised to occur under genetic and epigenetic contributions (Wang et al. 2017). Studies done within members of the same family suggest genetic factors may contribute towards predisposition and increased risk for the disease (Walsh et al. 2001; Albright et al. 2011), and research on twins with ME/CFS has provided evidence supporting a hereditary relation to ME/CFS (Buchwald et al. 2001).

Genome-wide association studies

Genome-wide association studies (GWAS) have the appeal of allowing for an ample genome

scan in an unbiased manner and without pre-existing hypotheses on which genes or loci may be involved in the disease. However, in order to achieve sufficient statistical power, GWAS usually require large sample sizes of tens of thousands of individuals, and ME/CFS studies often lack funding to recruit such numbers of participants. Consequently, few GWAS have been conducted in this field. Nevertheless, results have identified moderate associations between ME/CFS and single-nucleotide polymorphisms (SNPs) from immunologically relevant genes, suggesting that a potential genetic predisposition for immune dysregulation may exist (Smith et al. 2011; Schlauch et al. 2016).

It is worth mentioning that there are currently large-scale studies being conducted in ME/CFS, such as DecodeME (Devereux-Cooke et al. 2022). With an expected participation of over 25,000 individuals, this GWAS is anticipated to be the largest one to date. The study aims to enhance the ability to detect genetic variations of lower relative risk towards the disease, in the hope of identifying a genetic signature of ME/CFS (Dibble et al. 2020).

Candidate gene studies

Candidate gene studies have also proposed genetic polymorphisms associated with ME/CFS susceptibility. Examples of possible predisposing factors include association of autoimmunity-risk alleles (*CTLA4* rs3087243 and *PTPN22* rs2476601) in a subset of patients reporting an acute infection at the onset of the disease (Steiner et al. 2020), 21 SNPs as possible markers for subgroups of patients when compared with individuals with depression (Shimosako and Kerr 2014), and a link to regulatory pathways of immune neurotransmitter systems and human leukocyte antigen (HLA) (Wang et al. 2017).

Another focus of genetic research has been the HLA complex. This group of genes constitute the major histocompatibility complex (MHC) and is responsible for encoding a wide variety of cell surface markers, antigen-presenting molecules and other proteins involved in immune function (Murphy and Weaver 2017). HLAs are important to enable the immune system to differentiate self- from non-self-antigens, and they are highly polymorphic. This high level of polymorphism may lead to genetic variations that can influence an individual's susceptibility to autoimmune diseases. In fact, the link between HLAs and autoimmune diseases is well-

established (Trowsdale and Knight 2013; Cruz-Tapias et al. 2013; Matzaraki et al. 2017).

In the context of ME/CFS, Lande et al. (2020) found associations between alleles from class II gene *HLA-DQB1* and disease risk, and a follow-up study from the same group showed other independent associations with HLA class I and class II regions in ME/CFS (Hajdarevic et al. 2021).

Gene expression studies

Studies on gene expression have also identified genes related to immune dysfunction that may be implicated in predisposition and overall pathophysiology of ME/CFS (Vernon et al. 2002). A study conducted on adolescents diagnosed with ME/CFS reported downregulation of genes related to B cell differentiation and survival, and upregulation of genes related to antiviral responses and inflammation (Nguyen et al. 2017). These genetic patterns may be related to some of the symptoms experienced by affected individuals, including PEM and autonomic symptoms.

Epigenetics

Epigenetic modification, such as DNA methylation, may also play a role in the disease. DNA methylation mainly occurs in the cytosines of CpG dinucleotide sites across the genome and plays an important role in the regulation of gene expression (Moore et al. 2013). Studies on DNA methylation have identified differentially methylated CpG sites in genes related to the pathways related to cell signalling and immune response, neurological pathways, cellular metabolism, and kinase activity (Vega et al. 2014; Trivedi et al. 2018; Helliwell et al. 2020). For instance, Herrera et al. (2018) reported differentially methylated CpG sites associated with T cells ME/CFS patients. These results were in line with findings linking methylation in ME/CFS patients and functioning of CD4⁺ T cells (Brenu et al. 2014a) and suggest the involvement of genetic and epigenetic factors in the hypothesis of immune system dysregulation. Alternatively, other studies found methylated sites in genes associated with cellular metabolism, which could also suggest impairment in cellular energy production in ME/CFS (Vega et al. 2017). To investigate the differences seen within ME/CFS populations, authors have proposed that the stratification of patients could help to further corroborate the results (Vega et al. 2018).

Other disease mechanisms

Various alternative and complementary pathological mechanisms have been hypothesised in the context of ME/CFS. One such mechanism is metabolic dysfunction (Maya 2023), which is indicated by impaired mitochondrial functions (Nilsson et al. 2020) and abnormalities in metabolic pathways found in patients (Naviaux et al. 2016). Metabolic dysfunction is highly correlated with genes, age, family history, and external factors such as infections, microbiome, diet, and exercise.

There is also evidence that the autonomic nervous system is affected, with studies showing an increase in baseline and maximum heart rate on standing and tilting (Freeman and Komaroff 1997). This suggests that the sympathetic and parasympathetic functions of the nervous system may be altered (Goldstein 2013).

Additionally, life stressors have been shown to negatively impact the neuroendocrine circuits of stress (Stojanovich and Marisavljevich 2008; Bested and Marshall 2015). Longstanding or traumatic events can lead to changes in multiple systems described above, as well as the possible reactivation of different herpesviruses (Segerstrom and Miller 2004; Kuratsune and Watanabe 2008; Rivera et al. 2019).

1.1.4 Clinical manifestations and diagnosis

Apart from long-lasting fatigue, post-exertional malaise (PEM) is the hallmark symptom used in most case definitions and the primary distinction between ME/CFS and other fatigue-inducing illnesses. PEM leads to the worsening of symptoms in a disproportionate form after minimal physical, mental, or emotional activity levels, with prolonged recovery times that may last several days. Other key symptoms present in most case definitions include sleep disturbance or unrefreshing sleep (or both), cognitive impairment (commonly known as “brain fog”), orthostatic intolerance, muscular and joint pain, and headaches of a new type, pattern, or severity than before (Lim et al. 2020). The wider list of symptoms can be included in domains linked with possible mechanistic systems (Table 1.1).

To date, there is no definitive diagnostic test or biological marker for ME/CFS diagnosis

Table 1.1: Common domains and list of associated symptoms for ME/CFS. Source: Carruthers et al. (2003); Institute of Medicine (2015); Lacerda et al. (2017). PEM, post-exertional malaise; IBS, irritable bowel syndrome.

Domain	Associated symptoms
PEM	Marked physical or mental fatigue or exhaustion after minimal exertion or effort lasting >24 hours, fatigue or exhaustion after normal levels of activity lasting >24 hours, malaise after exertion or effort lasting >24 hours, worsening of symptoms after exertion or effort lasting >24 hours, pain after exertion or effort lasting >24 hours, and intolerance to exercise
Sleep	Sleep disturbances, unrefreshing sleep, insomnia or difficulties falling asleep, and non-restorative sleep
Neurocognitive	Brain fog or confusion, trouble concentrating, short-term memory problems, attention deficits, slow thinking, disorientation, loss of balance or unsteadiness of feet while standing up, poor coordination or unsteady movements, neck weakness, muscle discomfort or weakness, muscle twitching, tingling or numbness in arms or legs, unusual sensitivity to light or noise, temporary disturbance in eyesight, and difficulty in making decisions, retaining information, understanding things or thinking clearly, or finding or saying words
Autonomic	Dizziness while standing up, intolerance to standing up, feeling lightheaded, palpitations, palpitations while standing up, feeling sick or nauseous, dyspnea, bladder problems, pallor, and IBS symptoms
Immunological	Fever or chills, flu-like symptoms, frequent viral infections, tender glands in neck or armpit, sore throat, new sensitivities to food, medications, chemicals, or odours, stiffness in the mornings, and intolerance to alcohol
Neuroendocrine	Being unusually sweaty, unusually cold hands or feet, intolerance to extremes of heat or cold, decreased sexual interest or function, abnormal appetite or significant changes in weight, and worsening symptoms with stress
Pain	Headaches new, different or worse than before, migraine different or worse than before, pain in the chest or abdomen, pain in ≥2 joints without swelling or redness, joint pains moving to different joints without swelling or redness, and muscle pain

(Scheibenbogen et al. 2017). Consequently, the characterisation of the disease relies on the use of symptom-based case criteria that focus on the combination of symptoms while excluding any other known potential diseases that could explain fatigue and other major symptoms (Smith et al. 2014; Institute of Medicine 2015). These diagnostic criteria are often self-report assessments designed to screen for non-specific symptoms that may overlap with those other diseases and medical conditions (Brurberg et al. 2014).

Over 20 diagnostic criteria have been proposed for ME/CFS (Bayliss et al. 2014; Brurberg et al. 2014). These criteria have similarities and differences (Lim and Son 2020; Malato et al. 2021). Overall, all have a similar workflow of evaluation steps where suspected patients can be ruled out before the disease diagnosis (Figure 1.1). The lists of symptoms and exclusionary conditions assessed are what vary across the criteria (Lim and Son 2020).

The definitions more widely used are the Fukuda/CDC (CDC-1994, Fukuda et al. 1994), the Canadian Consensus Criteria (CCC-2003, Carruthers et al. 2003), the International Consensus Criteria (ICC-2011, Carruthers et al. 2011), and the Institute of Medicine Criteria (IOM-2015, Institute of Medicine 2015).

Briefly, the CDC-1994 requires a patient to display unexplained, persistent, or relapsing fatigue for at least six months, accompanied by at least four of eight additional fatigue-related symptoms (Fukuda et al. 1994). The symptoms are (1) impaired short-term memory or concentration, (2) sore throat, (3) tender cervical or axillary lymph nodes, (4) muscle pain (5) multi-joint pain without swelling or redness, (6) headaches that are new, different or worse, (7) unrefreshing sleep, and (8) PEM. This criterion is mainly used in research.

The CCC-2003 also requires the illness to persist for at least six months, with the hallmark symptoms being fatigue that substantially reduces the activity level, PEM, sleep disturbances, and muscle or multi-joint pain. Additionally, the CCC-2003 requires two or more neurological or cognitive symptoms, and at least one manifestation from two domains: autonomic, neuroendocrine, or immune (Carruthers et al. 2003).

The ICC-2011 is a modification of the previous CCC-2003 and focuses less on the prolonged and persistent characteristics of fatigue. As such, there is no need to maintain the symptoms for a minimum of six months; the diagnosis can be made immediately. Individuals are diagnosed

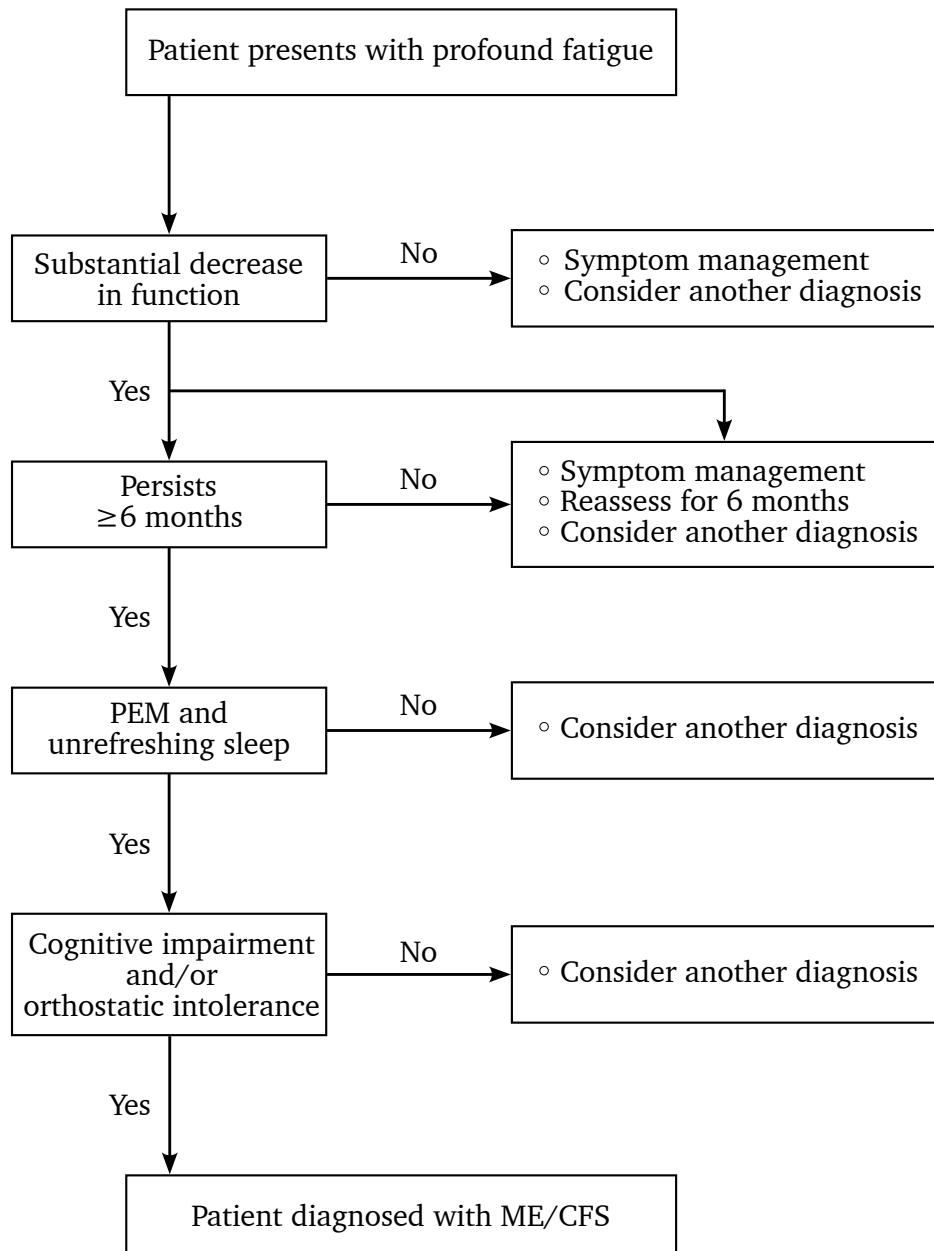


Figure 1.1: Example of a diagnostic flowchart for ME/CFS. Source: Institute of Medicine (2015, chap. 7). PEM, post-exertional malaise.

if fatigue results in at least a 50% reduction of their pre-morbid activity levels. Additionally, aside from post-exertional neuro-immune exhaustion (equivalent to PEM), ICC-2011 requires at least one neurological symptom, at least one immune, gastro-intestinal, or genitourinary symptom, and at least one symptom related to metabolism/energy production impairment (Carruthers et al. 2011).

The IOM-2015 was developed as a primary diagnostic tool for clinical purposes. It requires the presence of profound fatigue for six or more months, PEM, and sleep abnormalities as hallmark symptoms. It focuses on other specific symptoms impacting the patient's pre-illness activity levels and causing distressing levels of cognitive impairment and/or orthostatic intolerance (Institute of Medicine 2015).

Exclusion criteria are applied to all of the criteria mentioned above, with the exception of IOM-2015. These criteria include active medical conditions or disease processes that could explain the major symptoms of fatigue (e.g., untreated hypothyroidism or sleep apnea), psychiatric disorders, major depressive disorder, alcohol or other substance abuse, severe obesity (body mass index greater than 45 kg/m²) or eating disorders, and chronic viral infections. The list of exclusionary conditions also varies among case criteria, and efforts have been made to standardise the selected conditions across studies (Jason et al. 2023).

The lack of consistency across case definitions, in both inclusion and exclusion criteria, poses a major challenge in ME/CFS research (Jason et al. 2014; Nacul et al. 2017). The heterogeneity seen in patient cohorts, from the multiplicity of symptoms and use of different case definitions, inevitably contributes to the lack of replication observed in studies and hinders the identification of biomarkers and effective treatments (Nacul et al. 2019; Malato et al. 2021; 2023a). Moreover, ME/CFS might also represent a spectrum of diseases with similar clinical symptoms but potentially diverse mechanisms and pathways (Jason et al. 2005). These factors contribute to delays and possible misdiagnosis of patients, and it is expected that there is still a large number of non-diagnosed cases (Solomon and Reeves 2004; Bayliss et al. 2014).

1.1.5 Connection to other diseases

When investigating pathophysiological pathways or potential disease biomarkers in ME/CFS, many studies include healthy individuals as the control group (some of them self-reported participants). Conversely, comparing ME/CFS with other fatigue-inducing diseases has also proven to be of value. Firstly, as mentioned in the previous section, individuals with biologically identifiable diseases that explain most symptoms overlapping with ME/CFS, such as cancer or diabetes, can be excluded from the diagnosis (Carruthers et al. 2003). Secondly, exploring associations with other post-infectious and autoimmune diseases can be essential in understanding how to better characterise ME/CFS, helping to improve diagnosis or treatment strategies. ME/CFS is comparable to diseases such as multiple sclerosis or RA in terms of prevalence, long-term disability, and impact on the quality of life.

Multiple sclerosis

Multiple sclerosis (MS) can be included as disease control in ME/CFS studies. MS is an autoimmune disease of the central nervous system. It is characterised by the presence of perivenular inflammatory lesions, which cause the destruction of myelin sheaths surrounding nerve cell axons in the brain and spinal cord (Murphy and Weaver 2017; Dobson and Giovannoni 2019). Although the cause of MS is unknown, there are proposals for a combination of genetic and environmental risk factors, such as post-acute infections by herpesviruses like EBV (Handel et al. 2010; Bjornevik et al. 2022) and HHV6 (Engdahl et al. 2019), and vitamin D (Dobson and Giovannoni 2019).

MS patients exhibit symptoms overlapping with those of ME/CFS, including chronic and disabling fatigue, autonomic, and neurocognitive symptoms (Morris and Maes 2013; Gaber et al. 2014). However, there are distinct symptomatological patterns able to differentiate between the two conditions (Jason 2017; Domingues et al. 2023). Notably, Ohanian et al. (2016) reported that tender lymph nodes and flu-like symptoms could differentiate MS and ME/CFS individuals with an accuracy of approximately 81%. MS also serves as direct control in ME/CFS studies assessing the autoimmune hypothesis, and several disease pathways have been researched, such as immune response towards viral triggers (Loebel et al. 2017), immune

dysregulation and functioning (Ramos et al. 2016; Cliff et al. 2019), or mitochondrial dysregulation and muscle fatigue (Melvin et al. 2019). For instance, Cliff et al. (2019) compared ME/CFS subgroups with MS and found significant differences in immune-cell populations, including increased monocytes, dendritic cells, and CD4⁺ T cells, and decreased CD8⁺ T cells, which in turn were not observed when compared with healthy controls.

Rheumatoid arthritis

RA is a chronic inflammatory autoimmune disorder caused by autoreactive T cells and autoantibodies against antigens localised in joint synovium (Murphy and Weaver 2017). This results in the joint's inflammation and destruction, causing arthritis. The aetiology of RA remains unclear but it involves a combination of environmental and genetic factors, which also play a role in the disease severity (Majithia and Geraci 2007).

Similarly to MS, RA shares a considerable number of symptoms with ME/CFS. Both diseases cause malaise, widespread pain, and musculoskeletal symptoms that lead to functional impairment (Moss-Morris and Chalder 2003). Additionally, approximately 40% of RA patients meet the criteria for severe fatigue (Katz 2017).

Although not many studies include cohorts of ME/CFS and RA individuals, differences have been shown among the two illnesses. Regarding symptoms, when measurements are done using the same protocols (e.g., using the Disease Activity Score, DAS-28), RA patients generally report worse results in pain scores and tender and swollen joints. Studies have also shown evidence of neuroinflammation in the brain of ME/CFS patients and failed to replicate the results for RA under similar conditions (Mueller et al. 2020). Such differences suggest that immune activation in ME/CFS may have effects on the central nervous system, as opposed to RA, which seems to be a disease primarily of the peripheral joints (Mueller et al. 2020).

Other diseases

There are other illnesses that have been researched and show overlapping symptoms and mechanistic results to ME/CFS. For example, fibromyalgia (FM), which also does not have a biomarker, encompasses a heterogeneous population characterised by heightened central sensitivity to peripheral sensations (Kodner 2015). The overlap between FM and ME/CFS

has even led to discussions about whether the two illnesses represent distinct expressions of the same syndrome, and studies have analysed their differences (Abbi and Natelson 2012; Castro-Marrero et al. 2013). Vega et al. (2014) found a pattern of epigenetic modifications in ME/CFS related to immune response and discussed the possible biological difference between the two conditions by highlighting that epigenomic analyses in FM patients found differential methylation of genes associated with structural and nervous system development and neuron differentiation instead. Another key difference is that physical activity in FM patients can lead to positive outcomes and alleviate the experienced symptoms of pain, whereas in ME/CFS patients, exercise worsens PEM and related symptoms (Häuser et al. 2010; Kodner 2015).

Examples of other diseases mentioned in ME/CFS research are SLE (Sotzny et al. 2018), Sjögren's syndrome (Calabrese et al. 1994; Kim et al. 2023), Gulf War syndrome (Kang et al. 2003; Halpin et al. 2017), and Long Covid (Komaroff and Lipkin 2023; Gil et al. 2023). Notably, these and other diseases share a link to post-acute infection sequelae (Choutka et al. 2022). Associations between the latter condition and ME/CFS will be briefly discussed later in the thesis (Chapter 11).

1.1.6 Treatment and prognosis

As of now, there is no known cure for ME/CFS. The updated guidelines from the National Institute for Health and Care Excellence (NICE) state that no drug treatment has been found to be a safe and effective cure for ME/CFS (NICE Guideline [NG206] 2021). Therefore, treatments and pharmacological interventions are mostly recommended for symptom relief and management.

The primary symptoms focused for improvement include PEM, unrefreshing sleep and sleep habits, pain and discomfort in muscles and joints, headaches, dizziness or orthostatic intolerance, and memory and concentration problems. While anti-viral therapies and treatments for immune dysfunction, hypothalamic-pituitary-adrenal axis abnormalities, and autonomic or central nervous system dysfunction may be considered, their efficacy remains uncertain (Carruthers et al. 2003). Alternatively, non-pharmacological interventions and medications

may also be prescribed for the treatment and management of symptoms (Rowe et al. 2017). However, ME/CFS patients often exhibit hypersensitivity to standard medications given in the usual doses and can experience side effects or worsened symptoms (Carruthers et al. 2003).

In addition to direct action towards symptoms, lifestyle adjustments and self-help therapies are also focused in ME/CFS. Clinical support focuses on assisting patients in understanding how to cope with the disease, striking a balance between rest and activity to prevent the worsening of their symptoms (Carruthers et al. 2003; Rowe et al. 2017). Educating the patients and their support networks about the illness can also be important in fostering a sense of control over the symptoms, which has shown to be a positive long-term outcome (Cairns and Hotopf 2005). While adjusting to the disease, patients may experience depression, stress, and anxiety states, which can be addressed through supportive counselling and medical recommendations.

Ultimately, while both direct actions on specific symptoms and overall patient well-being can provide some relief and improve quality of life, the approaches are very case-dependent, and treatment programs are recommended to be individualised, with regular assessments to monitor progress and watch for the possible emergence of other illnesses (NICE Guideline [NG206] 2021).

The long-term prognosis for ME/CFS remains uncertain and varies among individuals. While some studies suggest better outcomes with older age (Ghali et al. 2022), the general trend is the reduction or improvement of symptoms rather than full recovery (Carruthers et al. 2011; Nacul et al. 2021).

1.1.7 Research challenges

ME/CFS has a complex aetiology and multifactorial nature, with signals of involvement from immune dysregulation, genetic predisposition, environmental factors, and potential viral infections (Rivera et al. 2019). This diversity of influencing factors contributes to the heterogeneity of observed symptoms.

The lack of an established diagnostic biomarker and use of different symptom-based case criteria for ME/CFS diagnostic may lead to the identification and inclusion of potentially mis-

diagnosed patients (Nacul et al. 2017). This is evidenced by the ample prevalence estimates and competing results proposed across various fields of research (see Section 1.1.3). Misdiagnosis can arise from the lack of agreement between case definitions when diagnosing distinct suspected cases (Nacul et al. 2017; Malato et al. 2021), but it can also occur when failing to diagnose a known disease that would otherwise exclude individuals from being possible cases of ME/CFS (Nacul et al. 2019; Malato et al. 2023a).

In this regard, several works have proposed the stratification of ME/CFS patients into specific subgroups as a way to minimise this transversal effect (Table 1.2). Diagnosed patients can be split by case definition, severity of symptoms, trigger for the disease, or age (Janal et al. 2006; Lewis et al. 2013; Hardcastle et al. 2014; Domingues et al. 2021). Alternatively, profiles related to immune and genomic subtypes have also been suggested (Kerr et al. 2007; 2008; Vega et al. 2018). Furthermore, duration and disease dynamics could also be linked with functional differences in ME/CFS, as patients might have distinct immune cell dysfunctions over time (Maya 2023). In this regard, a longitudinal study on cell cytotoxicity and cytokine secretion showed inconsistent levels of cytokines in the same individuals over different time-points (Brenu et al. 2012). Perhaps the identification of immune dysfunctional states (such as anergy, exhaustion, and senescence) within the immune cell populations would help to better understand the disease stage of the patients and help predict disease progression.

Likewise, the use of self-reported cases—in both patients and, at times, healthy controls—is also a problem. For example, a Polish study assessing the prevalence and characteristics of ME/CFS in a community identified 1400 individuals who believed to be suffering from ME/CFS, but only 69 individuals actually complied with the CDC-1994 case definition (Słomko et al. 2019).

In recent years, clinicians and researchers have coordinated efforts to define a recommended and standardised clinical diagnosis for ME/CFS, to be used in both clinical settings and research studies (Nacul et al. 2021). This involves the application of appropriate questionnaires, coupled with a comprehensive medical history, physical examination, functional tests and analyses, and appropriate differential diagnostics (Pheby et al. 2020; Steiner et al. 2023). The stratification of ME/CFS into more homogeneous clusters also has been proposed, as it

ensures more consistent results in both the diagnosis and sampled cohorts used in research (Jason et al. 2005). In fact, stratification of diagnosed patients by disease onset, disease duration, and sex is one of the strategies for the development of diagnostic biomarkers proposed by the European Network on ME/CFS (EUROMENE) Biomarker Landscape project (Scheibenbogen et al. 2017, Table 3).

Table 1.2: List of suggested proposals for patient stratification.

Stratification	Considerations	Examples of references
Presence/severity of symptoms	Compare subgroups for Absent vs. Present or Absent–Mild vs. Moderate–Severe	Landay et al. (1991); Hardcastle et al. (2014); Montoya et al. (2017); Cliff et al. (2019)
Subgroups of symptoms	Identification of similar clusters within specific domains (Table 1.1)	Asprusten et al. (2021); Chapter 7
Disease onset/trigger	Link with (auto)immune dysregulations and production of autoantibodies; Unknown/No infection vs. Infection	Szklarski et al. (2021); Domingues et al. (2021); Ruiz-Pablos et al. (2021); Sepúlveda et al. (2022)
Disease duration	Linked with disease progression stages; Prodromal vs. Early vs. Established disease	Brenu et al. (2012); Stoothoff (2017); Nacul et al. (2020)
Age	Differences in severity of experienced symptoms	Itoh et al. (2012); Lewis et al. (2013); Miike and Bell (2008)
Sex	Hormonal differences and possible predisposition towards the disease	(Pipper et al. 2023)

The lack of large-scale longitudinal study designs and overall large sample sizes are also a challenge for ME/CFS research. Smaller study cohorts, together with the possible misdiagnosis of patients, reduce the statistical power to test hypotheses and validate potential associations found (Malato et al. 2022a). Moreover, limited funding can also be a hindering factor for research. A study in the US estimated that ME/CFS poses an economic burden of 36 to 51

billion dollars annually (Jason and Mirin 2021), and studies in Europe propose similar values, adding that a modest 1% reduction in the overall burden of ME/CFS could deliver annual cost savings of approximately 400 million euros (McCrone et al. 2003; Pheby et al. 2020). Yet, ME/CFS receives comparatively less research funding than other chronic diseases with similar burden (Dimmock et al. 2016; Mirin et al. 2022).

In the last decades, biobanks specialised in ME/CFS have been created with the goal of enhancing biomedical research related to pathophysiology, biomarkers and therapeutic approaches. One such example is the United Kingdom ME/CFS biobank (UKMEB), where ME/CFS patients have been clinically evaluated and their diagnosis must be in accordance with specific case definitions (Lacerda et al. 2017; 2018). The UKMEB has also collected samples from other donors to serve as control groups, namely healthy individuals and MS patients. The data from different population groups collected by the UKMEB has been included in the analysis of this thesis. Specifically, the data related to symptoms (Chapter 2, Chapter 6, and Chapter 7) and antibody concentrations (Chapter 6 and Chapter 7).

1.2 SARS-CoV-2 and Covid-19

The SARS-CoV-2 virus is a positive-stranded RNA virus from the *Coronaviridae* family and *Betacoronavirus* genus that infects humans and other mammals and causes Covid-19, the respiratory and contagious disease responsible for the Covid-19 pandemic.

This airborne virus primarily spreads through close contact, via aerosols and respiratory droplets. The infection begins with the successful binding of the virus's spike proteins—the structural protein responsible for the “corona” naming—to the angiotensin-converting-enzyme 2 (ACE2) receptor, present at the host cell surface (Ge et al. 2013; Hoffmann et al. 2020). This enzyme and its implication in the disease will be more thoroughly discussed in a later chapter (Chapter 8). The incubation period varies from 2 to 10 days, with an estimated median of 5.8 days (Wei et al. 2022). The differences in incubation period mostly depend on the infected individuals but have generally been decreasing with newer variants overall (wild-type 5.2 days, Alpha 5 days, Beta 4.5 days, Delta 4.41 days, Omicron 3.42 days, Wu et al. 2022).

There are various factors that can influence Covid-19 transmission, such as population density, healthcare infrastructure, and employed public health measures (Halaji et al. 2021). Most infected individuals experience mild to moderate symptoms, but certain risk factors such as obesity, advanced age, and pre-existing health conditions like hypertension, diabetes, cancer, and immunosuppression can lead to more severe cases and complications (Meyerowitz et al. 2024). These complications may include acute respiratory distress syndrome, arrhythmia, acute lung injury, and critical cases can evolve into pneumonia and multiorgan failure (Wang et al. 2020; Lamers and Haagmans 2022).

Clinical diagnosis often relies on a combination of observed symptoms, epidemiological history, and laboratory tests for viral detection. The most common clinical signs are fever, chills, sore throat, new and persistent cough, or loss or change of sense of taste or smell. Other less frequent symptoms can be experienced, such as headaches, dyspnea, and general fatigue. There are also reports of high numbers of asymptomatic infections, which could have contributed to amplifying the outbreak through silent spread (World Health Organization 2023a). After the period of infection where individuals can propagate the virus, most eventually recover without

major complications or sequelae. However, the more acute cases may require hospitalisation, where high-flow therapy and ventilation are the more common treatments (Wang et al. 2020). Over time, effective vaccines were developed, which improved the overall individual response to the virus and reduced the disease spread.

1.2.1 Brief overview of the pandemic

The transmission of SARS-CoV-2 among humans started around December 2019 and initial cases were observed in the city of Wuhan, China (Bergeri et al. 2022). Unlike previous outbreaks of human coronavirus infection, such as the SARS outbreak in 2003 and the Middle East respiratory syndrome (MERS) in 2012, Covid-19 showed a higher reproductive rate and spread rapidly around the world (Liu et al. 2020). This rapid spread, coupled with a high death rate, evolved into causing millions of new daily cases and deaths, resulting in the World Health Organization (WHO) declaring it a public health emergency of international concern on January 30, 2020—a declaration that was only issued on five occasions (Wilder-Smith and Osman 2020).

During this period, the Covid-19 pandemic posed unprecedented challenges to the health-care systems globally. While specific vaccines were being researched, public health strategies proposed by governments ranged from less strict lockdown measures to the complete isolation of countries with travel restrictions (Wilder-Smith and Osman 2020). These measures, while aimed at containing transmission, also had significant social, economic, and political ramifications (Chu et al. 2020). The global impact of the pandemic underscored the need for collaborative efforts and communication on a global scale, resulting in extensive research into vaccines and therapeutic interventions.

The WHO declaration for public health emergency of international concern ended on May 5, 2023 (World Health Organization 2023b). As of early 2024, the pandemic has resulted in more than 774 million reported cases and more than 7 million deaths worldwide (World Health Organization 2023c).

1.2.2 Vaccines and vaccination

The rapid development and distribution of several highly effective SARS-CoV-2 vaccines is a story of success. Their deployment across the world brought hope and expectations for the remission of the virus, the end of the pandemic and a gradual return to a “new normal” (an expression widely used by the Portuguese media at the time). Various vaccines and vaccination methods have been approved and developed, including mRNA, adenovector, inactivated vector, and protein subunit vaccines (Ao et al. 2023).

Briefly, mRNA vaccines encode variants of the spike protein antigen. Depending on the vaccine, they may contain either the full-length spike glycoprotein or versions of the receptor-binding domain. These vaccines elicit effective responses from the adaptive immune system in terms of T cell activation and production of neutralising antibodies (Chaudhary et al. 2021). mRNA vaccines were first produced and implemented globally during the Covid-19 pandemic. They gained recognition at the time for being an innovative and fast approach to vaccine production and vaccination against infectious diseases. Some examples of mRNA vaccines include Pfizer–BioNTech (BNT162b1 and BNT162b2) and Moderna Biotech (mRNA-1273).

Adenovirus vector vaccines use engineered adenoviruses that cannot replicate (replication-incompetent) as delivery mechanism for the genetic material encoding the SARS-CoV-2 spike protein. These vaccines are known to elicit long-lasting immunity with one or two doses (Mendonça et al. 2021). Examples of this type of vaccine are the Oxford–AstraZeneca and the Johnson & Johnson (Janssen).

Inactivated vaccines employ whole virus particles that have been inactivated or killed as an immunogen (Jin et al. 2022). One example of such a vaccine implemented during the Covid-19 pandemic is the CoronaVac, also known as Sinovac. Albeit effective, these vaccines may have lower immunogenicity when compared to other types (Lim et al. 2021).

Protein subunit vaccines have been used for decades. They use specific immunogenic epitopes from the virus that are antigenic and confer strong humoral immunity (Rezaei and Nazari 2022). These were some of the more extensively researched and approved types of SARS-CoV-2 vaccines, with Novavax being the most well-known (Ao et al. 2023).

Vaccine implementation and vaccination programmes varied across countries and regions. The coverage rates depended on various factors such as vaccine availability, distribution infrastructure, and government policies (Irwin 2021; Blasioli et al. 2023). Overall, most countries implemented vaccination programmes in phases, initially prioritising high-risk groups such as healthcare workers, elderly individuals, and people with health conditions and risk factors mentioned above. Subsequent phases expanded the eligibility criteria to gradually include the broader population, usually stratified by age groups, occupation, and risk factors. One example of country-wide, phased vaccine implementation is Portugal (Figure 1.2). By the end of 2021, over 98% of residents above 12 years old had received at least the first dose of the vaccine.

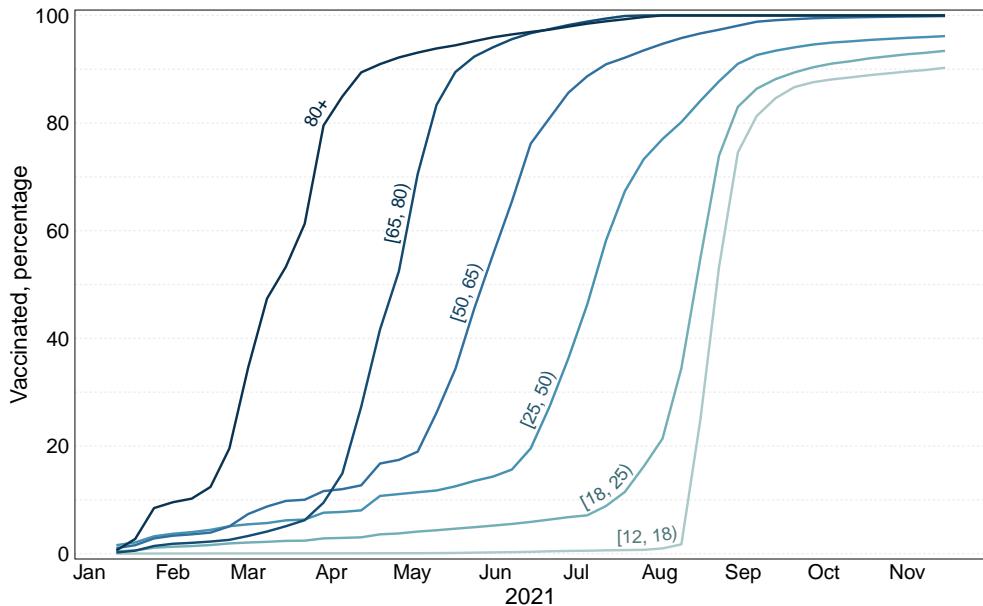


Figure 1.2: Total percentage of vaccine coverage among Portuguese residents by age groups. Available date ranges: Jan 11 2021 to Nov 15 2021. Source: National Directorate of Health (2021).

As of early 2024, 13.59 billion Covid-19 vaccine doses have been administered, and approximately 67% of the world population has received at least one Covid-19 vaccine dose (World Health Organization 2023c).

However, the virus continued to evolve rapidly through mutations under immune selective pressure, granting diversification into variants with distinct phenotypic characteristics for transmissibility, severity, and immune evasion (Markov et al. 2023). This led to continued

reports of new infections, even among the vaccinated population (Willyard 2023).

1.2.3 SARS-CoV-2 variants

RNA viruses, including SARS-CoV-2, exhibit a considerable rate of genomic mutation, leading to the possible emergence of more adapted variants capable of evading detection and neutralisation by the immune system (Markov et al. 2023). Particularly, the spike protein of these viruses undergoes frequent mutations, allowing for adaptations in the binding affinity towards the human ACE2 receptor (Singh and Yi 2021). With time and multiple recombinations, new subvariants of Covid-19 with a strong selective advantage became dominant (Figure 1.3). There is a myriad of SARS-CoV-2 lineages considered variants of concern (VOC), including Alpha (lineage B.1.1.7), Beta (B.1.351), Gamma (P1), and Delta (B.1.617.2). However, none has been more concerning than the variant Omicron (B.1.1.529), particularly its BA.5 subvariant (Cao et al. 2022).

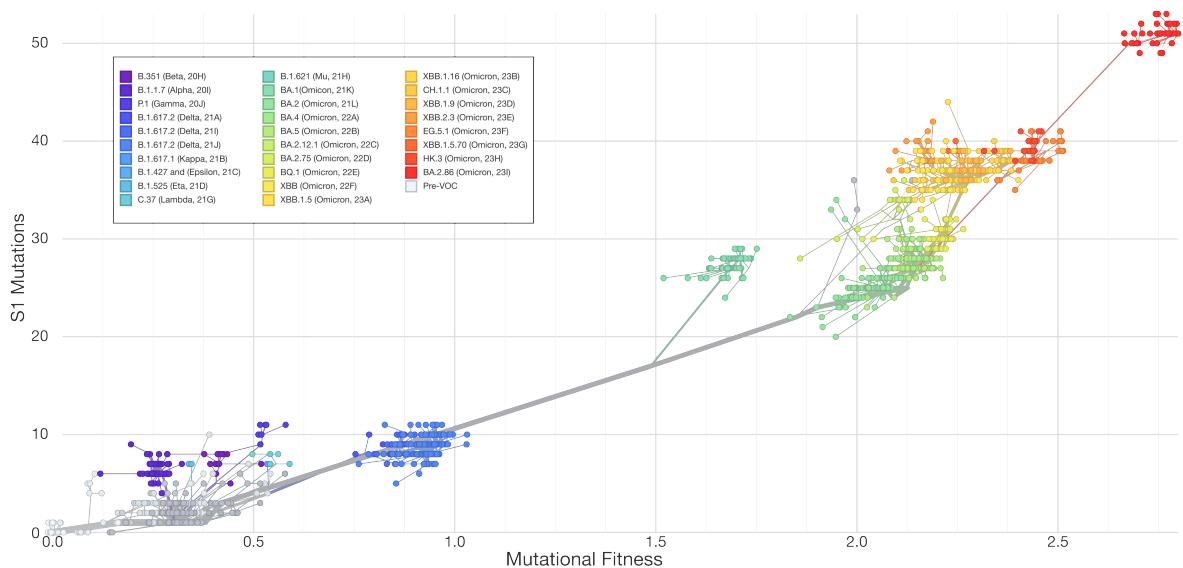


Figure 1.3: SARS-CoV-2 accumulation of amino acid mutations in the spike protein's S1 subunit as a function of their relative mutational fitness, shown for 2,322 European genomes sampled between December 2019 and December 2023. Generated by nextstrain.org (Hadfield et al. 2018).

Omicron

The Omicron variant was first detected at the end of 2021 in the Republic of Botswana. It quickly raised concerns due to its much higher capacity to cause reinfections compared to any other preceding variant (Pulliam et al. 2022). Before the start of 2022, confirmed cases of this variant were detected in various countries in Europe, North America, and Oceania, eventually becoming the dominant variant at a time when vaccines were already widely distributed.

Omicron and its subvariants have multiple mutations in the spike protein, granting them higher affinity to the ACE2 receptor (Starr et al. 2020). This increased affinity capacitates the subvariants to more effectively evade antibody binding and neutralisation from naturally- and vaccination-acquired immunity (Arora et al. 2022; Greaney et al. 2021).

Omicron BA.5 in Portugal

In Portugal, Omicron became dominant at the start of 2022 (Figure 1.4). Initially, most cases were related to the Omicron subvariant BA.1 (period of dominance: Jan 1 2022–Feb 6 2022), followed by BA.2 (period of dominance: Mar 27 2022–Apr 17 2022), with a slow transition period between the two (Malato et al. 2022b). Between January and February alone, Portugal documented 1.97 million infections, surpassing the total number of cases recorded during the entire epidemic until that period and highlighting the ongoing challenges posed by these variants.

By mid-2022, Portugal became one of the first countries to have Omicron BA.5 as a dominant variant (starting period of dominance: Jun 1 2022). Despite the majority of the population having received at least one vaccine dose (Figure 1.2) and with over 40% reporting a case of infection until that time (Malato et al. 2022b), the sudden rise in cases for this particular lineage suggested the ability to partially evade immunity from other variants, even BA.1 and BA.2. This was particularly concerning, as adapted vaccines under clinical trials were based on BA.1. The risks of infection and reinfection in vaccinated populations during BA.5 dominance will be studied in more detail in this thesis (Part III).

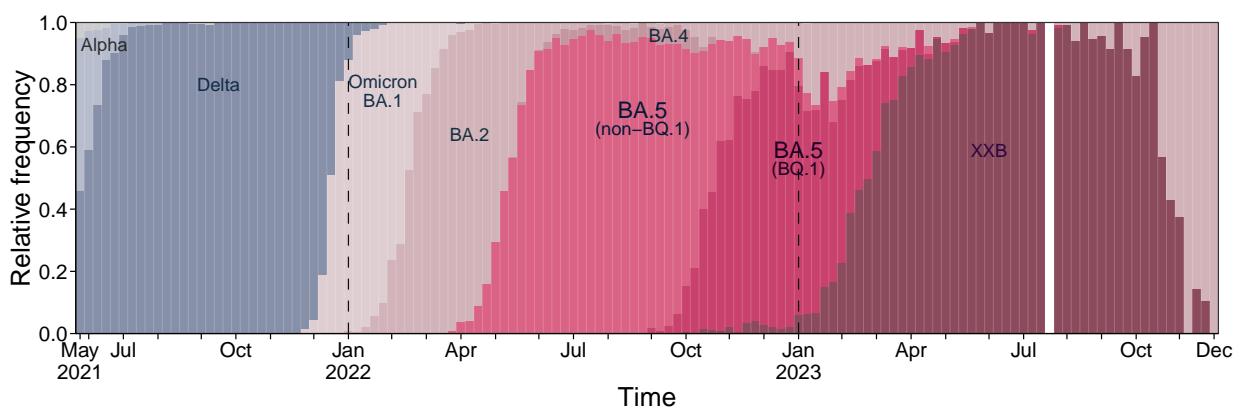


Figure 1.4: Weekly relative frequency of the most dominant SARS-CoV-2 variants in Portugal between May of 2021 (week 22) and December of 2023 (week 49). Gray-filled values indicate other variants. Dashed vertical lines indicate the start of 2022 and 2023. Columns left in white represent missing values on particular weeks. Source: Instituto Nacional de Saúde Doutor Ricardo Jorge (2023).

1.3 Aims of the thesis

The work presented in this thesis covers two topics: research strategies in ME/CFS (Part II) and risk of SARS-CoV-2 Omicron BA.5 subvariant in the context of highly vaccinated populations (Part III).

The research conducted in Part II aims to enhance reproducibility in studies related to ME/CFS and improve patient stratification to gain a better understanding of the illness. As such, the objectives are to investigate the diagnostic agreement among different case definitions and assess the statistical power to detect associations to candidate risk factors for the disease under the assumption of misdiagnosis. These approaches aim to help identify potential weaknesses in biomarker research related to patient misdiagnosis and improve the consistency of study results. Additionally, through stratification based on infection triggers and severity of symptoms from multiple domains, this thesis aims to study the link between herpesviruses infections and antibody responses in ME/CFS, using healthy individuals and MS patients as controls.

Concerning Part III, the research objectives are to understand and quantify the protection of SARS-CoV-2 vaccines and past infections from other VOC on the risk of Omicron BA.5 and study how this protection changed over time.

1.4 Outline of the thesis

As mentioned in Section 1.3, the overarching themes discussed in Part II encompass reproducibility in ME/CFS research and the stratification of diagnosed patients.

1. Diagnosis of ME/CFS is done through symptom-base criteria with multiple different case definitions that can diagnose different individuals as suspected cases. Additionally, an overlap of symptoms and severity has been described between ME/CFS-diagnosed individuals and other fatigue-inducing diseases. In Chapter 2, I used data from the specialised UKMEB to study the lack of diagnostic agreement between four of the most commonly used ME/CFS case definitions and the similarity of symptoms among patients and different control groups. The study also presents the initial ideas on patient misdiagnosis

- (or misclassification).
2. In Chapter 3 and Chapter 4, I extended the previous ideas of patient heterogeneity and simulated scenarios of case-control studies assuming possible inclusion of misdiagnosed patients to estimate the reduction in statistical power to detect an association to a candidate causal factor. The former chapter is a hypothetical application, extrapolating on the link between ME/CFS and viral infections to use data from three Covid-19 seroepidemiological studies. The latter describes the idea of misdiagnosis with more detail, performing simulations under different parametric conditions. Moreover, it analyses available data on candidate genetic and serological markers for ME/CFS and studies the reproducibility of the published results under the proposed assumption of patient misdiagnosis.
 3. There is a growing body of evidence linking EBV infection as a trigger to the pathogenesis of a subgroup of ME/CFS individuals, with possible antigen mimicry as the root of an autoimmune response. In Chapter 5, I re-analysed antibody-wide association analysis data from a previous study on IgG antibody responses against antigens derived from 14 EBV proteins (Loebel et al. 2017). Different regression models were built, relating antibody levels with both age and gender, with ME/CFS patients stratified into infection trigger-specific subgroups and compared with healthy controls.
 4. ME/CFS and MS share symptoms and have their onset linked with an acute viral infection. In Chapter 6, I analysed data from the UKMEB to understand the differences between the two diseases and their relationship with viral agents, assessing the symptomatic profiles and concentrations of IgG antibody responses to six different herpesviruses (CMV, EBV, HHV6, HSV1, HSV2, and VZV) from a population of MS controls and ME/CFS patients. Following the previous studies in this thesis, ME/CFS patients were stratified based on reported infection triggers. Furthermore, the study also tries to discriminate between the ME/CFS subgroups and MS, using different regression models.
 5. During ME/CFS diagnosis, various symptoms are assessed, and their combination can help clinicians infer which mechanistic domains might be exacerbated. In Chapter 7, I stratified patients into subgroups based on their symptomatological profiles of severity

related to seven different domains (Table 1.1) and studied the concentration of IgG antibody responses against the same herpesviruses as Chapter 6.

6. The SARS-CoV-2 entry into the human cells is mediated by the ACE2 receptors and complications associated with the disease have been linked with the down-regulation of this enzyme. Thus, individuals with a baseline ACE2 deficiency are potentially at an increased risk of developing Covid-19. To assess whether ME/CFS patients have increased susceptibility to Covid-19, in Chapter 8 I performed a meta-analysis of public CpG DNA methylation and gene expression data for ACE2 and its homologous ACE protein.

Part III was done throughout the second half of 2022, in the context of the Covid-19 pandemic and the rise SARS-CoV-2 Omicron BA.5 subvariant in highly vaccinated populations.

7. By mid-2022 it became important to understand the impact of previous infections on the risk of infection and reinfection posed by the Omicron BA.5 subvariant. In Chapter 9 I made use of the Portuguese Covid-19 registry and the national SARS-CoV-2 genetic surveillance data to identify the periods of dominance for previous VOC and Omicron subvariants (BA.1 and BA.2) and assess how infections during those periods (representing adapted vaccines to these dominant variants) was effective against BA.5.
8. Following the results of the previous chapter, in Chapter 10, I used an updated version of the same data to study how the hybrid immunity (vaccine + single BA.1/BA.2 infection) effectiveness of protection against a BA.5 infection decayed over time.

Most chapters presented in this thesis correspond to already published scientific articles. However, rather than displaying them by their chronological order of publication, the papers were included in a sequence based on their content.

In Part II, both studies from Chapter 2 (Malato et al. 2021) and Chapter 3 (Malato et al. 2022a) were published in the Proceedings of the Portuguese Statistical Society, whose publication is dependent on peer-review. The first study provides an introduction to the overall challenges related to agreement in ME/CFS diagnosis and patient heterogeneity, briefly mentioning the possible impact of misdiagnosis (described as “misclassification”). In fact, the ideas

on the impact of misdiagnosis began to take shape during the preparation of this work. The second study is an application to confirm and extend the previous ideas, using the notion of misdiagnosis in both patient diagnosis and serological tests. Ultimately, the concepts proposed in both articles were further developed and formalised, eventually being published in the article presented in Chapter 4 (Malato et al. 2023a).

The two chapters included in Part III (Chapter 9 and Chapter 10) were developed in parallel with the remaining project of the thesis. The resulting papers are correspondence articles that give an insight into the period of uncertainty regarding whether or not vaccines adapted to past lineages—particularly the adapted vaccine towards Omicron BA.1—would provide effective and enduring protection towards the new dominant Omicron BA.5 subvariant (Malato et al. 2022b; 2023b).

The original published articles are included *facsimile* in the Appendices (Part V).

To conclude, Chapter 11 (Part IV) provides a general discussion of the most important aspects of this thesis, with highlights on the main achievements and final remarks and potential future work.

Bibliography

- A New Clinical Entity? *The Lancet*, 267(6926):789–790, May 1956. ISSN 01406736. doi: 10.1016/S0140-6736(56)91252-1. URL <https://linkinghub.elsevier.com/retrieve/pii/S0140673656912521>.
- B. Abbi and B. H. Natelson. Is chronic fatigue syndrome the same illness as fibromyalgia: evaluating the 'single syndrome' hypothesis. *QJM : monthly journal of the Association of Physicians*, 106(1):3–9, Aug. 2012. doi: 10.1093/qjmed/hcs156.
- F. Albright, K. Light, A. Light, L. Bateman, and L. A. Cannon-Albright. Evidence for a heritable predisposition to Chronic Fatigue Syndrome. *BMC neurology*, 11:62, May 2011. ISSN 1471-2377. doi: 10.1186/1471-2377-11-62.
- D. Ao, X. He, J. Liu, and L. Xu. Strategies for the development and approval of COVID-19 vaccines and therapeutics in the post-pandemic period. *Signal Transduction and Targeted Therapy*, 8(1):1–17, Dec. 2023. ISSN 2059-3635. doi: 10.1038/s41392-023-01724-w. URL <https://www.nature.com/articles/s41392-023-01724-w>. Number: 1 Publisher: Nature Publishing Group.

- T. Aoki, H. Miyakoshi, Y. Usuda, and R. B. Herberman. Low NK syndrome and its relationship to chronic fatigue syndrome. *Clinical Immunology and Immunopathology*, 69(3):253–265, Dec. 1993. ISSN 0090-1229. doi: 10.1006/clin.1993.1178.
- M. E. Ariza. Commentary: Antibodies to Human Herpesviruses in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Patients. *Frontiers in Immunology*, 11, 2020. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2020.01400>.
- M. E. Ariza. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: The Human Herpesviruses Are Back! *Biomolecules*, 11(2):185, Feb. 2021. ISSN 2218-273X. doi: 10.3390/biom11020185. URL <https://www.mdpi.com/2218-273X/11/2/185>. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.
- P. Arora, A. Kempf, I. Nehlmeier, S. R. Schulz, A. Cossmann, M. V. Stankov, H.-M. Jäck, G. M. N. Behrens, S. Pöhlmann, and M. Hoffmann. Augmented neutralisation resistance of emerging omicron subvariants BA.2.12.1, BA.4, and BA.5. *The Lancet Infectious diseases*, 22(8):1117–1118, Aug. 2022. ISSN 1474-4457. doi: 10.1016/s1473-3099(22)00422-4. URL <https://europepmc.org/articles/PMC9239574>.
- T. T. Asprusten, L. Sletner, and V. B. B. Wyller. Are there subgroups of chronic fatigue syndrome? An exploratory cluster analysis of biological markers. *Journal of Translational Medicine*, 19(1), Jan. 2021. doi: 10.1186/s12967-021-02713-9.
- J. G. Ayres, N. Flint, E. G. Smith, W. S. Tunnicliffe, T. J. Fletcher, K. Hammond, D. Ward, and B. P. Marmion. Post-infection fatigue syndrome following Q fever. *QJM: monthly journal of the Association of Physicians*, 91(2):105–123, Feb. 1998. ISSN 1460-2725. doi: 10.1093/qjmed/91.2.105.
- I. J. Bakken, K. Tveito, N. Gunnes, S. Ghaderi, C. Stoltenberg, L. Trogstad, S. E. Håberg, and P. Magnus. Two age peaks in the incidence of chronic fatigue syndrome/myalgic encephalomyelitis: a population-based registry study from Norway 2008-2012. *BMC medicine*, 12:167, Oct. 2014. ISSN 1741-7015. doi: 10.1186/s12916-014-0167-5.
- G. R. Bantug, L. Galluzzi, G. Kroemer, and C. Hess. The spectrum of T cell metabolism in health and disease. *Nature Reviews Immunology*, 18(1):19–34, Jan. 2018. ISSN 1474-1741. doi: 10.1038/nri.2017.99. URL <https://www.nature.com/articles/nri.2017.99>. Number: 1 Publisher: Nature Publishing Group.
- M. Barreto, R. C. Ferreira, L. Lourenço, M. F. Moraes-Fontes, E. Santos, M. Alves, C. Carvalho, B. Martins, R. Andreia, J. a. F. Viana, C. Vasconcelos, L. Mota-Vieira, C. Ferreira, J. Demengeot, and A. M. Vicente. Low frequency of CD4+CD25+ Treg in SLE patients: a heritable trait associated with CTLA4 and TGF gene variants. *BMC Immunology*, 10(1):5, Jan. 2009. ISSN 1471-2172. doi: 10.1186/1471-2172-10-5. URL <https://doi.org/10.1186/1471-2172-10-5>.
- K. Bayliss, M. Goodall, A. Chisholm, B. Fordham, C. Chew-Graham, L. Riste, L. Fisher, K. Lovell, S. Peters, and A. Wearden. Overcoming the barriers to the diagnosis and management of chronic fatigue syndrome/ME in primary care: a meta synthesis of qualitative studies. *BMC Family Practice*, 15(1):44, Mar. 2014. ISSN

- 1471-2296. doi: 10.1186/1471-2296-15-44. URL <https://doi.org/10.1186/1471-2296-15-44>.
- P. O. Behan, W. M. Behan, and E. J. Bell. The postviral fatigue syndrome—an analysis of the findings in 50 cases. *The Journal of Infection*, 10(3):211–222, May 1985. ISSN 0163-4453. doi: 10.1016/s0163-4453(85)92488-0.
- I. Bergeri, H. C. Lewis, L. Subissi, A. Nardone, M. Valenciano, B. Cheng, K. Glonti, B. Williams, I.-O. O. Abejirinde, A. Simnceanu, A. Cassini, R. Grant, A. Rodriguez, A. Vicari, L. Al Ariqi, T. Azim, P. R. Wi-jesinghe, S. C. Rajatonirina, J. C. Okeibunor, L.-V. Le, M. Katz, A. Vaughan, P. Jorgensen, G. Freidl, R. Pe-body, and M. D. Van Kerkhove. Early epidemiological investigations: World Health Organization UNITY protocols provide a standardized and timely international investigation framework during the COVID-19 pandemic. *Influenza and Other Respiratory Viruses*, 16(1):7–13, 2022. ISSN 1750-2659. doi: 10.1111/irv.12915. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/irv.12915>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/irv.12915>.
- R. Bertinat, R. Villalobos-Labra, L. Hofmann, J. Blauensteiner, N. Sepúlveda, and F. Westermeier. Decreased NO production in endothelial cells exposed to plasma from ME/CFS patients. *Vascular Pharmacology*, 143:106953, Apr. 2022. ISSN 1537-1891. doi: 10.1016/j.vph.2022.106953. URL <https://www.sciencedirect.com/science/article/pii/S1537189122000027>.
- A. C. Bested and L. M. Marshall. Review of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: an evidence-based approach to diagnosis and management by clinicians. *Reviews on Environmental Health*, 30(4):223–249, 2015. ISSN 2191-0308. doi: 10.1515/reveh-2015-0026.
- K. Bjornevik, M. Cortese, B. C. Healy, J. Kuhle, M. J. Mina, Y. Leng, S. J. Elledge, D. W. Niebuhr, A. I. Scher, K. L. Munger, and A. Ascherio. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science (New York, N.Y.)*, 375(6578):296–301, Jan. 2022. ISSN 1095-9203. doi: 10.1126/science.abj8222.
- E. Blasioli, B. Mansouri, S. S. Tamvada, and E. Hassini. Vaccine Allocation and Distribution: A Review with a Focus on Quantitative Methodologies and Application to Equity, Hesitancy, and COVID-19 Pandemic. *Operations Research Forum*, 4(2):27, 2023. ISSN 2662-2556. doi: 10.1007/s43069-023-00194-8. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10028329/>.
- J. Blauensteiner, R. Bertinat, L. E. León, M. Riederer, N. Sepúlveda, and F. Westermeier. Altered endothelial dysfunction-related miRs in plasma from ME/CFS patients. *Scientific Reports*, 11(1):10604, May 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-89834-9. URL <https://www.nature.com/articles/s41598-021-89834-9>. Number: 1 Publisher: Nature Publishing Group.
- J. Blomberg, C.-G. Gottfries, A. Elfaitouri, M. Rizwan, and A. Rosén. Infection Elicited Autoimmunity and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: An Explanatory Model. *Frontiers in Immunology*, 9, Feb. 2018. doi: 10.3389/fimmu.2018.00229.
- J. Blomberg, M. Rizwan, A. Böhlin-Wiener, A. Elfaitouri, P. Julin, O. Zachrisson, A. Rosén, and C.-G. Gottfries. An-

- tibodies to Human Herpesviruses in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Patients. *Frontiers in Immunology*, 10, Aug. 2019. doi: 10.3389/fimmu.2019.01946.
- P. A. Bond and T. G. Dinan. Antibodies to Herpes Simplex Types 1 and 2 in Chronic Fatigue Syndrome. *Journal Of Chronic Fatigue Syndrome*, 13(1):35–40, Jan. 2006. ISSN 1057-3321. doi: 10.1300/J092v13n01_04. URL https://doi.org/10.1300/J092v13n01_04. Publisher: Taylor & Francis _eprint: https://doi.org/10.1300/J092v13n01_04.
- A. S. Bradley, B. Ford, and A. S. Bansal. Altered functional B cell subset populations in patients with chronic fatigue syndrome compared to healthy controls. *Clinical and Experimental Immunology*, 172(1):73–80, Apr. 2013. ISSN 0009-9104. doi: 10.1111/cei.12043. URL <https://doi.org/10.1111/cei.12043>.
- E. Brenu, D. Staines, and S. Marshall-Gradisnik. Methylation Profile of CD4+ T Cells in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *Journal of Clinical & Cellular Immunology*, 5(3):228–14, 2014a. ISSN 2155-9899. doi: 10.4172/2155-9899.1000228. URL <https://research-repository.griffith.edu.au/handle/10072/68935>. Accepted: 2017-05-03T16:10:59Z Publisher: Omics Publishing Group.
- E. W. Brenu, M. L. v. Driel, D. R. Staines, K. J. Ashton, S. B. Ramos, J. Keane, N. G. Klimas, and S. M. Marshall-Gradisnik. Immunological abnormalities as potential biomarkers in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *Journal of Translational Medicine*, 9(1):81, 2011. doi: 10.1186/1479-5876-9-81.
- E. W. Brenu, M. L. van Driel, D. R. Staines, K. J. Ashton, S. L. Hardcastle, J. Keane, L. Tajouri, D. Peterson, S. B. Ramos, and S. M. Marshall-Gradisnik. Longitudinal investigation of natural killer cells and cytokines in chronic fatigue syndrome/myalgic encephalomyelitis. *Journal of Translational Medicine*, 10:88, May 2012. ISSN 1479-5876. doi: 10.1186/1479-5876-10-88.
- E. W. Brenu, T. K. Huth, S. L. Hardcastle, K. Fuller, M. Kaur, S. Johnston, S. B. Ramos, D. R. Staines, and S. M. Marshall-Gradisnik. Role of adaptive and innate immune cells in chronic fatigue syndrome/myalgic encephalomyelitis. *International Immunology*, 26(4):233–242, Apr. 2014b. ISSN 0953-8178. doi: 10.1093/intimm/dxt068. URL <https://doi.org/10.1093/intimm/dxt068>.
- E. W. Brenu, S. Broadley, T. Nguyen, S. Johnston, S. Ramos, D. Staines, and S. Marshall-Gradisnik. A Preliminary Comparative Assessment of the Role of CD8+ T Cells in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis and Multiple Sclerosis. *Journal of Immunology Research*, 2016:e9064529, Jan. 2016. ISSN 2314-8861. doi: 10.1155/2016/9064529. URL <https://www.hindawi.com/journals/jir/2016/9064529/>. Publisher: Hindawi.
- G. Broderick, J. Fuite, A. Kreitz, S. D. Vernon, N. Klimas, and M. A. Fletcher. A formal analysis of cytokine networks in chronic fatigue syndrome. *Brain, Behavior, and Immunity*, 24(7):1209–1217, Oct. 2010. ISSN 1090-2139. doi: 10.1016/j.bbi.2010.04.012.
- K. G. Brurberg, M. S. Fønhus, L. Larun, S. Flottorp, and K. Malterud. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *BMJ Open*, 4(2):e003973, Feb. 2014. doi:

- 10.1136/bmjopen-2013-003973.
- D. Buchwald, R. Herrell, S. Ashton, M. Belcourt, K. Schmaling, P. Sullivan, M. Neale, and J. Goldberg. A twin study of chronic fatigue. *Psychosomatic Medicine*, 63(6):936–943, 2001. ISSN 0033-3174. doi: 10.1097/00006842-200111000-00012.
- E. Byrne. Idiopathic chronic fatigue and myalgia syndrome (myalgic encephalomyelitis): some thoughts on nomenclature and aetiology. *The Medical Journal of Australia*, 148(2):80–82, Jan. 1988. ISSN 0025-729X. doi: 10.5694/j.1326-5377.1988.tb104513.x.
- R. Cairns and M. Hotopf. A systematic review describing the prognosis of chronic fatigue syndrome. *Occupational Medicine*, 55(1):20–31, Jan. 2005. ISSN 0962-7480. doi: 10.1093/occmed/kqi013. URL <https://doi.org/10.1093/occmed/kqi013>.
- L. H. Calabrese, M. E. Davis, and W. S. Wilke. Chronic fatigue syndrome and a disorder resembling Sjögren's syndrome: preliminary report. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 18 Suppl 1:S28–31, Jan. 1994. ISSN 1058-4838. doi: 10.1093/clinids/18.supplement_1.s28.
- M. Caligiuri, C. Murray, D. Buchwald, H. Levine, P. Cheney, D. Peterson, A. Komaroff, and J. Ritz. Phenotypic and functional deficiency of natural killer cells in patients with chronic fatigue syndrome. *Journal of Immunology*, 139(10):3306–3313, 1987. ISSN 0022-1767.
- Y. Cao, A. Yisimayi, F. Jian, W. Song, T. Xiao, L. Wang, S. Du, J. Wang, Q. Li, X. Chen, Y. Yu, P. Wang, Z. Zhang, P. Liu, R. An, X. Hao, Y. Wang, J. Wang, R. Feng, H. Sun, L. Zhao, W. Zhang, D. Zhao, J. Zheng, L. Yu, C. Li, N. Zhang, R. Wang, X. Niu, S. Yang, X. Song, Y. Chai, Y. Hu, Y. Shi, L. Zheng, Z. Li, Q. Gu, F. Shao, W. Huang, R. Jin, Z. Shen, Y. Wang, X. Wang, J. Xiao, and X. S. Xie. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature*, 608(7923):593–602, Aug. 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04980-y. URL <https://www.nature.com/articles/s41586-022-04980-y>. Number: 7923 Publisher: Nature Publishing Group.
- B. M. Carruthers, A. K. Jain, K. L. D. Meirleir, D. L. Peterson, N. G. Klimas, A. M. Lerner, A. C. Bested, P. Flor-Henry, P. Joshi, A. C. P. Powles, J. A. Sherkey, and M. I. v. d. Sande. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Chronic Fatigue Syndrome*, 11(1):7–115, Jan. 2003. doi: 10.1300/j092v11n01_02.
- B. M. Carruthers, M. I. v. d. Sande, K. L. D. Meirleir, N. G. Klimas, G. Broderick, T. Mitchell, D. Staines, A. C. P. Powles, N. Speight, R. Vallings, L. Bateman, B. Baumgarten-Austrheim, D. S. Bell, N. Carlo-Stella, J. Chia, A. Darragh, D. Jo, D. Lewis, A. R. Light, S. Marshall-Gradisbik, I. Mena, J. A. Mikovits, K. Miwa, M. Murovska, M. L. Pall, and S. Stevens. Myalgic encephalomyelitis: International Consensus Criteria. *Journal of Internal Medicine*, 270(4):327–338, Aug. 2011. doi: 10.1111/j.1365-2796.2011.02428.x.
- J. Castro-Marrero, M. D. Cordero, N. Sáez-Francas, C. Jimenez-Gutierrez, F. J. Aguilar-Montilla, L. Aliste, and J. Alegre-Martin. Could mitochondrial dysfunction be a differentiating marker between chronic fatigue syndrome and fibromyalgia? *Antioxidants & Redox Signaling*, 19(15):1855–1860, Nov. 2013. ISSN 1557-7716.

- doi: 10.1089/ars.2013.5346.
- J. Castro-Marrero, M. Faro, L. Aliste, N. Sáez-Francàs, N. Calvo, A. Martínez-Martínez, T. F. de Sevilla, and J. Alegre. Comorbidity in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis: A Nationwide Population-Based Cohort Study. *Psychosomatics*, 58(5):533–543, Sept. 2017. ISSN 0033-3182. doi: 10.1016/j.psym.2017.04.010. URL <https://www.sciencedirect.com/science/article/pii/S0033318217301184>.
- N. Chaudhary, D. Weissman, and K. A. Whitehead. mRNA vaccines for infectious diseases: principles, delivery and clinical translation. *Nature Reviews. Drug Discovery*, 20(11):817–838, 2021. ISSN 1474-1776. doi: 10.1038/s41573-021-00283-5. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8386155/>.
- J. Choutka, V. Jansari, M. Hornig, and A. Iwasaki. Unexplained post-acute infection syndromes. *Nature Medicine*, 28(5):911–923, May 2022. ISSN 1546-170X. doi: 10.1038/s41591-022-01810-6. URL <https://www.nature.com/articles/s41591-022-01810-6>. Number: 5 Publisher: Nature Publishing Group.
- I. Y.-H. Chu, P. Alam, H. J. Larson, and L. Lin. Social consequences of mass quarantine during epidemics: a systematic review with implications for the COVID-19 response. *Journal of Travel Medicine*, 27(7):taaa192, Nov. 2020. ISSN 1708-8305. doi: 10.1093/jtm/taaa192. URL <https://doi.org/10.1093/jtm/taaa192>.
- L. Chu, I. J. Valencia, D. W. Garvert, and J. G. Montoya. Onset Patterns and Course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Pediatrics*, 7, Feb. 2019. doi: 10.3389/fped.2019.00012.
- J. M. Cliff, E. C. King, J.-S. Lee, N. Sepúlveda, A.-S. Wolf, C. Kingdon, E. Bowman, H. M. Dockrell, L. Nacul, E. Lacerda, and E. M. Riley. Cellular Immune Function in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Frontiers in Immunology*, 10, Apr. 2019. doi: 10.3389/fimmu.2019.00796. URL <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2019.00796/full>.
- P. Cruz-Tapias, J. Castiblanco, and J.-M. Anaya. HLA Association with Autoimmune Diseases. In *Autoimmunity: From Bench to Bedside [Internet]*. El Rosario University Press, July 2013. URL <https://www.ncbi.nlm.nih.gov/books/NBK459459/>.
- M. Curriu, J. Carrillo, M. Massanella, J. Rigau, J. Alegre, J. Puig, A. M. Garcia-Quintana, J. Castro-Marrero, E. Negredo, B. Clotet, C. Cabrera, and J. Blanco. Screening NK-, B- and T-cell phenotype and function in patients suffering from Chronic Fatigue Syndrome. *Journal of Translational Medicine*, 11(1):68, 2013. doi: 10.1186/1479-5876-11-68.
- A. Devereux-Cooke, S. Leary, S. J. McGrath, E. Northwood, A. Redshaw, C. Shepherd, P. Stacey, C. Tripp, J. Wilson, M. Mar, D. Boobyer, S. Bromiley, S. Chowdhury, C. Dransfield, M. Almas, O. Almelid, D. Buchanan, D. Garcia, J. Ireland, S. M. Kerr, I. Lewis, E. McDowall, M. Migdal, P. Murray, D. Perry, C. P. Ponting, V. Vitart, and J. C. Wolfe. DecodeME: community recruitment for a large genetics study of myalgic encephalomyelitis / chronic fatigue syndrome. *BMC Neurology*, 22(1):269, July 2022. ISSN 1471-2377. doi: 10.1186/s12883-022-02763-6. URL <https://doi.org/10.1186/s12883-022-02763-6>.
- J. J. Dibble, S. J. McGrath, and C. P. Ponting. Genetic risk factors of ME/CFS: a critical review. *Human Molecular Genetics*, 31(18):3623–3638, Oct. 2022. ISSN 0964-098X. doi: 10.1093/hmg/adv310. URL <https://doi.org/10.1093/hmg/adv310>.

- Genetics*, 29(R1):R117–R124, Sept. 2020. ISSN 0964-6906. doi: 10.1093/hmg/ddaa169. URL <https://doi.org/10.1093/hmg/ddaa169>.
- M. E. Dimmock, A. A. Mirin, and L. A. Jason. Estimating the disease burden of ME/CFS in the United States and its relation to research funding. *Journal of Medicine and Therapeutics*, 1(1), 2016. doi: 10.15761/jmt.1000102.
- R. Dobson and G. Giovannoni. Multiple sclerosis - a review. *European Journal of Neurology*, 26(1):27–40, Jan. 2019. ISSN 1468-1331. doi: 10.1111/ene.13819.
- T. D. Domingues, A. D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, F. Westermeier, C. Scheibenbogen, J. M. Cliff, L. Nacul, E. M. Lacerda, H. Mourão, and N. Sepúlveda. Herpesviruses Serology Distinguishes Different Subgroups of Patients From the United Kingdom Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Biobank. *Frontiers in Medicine*, 8, July 2021. doi: 10.3389/fmed.2021.686736.
- T. D. Domingues, J. a. Malato, A. D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, P. Bicek, L. Graça, H. Mourão, C. Scheibenbogen, F. Westermeier, L. Nacul, J. M. Cliff, E. Lacerda, and N. Sepúlveda. Association analysis between symptomatology and herpesvirus IgG antibody concentrations in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and multiple sclerosis. *Heliyon*, 9(7):e18250, July 2023. ISSN 2405-8440. doi: 10.1016/j.heliyon.2023.e18250. URL <https://www.sciencedirect.com/science/article/pii/S2405844023054580>.
- E. Engdahl, R. Gustafsson, J. Huang, M. Biström, I. Lima Bomfim, P. Stridh, M. Khademi, N. Brenner, J. Butt, A. Michel, D. Jons, M. Hortlund, L. Alonso-Magdalena, A. K. Hedström, L. Flamand, M. Ihira, T. Yoshikawa, O. Andersen, J. Hillert, L. Alfredsson, T. Waterboer, P. Sundström, T. Olsson, I. Kockum, and A. Fogdell-Hahn. Increased Serological Response Against Human Herpesvirus 6A Is Associated With Risk for Multiple Sclerosis. *Frontiers in Immunology*, 10, 2019. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02715>.
- W. Eriksen. ME/CFS, case definition, and serological response to Epstein-Barr virus. A systematic literature review. *Fatigue: Biomedicine, Health & Behavior*, 6(4):220–234, Oct. 2018. ISSN 2164-1846. doi: 10.1080/21641846.2018.1503125. URL <https://doi.org/10.1080/21641846.2018.1503125>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/21641846.2018.1503125>.
- F. Estévez-López, K. Mudie, X. Wang-Steverding, I. J. Bakken, A. Ivanovs, J. Castro-Marrero, L. Nacul, J. Alegre, P. Zalewski, J. Słomko, E. B. Strand, D. Pheby, E. Shikova, L. Lorusso, E. Capelli, S. Sekulic, C. Scheibenbogen, N. Sepúlveda, M. Murovska, and E. Lacerda. Systematic Review of the Epidemiological Burden of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Across Europe: Current Evidence and EUROMENE Research Recommendations for Epidemiology. *Journal of Clinical Medicine*, 9(5):1557, May 2020. doi: 10.3390/jcm9051557.
- B. Evengård and N. Klimas. Chronic Fatigue Syndrome. *Drugs*, 62(17):2433–2446, Dec. 2002. ISSN 1179-1950. doi: 10.2165/00003495-200262170-00003. URL <https://doi.org/10.2165/00003495-200262170-00003>.

00003495-200262170-00003.

- M. A. Fletcher, X. R. Zeng, K. Maher, S. Levis, B. Hurwitz, M. Antoni, G. Broderick, and N. G. Klimas. Biomarkers in chronic fatigue syndrome: evaluation of natural killer cell function and dipeptidyl peptidase IV/CD26. *PloS One*, 5(5):e10817, May 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0010817.
- O. Fluge, O. Bruland, K. Risa, A. Storstein, E. K. Kristoffersen, D. Sapkota, H. Næss, O. Dahl, H. Nyland, and O. Mella. Benefit from B-lymphocyte depletion using the anti-CD20 antibody rituximab in chronic fatigue syndrome. A double-blind and placebo-controlled study. *PloS One*, 6(10):e26358, 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0026358.
- O. Fluge, I. G. Rekeland, K. Lien, H. Thürmer, P. C. Borchgrevink, C. Schäfer, K. Sørland, J. Aßmus, I. Ktoridou-Valen, I. Herder, M. E. Gotaas, O. Kvammen, K. A. Baranowska, L. M. L. J. Bohnen, S. S. Martinsen, A. E. Lonar, A.-E. H. Solvang, A. E. S. Gya, O. Bruland, K. Risa, K. Alme, O. Dahl, and O. Mella. B-Lymphocyte Depletion in Patients With Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. 170(9):585, Apr. 2019. doi: 10.7326/m18-1451.
- A. Fonseca, M. Szysz, H. T. Ly, C. Cordeiro, and N. Sepúlveda. IgG Antibody Responses to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Their Effective Potential for Disease Diagnosis and Pathological Antigenic Mimicry. *Medicina*, 60(1):161, Jan. 2024. ISSN 1010-660X. doi: 10.3390/medicina60010161. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10820613/>.
- R. Freeman and A. L. Komaroff. Does the chronic fatigue syndrome involve the autonomic nervous system? *The American Journal of Medicine*, 102(4):357–364, Apr. 1997. ISSN 0002-9343. doi: 10.1016/s0002-9343(97)00087-9.
- K. Fukuda, S. E. Straus, I. Hickie, M. C. Sharpe, J. G. Dobbins, and A. Komaroff. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Annals of Internal Medicine*, 121(12):953, Dec. 1994. doi: 10.7326/0003-4819-121-12-199412150-00009.
- T. A.-Z. K. Gaber, W. W. Oo, and H. Ringrose. Multiple Sclerosis/Chronic Fatigue Syndrome overlap: When two common disorders collide. *NeuroRehabilitation*, 35:529–534, 2014. ISSN 18786448, 10538135. doi: 10.3233/NRE-141146.
- A. G. Gabibov, A. A. Belogurov Jr., Y. A. Lomakin, M. Y. Zakharova, M. E. Avakyan, V. V. Dubrovskaya, I. V. Smirnov, A. S. Ivanov, A. A. Molnar, V. E. Gurtsevitch, S. V. Diduk, K. V. Smirnova, B. Avalle, S. N. Sharanova, A. Tramontano, A. Friboulet, A. N. Boyko, N. A. Ponomarenko, and N. V. Tikunova. Combinatorial antibody library from multiple sclerosis patients reveals antibodies that cross-react with myelin basic protein and EBV antigen. *The FASEB Journal*, 25(12):4211–4221, 2011. ISSN 1530-6860. doi: 10.1096/fj.11-190769. URL <https://onlinelibrary.wiley.com/doi/abs/10.1096/fj.11-190769>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1096/fj.11-190769>.
- J. F. Galpine and C. Brady. Benign myalgic encephalomyelitis. *Lancet (London, England)*, 272(6972):757–758,

- Apr. 1957. ISSN 0140-6736. doi: 10.1016/s0140-6736(57)91024-3.
- X.-Y. Ge, J.-L. Li, X.-L. Yang, A. A. Chmura, G. Zhu, J. H. Epstein, J. K. Mazet, B. Hu, W. Zhang, C. Peng, Y.-J. Zhang, C.-M. Luo, B. Tan, N. Wang, Y. Zhu, G. Crameri, S.-Y. Zhang, L.-F. Wang, P. Daszak, and Z.-L. Shi. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*, 503(7477):535–538, Nov. 2013. ISSN 1476-4687. doi: 10.1038/nature12711. URL <https://www.nature.com/articles/nature12711>. Number: 7477 Publisher: Nature Publishing Group.
- A. Ghali, C. Lacout, J.-O. Fortrat, K. Depres, M. Ghali, and C. Lavigne. Factors Influencing the Prognosis of Patients with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Diagnostics*, 12(10):2540, Oct. 2022. ISSN 2075-4418. doi: 10.3390/diagnostics12102540. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9600584/>.
- A. Gil, G. E. Hoag, J. P. Salerno, M. Hornig, N. Klimas, and L. K. Selin. Identification of CD8 T-cell dysfunction associated with symptoms in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and Long COVID and treatment with a nebulized antioxidant/anti-pathogen agent in a retrospective case series. *Brain, Behavior, & Immunity - Health*, page 100720, Dec. 2023. ISSN 2666-3546. doi: 10.1016/j.bbih.2023.100720. URL <https://www.sciencedirect.com/science/article/pii/S2666354623001345>.
- J. Goldstein. *Betrayal by the Brain: The Neurologic Basis of Chronic Fatigue Syndrome, Fibromyalgia Syndrome, and Related Neural Network*. Routledge, New York, Aug. 2013. ISBN 978-1-315-06136-8. doi: 10.4324/9781315061368.
- A. J. Greaney, T. N. Starr, P. Gilchuk, S. J. Zost, E. Binshtain, A. N. Loes, S. K. Hilton, J. Huddleston, R. Eguia, K. H. D. Crawford, A. S. Dingens, R. S. Nargi, R. E. Sutton, N. Suryadevara, P. W. Rothlauf, Z. Liu, S. P. J. Whelan, R. H. Carnahan, J. E. Crowe, and J. D. Bloom. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host & Microbe*, 29(1):44–57.e9, Jan. 2021. ISSN 1931-3128. doi: 10.1016/j.chom.2020.11.007. URL <https://www.sciencedirect.com/science/article/pii/S1931312820306247>.
- J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, and R. A. Neher. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, Dec. 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty407. URL <https://doi.org/10.1093/bioinformatics/bty407>.
- M. Haffke, H. Freitag, G. Rudolf, M. Seifert, W. Doehner, N. Scherbakov, L. Hanitsch, K. Wittke, S. Bauer, F. Konietzschke, F. Paul, J. Bellmann-Strobl, C. Kedor, C. Scheibenbogen, and F. Sotzny. Endothelial dysfunction and altered endothelial biomarkers in patients with post-COVID-19 syndrome and chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 20(1):138, Mar. 2022. ISSN 1479-5876. doi: 10.1186/s12967-022-03346-2. URL <https://doi.org/10.1186/s12967-022-03346-2>.
- R. Hajdarevic, A. Lande, I. Rekeland, A. Rydland, E. B. Strand, D. D. Sosa, L. E. Creary, O. Mella, T. Egeland, O. D.

- Saugstad, O. Fluge, B. A. Lie, and M. K. Viken. Fine mapping of the major histocompatibility complex (MHC) in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) suggests involvement of both HLA class I and class II loci. *Brain, Behavior, and Immunity*, 98:101–109, Nov. 2021. ISSN 0889-1591. doi: 10.1016/j.bbi.2021.08.219. URL <https://www.sciencedirect.com/science/article/pii/S0889159121005092>.
- M. Halaji, M. Heiat, N. Faraji, and R. Ranjbar. Epidemiology of COVID-19: An updated review. *Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences*, 26:82, Sept. 2021. ISSN 1735-1995. doi: 10.4103/jrms.JRMS_506_20. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8548902/>.
- P. Halpin, M. V. Williams, N. G. Klimas, M. A. Fletcher, Z. Barnes, and M. E. Ariza. Myalgic encephalomyelitis/chronic fatigue syndrome and gulf war illness patients exhibit increased humoral responses to the herpesviruses-encoded dUTPase: Implications in disease pathophysiology. *Journal of Medical Virology*, 89(9):1636–1645, 2017. ISSN 1096-9071. doi: 10.1002/jmv.24810. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.24810>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.24810>.
- A. E. Handel, A. J. Williamson, G. Disanto, L. Handunnetthi, G. Giovannoni, and S. V. Ramagopalan. An updated meta-analysis of risk of multiple sclerosis following infectious mononucleosis. *PLoS One*, 5(9):e12496, Sept. 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0012496.
- S. Hardcastle, E. Brenu, S. Johnston, T. Nguyen, T. Huth, and M. Kaur. Analysis of the relationship between immune dysfunction and symptom severity in patients with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (CFS/ME). *J Clin Cell Immunol*, 5:2, Jan. 2014.
- D. Harley, D. Bossingham, D. M. Purdie, N. Pandeya, and A. C. Sleigh. Ross River virus disease in tropical Queensland: evolution of rheumatic manifestations in an inception cohort followed for six months. *The Medical Journal of Australia*, 177(7):352–355, Oct. 2002. ISSN 0025-729X. doi: 10.5694/j.1326-5377.2002.tb04836.x.
- W. Häuser, P. Klose, J. Langhorst, B. Moradi, M. Steinbach, M. Schiltenwolf, and A. Busch. Efficacy of different types of aerobic exercise in fibromyalgia syndrome: a systematic review and meta-analysis of randomised controlled trials. *Arthritis Research & Therapy*, 12(3):R79, May 2010. ISSN 1478-6354. doi: 10.1186/ar3002. URL <https://doi.org/10.1186/ar3002>.
- S. Havervall, A. Rosell, M. Phillipson, S. M. Mangsbo, P. Nilsson, S. Hofer, and C. Thålin. Symptoms and Functional Impairment Assessed 8 Months After Mild COVID-19 Among Health Care Workers. *JAMA*, 325(19):2015–2016, May 2021. ISSN 0098-7484. doi: 10.1001/jama.2021.5612. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8027932/>.
- A. M. Helliwell, E. C. Sweetman, P. A. Stockwell, C. D. Edgar, A. Chatterjee, and W. P. Tate. Changes in DNA methylation profiles of myalgic encephalomyelitis/chronic fatigue syndrome patients reflect systemic dysfunctions. *Clinical Epigenetics*, 12(1):167, Nov. 2020. ISSN 1868-7083. doi: 10.1186/s13148-020-00960-z. URL <https://doi.org/10.1186/s13148-020-00960-z>.

- D. A. Henderson and A. Shelokov. Epidemic neuromyasthenia; clinical syndrome. *The New England Journal of Medicine*, 260(15):757–764 contd, Apr. 1959. ISSN 0028-4793. doi: 10.1056/NEJM195904092601506.
- S. Herrera, W. C. de Vega, D. Ashbrook, S. D. Vernon, and P. O. McGowan. Genome-epigenome interactions associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Epigenetics*, 13(12):1174–1190, Dec. 2018. ISSN 1559-2294. doi: 10.1080/15592294.2018.1549769. URL <https://doi.org/10.1080/15592294.2018.1549769>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/15592294.2018.1549769>.
- I. Hickie, T. Davenport, D. Wakefield, U. Vollmer-Conna, B. Cameron, S. D. Vernon, W. C. Reeves, and A. Lloyd. Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study. *BMJ (Clinical research ed.)*, 333(7568):575, Sept. 2006. doi: 10.1136/bmj.38933.585764.ae.
- M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten, and S. Pöhlmann. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2):271–280.e8, Apr. 2020. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2020.02.052. URL [https://www.cell.com/cell/abstract/S0092-8674\(20\)30229-4](https://www.cell.com/cell/abstract/S0092-8674(20)30229-4). Publisher: Elsevier.
- G. P. Holmes, J. E. Kaplan, N. M. Gantz, A. L. Komaroff, L. B. Schonberger, S. E. Straus, J. F. Jones, R. E. Dubois, C. Cunningham-Rundles, S. Pahwa, G. Tosato, L. S. Zegans, D. T. Purtilo, N. Brown, R. T. Schooley, and I. Brus. Chronic Fatigue Syndrome: A Working Case Definition. *Annals of Internal Medicine*, 108(3):387–389, Mar. 1988. ISSN 0003-4819. doi: 10.7326/0003-4819-108-3-387. URL <https://www.acpjournals.org/doi/10.7326/0003-4819-108-3-387>. Publisher: American College of Physicians.
- M. Hornig, J. G. Montoya, N. G. Klimas, S. Levine, D. Felsenstein, L. Bateman, D. L. Peterson, C. G. Gottschalk, A. F. Schultz, X. Che, M. L. Eddy, A. L. Komaroff, and W. I. Lipkin. Distinct plasma immune signatures in ME/CFS are present early in the course of illness. *Science Advances*, 1(1):e1400121, Feb. 2015. ISSN 2375-2548. doi: 10.1126/sciadv.1400121.
- C. A. Horwitz, W. Henle, G. Henle, H. Rudnick, and E. Latts. Long-term serological follow-up of patients for Epstein-Barr virus after recovery from infectious mononucleosis. *The Journal of Infectious Diseases*, 151(6): 1150–1153, June 1985. ISSN 0022-1899. doi: 10.1093/infdis/151.6.1150.
- B. M. Hyde. *The nightingale, myalgic encephalomyelitis (ME) definition*. The Nightingale Research Foundation, Ottawa, Canada, Jan. 2007. URL <https://www.investinme.org/Documents/PDFdocuments/Byron%20Hyde%20Definition%20Booklet.pdf>.
- Institute of Medicine. *Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness*. The National Academies Press, Washington, DC, 2015. ISBN 978-0-309-31689-7. doi: 10.17226/19012. URL <https://www.nap.edu/catalog/19012/beyond-myalgic-encephalomyelitischronic-fatigue-syndrome-redefining-an-illness>.
- Instituto Nacional de Saúde Doutor Ricardo Jorge. Genetic diversity of the novel coronavirus SARS-CoV-2

- (COVID-19) in Portugal, 2023. URL <https://insaflu.insa.pt/covid19>.
- A. Irwin. What it will take to vaccinate the world against COVID-19. *Nature*, 592(7853):176–178, Mar. 2021. doi: 10.1038/d41586-021-00727-3. URL <https://www.nature.com/articles/d41586-021-00727-3>. Bandiera_abtest: a Cg_type: News Explainer Number: 7853 Publisher: Nature Publishing Group Subject_term: SARS-CoV-2, Policy, Government, Vaccines.
- R. Isaacs. Chronic infectious mononucleosis. *Blood*, 3(8):858–861, Aug. 1948. ISSN 0006-4971.
- Y. Itoh, T. Shigemori, T. Igarashi, and Y. Fukunaga. Fibromyalgia and chronic fatigue syndrome in children. *Pediatrics International: Official Journal of the Japan Pediatric Society*, 54(2):266–271, Apr. 2012. ISSN 1442-200X. doi: 10.1111/j.1442-200X.2011.03514.x.
- M. N. Janal, D. S. Ciccone, and B. H. Natelson. Sub-typing CFS patients on the basis of ‘minor’ symptoms. *Biological Psychology*, 73(2):124–131, Aug. 2006. ISSN 0301-0511. doi: 10.1016/j.biopsych.2006.01.003. URL <https://www.sciencedirect.com/science/article/pii/S0301051106000214>.
- L. Jason and A. Mirin. Updating the National Academy of Medicine ME/CFS prevalence and economic impact figures to account for population growth and inflation. *Fatigue: Biomedicine, Health & Behavior*, 9(1):9–13, Jan. 2021. ISSN 2164-1846. doi: 10.1080/21641846.2021.1878716. URL <https://doi.org/10.1080/21641846.2021.1878716>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/21641846.2021.1878716>.
- L. A. Jason. Differentiating Multiple Sclerosis from Myalgic Encephalomyelitis and Chronic Fatigue Syndrome. *Insights in Biomedicine*, 02(02), 2017. doi: 10.21767/2572-5610.10027.
- L. A. Jason, K. Corradi, S. Torres-Harding, R. R. Taylor, and C. King. Chronic Fatigue Syndrome: The Need for Subtypes. *Neuropsychology Review*, 15(1):29–58, Mar. 2005. doi: 10.1007/s11065-005-3588-2.
- L. A. Jason, M. Sunquist, A. Brown, M. Evans, S. D. Vernon, J. Furst, and V. Simonis. Examining case definition criteria for chronic fatigue syndrome and myalgic encephalomyelitis. *Fatigue : biomedicine, health & behavior*, 2(1):40–56, Jan. 2014. ISSN 2164-1846. doi: 10.1080/21641846.2013.862993. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3912876/>.
- L. A. Jason, S. Ravichandran, B. Z. Katz, B. H. Natelson, and H. F. Bonilla. Establishing a consensus on ME/CFS exclusionary illnesses. *Fatigue: Biomedicine, Health & Behavior*, 11(1):1–13, Jan. 2023. ISSN 2164-1846, 2164-1862. doi: 10.1080/21641846.2022.2150487. URL <https://www.tandfonline.com/doi/full/10.1080/21641846.2022.2150487>.
- L. Jin, Z. Li, X. Zhang, J. Li, and F. Zhu. CoronaVac: A review of efficacy, safety, and immunogenicity of the inactivated vaccine against SARS-CoV-2. *Human Vaccines & Immunotherapeutics*, 18(6):2096970, Nov. 2022. ISSN 2164-554X. doi: 10.1080/21645515.2022.2096970.
- H. K. Kang, B. H. Natelson, C. M. Mahan, K. Y. Lee, and F. M. Murphy. Post-traumatic stress disorder and chronic fatigue syndrome-like illness among Gulf War veterans: a population-based survey of 30,000 veterans.

- American Journal of Epidemiology*, 157(2):141–148, Jan. 2003. ISSN 0002-9262. doi: 10.1093/aje/kwf187.
- F. Kasimir, D. Toomey, Z. Liu, A. C. Kaiping, M. E. Ariza, and B. K. Prusty. Tissue specific signature of HHV-6 infection in ME/CFS. *Frontiers in Molecular Biosciences*, 9, 2022. ISSN 2296-889X. URL <https://www.frontiersin.org/articles/10.3389/fmolb.2022.1044964>.
- P. Katz. Fatigue in Rheumatoid Arthritis. *Current Rheumatology Reports*, 19(5):25, May 2017. ISSN 1534-6307. doi: 10.1007/s11926-017-0649-5.
- J. R. Kerr, B. Burke, R. Petty, J. Gough, D. Fear, D. L. Matthey, J. S. Axford, A. G. Dagleish, and D. J. Nutt. Seven genomic subtypes of chronic fatigue syndrome/myalgic encephalomyelitis: a detailed analysis of gene networks and clinical phenotypes. *Journal of Clinical Pathology*, 61(6):730–739, Nov. 2007. doi: 10.1136/jcp.2007.053553.
- J. R. Kerr, R. Petty, B. Burke, J. Gough, D. Fear, L. I. Sinclair, D. L. Matthey, S. C. M. Richards, J. Montgomery, D. A. Baldwin, P. Kellam, T. J. Harrison, G. E. Griffin, J. Main, D. Enlander, D. J. Nutt, and S. T. Holgate. Gene Expression Subtypes in Patients with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *The Journal of Infectious Diseases*, 197(8):1171–1184, Apr. 2008. ISSN 0022-1899. doi: 10.1086/533453. URL <https://doi.org/10.1086/533453>.
- L. Kim, C. Kedor, F. Buttigereit, H. Heidecke, D. Schaumburg, and C. Scheibenbogen. Characterizing Sjögren-Associated Fatigue: A Distinct Phenotype from ME/CFS. *Journal of Clinical Medicine*, 12(15):4994, July 2023. ISSN 2077-0383. doi: 10.3390/jcm12154994. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10419548/>.
- N. G. Klimas, F. R. Salvato, R. Morgan, and M. A. Fletcher. Immunologic abnormalities in chronic fatigue syndrome. *Journal of Clinical Microbiology*, 28(6):1403–1410, 1990. ISSN 0095-1137. URL <https://jcm.asm.org/content/28/6/1403>.
- C. Kodner. Common Questions About the Diagnosis and Management of Fibromyalgia. *American Family Physician*, 91(7):472–478, Apr. 2015. URL <https://www.aafp.org/pubs/afp/issues/2015/0401/p472.html>.
- A. L. Komaroff and W. I. Lipkin. ME/CFS and Long COVID share similar symptoms and biological abnormalities: road map to the literature. *Frontiers in Medicine*, 10:1187163, June 2023. ISSN 2296-858X. doi: 10.3389/fmed.2023.1187163. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10278546/>.
- K. Kostov. The Causal Relationship between Endothelin-1 and Hypertension: Focusing on Endothelial Dysfunction, Arterial Stiffness, Vascular Remodeling, and Blood Pressure Regulation. *Life*, 11(9):986, Sept. 2021. ISSN 2075-1729. doi: 10.3390/life11090986. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8472034/>.
- H. Kuratsune and Y. Watanabe. Chronic Fatigue Syndrome. In Y. Watanabe, B. Evengård, B. H. Natelson, L. A. Jason, and H. Kuratsune, editors, *Fatigue Science for Human Health*, pages 67–88. Springer Japan, Tokyo, 2008. ISBN 978-4-431-73464-2. doi: 10.1007/978-4-431-73464-2_6. URL https://doi.org/10.1007/978-4-431-73464-2_6.

- 978-4-431-73464-2_6.**
- E. M. Lacerda, E. W. Bowman, J. M. Cliff, C. C. Kingdon, E. C. King, J.-S. Lee, T. G. Clark, H. M. Dockrell, E. M. Riley, H. Curran, and L. Nacul. The UK ME/CFS Biobank for biomedical research on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) and Multiple Sclerosis. *Open Journal of Bioresources*, 4(1):4, Feb. 2017. ISSN 2056-5542. doi: 10.5334/ojb.28. URL <https://openbioreources.metajnl.com/articles/10.5334/ojb.28>. Number: 1 Publisher: Ubiquity Press.
- E. M. Lacerda, K. Mudie, C. C. Kingdon, J. D. Butterworth, S. O'Boyle, and L. Nacul. The UK ME/CFS Biobank: A Disease-Specific Biobank for Advancing Clinical Research Into Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Neurology*, 9:1026, 2018. ISSN 1664-2295. doi: 10.3389/fneur.2018.01026.
- E. M. Lacerda, C. McDermott, C. C. Kingdon, J. Butterworth, J. M. Cliff, and L. Nacul. Hope, disappointment and perseverance: Reflections of people with Myalgic encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) and Multiple Sclerosis participating in biomedical research. A qualitative focus group study. *Health Expectations*, 22(3):373–384, 2019. ISSN 1369-7625. doi: 10.1111/hex.12857. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/hex.12857>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/hex.12857>.
- M. M. Lamers and B. L. Haagmans. SARS-CoV-2 pathogenesis. *Nature Reviews Microbiology*, 20(5):270–284, May 2022. ISSN 1740-1534. doi: 10.1038/s41579-022-00713-0. URL <https://www.nature.com/articles/s41579-022-00713-0>. Number: 5 Publisher: Nature Publishing Group.
- K. Lan and M.-H. Luo. Herpesviruses: epidemiology, pathogenesis, and interventions. *Virologica Sinica*, 32(5): 347–348, Oct. 2017. ISSN 1674-0769. doi: 10.1007/s12250-017-4108-2. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6598912/>.
- A. Landay, E. Lennette, C. Jessop, and J. Levy. Chronic fatigue syndrome: clinical condition associated with immune activation. *The Lancet*, 338(8769):707–712, Sept. 1991. ISSN 01406736. doi: 10.1016/0140-6736(91)91440-6. URL <https://linkinghub.elsevier.com/retrieve/pii/0140673691914406>.
- A. Lande, O. Fluge, E. B. Strand, S. T. Flåm, D. D. Sosa, O. Mella, T. Egeland, O. D. Saugstad, B. A. Lie, and M. K. Viken. Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Scientific Reports*, 10(1):5267, Mar. 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-62157-x. URL <https://www.nature.com/articles/s41598-020-62157-x>. Number: 1 Publisher: Nature Publishing Group.
- J.-S. Lee, E. M. Lacerda, L. Nacul, C. C. Kingdon, J. Norris, S. O'Boyle, C. h. Roberts, L. Palla, E. M. Riley, and J. M. Cliff. Salivary DNA Loads for Human Herpesviruses 6 and 7 Are Correlated With Disease Phenotype in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Medicine*, 8:656692, Aug. 2021. ISSN 2296-858X. doi: 10.3389/fmed.2021.656692. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8378328/>.

- A. M. Lerner, S. H. Beqaj, R. G. Deeter, and J. T. Fitzgerald. IgM serum antibodies to human cytomegalovirus nonstructural gene products p52 and CM2(UL44 and UL57) are uniquely present in a subset of patients with chronic fatigue syndrome. *In Vivo (Athens, Greece)*, 16(3):153–159, 2002. ISSN 0258-851X.
- I. Lewis, J. Pairman, G. Spickett, and J. L. Newton. Is Chronic Fatigue Syndrome in Older Patients a different disease? – a clinical cohort study. *European Journal of Clinical Investigation*, 43(3):302–308, 2013. ISSN 1365-2362. doi: 10.1111/eci.12046. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/eci.12046>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/eci.12046>.
- E.-J. Lim and C.-G. Son. Review of case definitions for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 18(1):289, Dec. 2020. ISSN 1479-5876. doi: 10.1186/s12967-020-02455-0. URL <https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-020-02455-0>.
- E.-J. Lim, Y.-C. Ahn, E.-S. Jang, S.-W. Lee, S.-H. Lee, and C.-G. Son. Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *Journal of Translational Medicine*, 18(1), Feb. 2020. doi: 10.1186/s12967-020-02269-0.
- W. W. Lim, L. Mak, G. M. Leung, B. J. Cowling, and M. Peiris. Comparative immunogenicity of mRNA and inactivated vaccines against COVID-19. *The Lancet. Microbe*, 2(9):e423, Sept. 2021. ISSN 2666-5247. doi: 10.1016/S2666-5247(21)00177-4.
- Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 27(2):taaa021, Mar. 2020. ISSN 1708-8305. doi: 10.1093/jtm/taaa021. URL <https://doi.org/10.1093/jtm/taaa021>.
- M. Loebel, P. Grabowski, H. Heidecke, S. Bauer, L. G. Hanitsch, K. Wittke, C. Meisel, P. Reinke, H.-D. Volk, O. Fluge, O. Mella, and C. Scheibenbogen. Antibodies to adrenergic and muscarinic cholinergic receptors in patients with Chronic Fatigue Syndrome. *Brain, Behavior, and Immunity*, 52:32–39, Feb. 2016. ISSN 0889-1591. doi: 10.1016/j.bbi.2015.09.013. URL <https://www.sciencedirect.com/science/article/pii/S0889159115300209>.
- M. Loebel, M. Eckey, F. Sotzny, E. Hahn, S. Bauer, P. Grabowski, J. Zerweck, P. Holenya, L. G. Hanitsch, K. Wittke, P. Borchmann, J.-U. Rüffer, F. Hiepe, K. Ruprecht, U. Behrends, C. Meindl, H.-D. Volk, U. Reimer, and C. Scheibenbogen. Serological profiling of the EBV immune response in Chronic Fatigue Syndrome using a peptide microarray. *PLOS ONE*, 12(6):e0179124, Dec. 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0179124. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0179124>. Publisher: Public Library of Science.
- L. Lorusso, S. V. Mikhaylova, E. Capelli, D. Ferrari, G. K. Ngonga, and G. Ricevuti. Immunological aspects of chronic fatigue syndrome. *Autoimmunity Reviews*, 8(4):287–291, Feb. 2009. doi: 10.1016/j.autrev.2008.08.003.
- J. D. Lünemann, I. Jelcić, S. Roberts, A. Lutterotti, B. Tackenberg, R. Martin, and C. Münz. EBNA1-specific

T cells from patients with multiple sclerosis cross react with myelin antigens and co-produce IFN-gamma and IL-2. *The Journal of Experimental Medicine*, 205(8):1763–1773, Aug. 2008. ISSN 1540-9538. doi: 10.1084/jem.20072397.

M. Maes. A new case definition of Neuro-Inflammatory and Oxidative Fatigue (NIOF), a neuroprogressive disorder, formerly known as chronic fatigue syndrome or Myalgic Encephalomyelitis: results of multivariate pattern recognition methods and external validation by neuro-immune biomarkers. *Neuro Endocrinology Letters*, 36 (4):320–329, 2015. ISSN 0172-780X.

P Magnus, N. Gunnes, K. Tveito, I. J. Bakken, S. Ghaderi, C. Stoltenberg, M. Hornig, W. I. Lipkin, L. Trogstad, and S. E. Håberg. Chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) is associated with pandemic influenza infection, but not with an adjuvanted pandemic influenza vaccine. *Vaccine*, 33(46):6173–6177, Nov. 2015. ISSN 1873-2518. doi: 10.1016/j.vaccine.2015.10.018.

V. Majithia and S. A. Geraci. Rheumatoid Arthritis: Diagnosis and Management. *The American Journal of Medicine*, 120(11):936–939, Nov. 2007. ISSN 0002-9343, 1555-7162. doi: 10.1016/j.amjmed.2007.04.005. URL [https://www.amjmed.com/article/S0002-9343\(07\)00361-0/abstract](https://www.amjmed.com/article/S0002-9343(07)00361-0/abstract). Publisher: Elsevier.

J. a. Malato, L. Graça, L. Nacul, E. M. Lacerda, , and N. Sepúlveda. Statistical challenges of investigating a disease with a complex diagnosis. In P Milheiro, A. Pacheco, B. de Sousa, I. F. Alves, I. Pereira, M. J. a. Polidoro, and S. Ramos, editors, *Estatística: Desafios transversais às ciências com dados*, Eds. Sociedade portuguesa de estatística, pages 153–167. Sociedade Portuguesa de Estatística, 2021. ISBN 978-972-8890-47-6. URL <https://www.spestatistica.pt/storage/app/uploads/public/609/28f/6d0/60928f6d08a0c016386627.pdf>.

J. a. Malato, L. Graça, and N. Sepúlveda. Impact of Misclassification and Imperfect Serological Tests in Association Analyses of ME/CFS Applied to COVID-19 Data. In R. Bispo, L. Henriques-Rodrigues, R. Alpizar-Jara, and M. de Carvalho, editors, *Recent Developments in Statistics and Data Science*, Springer Proceedings in Mathematics & Statistics, pages 215–225, Cham, 2022a. Springer International Publishing. ISBN 978-3-031-12766-3. doi: 10.1007/978-3-031-12766-3_15.

J. a. Malato, R. M. Ribeiro, P. P. Leite, P. Casaca, E. Fernandes, C. Antunes, V. R. Fonseca, M. C. Gomes, and L. Graca. Risk of BA.5 Infection among Persons Exposed to Previous SARS-CoV-2 Variants. *New England Journal of Medicine*, 387(10):953–954, Sept. 2022b. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMc2209479. URL <http://www.nejm.org/doi/10.1056/NEJMc2209479>.

J. a. Malato, L. Graça, and N. Sepúlveda. Impact of Misdiagnosis in Case-Control Studies of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Diagnostics*, 13(3):531, Jan. 2023a. ISSN 2075-4418. doi: 10.3390/diagnostics13030531. URL <https://www.mdpi.com/2075-4418/13/3/531>. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

J. a. Malato, R. M. Ribeiro, E. Fernandes, P. P. Leite, P. Casaca, C. Antunes, V. R. Fonseca, M. C. Gomes, and

- L. Graca. Stability of hybrid versus vaccine immunity against BA.5 infection over 8 months. *The Lancet Infectious Diseases*, 23(2):148–150, Feb. 2023b. ISSN 14733099. doi: 10.1016/S1473-3099(22)00833-7. URL <https://linkinghub.elsevier.com/retrieve/pii/S1473309922008337>.
- P. V. Markov, M. Ghafari, M. Beer, K. Lythgoe, P. Simmonds, N. I. Stilianakis, and A. Katzourakis. The evolution of SARS-CoV-2. *Nature Reviews Microbiology*, 21(6):361–379, June 2023. ISSN 1740-1534. doi: 10.1038/s41579-023-00878-2. URL <https://www.nature.com/articles/s41579-023-00878-2>. Number: 6 Publisher: Nature Publishing Group.
- V. Matzarakis, V. Kumar, C. Wijmenga, and A. Zhernakova. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biology*, 18(1):76, Apr. 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1207-1. URL <https://doi.org/10.1186/s13059-017-1207-1>.
- J. Maya. Surveying the Metabolic and Dysfunctional Profiles of T Cells and NK Cells in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *International Journal of Molecular Sciences*, 24(15):11937, July 2023. ISSN 1422-0067. doi: 10.3390/ijms241511937. URL <https://www.mdpi.com/1422-0067/24/15/11937>.
- P. McCrone, L. Darbshire, L. Ridsdale, and P. Seed. The economic cost of chronic fatigue and chronic fatigue syndrome in UK primary care. *Psychological Medicine*, 33(2):253–261, Feb. 2003. ISSN 1469-8978, 0033-2917. doi: 10.1017/S0033291702006980. URL <https://www.cambridge.org/core/journals/psychological-medicine/article/economic-cost-of-chronic-fatigue-and-chronic-fatigue-syndrome-in-uk-primary-care/3556478BD17F4CA3EA85ED1870997C52>. Publisher: Cambridge University Press.
- C. P. McEvedy and A. W. Beard. Concept of Benign Myalgic Encephalomyelitis. *Br Med J*, 1(5687):11–15, Jan. 1970. ISSN 0007-1447, 1468-5833. doi: 10.1136/bmj.1.5687.11. URL <https://www.bmjjournals.org/content/1/5687/11>. Publisher: British Medical Journal Publishing Group Section: Papers and Originals.
- A. Melvin, E. Lacerda, H. M. Dockrell, S. O’Rahilly, and L. Nacul. Circulating levels of GDF15 in patients with myalgic encephalomyelitis/chronic fatigue syndrome. *Journal of Translational Medicine*, 17(1):409, Dec. 2019. ISSN 1479-5876. doi: 10.1186/s12967-019-02153-6. URL <https://doi.org/10.1186/s12967-019-02153-6>.
- S. A. Mendonça, R. Lorincz, P. Boucher, and D. T. Curiel. Adenoviral vector vaccine platforms in the SARS-CoV-2 pandemic. *NPJ vaccines*, 6(1):97, Aug. 2021. ISSN 2059-0105. doi: 10.1038/s41541-021-00356-x.
- F. Mensah, A. Bansal, S. Berkovitz, A. Sharma, V. Reddy, M. J. Leandro, and G. Cambridge. Extended B cell phenotype in patients with myalgic encephalomyelitis/chronic fatigue syndrome: a cross-sectional study. *Clinical and Experimental Immunology*, 184(2):237–247, May 2016. ISSN 0009-9104. doi: 10.1111/cei.12749. URL <https://doi.org/10.1111/cei.12749>.
- E. A. Meyerowitz, J. Scott, A. Richterman, V. Male, and M. Cevik. Clinical course and management of COVID-19 in the era of widespread population immunity. *Nature Reviews Microbiology*, 22(2):75–88, Feb. 2024.

- ISSN 1740-1534. doi: 10.1038/s41579-023-01001-1. URL <https://www.nature.com/articles/s41579-023-01001-1>. Number: 2 Publisher: Nature Publishing Group.
- T. Miike and D. S. Bell. Chronic Fatigue Syndrome in Childhood and Adolescence. In Y. Watanabe, B. Evengård, B. H. Natelson, L. A. Jason, and H. Kuratsune, editors, *Fatigue Science for Human Health*, pages 153–171. Springer Japan, Tokyo, 2008. ISBN 978-4-431-73464-2. doi: 10.1007/978-4-431-73464-2_12. URL https://doi.org/10.1007/978-4-431-73464-2_12.
- A. A. Mirin, M. E. Dimmock, and L. A. Jason. Updated ME/CFS prevalence estimates reflecting post-COVID increases and associated economic costs and funding implications. *Fatigue: Biomedicine, Health & Behavior*, 10(2):83–93, Apr. 2022. ISSN 2164-1846. doi: 10.1080/21641846.2022.2062169. URL <https://doi.org/10.1080/21641846.2022.2062169>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/21641846.2022.2062169>.
- J. G. Montoya, T. H. Holmes, J. N. Anderson, H. T. Maecker, Y. Rosenberg-Hasson, I. J. Valencia, L. Chu, J. W. Younger, C. M. Tato, and M. M. Davis. Cytokine signature associated with disease severity in chronic fatigue syndrome patients. *Proceedings of the National Academy of Sciences*, 114(34):E7150–E7158, July 2017. doi: 10.1073/pnas.1710519114.
- L. D. Moore, T. Le, and G. Fan. DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, 38(1):23–38, Jan. 2013. ISSN 1740-634X. doi: 10.1038/npp.2012.112. URL <https://www.nature.com/articles/npp2012112>. Number: 1 Publisher: Nature Publishing Group.
- G. Morris and M. Maes. Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics. *BMC Medicine*, 11(1), Sept. 2013. doi: 10.1186/1741-7015-11-205.
- R. Moss-Morris and T. Chalder. Illness perceptions and levels of disability in patients with chronic fatigue syndrome and rheumatoid arthritis. *Journal of Psychosomatic Research*, 55(4):305–308, Oct. 2003. ISSN 0022-3999. doi: 10.1016/s0022-3999(03)00013-8.
- C. Mueller, J. C. Lin, H. H. Thannickal, A. Daredia, T. S. Denney, R. Beyers, and J. W. Younger. No evidence of abnormal metabolic or inflammatory activity in the brains of patients with rheumatoid arthritis: results from a preliminary study using whole-brain magnetic resonance spectroscopic imaging (MRSI). *Clinical Rheumatology*, 39(6):1765–1774, 2020. ISSN 0770-3198. doi: 10.1007/s10067-019-04923-5. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7237391/>.
- K. M. Murphy and C. Weaver. *Janeway's immunobiology; 9th edition*. Garland Science, Taylor & Francis Group, New York, 2017. ISBN 978-0-8153-4551-0. URL <https://bib-pubdb1.desy.de/record/442324>. Pages: xx, 904 pages : illustrations, diagrams tex.ddc: 616.07907 tex.othercontributors: Mowat, Allan and Berg, Leslie and Chaplin, David and Janeway, Charles and Travers, Paul and Walport, Mark tex.reportid: PUBDB-2020-02992 tex.shelfmark: 616.079 Mur tex.typ: PUB:(DE-HGF)3.

- L. Nacul, C. C. Kingdon, E. W. Bowman, H. Curran, and E. M. Lacerda. Differing case definitions point to the need for an accurate diagnosis of myalgic encephalomyelitis/chronic fatigue syndrome. *Fatigue : biomedicine, health & behavior*, 5(1):1–4, 2017. ISSN 2164-1846. doi: 10.1080/21641846.2017.1273863. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5730342/>.
- L. Nacul, E. M. Lacerda, C. C. Kingdon, H. Curran, and E. W. Bowman. How have selection bias and disease misclassification undermined the validity of myalgic encephalomyelitis/chronic fatigue syndrome studies? *Journal of Health Psychology*, 24(12):1765–1769, Oct. 2019. ISSN 1461-7277. doi: 10.1177/1359105317695803.
- L. Nacul, S. O’Boyle, L. Palla, F. E. Nacul, K. Mudie, C. C. Kingdon, J. M. Cliff, T. G. Clark, H. M. Dockrell, and E. M. Lacerda. How Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) Progresses: The Natural History of ME/CFS. *Frontiers in Neurology*, 11, 2020. ISSN 1664-2295. doi: <https://doi.org/10.3389/fneur.2020.00826>. URL <https://www.frontiersin.org/articles/10.3389/fneur.2020.00826>.
- L. Nacul, F. J. Authier, C. Scheibenbogen, L. Lorusso, I. B. Helland, J. A. Martin, C. A. Sirbu, A. M. Mengshoel, O. Polo, U. Behrends, H. Nielsen, P. Grabowski, S. Sekulic, N. Sepulveda, F. Estévez-López, P. Zalewski, D. F. H. Pheby, J. Castro-Marrero, G. K. Sakkas, E. Capelli, I. Brundslund, J. Cullinan, A. Krumina, J. Bergquist, M. Murovska, R. C. W. Vermuelen, and E. M. Lacerda. European Network on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (EUROMENE): Expert Consensus on the Diagnosis, Service Provision, and Care of People with ME/CFS in Europe. *Medicina*, 57(5):510, May 2021. ISSN 1648-9144. doi: 10.3390/medicina57050510. URL <https://www.mdpi.com/1648-9144/57/5/510>.
- L. C. Nacul, E. M. Lacerda, D. Pheby, P. Campion, M. Molokhia, S. Fayyaz, J. C. Leite, F. Poland, A. Howe, and M. L. Drachler. Prevalence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in three regions of England: a repeated cross-sectional study in primary care. *BMC Medicine*, 9(1):91, Dec. 2011. ISSN 1741-7015. doi: 10.1186/1741-7015-9-91. URL <http://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-9-91>.
- B. Natelson, M. Haghghi, and N. Ponzio. Evidence for the presence of immune dysfunction in chronic fatigue syndrome. *Clinical and Diagnostic Laboratory Immunology*, 9(4):747–752, 2002. ISSN 1071-412X. doi: 10.1128/CDLI.9.4.747-752.2002.
- National Directorate of Health. COVID-19 Vaccination Report, 2021. URL <https://covid19.min-saude.pt/relatorio-vacinacao/>.
- R. K. Naviaux, J. C. Naviaux, K. Li, A. T. Bright, W. A. Alaynick, L. Wang, A. Baxter, N. Nathan, W. Anderson, and E. Gordon. Metabolic features of chronic fatigue syndrome. *Proceedings of the National Academy of Sciences*, 113(37):E5472–E5480, Sept. 2016. doi: 10.1073/pnas.1607571113. URL <https://www.pnas.org/doi/full/10.1073/pnas.1607571113>. Publisher: Proceedings of the National Academy of Sciences.
- C. B. Nguyen, L. Alsøe, J. M. Lindvall, D. Sulheim, E. Fagermoen, A. Winger, M. Kaarbø, H. Nilsen, and V. B. Wyller. Whole blood gene expression in adolescent chronic fatigue syndrome: an exploratory cross-sectional study

- suggesting altered B cell differentiation and survival. *Journal of Translational Medicine*, 15(1):102, May 2017. ISSN 1479-5876. doi: 10.1186/s12967-017-1201-0.
- NICE Guideline [NG206]. Myalgic encephalomyelitis (or encephalopathy)/chronic fatigue syndrome: diagnosis and management | Guidance, Oct. 2021. URL <https://www.nice.org.uk/guidance/ng206>. Publisher: NICE.
- I. Nilsson, J. Palmer, E. Apostolou, C.-G. Gottfries, M. Rizwan, C. Dahle, and A. Rosén. Metabolic Dysfunction in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Not Due to Anti-mitochondrial Antibodies. *Frontiers in Medicine*, 7, 2020. ISSN 2296-858X. URL <https://www.frontiersin.org/articles/10.3389/fmed.2020.00108>.
- N. Noor, I. Urts, A. Degueure, L. Rando, V. Kata, E. M. Cornett, A. D. Kaye, F. Imani, M. Narimani-Zamanabadi, G. Varrassi, and O. Viswanath. A Comprehensive Update of the Current Understanding of Chronic Fatigue Syndrome. *Anesthesiology and Pain Medicine*, 11(3):e113629, June 2021. ISSN 2228-7523. doi: 10.5812/aapm.113629. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8438707/>.
- D. Ohanian, A. Brown, M. Sunquist, J. Furst, L. Nicholson, L. Klebek, and L. A. Jason. Identifying Key Symptoms Differentiating Myalgic Encephalomyelitis and Chronic Fatigue Syndrome from Multiple Sclerosis. *Neurology (E-Cronicon)*, 4(2):41–45, 2016.
- H. Ono, W. Sato, and T. Yamamura. Dysregulation of T and B cells in myalgic encephalomyelitis/chronic fatigue syndrome. *Journal of the Neurological Sciences*, 381:899–900, Oct. 2017. ISSN 0022-510X, 1878-5883. doi: 10.1016/j.jns.2017.08.2533. URL [https://www.jns-journal.com/article/S0022-510X\(17\)33030-7/fulltext](https://www.jns-journal.com/article/S0022-510X(17)33030-7/fulltext). Publisher: Elsevier.
- R. Patarca. Cytokines and chronic fatigue syndrome. *Annals of the New York Academy of Sciences*, 933:185–200, Mar. 2001. ISSN 0077-8923. doi: 10.1111/j.1749-6632.2001.tb05824.x.
- M. Pedersen, T. T. Asprusten, K. Godang, T. M. Leegaard, L. T. Osnes, E. Skovlund, T. Tjade, M. G. Øie, and V. B. B. Wyller. Predictors of chronic fatigue in adolescents six months after acute Epstein-Barr virus infection: A prospective cohort study. *Brain, Behavior, and Immunity*, 75:94–100, Jan. 2019. ISSN 1090-2139. doi: 10.1016/j.bbi.2018.09.023.
- D. F. H. Pheby, D. Araja, U. Berkis, E. Brenna, J. Cullinan, J.-D. d. Korwin, L. Gitto, D. A. Hughes, R. M. Hunter, D. Trepel, and X. Wang-Steverding. The Development of a Consistent Europe-Wide Approach to Investigating the Economic Impact of Myalgic Encephalomyelitis (ME/CFS): A Report from the European Network on ME/CFS (EUROMENE). *Healthcare*, 8(2):88, Apr. 2020. doi: 10.3390/healthcare8020088.
- C. Pipper, L. Bliem, L. León, D. Mennickent, C. Bodner, E. Guzmán-Gutierrez, M. Stingl, E. Untersmayr-Elsenhuber, B. Wagner, R. Bertinat, N. Sepúlveda, and F. Westermeier. Sex and disease severity-based analysis of steroid hormones in ME/CFS. preprint, In Review, Oct. 2023. URL <https://www.researchsquare.com/article/rs-3428946/v1>.

- J. S. Pober and W. C. Sessa. Evolving functions of endothelial cells in inflammation. *Nature Reviews Immunology*, 7(10):803–815, Oct. 2007. ISSN 1474-1741. doi: 10.1038/nri2171. URL <https://www.nature.com/articles/nri2171>. Number: 10 Publisher: Nature Publishing Group.
- T. Pradeu and E. L. Cooper. The danger theory: 20 years later. *Frontiers in Immunology*, 3, 2012. ISSN 1664-3224. URL <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2012.00287>.
- Prevail III Study Group. A Longitudinal Study of Ebola Sequelae in Liberia. *New England Journal of Medicine*, 380(10):924–934, Mar. 2019. ISSN 0028-4793. doi: 10.1056/NEJMoa1805435. URL <https://doi.org/10.1056/NEJMoa1805435>. Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMoa1805435>.
- C. Pritchard. Fibrositis and the chronic fatigue syndrome. *Annals of Internal Medicine*, 108(6):906, June 1988. ISSN 0003-4819. doi: 10.7326/0003-4819-108-6-906_1.
- J. R. C. Pulliam, C. van Schalkwyk, N. Govender, A. von Gottberg, C. Cohen, M. J. Groome, J. Dushoff, K. Mlisana, and H. Moultrie. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science*, 376(6593):eabn4947, Mar. 2022. doi: 10.1126/science.abn4947. URL <https://www.science.org/doi/10.1126/science.abn4947>. Publisher: American Association for the Advancement of Science.
- O. L. Quintero, M. J. Amador-Patarroyo, G. Montoya-Ortiz, A. Rojas-Villarraga, and J.-M. Anaya. Autoimmune disease and gender: plausible mechanisms for the female predominance of autoimmunity. *Journal of Autoimmunity*, 38(2-3):J109–119, May 2012. ISSN 1095-9157. doi: 10.1016/j.jaut.2011.10.003.
- S. Ramos, E. Brenu, S. Broadley, R. Kwiatek, J. Ng, T. Nguyen, S. Freeman, D. Staines, and S. Marshall-Gradisnik. Regulatory T, natural killer T and \upgamma\updelta T cells in multiple sclerosis and chronic fatigue syndrome/myalgic encephalomyelitis: a comparison. *Asian Pacific Journal of Allergy and Immunology*, 2016. doi: 10.12932/ap0733.
- A. M. Ramsay and E. O'sullivan. Encephalomyelitis simulating poliomyelitis. *Lancet (London, England)*, 270 (6926):761–764, May 1956. ISSN 0140-6736. doi: 10.1016/s0140-6736(56)91234-x.
- S. Rasa, Z. Nora-Krukle, N. Henning, E. Eliassen, E. Shikova, T. Harrer, C. Scheibenbogen, M. Murovska, B. K. Prusty, and the European Network on ME/CFS (EUROMENE). Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 16(1):268, Oct. 2018. ISSN 1479-5876. doi: 10.1186/s12967-018-1644-y. URL <https://doi.org/10.1186/s12967-018-1644-y>.
- M. Rezaei and M. Nazari. New Generation Vaccines for COVID-19 Based on Peptide, Viral Vector, Artificial Antigen Presenting Cell, DNA or mRNA. *Avicenna Journal of Medical Biotechnology*, 14(1):30–36, 2022. ISSN 2008-2835. doi: 10.18502/ajmb.v14i1.8167. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9017465/>.

- J. L. Rivas, T. Palencia, G. Fernández, and M. García. Association of T and NK Cell Phenotype With the Diagnosis of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Frontiers in Immunology*, 9, May 2018. doi: 10.3389/fimmu.2018.01028.
- M. C. Rivera, C. Mastronardi, C. Silva-Aldana, M. Arcos-Burgos, and B. Lidbury. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Comprehensive Review. *Diagnostics*, 9(3):91, Aug. 2019. doi: 10.3390/diagnostics9030091.
- M. Rojas, P. Restrepo-Jiménez, D. M. Monsalve, Y. Pacheco, Y. Acosta-Ampudia, C. Ramírez-Santana, P. S. C. Leung, A. A. Ansari, M. E. Gershwin, and J.-M. Anaya. Molecular mimicry and autoimmunity. *Journal of Autoimmunity*, 95:100–123, Dec. 2018. ISSN 1095-9157. doi: 10.1016/j.jaut.2018.10.012.
- B. V. Rooney, B. E. Crucian, D. L. Pierson, M. L. Laudenslager, and S. K. Mehta. Herpes Virus Reactivation in Astronauts During Spaceflight and Its Application on Earth. *Frontiers in Microbiology*, 10, 2019. ISSN 1664-302X. URL <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00016>.
- P. C. Rowe. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Trial Fails to Confirm Earlier Observations of Rituximab's Effectiveness. *Annals of Internal Medicine*, 170(9):656–657, May 2019. ISSN 0003-4819. doi: 10.7326/M19-0643. URL <https://www.acpjournals.org/doi/10.7326/M19-0643>. Publisher: American College of Physicians.
- P. C. Rowe, R. A. Underhill, K. J. Friedman, A. Gurwitt, M. S. Medow, M. S. Schwartz, N. Speight, J. M. Stewart, R. Vallings, and K. S. Rowe. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Diagnosis and Management in Young People: A Primer. *Frontiers in Pediatrics*, 5:121, 2017. ISSN 2296-2360. doi: 10.3389/fped.2017.00121.
- M. Ruiz-Pablos, B. Paiva, R. Montero-Mateo, N. Garcia, and A. Zabaleta. Epstein-Barr Virus and the Origin of Myalgic Encephalomyelitis or Chronic Fatigue Syndrome. *Frontiers in Immunology*, 12, 2021. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2021.656797>.
- C. Scheibenbogen, H. Freitag, J. Blanco, E. Capelli, E. Lacerda, J. Authier, M. Meeus, J. C. Marrero, Z. Nora-Kruk, E. Oltra, and others. The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE. *Journal of translational medicine*, 15(1):162, 2017. doi: <https://doi.org/10.1186/s12967-017-1263-z>.
- N. Scherbakov, M. Szklarski, J. Hartwig, F. Sotzny, S. Lorenz, A. Meyer, P. Grabowski, W. Doehner, and C. Scheibenbogen. Peripheral endothelial dysfunction in myalgic encephalomyelitis/chronic fatigue syndrome. *ESC heart failure*, 7(3):1064–1071, June 2020. ISSN 2055-5822. doi: 10.1002/ehf2.12633.
- K. A. Schlauch, S. F. Khaiboullina, K. L. De Meirleir, S. Rawat, J. Petereit, A. A. Rizvanov, N. Blatt, T. Mijatovic, D. Kulick, A. Palotás, and V. C. Lombardi. Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. *Translational Psychiatry*, 6(2):e730–e730, Feb. 2016. ISSN 2158-3188. doi: 10.1038/tp.2015.208. URL <https://www.nature.com/articles/>

- [tp2015208](#). Number: 2 Publisher: Nature Publishing Group.
- S. C. Segerstrom and G. E. Miller. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological Bulletin*, 130(4):601–630, July 2004. ISSN 0033-2909. doi: 10.1037/0033-2909.130.4.601.
- N. Sepúlveda. Impact of genetic variation on the molecular mimicry between Anoctamin-2 and Epstein-Barr virus nuclear antigen 1 in Multiple Sclerosis. *Immunology Letters*, 238:29–31, Oct. 2021. ISSN 1879-0542. doi: 10.1016/j.imlet.2021.07.007.
- N. Sepúlveda, J. Carneiro, E. Lacerda, and L. Nacul. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome as a Hyper-Regulated Immune System Driven by an Interplay Between Regulatory T Cells and Chronic Human Herpesvirus Infections. *Frontiers in Immunology*, 10, 2019. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02684>.
- N. Sepúlveda, J. a. Malato, F. Sotzny, A. D. Grabowska, A. Fonseca, C. Cordeiro, L. Graça, P. Biecek, U. Behrends, J. Mautner, F. Westermeier, E. M. Lacerda, and C. Scheibenbogen. Revisiting IgG Antibody Reactivity to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Its Potential Application to Disease Diagnosis. *Frontiers in Medicine*, 9, 2022. ISSN 2296-858X. doi: 10.3389/fmed.2022.921101. URL <https://www.frontiersin.org/articles/10.3389/fmed.2022.921101>.
- N. A. Shadick, C. B. Phillips, E. L. Logigan, A. C. Steere, R. F. Kaplan, V. P. Berardi, P. H. Duray, M. G. Larson, E. A. Wright, K. S. Ginsburg, J. N. Katz, and M. H. Liang. The long-term clinical outcomes of Lyme disease. A population-based retrospective cohort study. *Annals of Internal Medicine*, 121(8):560–567, Oct. 1994. ISSN 0003-4819. doi: 10.7326/0003-4819-121-8-199410150-00002.
- M. C. Sharpe, L. C. Archard, J. E. Banatvala, L. K. Borysiewicz, A. W. Clare, A. David, R. H. Edwards, K. E. Hawton, H. P. Lambert, and R. J. Lane. A report–chronic fatigue syndrome: guidelines for research. *Journal of the Royal Society of Medicine*, 84(2):118–121, Feb. 1991. ISSN 0141-0768. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1293107/>.
- N. Shimosako and J. R. Kerr. Use of single-nucleotide polymorphisms (SNPs) to distinguish gene expression subtypes of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *Journal of Clinical Pathology*, 67 (12):1078–1083, Sept. 2014. doi: 10.1136/jclinpath-2014-202597.
- D. Singh and S. V. Yi. On the origin and evolution of SARS-CoV-2. *Experimental & Molecular Medicine*, 53(4): 537–547, Apr. 2021. ISSN 2092-6413. doi: 10.1038/s12276-021-00604-z. URL <https://www.nature.com/articles/s12276-021-00604-z>. Number: 4 Publisher: Nature Publishing Group.
- J. Słomko, J. L. Newton, S. Kujawski, M. Tafil-Klawe, J. Klawe, D. Staines, S. Marshall-Gradisnik, and P. Zalewski. Prevalence and characteristics of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) in Poland: a cross-sectional study. *BMJ open*, 9(3):e023955, Mar. 2019. ISSN 2044-6055. doi: 10.1136/bmjopen-2018-023955.

- A. K. Smith, H. Fang, T. Whistler, E. R. Unger, and M. S. Rajeevan. Convergent Genomic Studies Identify Association of GRIK2 and NPAS2 with Chronic Fatigue Syndrome. *Neuropsychobiology*, 64(4):183–194, 2011. doi: 10.1159/000326692.
- M. E. B. Smith, H. D. Nelson, E. Haney, M. Pappas, M. Daeges, N. Wasson, and M. McDonagh. Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Evidence report/technology assessment*, (219):1–433, Dec. 2014. ISSN 1530-4396. doi: 10.23970/ahrqepcerta219. URL <https://doi.org/10.23970/AHRQEPCERTA219>.
- L. Solomon and W. C. Reeves. Factors Influencing the Diagnosis of Chronic Fatigue Syndrome. *Archives of Internal Medicine*, 164(20):2241–2245, Nov. 2004. ISSN 0003-9926. doi: 10.1001/archinte.164.20.2241. URL <https://doi.org/10.1001/archinte.164.20.2241>.
- M. Sospedra, Y. Zhao, H. zur Hausen, P. A. Muraro, C. Hamashin, E.-M. de Villiers, C. Pinilla, and R. Martin. Recognition of conserved amino acid motifs of common viruses and its role in autoimmunity. *PLoS pathogens*, 1(4):e41, Dec. 2005. ISSN 1553-7366. doi: 10.1371/journal.ppat.0010041.
- F. Sotzny, J. Blanco, E. Capelli, J. Castro-Marrero, S. Steiner, M. Murovska, and C. Scheibenberg. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome – Evidence for an autoimmune disease. *Autoimmunity Reviews*, 17(6):601–609, June 2018. doi: 10.1016/j.autrev.2018.01.009.
- T. N. Starr, A. J. Greaney, S. K. Hilton, D. Ellis, K. H. D. Crawford, A. S. Dingens, M. J. Navarro, J. E. Bowen, M. A. Tortorici, A. C. Walls, N. P. King, D. Veesler, and J. D. Bloom. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*, 182(5):1295–1310.e20, Sept. 2020. ISSN 0092-8674. doi: 10.1016/j.cell.2020.08.012. URL <https://www.sciencedirect.com/science/article/pii/S0092867420310035>.
- S. Steiner, S. C. Becker, J. Hartwig, F. Sotzny, S. Lorenz, S. Bauer, M. Löbel, A. B. Stittrich, P. Grabowski, and C. Scheibenberg. Autoimmunity-Related Risk Variants in PTPN22 and CTLA4 Are Associated With ME/CFS With Infectious Onset. *Frontiers in Immunology*, 11, 2020. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2020.00578>.
- S. Steiner, A. Fehrer, F. Hoheisel, S. Schoening, A. Aschenbrenner, N. Babel, J. Bellmann-Strobl, C. Finke, O. Fluge, L. Froehlich, A. Goebel, B. Grande, J.-P. Haas, B. Hohberger, L. A. Jason, A. L. Komaroff, E. Lacerda, M. Liebl, A. Maier, O. Mella, L. Nacul, F. Paul, B. K. Prusty, C. Puta, G. Riemekasten, W. Ries, P. C. Rowe, B. Sawitzki, Y. Shoenfeld, J. L. Schultze, M. Seifert, N. Sepúlveda, F. Sotzny, E. Stein, M. Stingl, F. Ufer, C. Veauthier, F. Westermeier, K. Wirth, B. Wolfarth, P. Zalewski, U. Behrends, and C. Scheibenberg. Understanding, diagnosing, and treating Myalgic encephalomyelitis/chronic fatigue syndrome – State of the art: Report of the 2nd international meeting at the Charité Fatigue Center. *Autoimmunity Reviews*, 22(11):103452, Nov. 2023. ISSN 1568-9972. doi: 10.1016/j.autrev.2023.103452. URL <https://www.sciencedirect.com/science/article/pii/S1568997223001866>.

- C. M. Steyers and F. J. Miller. Endothelial dysfunction in chronic inflammatory diseases. *International Journal of Molecular Sciences*, 15(7):11324–11349, June 2014. ISSN 1422-0067. doi: 10.3390/ijms150711324.
- L. Stojanovich and D. Marisavljevich. Stress as a trigger of autoimmune disease. *Autoimmunity Reviews*, 7(3): 209–213, 2008. ISSN 1568-9972. doi: 10.1016/j.autrev.2007.11.007.
- J. Stoothoff. Subtyping Patients with Myalgic Encephalomyelitis (ME) and Chronic Fatigue Syndrome (CFS) By Course of Illness. *Journal of Biosensors, Biomarkers and Diagnostics*, 2(1):1–9, June 2017. doi: 10.15226/2575-6303/2/1/00113.
- B. Sundaresan, F. Shirafkan, K. Ripperger, and K. Rattay. The Role of Viral Infections in the Onset of Autoimmune Diseases. *Viruses*, 15(3):782, Mar. 2023. ISSN 1999-4915. doi: 10.3390/v15030782.
- M. Szklarski, H. Freitag, S. Lorenz, S. C. Becker, F. Sotzny, S. Bauer, J. Hartwig, H. Heidecke, K. Wittke, C. Kedor, L. G. Hanitsch, P. Grabowski, N. Sepúlveda, and C. Scheibenbogen. Delineating the Association Between Soluble CD26 and Autoantibodies Against G-Protein Coupled Receptors, Immunological and Cardiovascular Parameters Identifies Distinct Patterns in Post-Infectious vs. Non-Infection-Triggered Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Immunology*, 12, Apr. 2021. doi: 10.3389/fimmu.2021.644548.
- K. Tengvall, J. Huang, C. Hellström, P. Kammer, M. Biström, B. Ayoglu, I. Lima Bomfim, P. Stridh, J. Butt, N. Brenner, A. Michel, K. Lundberg, L. Padyukov, I. E. Lundberg, E. Svenungsson, I. Ernberg, S. Olafsson, A. T. Dilthey, J. Hillert, L. Alfredsson, P. Sundström, P. Nilsson, T. Waterboer, T. Olsson, and I. Kockum. Molecular mimicry between Anoctamin 2 and Epstein-Barr virus nuclear antigen 1 associates with multiple sclerosis risk. *Proceedings of the National Academy of Sciences of the United States of America*, 116(34):16955–16960, Aug. 2019. ISSN 1091-6490. doi: 10.1073/pnas.1902623116.
- The Medical Staff Of The Royal Free Hospital. An Outbreak of Encephalomyelitis in the Royal Free Hospital Group, London, in 1955. *British Medical Journal*, 2(5050):895–904, Oct. 1957. ISSN 0007-1447. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1962472/>.
- J. Theorell, I. Bileviciute-Ljungar, B. Tesi, H. Schlums, M. S. Johnsgaard, B. Asadi-Azarbaijani, E. Bolle Strand, and Y. T. Bryceson. Unperturbed Cytotoxic Lymphocyte Phenotype and Function in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Patients. *Frontiers in Immunology*, 8, 2017. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2017.00723>.
- A. Tomoda, T. Joudoi, E.-M. Rabab, T. Matsumoto, T. Park, and T. Miike. Cytokine production and modulation: Comparison of patients with chronic fatigue syndrome and normal controls. *Psychiatry Research*, 134(1): 101–104, 2005. ISSN 0165-1781. doi: 10.1016/j.psychres.2005.01.002.
- M. S. Trivedi, E. Oltra, L. Sarria, N. Rose, V. Beljanski, M. A. Fletcher, N. G. Klimas, and L. Nathanson. Identification of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome-associated DNA methylation patterns. *PLOS ONE*, 13(7):e0201066, July 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0201066. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201066>. Publisher: Public

- Library of Science.
- J. Trowsdale and J. C. Knight. Major Histocompatibility Complex Genomics and Human Disease. *Annual Review of Genomics and Human Genetics*, 14(1):301–323, 2013. doi: 10.1146/annurev-genom-091212-153455. URL <https://doi.org/10.1146/annurev-genom-091212-153455>. _eprint: <https://doi.org/10.1146/annurev-genom-091212-153455>.
- S.-Y. Tsai, T.-Y. Yang, H.-J. Chen, C.-S. Chen, W.-M. Lin, W.-C. Shen, C.-N. Kuo, and C.-H. Kao. Increased risk of chronic fatigue syndrome following herpes zoster: a population-based study. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*, 33(9):1653–1659, Sept. 2014. ISSN 1435-4373. doi: 10.1007/s10096-014-2095-x.
- R. A. Underhill. Myalgic encephalomyelitis, chronic fatigue syndrome: An infectious disease. *Medical Hypotheses*, 85(6):765–773, Dec. 2015. ISSN 1532-2777. doi: 10.1016/j.mehy.2015.10.011.
- A. R. Valdez, E. E. Hancock, S. Adebayo, D. J. Kiernicki, D. Proskauer, J. R. Attewell, L. Bateman, A. DeMaria, C. W. Lapp, P. C. Rowe, and C. Proskauer. Estimating Prevalence, Demographics, and Costs of ME/CFS Using Large Scale Medical Claims Data and Machine Learning. *Frontiers in Pediatrics*, 6:412, 2018. ISSN 2296-2360. doi: 10.3389/fped.2018.00412.
- W. C. d. Vega, S. D. Vernon, and P. O. McGowan. DNA Methylation Modifications Associated with Chronic Fatigue Syndrome. *PLoS ONE*, 9(8):e104757, Aug. 2014. doi: 10.1371/journal.pone.0104757.
- W. C. d. Vega, S. Herrera, S. D. Vernon, and P. O. McGowan. Epigenetic modifications and glucocorticoid sensitivity in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *BMC Medical Genomics*, 10(1), Feb. 2017. doi: 10.1186/s12920-017-0248-3.
- W. C. d. Vega, L. Erdman, S. D. Vernon, A. Goldenberg, and P. O. McGowan. Integration of DNA methylation & health scores identifies subtypes in myalgic encephalomyelitis/chronic fatigue syndrome. *Epigenomics*, 10(5):539–557, May 2018. doi: 10.2217/epi-2017-0150.
- S. D. Vernon, E. R. Unger, I. M. Dimulescu, M. Rajeevan, and W. C. Reeves. Utility of the Blood for Gene Expression Profiling and Biomarker Discovery in Chronic Fatigue Syndrome. *Disease Markers*, 18(4):193–199, 2002. ISSN 0278-0240. doi: 10.1155/2002/892374. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3851413/>.
- A. Vincent, D. J. Brimmer, M. O. Whipple, J. F. Jones, R. Boneva, B. D. Lahr, E. Maloney, J. L. St Sauver, and W. C. Reeves. Prevalence, incidence, and classification of chronic fatigue syndrome in Olmsted County, Minnesota, as estimated using the Rochester Epidemiology Project. *Mayo Clinic Proceedings*, 87(12):1145–1152, Dec. 2012. ISSN 1942-5546. doi: 10.1016/j.mayocp.2012.08.015.
- C. M. Walsh, N. Z. Zainal, S. J. Middleton, and E. S. Paykel. A family history study of chronic fatigue syndrome. *Psychiatric Genetics*, 11(3):123–128, Sept. 2001. ISSN 0955-8829. doi: 10.1097/00041444-200109000-00003.
- D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, and

- Z. Peng. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA*, 323(11):1061–1069, Mar. 2020. ISSN 1538-3598. doi: 10.1001/jama.2020.1585.
- T. Wang, J. Yin, A. H. Miller, and C. Xiao. A systematic review of the association between fatigue and genetic polymorphisms. *Brain, Behavior, and Immunity*, 62:230–244, May 2017. ISSN 0889-1591. doi: 10.1016/j.bbi.2017.01.007. URL <https://www.sciencedirect.com/science/article/pii/S0889159117300077>.
- Y. Wei, L. Wei, Y. Liu, L. Huang, S. Shen, R. Zhang, J. Chen, Y. Zhao, H. Shen, and F. Chen. Comprehensive estimation for the length and dispersion of COVID-19 incubation period: a systematic review and meta-analysis. *Infection*, 50(4):803–813, Aug. 2022. ISSN 1439-0973. doi: 10.1007/s15010-021-01682-x. URL <https://doi.org/10.1007/s15010-021-01682-x>.
- C. C. Whitacre. Sex differences in autoimmune disease. *Nature Immunology*, 2(9):777–780, Sept. 2001. ISSN 1529-2916. doi: 10.1038/ni0901-777. URL <https://www.nature.com/articles/ni0901-777>. Number: 9 Publisher: Nature Publishing Group.
- A. Wilder-Smith and S. Osman. Public health emergencies of international concern: a historic overview. *Journal of Travel Medicine*, 27(8):taaa227, Dec. 2020. ISSN 1708-8305. doi: 10.1093/jtm/taaa227. URL <https://doi.org/10.1093/jtm/taaa227>.
- C. Willyard. Are repeat COVID infections dangerous? What the science says. *Nature*, 616(7958):650–652, Apr. 2023. doi: 10.1038/d41586-023-01371-9. URL <https://www.nature.com/articles/d41586-023-01371-9>. Bandiera_abtest: a Cg_type: News Feature Number: 7958 Publisher: Nature Publishing Group Subject_term: SARS-CoV-2, Infection, Immunology.
- K. Wirth and C. Scheibenbogen. A Unifying Hypothesis of the Pathophysiology of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS): Recognitions from the finding of autoantibodies against β 2-adrenergic receptors. *Autoimmunity Reviews*, 19(6):102527, June 2020. ISSN 1873-0183. doi: 10.1016/j.autrev.2020.102527.
- W. Wojcik, D. Armstrong, and R. Kanaan. Chronic fatigue syndrome: Labels, meanings and consequences. *Journal of Psychosomatic Research*, 70(6):500–504, June 2011. ISSN 0022-3999. doi: 10.1016/j.jpsychores.2011.02.002. URL <https://www.sciencedirect.com/science/article/pii/S0022399911000456>.
- E. Wood, K. H. Hall, and W. Tate. Role of mitochondria, oxidative stress and the response to antioxidants in myalgic encephalomyelitis/chronic fatigue syndrome: A possible approach to SARS-CoV-2 'long-haulers'? *Chronic Diseases and Translational Medicine*, 7(1):14–26, Mar. 2021. ISSN 2589-0514. doi: 10.1016/j.cdtm.2020.11.002.
- World Health Organization. Coronavirus disease (COVID-19), Sept. 2023a. URL [https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-\(covid-19\)](https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-(covid-19)).
- World Health Organization. Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic, 2023b. URL <https://www.who.int/news/item/>

[05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic.](#)

World Health Organization. WHO Coronavirus (COVID-19) dashboard, 2023c. URL <https://data.who.int/dashboards/covid19/cases>.

Y. Wu, L. Kang, Z. Guo, J. Liu, M. Liu, and W. Liang. Incubation Period of COVID-19 Caused by Unique SARS-CoV-2 Strains: A Systematic Review and Meta-analysis. *JAMA Network Open*, 5(8):e2228008, Aug. 2022. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2022.28008. URL <https://doi.org/10.1001/jamanetworkopen.2022.28008>.

S. Yan, K. Kotschenreuther, S. Deng, and D. M. Kofler. Regulatory T cells in rheumatoid arthritis: functions, development, regulation, and therapeutic potential. *Cellular and Molecular Life Sciences*, 79(10):533, 2022. ISSN 1420-682X. doi: 10.1007/s00018-022-04563-0. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9522664/>.

M. B. Yunus. Fibromyalgia syndrome: new research on an old malady. *BMJ (Clinical research ed.)*, 298(6672): 474–475, Feb. 1989. ISSN 0959-8138. doi: 10.1136/bmj.298.6672.474.

Part II

**UNCERTAINTY AND
STRATIFICATION IN ME/CFS**

Chapter 2

Statistical challenges of investigating a disease with a complex diagnosis

J. Malato, L. Graça, L. Nacul, E. M. Lacerda, and N. Sepúlveda. Statistical challenges of investigating a disease with a complex diagnosis. In P. Milheiro, A. Pacheco, B. de Sousa, I. F. Alves, I. Pereira, M. J. Polidoro, and S. Ramos, editors, *Estatística: Desafios transversais às ciências com dados*, Eds. Sociedade Portuguesa de Estatística, pages 153–167. Sociedade Portuguesa de Estatística, 2021. ISBN 978-972-8890-47-6. URL <https://www.spestatistica.pt/storage/app/uploads/public/609/28f/6d0/60928f6d08a0c016386627.pdf>.

Abstract

Given the absence of a disease-specific biomarker, there are more than 20 symptoms-based case definitions of myalgic encephalomyelitis/chronic fatigue syndrome. As a consequence, the diagnosis for a given patient could vary from one case definition to another. In this context, we analyse data from a biobank dedicated to this disease in order to study the agreement between different case definitions, the similarity between symptom's profile among all participants including healthy controls and patients with multiple sclerosis. We also investigate the impact of patients' misclassification on a hypothetical association analysis using data simulation.

Keywords: multidimensional scaling; cluster analysis; misclassification; Cohen's κ coefficient; Jacard's similarity index.

2.1 Introduction

Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) is a complex disease whose patients manifest unexplained fatigue lasting for more than six months (Fukuda et al. 1994) or suffer from post-exertional malaise that is not alleviated by rest (Carruthers et al. 2003).

Disease prevalence has been estimated between 0.4% and 1.0% affecting six women to one man (Morris and Maes 2013). The underlying pathological mechanisms remain poorly understood, but they are often associated with environmental stressors, including severe viral infections (Rasa et al. 2018).

Until now there is no accurate biomarker for disease diagnosis. To overcome this problem, researchers and clinicians altogether have proposed more than 20 different case definitions based on patients' symptomatology while excluding known diseases that could explain the fatigue reported by suspected cases (Brurberg et al. 2014). As a consequence, the diagnosis for a given patient can vary from one case definition to another. Therefore, research from ME/CFS could be affected by the inclusion of false positive cases in the respective data.

In the present paper, we discuss the problem of diagnosing ME/CFS using data from the United Kingdom ME/CFS Biobank (UKMEB). With this purpose, we first introduce the biobank and its data. We then assess the agreement between 4 common case definitions of ME/CFS in 275 suspected cases belonging to the UKMEB. We then estimate the similarity between symptom's severity profiles from suspected cases, patients with multiple sclerosis, and healthy controls. We also study the impact of patients' misclassification on the statistical power of a hypothetical association analysis. Finally, we conclude this paper with some final remarks.

2.2 The UKMEB

The UKMEB refers to a large data set of suspected cases of ME/CFS, healthy controls, and patients with multiple sclerosis included as an additional control group (Lacerda et al. 2018). In terms of recruitment, suspected cases were identified in different institutions across the National Health Service from the United Kingdom and then referred to the CureMe group, a dedicated clinical research team based in the London School of Hygiene & Tropical Medicine and responsible for recruiting, managing, and curating the biobank. For this paper, the data set under analysis consists of a total of 523 participants divided into 275 suspected cases of ME/CFS, 136 healthy controls, and 112 patients with multiple sclerosis.

2.3 Diagnostic agreement analysis

After patients' referral for a possible integration in the biobank, suspected cases were comprehensively evaluated according to four case definitions of ME/CFS: Centre for Disease Control criteria (CDC-1994) (Fukuda et al. 1994), Canadian Consensus Criteria (CCC-2003) (Carruthers et al. 2003), Institute of Medicine Criteria (IOM-2015) (Institute of Medicine 2015), and International Consensus Criteria (ICC-2011) (Carruthers et al. 2011). The CDC-1994 requires the patients to have unexplained fatigue for at least 6 months and at least four out of eight fatigue-related symptoms. The IOM-2015 is typically used by general practitioners and it requires the patients to show at least three main symptoms such as profound fatigue, post-exertional malaise, and unrefreshing sleep. The CCC-2003 requires the patients to manifest four or more fatigue specific symptoms, at least two neurological or cognitive ones, and at least one autoimmune, neuroendocrine, or immune symptom. Finally, the ICC-2011 is more focused on neuro-immune and cognitive symptoms, and on the inability to produce sufficient energy on demand (post-exertional neuroimmune exhaustion).

There were 269 (97.8%), 233 (84.7%), 229 (83.3%) and 213 (77.5%) out of 275 suspected cases whose symptoms agreed with CDC-1994, IOM-2015, CCC-2003, and ICC-2011, respectively (Table 2.1). This finding suggests that the general practitioners who referred the suspected cases to a possible integration in the biobank made their diagnosis based on the CDC-1994. Unsurprisingly, only 62.9% of the suspected cases ($n = 173$) had a positive diagnosis across all the four case definitions. Therefore, the remaining suspected cases had at least one negative diagnosis.

It is worth noting that there were no suspected cases who had a negative diagnosis across all case definitions. There were also three individuals whose symptoms agreed with ICC-2011 only, IOM-2015 only, or both criteria. These individuals were considered to be fatigued but non-ME/CFS patients given that they did not agree with either the CDC-1994 or the CCC-2003 as recommended for ME/CFS research (Pheby et al. 2020).

To better understand the agreement between diagnostic outcomes obtained from different case definitions, we used the Jaccard's similarity index, J (Gower and Warrens 2014). Note

Table 2.1: Frequency of suspected cases of ME/CFS according to their diagnostic outcomes using different case definitions. Percentages in the last row indicate the proportion of diagnosed cases by each case definition.

Case definition				N	% of total suspected cases
CDC-1994	IOM-2015	CCC-2003	ICC-2011		
+	+	+	+	173	62.9
+	+	+	-	32	11.6
+	-	+	+	16	5.8
+	+	-	+	16	5.8
+	-	-	-	14	5.1
+	+	-	-	10	3.6
+	-	+	-	5	1.8
+	-	-	+	3	1.1
-	-	+	+	3	1.1
-	-	-	+	1	0.4
-	+	-	-	1	0.4
-	+	-	+	1	0.4
97.8%	84.7%	83.3%	77.5%	275	100%

that this index is usually a measure used to compare objects with shared attributes. Here we instead applied this index to compare attributes themselves. For a pair of case definitions (C_i, C_j) , this index was estimated as

$$J(C_i, C_j) = \frac{S}{S_i + S_j - S}, \quad i, j = 1, \dots, 4, \quad (2.1)$$

where S_i and S_j are the number of suspected cases with a positive diagnosis by C_i and C_j , respectively, and S is the number of suspected cases with a positive diagnosis by both criteria. In theory, the index is defined between 0 and 1 (i.e., no and full agreement between C_i and C_j across all individuals, respectively).

The estimates of this index ranged from 0.752 (IOM-2015 versus ICC-2011) to 0.876 (CDC-1994 versus IOM-2015; CDC-1994 versus CCC-2003) (Table 2.2). The estimates showed the stringency and differences in scope of each case definition. In addition, these estimates showed that, even if the general practitioners applied two different case definitions of ME/CFS in their

diagnosis, there could still be a fraction of suspected cases where the respective diagnostic outcomes might not agree with each other.

Table 2.2: Estimates of the Jaccard's similarity index for the four case definitions of ME/CFS using data from the UKMEB.

	CDC-1994	IOM-2015	CCC-2003	ICC-2011
CDC-1994	1.000	0.876	0.876	0.760
IOM-2005	0.876	1.000	0.840	0.752
CCC-2003	0.876	0.840	1.000	0.753
ICC-2011	0.760	0.752	0.753	1.000

2.4 Symptoms' similarity analysis

A major advantage of using data from the UKMEB is the comprehensive symptom's characterisation of all study participants. In particular, each participant had to report the severity of 57 symptoms occurred a month before data collection. Severity of each symptom was categorised into absence, mild, moderate, and severe. These invaluable data were then analysed to assess the similarity of all participants in terms of their symptom's severity profile. With this purpose, we first computed all possible 4×4 contingency tables resulting from cross-tabulating the symptom's severity data for any given pair of participants (i, j) , $i, j = 1, \dots, 523$. We then calculated a similarity matrix between any given pair of individuals by estimating the Cohen's κ coefficient (Agresti and Kateri 2011) in the corresponding 4×4 contingency tables, that is,

$$\kappa_{ij} = \frac{\sum_{k=1}^4 p_{ij,kk} - \sum_{k=1}^4 p_{ij,k} \cdot p_{ij,k}}{1 - \sum_{k=1}^4 p_{ij,k} \cdot p_{ij,k}}, \quad (2.2)$$

where $k = 1, \dots, 4$, $p_{ij,kk}$ is the proportion of symptoms with severity k reported by both individuals i and j , $p_{ij,k \cdot}$ is the proportion of symptoms with severity k reported by individual i , and $p_{ij,\cdot k}$ is the proportion of symptoms with severity k reported by individual j . The resulting similarity matrix was then analysed by classical multidimensional scaling (MDS; Figure 2.1A) and hierarchical cluster analysis using complete linkage (Figure 2.1B).

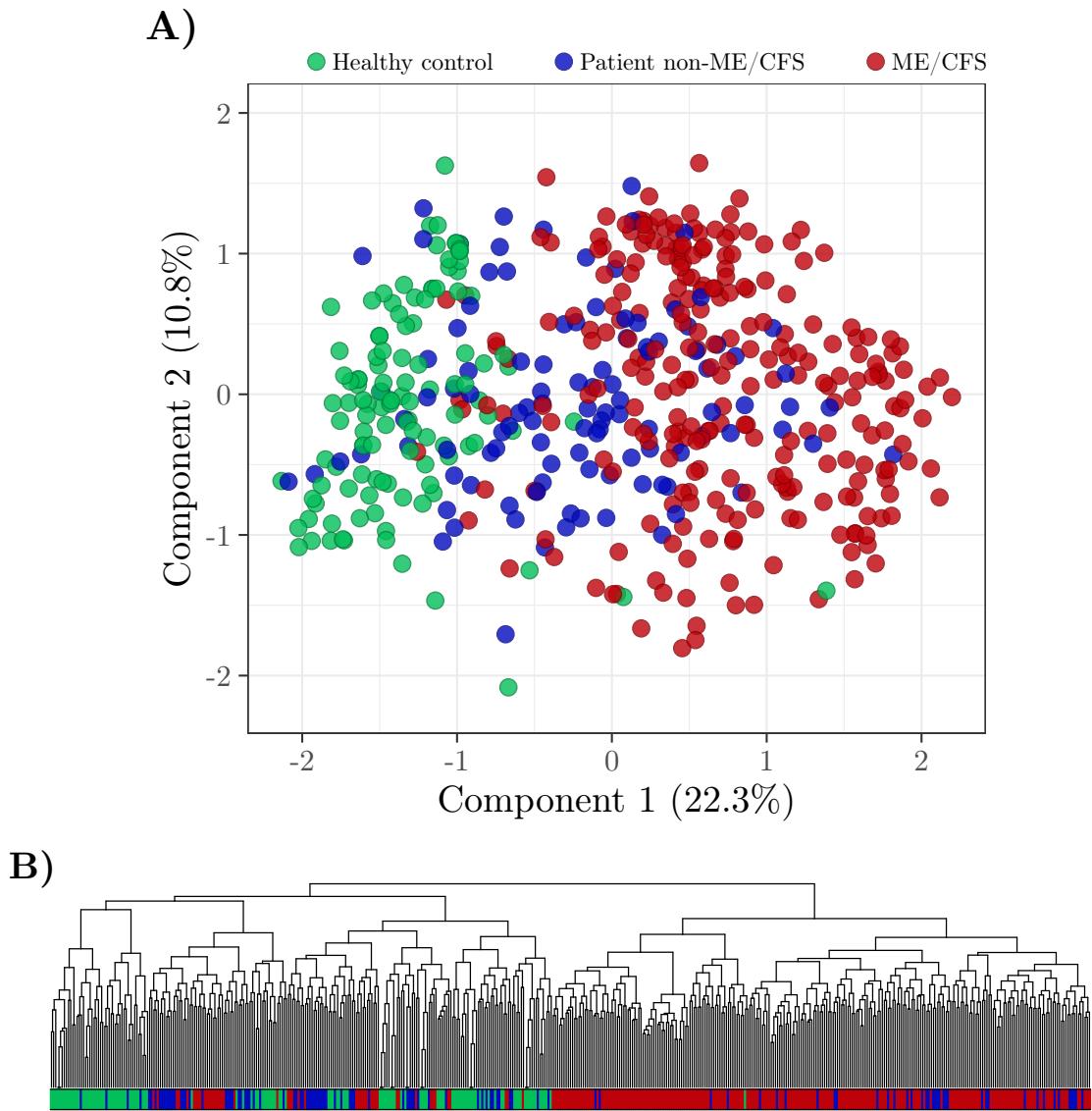


Figure 2.1: Symptom's similarity analysis based on the Cohen's κ coefficient: classical multidimensional scaling (A); dendrogram of hierarchical clustering analysis based on complete linkage (B) where the colour coding at the bottom is the same shown in (A).

With respect to the classical MDS, the first two components could explain 33.1% of the total inertia (Figure 2.1A). More importantly, the first component clearly discriminated healthy controls from suspected cases of ME/CFS. In the same component, patients with multiple sclerosis and the three fatigued non-ME/CFS cases were located between these two groups with some overlap. As expected, healthy participants were the most homogeneous cohort due to an absence or, at most, mild severity of the different symptoms. In contrast, the suspected

cases of ME/CFS consisted of a diverse group as evidenced by their wide spread in the plot. Interestingly, a few suspected cases of ME/CFS had symptom's severity profiles similar to the ones from healthy controls. In agreement with these observations, the hierarchical cluster analysis revealed that some suspected cases of ME/CFS could be placed in clusters together with healthy controls and patients with multiple sclerosis (Figure 2.1B); a detailed analysis on the optimal number of clusters will be done elsewhere. Therefore, it was reasonable to assume that some of the suspected cases of ME/CFS, although agreeing with CDC-1994 or CCC-2003, could be in fact true cases of another disease, as discussed Nacul et al. (2019).

2.5 Impact of misclassification on an association analysis

Given the possibility of patients' misclassification, we performed a small simulation study to assess the reduction of statistical power attributed to this issue in the context of an association analysis. With this purpose, we simulated data from a case-control study with the aim to investigate a hypothetical association of a binary exposure variable (exposed versus not exposed) with ME/CFS. In this scenario, the observable data could be summarised by a 2×2 frequency table whose sampling distribution was given by the following product of two Binomial distributions,

$$f(x_0, x_1 | n_0, n_1; \theta_0, \theta_1) = \prod_{i=0,1} \binom{n_i}{x_i} \theta_i^{x_i} (1 - \theta_i)^{n_i - x_i}, \quad (2.3)$$

where x_0 and x_1 are the frequencies of exposed healthy controls and suspected cases, respectively, n_0 and n_1 are the associated sample sizes, and θ_0 and θ_1 are the corresponding probabilities of exposure in healthy controls and suspected cases.

To study the impact of a potential misclassification of suspected cases on the detection of a possible association, four main assumptions were considered for the simulated data: (i) suspected cases could be divided into apparent (or false positive) cases and true positive cases of ME/CFS; (ii) the apparent cases were deemed equivalent to healthy controls in terms of

degree of exposure, i.e., the probability of exposure in these individuals was given by θ_0 ; (iii) there was an overall misclassification rate, γ , for the suspected cases; and (iv) misclassification was only dependent on the true clinical status of each suspected case. Under the assumption (ii) and the law of total probability, the probability of exposure associated with suspected cases could be written as

$$\theta_1 = \gamma\theta_0 + (1 - \gamma)\theta_1^*, \quad (2.4)$$

where θ_1^* is the probability of exposed true cases.

We then studied the power of rejecting the null hypothesis of lack of association (i.e., H_0 : odds ratio = 1) by the Pearson's χ^2 test for independence, when considering this simple misclassification scenario. Similar investigation could have been done using Fisher's exact test instead. With this purpose, we used simulation to estimate the number of times that H_0 could be rejected at a significance level of 5%.

We augmented the observable 2×2 frequency table where the suspected cases were subdivided into apparent and true positive cases (Table 2.3). In this case, we simulated data from healthy controls according to the Binomial distribution with a sample size of n_0 individuals and probability of success θ_0 . With respect to the suspected cases, we simulated data from a Multinomial distribution with a sample size of n_1 individuals and probability vector given by the probabilities shown in Table 2.3. Note that, given assumption (iv), the associated Multinomial distribution could be decomposed into the following Binomial distribution

$$n_{1,m}|n_1; \gamma \rightsquigarrow \text{Bin}(n_1, \gamma), \quad (2.5)$$

referring to how many individuals were hypothetically misclassified as true positive cases, and two Binomial distributions conditional to $n_{1,m}$

$$X_{1,F}|n_{1,m}; \theta_0 \rightsquigarrow \text{Bin}(n_{1,m}, \theta_0), \quad (2.6)$$

and

$$X_{1,T}|n_1 - n_{1,m}; \theta_1^* \rightsquigarrow \text{Bin}(n_1 - n_{1,m}, \theta_1^*), \quad (2.7)$$

where $X_{1,F}$ and $X_{1,T}$ were the random variables referring to the number of exposed false positive and true positive cases, respectively.

Table 2.3: Augmented version of the observable 2×2 frequency table and the respective probabilities under a Binomial and a Multinomial distribution for healthy controls and suspected cases, respectively.

Exposure	Healthy Controls	Suspected Cases	
		False positive cases	True positive cases
1	θ_0	$\theta_0\gamma$	$\theta_1^*(1-\gamma)$
0	$1-\theta_0$	$(1-\theta_0)\gamma$	$(1-\theta_1^*)(1-\gamma)$

For illustrative purposes, we performed our simulation study with $n_0 = n_1 = 100$, $\theta_0 = 0.25$, and $\theta_1^* = 0.35$. According to this parameter specification, the odds ratio of true positive cases versus healthy controls was 1.62, a low but reasonable value for a putative association with ME/CFS, given that there is no disease-specific biomarker. To estimate the power of rejecting H_0 , we generated 10,000 data sets for each value of γ , ranging from 0 (no misclassification) to 1 (full misclassification) with a lag of 0.01. In each data set, H_0 was rejected if the p-value of the Pearson's χ^2 test was less than 0.05. For a given parameter set, power was finally estimated as the proportion of simulated data sets in which H_0 was rejected.

As expected, the estimated power decreased with the misclassification rate γ (Figure 2.2). As a control scenario, when all suspected cases were considered to be false positives ($\gamma = 1$) and therefore the data sets were simulated from H_0 , the corresponding power was estimated at 5%, the significance level specified for the Pearson's χ^2 test. In opposition, when the suspected cases were all considered true positive cases ($\gamma = 0$), the power to detect a hypothetical association was estimated at 34%. This low power simply reflected the limited sample size to detect a weak association between exposure and the disease. In a less extreme case of misclassification, $\gamma = 10\%$ implied an estimated power of 29%, which reflected a decrease in 14.7% of the power estimated for the scenario with no misclassification.

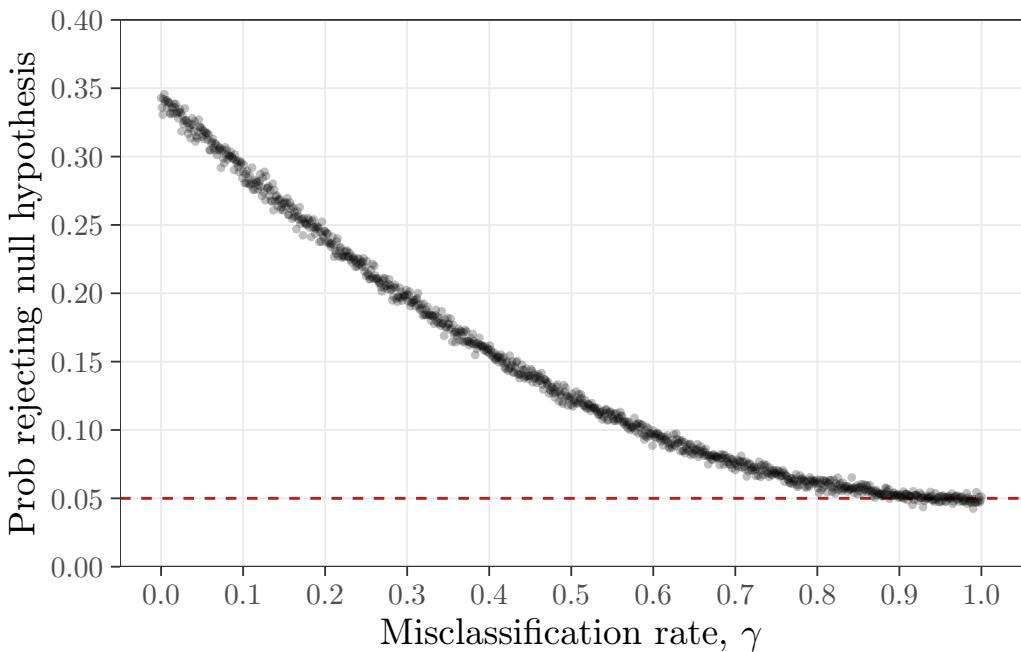


Figure 2.2: Estimated probability of rejecting H_0 (i.e., lack of association) as function of the misclassification rate γ .

2.6 Concluding remarks

In summary, our analysis showed that suspected cases of ME/CFS from the UKMEB did not fully agree with four main case definitions of the disease. In addition, some of these suspected cases showed symptom's severity profiles similar to healthy controls and patients with multiple sclerosis. These findings demonstrated the difficulty of diagnosing ME/CFS based on symptoms' assessment alone. To overcome this and other difficulties, there are currently efforts for a stronger collaboration among European researchers for accelerating the discovery of an objective disease-specific biomarker (Scheibenbogen et al. 2017). However, joint efforts for biomarker discovery are very likely to suffer from limited statistical power due to a possible misclassification of the suspected cases. A possible solution to this problem is to take into account for misclassification in the respective statistical analysis. Such a solution is also problematic because modelling misclassification leads to an eventual problem of overparameterisation. From a frequentist standpoint, overparameterization could be avoided by fixing the misclassification rate in a reasonable estimate for the sensitivity of the diagnostic test. A more elegant way of

doing so is to use Bayesian analysis where the prior information about the misclassification rate takes the form of a probability distribution. However, both frequentist and Bayesian solutions show a main hurdle for their implementation in the research of ME/CFS. Given the lack of a disease biomarker, it is unclear which reasonable value or probability distribution to choose for the sensitivity of current diagnostic tools of ME/CFS.

As a final remark, our formulation of the misclassification problem assumed that misclassification is only dependent on the true clinical status of the suspected cases. In practice, it is very likely that misclassification is dependent on the symptoms' severity profile of a given individual, or at least dependent on a given set of covariates. If so, Paulino et al. (2003) provided a Bayesian solution for modelling misclassification in this scenario. Given its technical complexity, we envision some difficulties in a wide application of this statistical solution by researchers of ME/CFS who are typically not trained in such advanced statistical methodology. To overcome this potential problem, we recommend a strong collaboration between these researchers and biostatisticians who have in principle the technical skills needed.

Bibliography

- A. Agresti and M. Kateri. Categorical Data Analysis. In M. Lovric, editor, *International Encyclopedia of Statistical Science*, pages 206–208. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_161. URL http://link.springer.com/10.1007/978-3-642-04898-2_161.
- K. G. Brurberg, M. S. Fønhus, L. Larun, S. Flottorp, and K. Malterud. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *BMJ Open*, 4(2):e003973, Feb. 2014. doi: 10.1136/bmjopen-2013-003973.
- B. M. Carruthers, A. K. Jain, K. L. D. Meirleir, D. L. Peterson, N. G. Klimas, A. M. Lerner, A. C. Bested, P. Flor-Henry, P. Joshi, A. C. P. Powles, J. A. Sherkey, and M. I. v. d. Sande. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Chronic Fatigue Syndrome*, 11(1):7–115, Jan. 2003. doi: 10.1300/j092v11n01_02.
- B. M. Carruthers, M. I. v. d. Sande, K. L. D. Meirleir, N. G. Klimas, G. Broderick, T. Mitchell, D. Staines, A. C. P. Powles, N. Speight, R. Vallings, L. Bateman, B. Baumgarten-Austrheim, D. S. Bell, N. Carlo-Stella, J. Chia, A. Darragh, D. Jo, D. Lewis, A. R. Light, S. Marshall-Gradisbik, I. Mena, J. A. Mikovits, K. Miwa, M. Murovska,

- M. L. Pall, and S. Stevens. Myalgic encephalomyelitis: International Consensus Criteria. *Journal of Internal Medicine*, 270(4):327–338, Aug. 2011. doi: 10.1111/j.1365-2796.2011.02428.x.
- K. Fukuda, S. E. Straus, I. Hickie, M. C. Sharpe, J. G. Dobbins, and A. Komaroff. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Annals of Internal Medicine*, 121(12):953, Dec. 1994. doi: 10.7326/0003-4819-121-12-199412150-00009.
- J. C. Gower and M. J. Warrens. Similarity, dissimilarity, and distance, measures of. *Wiley StatsRef: Statistics Reference Online*, pages 1–11, 2014.
- Institute of Medicine. *Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness*. The National Academies Press, Washington, DC, 2015. ISBN 978-0-309-31689-7. doi: 10.17226/19012. URL <https://www.nap.edu/catalog/19012/beyond-myalgic-encephalomyelitischronic-fatigue-syndrome-redefining-an-illness>.
- E. M. Lacerda, K. Mudie, C. C. Kingdon, J. D. Butterworth, S. O’Boyle, and L. Nacul. The UK ME/CFS Biobank: A Disease-Specific Biobank for Advancing Clinical Research Into Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Neurology*, 9:1026, 2018. ISSN 1664-2295. doi: 10.3389/fneur.2018.01026.
- G. Morris and M. Maes. Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics. *BMC Medicine*, 11(1), Sept. 2013. doi: 10.1186/1741-7015-11-205.
- L. Nacul, E. M. Lacerda, C. C. Kingdon, H. Curran, and E. W. Bowman. How have selection bias and disease misclassification undermined the validity of myalgic encephalomyelitis/chronic fatigue syndrome studies? *Journal of Health Psychology*, 24(12):1765–1769, Oct. 2019. ISSN 1461-7277. doi: 10.1177/1359105317695803.
- C. D. Paulino, P. Soares, and J. Neuhaus. Binomial Regression with Misclassification. *Biometrics. Journal of the International Biometric Society*, 59(3):670–675, Sept. 2003. doi: 10.1111/1541-0420.00077.
- D. F. H. Pheby, D. Araja, U. Berkis, E. Brenna, J. Cullinan, J.-D. d. Korwin, L. Gitto, D. A. Hughes, R. M. Hunter, D. Trepel, and X. Wang-Steverding. The Development of a Consistent Europe-Wide Approach to Investigating the Economic Impact of Myalgic Encephalomyelitis (ME/CFS): A Report from the European Network on ME/CFS (EUROMENE). *Healthcare*, 8(2):88, Apr. 2020. doi: 10.3390/healthcare8020088.
- S. Rasa, Z. Nora-Krukla, N. Henning, E. Eliassen, E. Shikova, T. Harrer, C. Scheibenbogen, M. Murovska, B. K. Prusty, and the European Network on ME/CFS (EUROMENE). Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 16(1):268, Oct. 2018. ISSN 1479-5876. doi: 10.1186/s12967-018-1644-y. URL <https://doi.org/10.1186/s12967-018-1644-y>.
- C. Scheibenbogen, H. Freitag, J. Blanco, E. Capelli, E. Lacerda, J. Authier, M. Meeus, J. C. Marrero, Z. Nora-Krukla, E. Oltra, and others. The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE. *Journal of translational medicine*, 15(1):162, 2017. doi: <https://doi.org/10.1186/s12967-017-1263-z>.

Chapter 3

Impact of misclassification and imperfect serological tests in association analyses of ME/CFS applied to Covid-19 data

J. Malato, L. Graça, and N. Sepúlveda. Impact of misclassification and imperfect serological tests in association analyses of ME/CFS applied to COVID-19 data. In R. Bispo, L. Henriques-Rodrigues, R. Alpizar-Jara, and M. de Carvalho, editors, *Recent Developments in Statistics and Data Science*, Springer Proceedings in Mathematics & Statistics, pages 215–225, Cham, 2022. Springer International Publishing. ISBN 978-3-031-12766-3. doi: 10.1007/978-3-031-12766-315.

Abstract

The diagnosis of ME/CFS is problematic due to the absence of a disease specific biomarker. As such, it is conducted under uncertainty using symptom-based criteria and the exclusion of known diseases. The possibility of misdiagnosing patients reduces the power to detect new and previously identified factors that can be associated with the disease. To investigate this problem, we previously conducted a simulation study to estimate the power of case-control association studies as a function of the misdiagnosed rate. Here we extended this simulation study to the more general situation where there is also the possibility of having misclassification in a binary factor related to a previous exposure to a given infection. Given the suggested link between ME/CFS and past viral infections including SARS-CoV-2 (that causes Covid-19), we performed the simulation study in the specific context of serological testing of this new coronavirus using published data from Portuguese, Spanish and Iranian seroepidemiological studies.

Keywords: misclassification; simulation; power studies; serology; myalgic encephalomyelitis/chronic fatigue.

3.1 Introduction

Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) is one example of a complex disease with uncertainty in its diagnosis (Nacul et al. 2017). Patients diagnosed with this debilitating disorder manifest heterogeneous symptoms such as unexplained long-lasting fatigue (Fukuda et al. 1994), post-exertional malaise that arises after slight physical, or mental effort and is not alleviated by rest (Carruthers et al. 2003), accompanied by other symptoms. Its prevalence is estimated between 0.4% and 1% depending on the population, affecting more women than men, at a 6:1 ratio (Morris and Maes 2013; Lim et al. 2020).

The aetiology of ME/CFS has been proved difficult to determine. Different reported factors such as acute infections, genetic predisposition, or environmental stressors can serve as triggers for the disease onset (Lacerda et al. 2019; Chu et al. 2019). Moreover there is no biomarker, or combination of biomarkers, that characterise this heterogeneous disease, which ultimately leave its diagnosis to be mostly done on the basis of specific symptoms and exclusion of other diseases (Smith et al. 2014). This further increases the uncertainty surrounding an objective diagnosis, which has resulted in more than 20 symptom-based criteria currently used to clinically diagnose ME/CFS (Brurberg et al. 2014). Despite proposed protocols for criteria standardisation in ME/CFS research (Pheby et al. 2020), distinct studies will inevitably define the cohort of patients differently, potentially with conflicting results (Nacul et al. 2017). This inherent level of misclassification—non-ME/CFS patients being incorrectly diagnosed as such—amongst ME/CFS cohorts has already been described in a study characterising the genome of suspected patients (Brown et al. 2021) and should be taken into account in order to minimise the negative effects on association studies (Malato et al. 2021).

Despite the pathomechanisms of ME/CFS remaining unknown, the disease has been described as having an autoimmune onset (Lorusso et al. 2009; Sotzny et al. 2018). This immune dysregulation often occurs after exposure to an acute viral infection (Rasa et al. 2018; Blomberg et al. 2018; Chu et al. 2019), with multiple association studies relating the exposure to viruses as trigger for ME/CFS development (Bansal et al. 2012; Sotzny et al. 2018). Serological surveys have thus been conducted to better understand the role of distinct viruses

in this disease. However, so far there have not been replicable confirmed associations. Possible arguments for this can be the disparate cohorts of (inherently misclassified) suspected ME/CFS patients used and other factors related to study design such as the low sample sizes used (Scheibenbogen et al. 2017) or the further stratification of patients into different subtypes (Jason et al. 2005). Additionally, the serological tests used to assess exposure/non-exposure to the viruses are based on predetermined and arbitrary cutoff values to determine seropositive individuals (Scheibenbogen et al. 2017). This important but often overlooked aspect can potentially add an additional layer to the misclassification on ME/CFS, with impacts on the studies' reproducibility (Domingues et al. 2021a).

Previously, we studied the dissimilarity between different symptom diagnosis criteria and simulated the impacts of misclassification in a single scenario of potential misdiagnosis of suspected patients (Malato et al. 2021). In the present paper we extended the proposed ideas on misclassification and studied its impact on the statistical power of serology hypothetical association studies. More recently, studies have related ME/CFS and the chronic post-viral syndrome developed after infection by the SARS-CoV-2 virus, responsible for the Covid-19 pandemic (Komaroff and Bateman 2021). Despite the need for more extensive research on this topic, studies have reported that subset of patients following Covid-19 infection can develop a chronic syndrome that fulfils ME/CFS diagnostic criteria (Kedor et al. 2021). For illustrative purposes we extrapolated on the idea that there is in fact an association between Covid-19 and ME/CFS onset, however mild ($1.25 \leq \text{odds ratio} \leq 2.0$), and simulated multiple case-control association studies with different sample sizes, using results for seroprevalence surveys from three countries: Portugal (Kislaya et al. 2021), Spain (Pollán et al. 2020), and Iran (Khalagi et al. 2021). For each serology study, we hypothesised on the impact of misclassification, also accounting for the estimated levels of sensitivity and specificity.

3.2 Simulation study

3.2.1 Mathematical formulation of the problem

Following-up on the reported ideas on misclassification (Malato et al. 2021), the goal of the proposed hypothetical study was to assess the association of a binary exposure outcome (as exposed versus non-exposed) after a serological survey for Covid-19 with ME/CFS. This was accomplished by comparing a cohort of sampled patients suspected of ME/CFS to a cohort of sampled matched healthy controls. The sampling distribution of the designed case-control study was then, the product of two Binomial distributions given by the number of sampled individuals from the two cohorts, n_0 and n_1 , respectively for healthy controls and suspected ME/CFS patients, and the probability of exposure to the virus, θ_0 and θ_1 , respectively; with x_0 and x_1 being the observed frequencies of exposed healthy controls and suspected ME/CFS (Malato et al. 2021). Altogether, the sampled populations can be summarised by a 2×2 frequency table that presents different outlines depending on the described parameters n_i and θ_i , $i = \{0, 1\}$. Testing the null hypothesis for lack of association to ME/CFS (i.e., $H_0 : \theta_0 = \theta_1$) was done through the Pearson's χ^2 test for independence. After testing, H_0 was rejected if the p-value for the Pearson's χ^2 test was less than the prespecified level of significance of 5%. Through simulation, and by repeating the inference multiple times under the same conditions, the power of the study was estimated as the overall proportion in which H_0 was rejected.

Previously (Malato et al. 2021), to account for the inherent misclassification as a diluting effect for the detection of a potential association, four assumptions were considered for the ME/CFS cohort: (i) sampled suspected ME/CFS cases can be divided into apparent (false positives) and true positive cases; (ii) the misclassified apparent cases are considered healthy controls, in the sense that they share the same probability of exposure to Covid-19, θ_0 ; (iii) there is an overall misclassification rate, γ , creating the two distinct possibilities of apparent and true cases within the cohort for suspected cases; and (iv) this misclassification rate is only dependent on the true clinical status of each of the suspected cases. Under the assumption (ii) and the law of total probability, the probability of exposure associated with the suspected

cases was written as

$$\theta_1 = \gamma\theta_0 + (1 - \gamma)\theta_1^*, \quad (3.1)$$

where θ_1^* is the exposure probability of true ME/CFS cases.

However, this analysis does not account for the sensitivity and specificity of a serology test if the exposure to a given infection is determined this way. Therefore, four additional assumptions were considered for this study, with effects transversal to all data sets: (v) for each serology test performed, individuals can only be classified as seropositive or seronegative—in opposition to serology tests where there are more than two possible outcomes; (vi) the levels of sensitivity, π_{se} , and specificity, π_{sp} , respectively determine the accuracy of a test to identify truly exposed and truly non-exposed individuals; (vii) these parameters related to the performance of the serology test create a category of undetected false positives and false negative for individuals poorly measured by the serology assessment; and (viii) the binary exposure outcomes given by π_{se} and π_{sp} are independent from the assessed cohort. Under these assumptions, the probability of exposure for suspected cases from Equation (3.1) can be extended to

$$\theta_1 = \pi_{se}\gamma\theta_0 + (1 - \pi_{sp})\gamma(1 - \theta_0) + \pi_{se}(1 - \gamma)\theta_1^* + (1 - \pi_{sp})(1 - \gamma)(1 - \theta_1^*). \quad (3.2)$$

Under the eight assumptions, the observable 2×2 frequency table can be augmented, as the cohort for suspected ME/CFS is divided into apparent and true cases based on the misclassification rate, γ , and with sensitivity and specificity, respectively π_{se} and π_{sp} , defining the serology tests' overall accuracy to determine the seropositive (either true positive or false positive) and seronegative (both true and false negative) populations on both cohorts (Table 3.1)¹.

3.2.2 Parameterisation using real-word data

As example of real-life application, we looked at data from three distinct seroepidemiologic surveys: Portugal (Kislaya et al. 2021), Spain (Pollán et al. 2020), and Iran (Khalagi et al. 2021).

¹Instead of the Pearson's χ^2 test, an analogous investigation could also been proposed using the Fisher's exact test to assess the null hypothesis for lack of association. Equation (3.2) includes parameters related to the accuracy of serology tests; based on this formulation, one can obtain Equation (3.1) by simply assuming $\pi_{se} = \pi_{sp} = 1$.

Table 3.1: Augmented version of the observable 2×2 frequency table in the case-control association study scenario with possible misclassification of suspected ME/CFS cases (into apparent and true cases) and existence of false positive and false negative serological outcomes observed from serology tests done to assess exposure (confirmed by the true exposure indicator columns, with E for exposed individuals and \bar{E} for non-exposed).

Observed test outcome	True exposure indicator	Controls	Suspected cases	
			(Apparent)	(True)
Seropositive	E	$\pi_{se}\theta_0$	$\pi_{se}\gamma\theta_0$	$\pi_{se}(1-\gamma)\theta_1^*$
	\bar{E}	$(1-\pi_{sp})(1-\theta_0)$	$(1-\pi_{sp})\gamma(1-\theta_0)$	$(1-\pi_{sp})(1-\gamma)(1-\theta_1^*)$
Seronegative	E	$(1-\pi_{se})\theta_0$	$(1-\pi_{se})\gamma\theta_0$	$(1-\pi_{se})(1-\gamma)\theta_1^*$
	\bar{E}	$\pi_{sp}(1-\theta_0)$	$\pi_{sp}\gamma(1-\theta_0)$	$\pi_{sp}(1-\gamma)(1-\theta_1^*)$

The studies occurred between April and August 2020 and applied similar methods of estimation of their populations' seroprevalence. Also, all surveys presented information regarding the sensitivity and specificity estimates for the serology tests performed. The estimated values for the mentioned parameters in each survey are presented in Table 3.2.

Table 3.2: Parameter values used in the study, where the probability of exposure to the virus and the sensitivity and specificity of the serology test are given by θ_0 , π_{se} , and π_{sp} , respectively.

Reference	Country	θ_0	π_{se}	π_{sp}
Kislaya et al. (2021)	Portugal	0.025	0.95	0.98
Pollán et al. (2020)	Spain	0.050	0.80	0.98
Khalagi et al. (2021)	Iran	0.150	0.75	0.98

For the purpose of the study, we assumed the existence of an association between exposure to Covid-19 and ME/CFS onset. Despite few evidences thus far due to the novelty of the topic, some studies have mentioned this association based on the idea of immune dysregulation, linking the development of post-Covid-19 chronic symptoms with the autoimmune proposal for ME/CFS (Sotzny et al. 2018). Since there are no biomarkers for ME/CFS diagnosis, we defined the association as a mild relation with three possible values of the overall true odds ratio, $\Delta_T = \{1.25, 1.5, 2\}$. Based on the values of θ_0 from the three surveys and the proposed Δ_T , the probability of exposure on true ME/CFS cases was determined by

$$\theta_1^* = \frac{\theta_0 \Delta_T}{1 + \theta_0 (\Delta_T - 1)} . \quad (3.3)$$

3.2.3 Simulation structure

The impact of inherent misclassification on the hypothetical case-control association studies was assessed through multiple simulations on different parametric values for θ_0 , π_{se} , π_{sp} , in accordance to each serological survey (Table 3.2), and Δ_T . For each combination of θ_0 and Δ_T , parameters θ_1 and θ_1^* were calculated from Equation (3.2) and Equation (3.3), respectively. To illustrate how sample sizes also influence the overall power of a study, we performed our simulations considering cohort sample sizes of $n_0 = n_1 = \{100, 250, 500, 1000, 2500, 5000\}$.

To assess the power of rejecting H_0 , 10,000 data sets were generated for each value of γ , ranging from 0 (no misclassification) to 1 (no true ME/CFS patients in the cohort for suspected cases) with a lag of 0.01. As previously mentioned, H_0 was rejected at each data set if the p-value from the Pearson's χ^2 test was less than the usual level of significance. Finally, for each parameter set, power was estimated as the proportion of simulated data sets in which H_0 was rejected. All simulations and analyses were done using R statistical software, version 4.1.0 (R Core Team 2020), using our own scripts, available upon request².

3.3 Simulation results

As expected, the estimated power to detect the hypothetical association decreased with misclassification rate (Figure 3.1). Looking at the extreme cases, the estimated power was highest when no misclassification was considered and all suspected ME/CFS cases were considered to be true positives ($\gamma = 0$). Irrespective of the scenario, as misclassification increases, the overall power is reduced towards 5% at the opposite most extreme value ($\gamma = 1$)—i.e., the significance level specified for the Pearson's χ^2 test.

Along with gradually increasing the misclassification of suspected patients, the power to detect an association was estimated by varying the values of probability of exposure in healthy controls, θ_0 , and sensitivity of the serology test, π_{se} , for each country serological scenario. In all three illustrated scenarios, the specificity was the same and estimated at $\pi_{sp} = 0.98$ (Table 3.2).

²For the purpose of study consistency, the estimated seroprevalence values published on the serological surveys were considered as the probability of exposure in the cohort of matched healthy controls, θ_0 .

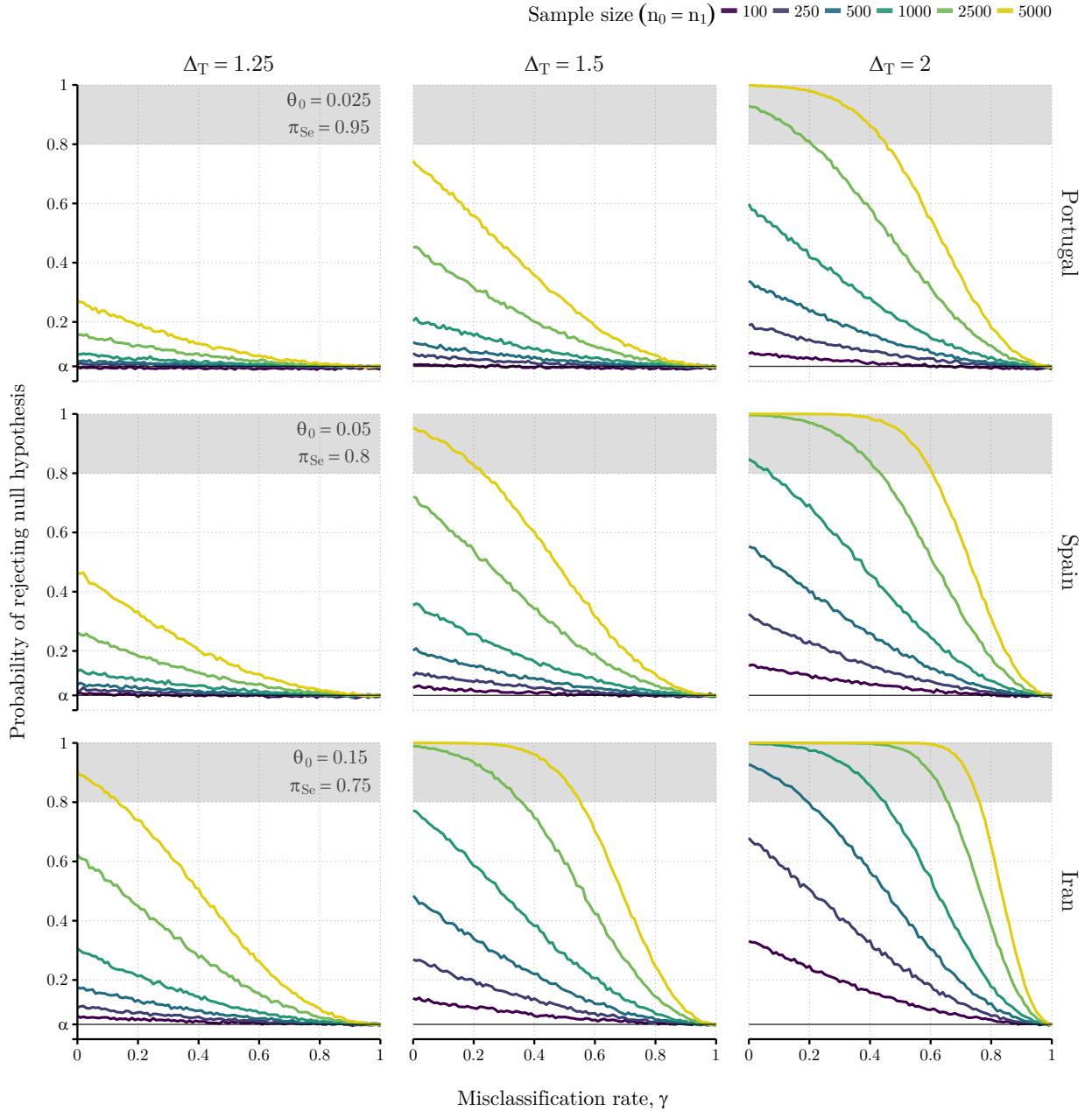


Figure 3.1: Probabilities of rejecting the null hypothesis, i.e., absence of association between the two populations as function of the misclassification rate. Each column represents the values attributed to the true odds ratio for Covid-19 exposure and true ME/CFS, assessed between true positive cases and healthy controls. Each row indicates a country serologic survey, with distinct values of θ_0 and π_{Se} identified in the first column of each survey, and fixed $\pi_{sp} = 0.98$ across all simulations. Power analysis was estimated for different cohort sample sizes of 100, 250, 500, 1000, 2500, and 5000 individuals ($n_0 = n_1$), represented by the lines of different colours in each scenario. Gray filled area indicates scenarios where the probability of rejecting the null hypothesis is above 80%. Dark horizontal line indicates the level of significance used, $\alpha = 0.05$.

Overall, only sample sizes of $n_i \geq 500$ individuals were able to reach a power of at least 80%—the specified power threshold to identify what can be considered as having acceptable reproducibility level (Table 3.3). Simulations with larger sample sizes granted a consistency to the reproducibility of the studies, with power remaining above the defined threshold for higher values of misclassification. Similarly, higher values of Δ_T also affected positively the overall power of each study (Table 3.3).

The Portuguese survey had the smallest estimate for the probability of virus exposure and the highest sensitivity (Kislaya et al. 2021). For this scenario, only higher association values of $\Delta_T = 2$ and cohort sample sizes of 2500 and 5000 reached the power of at least 80%. Under these parametric conditions, the acceptable level of reproducibility was observable at $\gamma \leq 0.45$.

Compared to Portugal's results, the Spain's survey had θ_0 increased and the π_{se} decreased (Pollán et al. 2020). In this scenario there was an increase on the reproducibility, with more studies surpassing the 80% threshold. Nonetheless, this only occurred for sample sizes of $n_i \geq 1000$.

Lastly, Iran's survey (Khalagi et al. 2021) had the highest estimate for θ_0 and lowest estimates for π_{se} . Despite the lower sensitivity, simulations under these parameters had higher power for the same sample sizes than the other two scenarios (Figure 3.1 and Table 3.3).

3.4 Discussion

Focusing on ME/CFS, our simulation results showed how misclassification of patients poses an impact on the ability to consistently recognise true associations to a triggering viral exposure, prior to the disease onset. While still researching for biomarkers able discriminate the disease, the power is very likely to suffer from limited statistical power due to possible misclassification of the suspected ME/CFS cases. The proposed solution to this problem is to take into account for misclassification in the respective statistical analysis.

The results evidenced how increasing a study's sample size can increase its power. Until now, misclassification studies mostly focused on identifying the extent of misdiagnosed of patients when using distinct diagnosis criteria, not particularly looking at sample sizes (Malato

Table 3.3: Maximum values of misclassification rate, γ , that maintain power if at least 80% to reject the null hypothesis of lack of association, for different values of true odds ratio, Δ_T , country of serological survey, and sample sizes, n_i , $i = (0, 1)$. Cells with no value indicate the inability to reach the power threshold between cohort, even at $\gamma = 0$.

Country \ Δ_T	1.25	1.50	2.00	n_i
Portugal	—	—	—	100
	—	—	—	
	—	—	—	
Spain	—	—	—	250
	—	—	—	
	—	—	—	
Iran	—	—	—	500
	—	—	—	
	—	—	0.19	
Portugal	—	—	—	1000
	—	—	0.07	
	—	—	0.45	
Spain	—	—	0.20	2500
	—	—	0.43	
	—	0.35	0.65	
Iran	—	—	0.45	5000
	—	0.22	0.60	
	0.13	0.55	0.76	

et al. 2021). With ME/CFS research being usually underfunded (Dimmock et al. 2016; Mirin et al. 2020), case-control studies are frequently performed on sample sizes below 250 patients. This allows for potential sporadic associations that ultimately cannot be replicated in follow-up studies. Throughout efforts to raise awareness and laboratory collaborations, studies have been increasing their sampled populations. After all, our study showed that under the parameterised conditions, only cohorts with samples above 500 individuals were able to consistently reject the null hypothesis under some levels of misclassification (Table 3.3).

A more in depth study would be required to pose a more general conclusion on the influence in power caused by prevalence of exposure and the sensitivity and specificity of the serology test.

One can argue that increasing the prevalence will make for better comparisons between cohorts through the Pearson's χ^2 test for independence, as it might improve the frequency distributions across the 2×2 contingency table cells. Whereas, sensitivity and specificity will produce a lessened effect, as serology tests keep improving—but still impactful, if not from the estimated π_{se} and π_{sp} , then because the majority of serological cutoff values for seropositivity used arise from inherently arbitrary choices if the researchers and manufacturers of the serology tests (Domingues et al. 2021a;b). Nevertheless, diagnostic accuracy is still of extremely importance in the evaluation of medical diagnostic tests and should be taken into account when replication of a study—in this case, a scenario of a serology study—is necessary.

This hypothetical study was done in the context of the recent Covid-19 pandemic and the association of the long-term symptoms caused by the SARS-CoV-2 virus and ME/CFS diagnosis. With the lack of extensive information on this Covid-19 exposure-ME/CFS diagnosis relation premise, parameter Δ_T was defined within low-to-mild values as to not profoundly influence the simulation results. As more studies and serological surveys are published on the matter, focusing different populations or even focusing on serology tests for different specific antibodies against Covid-19, one could better parameterise the simulation study.

ME/CFS is a complex disease and there is still lack of understanding to the extension of the disease's aetiology and pathophysiology. Even under these uncertainties, accepting and accounting for a level of patient misclassification—however small—in association studies might help to improve the study designs and increase scientific reproducibility. Ultimately, the ability to replicate and reproduce the results proposed by a study is one of the most important aspects in research, and consistent results are what allows ideas to become postulates, continuously driving science forwards.

Bibliography

- A. S. Bansal, A. S. Bradley, K. N. Bishop, S. Kiani-Alikhan, and B. Ford. Chronic fatigue syndrome, the immune system and viral infection. 26(1):24–31, Jan. 2012. doi: 10.1016/j.bbi.2011.06.016.
- J. Blomberg, C.-G. Gottfries, A. Elfaitouri, M. Rizwan, and A. Rosén. Infection Elicited Autoimmunity and Myalgic

- Encephalomyelitis/Chronic Fatigue Syndrome: An Explanatory Model. *Frontiers in Immunology*, 9, Feb. 2018. doi: 10.3389/fimmu.2018.00229.
- D. Brown, C. Birch, J. Younger, and E. Worthey. ME/CFS: whole genome sequencing uncovers a misclassified case of glycogen storage disease type 13 previously diagnosed as ME/CFS. *Molecular Genetics and Metabolism*, 132:S194–S195, Apr. 2021. doi: 10.1016/s1096-7192(21)00388-7.
- K. G. Brurberg, M. S. Fønhus, L. Larun, S. Flottorp, and K. Malterud. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *BMJ Open*, 4(2):e003973, Feb. 2014. doi: 10.1136/bmjopen-2013-003973.
- B. M. Carruthers, A. K. Jain, K. L. D. Meirleir, D. L. Peterson, N. G. Klimas, A. M. Lerner, A. C. Bested, P. Flor-Henry, P. Joshi, A. C. P. Powles, J. A. Sherkey, and M. I. v. d. Sande. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Chronic Fatigue Syndrome*, 11(1):7–115, Jan. 2003. doi: 10.1300/j092v11n01_02.
- L. Chu, I. J. Valencia, D. W. Garvert, and J. G. Montoya. Onset Patterns and Course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Pediatrics*, 7, Feb. 2019. doi: 10.3389/fped.2019.00012.
- M. E. Dimmock, A. A. Mirin, and L. A. Jason. Estimating the disease burden of ME/CFS in the United States and its relation to research funding. *Journal of Medicine and Therapeutics*, 1(1), 2016. doi: 10.15761/jmt.1000102.
- T. D. Domingues, A. D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, F. Westermeier, C. Scheibenbogen, J. M. Cliff, L. Nacul, E. M. Lacerda, H. Mourão, and N. Sepúlveda. Herpesviruses Serology Distinguishes Different Subgroups of Patients From the United Kingdom Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Biobank. *Frontiers in Medicine*, 8, July 2021a. doi: 10.3389/fmed.2021.686736.
- T. D. Domingues, H. Mourão, and N. Sepúlveda. Analysis of antibody data using Finite Mixture Models based on Scale Mixtures of Skew-Normal distributions, Mar. 2021b. URL <https://www.medrxiv.org/content/10.1101/2021.03.08.21252807v1>. ISSN: 2125-2807 Pages: 2021.03.08.21252807.
- K. Fukuda, S. E. Straus, I. Hickie, M. C. Sharpe, J. G. Dobbins, and A. Komaroff. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Annals of Internal Medicine*, 121(12):953, Dec. 1994. doi: 10.7326/0003-4819-121-12-199412150-00009.
- L. A. Jason, K. Corradi, S. Torres-Harding, R. R. Taylor, and C. King. Chronic Fatigue Syndrome: The Need for Subtypes. *Neuropsychology Review*, 15(1):29–58, Mar. 2005. doi: 10.1007/s11065-005-3588-2.
- C. Kedor, H. Freitag, L. Meyer-Arndt, K. Wittke, T. Zoller, F. Steinbeis, M. Haffke, G. Rudolf, B. Heidecker, H. D. Volk, C. Skurk, F. Paul, J. Bellmann-Strobl, and C. Scheibenbogen. Chronic COVID-19 Syndrome and Chronic Fatigue Syndrome (ME/CFS) following the first pandemic wave in Germany – a first analysis of a prospective observational study. Feb. 2021. doi: 10.1101/2021.02.06.21249256.
- K. Khalagi, S. Gharibzadeh, D. Khalili, M. A. Mansournia, S. M. Samiee, S. Aghamohamadi, M. M.-M.-A. Roodaki, S. M. Hashemi, K. Tayeri, H. N. Tabar, K. Azadmanesh, J. S. Tabrizi, K. Mohammad, F. Hajipour, S. Namaki, A. Raeisi, and A. Ostovar. Prevalence of COVID-19 in Iran: results of the first survey of the Iranian COVID-19

- Serological Surveillance programme. 27(11):1666–1671, Nov. 2021. doi: 10.1016/j.cmi.2021.06.002.
- I. Kislaya, P. Gonçalves, M. Barreto, R. D. Sousa, A. C. Garcia, R. Matos, R. Guiomar, A. P. Rodrigues, and O. b. o. I.-. Group. Seroprevalence of SARS-CoV-2 Infection in Portugal in May-July 2020: Results of the First National Serological Survey (ISNCOVID-19). 34(2):87, Feb. 2021. doi: 10.20344/amp.15122.
- A. L. Komaroff and L. Bateman. Will COVID-19 Lead to Myalgic Encephalomyelitis/Chronic Fatigue Syndrome? *Frontiers in Medicine*, 7, Jan. 2021. doi: 10.3389/fmed.2020.606824.
- E. M. Lacerda, K. Geraghty, C. C. Kingdon, L. Palla, and L. Nacul. A logistic regression analysis of risk factors in ME/CFS pathogenesis. *BMC Neurology*, 19(1), Nov. 2019. doi: 10.1186/s12883-019-1468-2.
- E.-J. Lim, Y.-C. Ahn, E.-S. Jang, S.-W. Lee, S.-H. Lee, and C.-G. Son. Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *Journal of Translational Medicine*, 18(1), Feb. 2020. doi: 10.1186/s12967-020-02269-0.
- L. Lorusso, S. V. Mikhaylova, E. Capelli, D. Ferrari, G. K. Ngonga, and G. Ricevuti. Immunological aspects of chronic fatigue syndrome. *Autoimmunity Reviews*, 8(4):287–291, Feb. 2009. doi: 10.1016/j.autrev.2008.08.003.
- J. a. Malato, L. Graça, L. Nacul, E. M. Lacerda, , and N. Sepúlveda. Statistical challenges of investigating a disease with a complex diagnosis. In P. Milheiro, A. Pacheco, B. de Sousa, I. F. Alves, I. Pereira, M. J. a. Polidoro, and S. Ramos, editors, *Estatística: Desafios transversais às ciências com dados*, Eds. Sociedade portuguesa de estatística, pages 153–167. Sociedade Portuguesa de Estatística, 2021. ISBN 978-972-8890-47-6. URL <https://www.spestatistica.pt/storage/app/uploads/public/609/28f/6d0/60928f6d08a0c016386627.pdf>.
- A. A. Mirin, M. E. Dimmock, and L. A. Jason. Research update: The relation between ME/CFS disease burden and research funding in the USA. *Work (Reading, Mass.)*, 66:277–282, 2020. ISSN 10519815, 18759270. doi: 10.3233/WOR-203173.
- G. Morris and M. Maes. Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics. *BMC Medicine*, 11(1), Sept. 2013. doi: 10.1186/1741-7015-11-205.
- L. Nacul, C. C. Kingdon, E. W. Bowman, H. Curran, and E. M. Lacerda. Differing case definitions point to the need for an accurate diagnosis of myalgic encephalomyelitis/chronic fatigue syndrome. *Fatigue : biomedicine, health & behavior*, 5(1):1–4, 2017. ISSN 2164-1846. doi: 10.1080/21641846.2017.1273863. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5730342/>.
- D. F. H. Pheby, D. Araja, U. Berkis, E. Brenna, J. Cullinan, J.-D. d. Korwin, L. Gitto, D. A. Hughes, R. M. Hunter, D. Trepel, and X. Wang-Steverding. The Development of a Consistent Europe-Wide Approach to Investigating the Economic Impact of Myalgic Encephalomyelitis (ME/CFS): A Report from the European Network on ME/CFS (EUROMENE). *Healthcare*, 8(2):88, Apr. 2020. doi: 10.3390/healthcare8020088.
- M. Pollán, B. Pérez-Gómez, R. Pastor-Barriuso, J. Oteo, M. A. Hernán, M. Pérez-Olmeda, J. L. Sanmartín,

- A. Fernández-García, I. Cruz, N. F. d. Larrea, M. Molina, F. Rodríguez-Cabrera, M. Martín, P. Merino-Amador, J. L. Paniagua, J. F. Muñoz-Montalvo, F. Blanco, R. Yotti, F. Blanco, R. G. Fernández, M. Martín, S. M. Navarro, M. Molina, J. F. Muñoz-Montalvo, M. S. Hernández, J. L. Sanmartín, M. Cuenca-Estrella, R. Yotti, J. L. Paniagua, N. F. d. Larrea, P. Fernández-Navarro, R. Pastor-Barriuso, B. Pérez-Gómez, M. Pollán, A. Avellón, G. Fedele, A. Fernández-García, J. O. Iglesias, M. T. P. Olmeda, I. Cruz, M. E. F. Martinez, F. D. Rodríguez-Cabrera, M. A. Hernán, S. P. Fernández, J. M. R. Aguirre, J. M. N. Marí, B. P. Borrás, A. B. P. Jiménez, M. Rodríguez-Iglesias, A. M. C. Gascón, M. L. L. Alcaine, I. D. Suárez, O. S. Álvarez, M. R. Pérez, M. C. Sanchís, C. J. V. Gomila, L. C. Saladrigas, A. H. Fernández, A. Oliver, E. C. Feliciano, M. N. G. Quintana, J. M. B. Fernández, M. A. H. Betancor, M. H. Febles, L. M. Martín, L.-M. L. López, T. U. Miota, I. D. B. Población, M. S. C. Pérez, M. N. V. Fernández, T. M. Enríquez, M. V. Arranz, M. D.-G. González, I. Fernández-Natal, G. M. Lobón, J. L. M. Bellido, P. Ciruela, A. M. i. Casals, M. D. Botías, M. A. M. Maeso, D. P. d. Campo, A. F. d. Castro, R. L. Ramírez, M. F. E. Retamosa, M. R. González, M. S. B. Lobeiras, A. F. Losada, A. Aguilera, G. Bou, Y. Caro, N. Marauri, L. M. S. Blanco, I. d. C. González, M. H. Pascual, R. A. Fernández, P. Merino-Amador, N. C. Castro, A. T. Lizcano, C. R. Almagro, M. S. Hernández, N. A. Elizaga, M. E. Sanz, C. E. Baquedano, A. B. Bascaran, S. I. Tamayo, L. E. Otazua, R. B. Benarroch, J. L. Flores, and A. V. d. l. Villa. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. 396(10250):535–544, Aug. 2020. doi: 10.1016/s0140-6736(20)31483-5.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.
- S. Rasa, Z. Nora-Krukle, N. Henning, E. Eliassen, E. Shikova, T. Harrer, C. Scheibenbogen, M. Murovska, B. K. Prusty, and the European Network on ME/CFS (EUROMENE). Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 16(1):268, Oct. 2018. ISSN 1479-5876. doi: 10.1186/s12967-018-1644-y. URL <https://doi.org/10.1186/s12967-018-1644-y>.
- C. Scheibenbogen, H. Freitag, J. Blanco, E. Capelli, E. Lacerda, J. Authier, M. Meeus, J. C. Marrero, Z. Nora-Krukle, E. Oltra, and others. The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE. *Journal of translational medicine*, 15(1):162, 2017. doi: <https://doi.org/10.1186/s12967-017-1263-z>.
- M. E. B. Smith, H. D. Nelson, E. Haney, M. Pappas, M. Daeges, N. Wasson, and M. McDonagh. Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Evidence report/technology assessment*, (219):1–433, Dec. 2014. ISSN 1530-4396. doi: 10.23970/ahrqepcerta219. URL <https://doi.org/10.23970/AHQEPCERTA219>.
- F. Sotzny, J. Blanco, E. Capelli, J. Castro-Marrero, S. Steiner, M. Murovska, and C. Scheibenbogen. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome – Evidence for an autoimmune disease. *Autoimmunity Reviews*, 17(6):601–609, June 2018. doi: 10.1016/j.autrev.2018.01.009.

Chapter 4

Impact of misdiagnosis in case-control studies of ME/CFS

J. Malato, L. Graça, and N. Sepúlveda. Impact of misdiagnosis in case-control studies of myalgic encephalomyelitis/chronic fatigue syndrome. *Diagnostics*. 2023; 13(3):531. doi: <https://doi.org/10.3390/diagnostics13030531>.

Abstract

Misdiagnosis of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) can occur when different case definitions are used by clinicians (relative misdiagnosis) or when failing the genuine diagnosis of another disease (misdiagnosis in a strict sense). This problem translates to a recurrent difficulty in reproducing research findings. To tackle this problem, we simulated data from case-control studies under misdiagnosis in a strict sense. We then estimated the power to detect a genuine association between a potential causal factor and ME/CFS. A minimum power of 80% was obtained for studies with more than 500 individuals per study group. When the simulation study was extended to the situation where the potential causal factor could not be determined perfectly (e.g., seropositive/seronegative in serological association studies), the minimum power of 80% could only be achieved in studies with more than 1000 individuals per group. In conclusion, current ME/CFS studies have suboptimal power under the assumption of misdiagnosis. This power can be improved by increasing the overall sample size using multi-centric studies, reporting the excluded illnesses and their exclusion criteria, or focusing on a homogeneous cohort of ME/CFS patients with a specific pathological mechanism where the chance of misdiagnosis is reduced.

Keywords: misdiagnosis; misclassification; association studies; simulation; statistical power; ME/CFS

4.1 Introduction

Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) is a heterogeneous disease whose hallmark symptom is unexplained persistent fatigue (Fukuda et al. 1994) or post-exertional malaise upon minimal physical or mental effort (Carruthers et al. 2003). Disease heterogeneity derives from the coexistence of multiple pathological mechanisms in the same patient. Examples of these mechanisms are leaky gut (König et al. 2022), the presence of deleterious autoantibodies (Wirth and Scheibenbogen 2020), oxidative stress (Castro-Marrero et al. 2013; Wood et al. 2021), persisting viral infections (Rasa et al. 2018; Sepúlveda et al. 2019), and severe longstanding stress (Rivera et al. 2019). Unsurprisingly, research efforts to find a biomarker for disease diagnosis have failed over the years.

Current diagnosis of ME/CFS is performed via multiple polythetic disease definitions where some but not all core symptoms should be present in a suspected case (Smith et al. 2014). A differential diagnosis should also be made by excluding known diseases that could explain fatigue and other major symptoms (e.g., multiple sclerosis and diabetes). Given the multiplicity of existing disease definitions, it is possible to diagnose a suspected case of ME/CFS by a consensual case definition but not by an alternative one (Malato et al. 2021). This situation is here referred to as relative misdiagnosis because it is only admissible when considering the outcome of a given case definition relative to the one from another case definition. This type of misdiagnosis is typically present when comparing or combining data from studies using different case definitions. Given that consensual definitions for ME/CFS are both difficult to find and suboptimal to patient/control discrimination (Jason et al. 2014), some efforts were made to investigate empirical approaches to ME/CFS diagnosis (Reeves et al. 2005; Jason et al. 2015; Conroy et al. 2023).

Misdiagnosis, in a strict sense, arises from the situation where ME/CFS-diagnosed individuals, irrespective of the case definition, are genuine patients of another disease. This has been illustrated in a patient initially diagnosed with ME/CFS but was found to have a rare autosomal adult-onset disorder (Brown et al. 2021). This misdiagnosis can result from random fluctuations in the natural, pathological process of the exclusionary disease (e.g., low-graded

remitting/relapsing multiple sclerosis). It can also emerge from limited resources to run the battery of tests necessary to exclude all known diseases that could explain fatigue; for example, not performing whole-genome sequencing to exclude rare genetic diseases. There is also ambiguity around the exclusionary criteria themselves, which leaves clinicians unsure of what illnesses should be actually excluded (Jason et al. 2023). Therefore, this type of misdiagnosis seems inevitably present in ME/CFS studies (Nacul et al. 2019).

In this paper, we performed a simulation study to determine the statistical power of detecting associations with ME/CFS under misdiagnosis in a strict sense; relative misdiagnosis is beyond the scope of this paper because it is more related to a discussion about the different case definitions, as made elsewhere (Brurberg et al. 2014; Lim et al. 2020). We also investigated the impact of imperfect sensitivity/specificity for the presence of a given antibody that could be causing ME/CFS. Finally, we extended our analysis to discuss the statistical power of two published studies (Cliff et al. 2019; Steiner et al. 2020).

4.2 Statistical methodology

4.2.1 Formulation of the problem

Let us assume a typical case-control study in which diagnosed ME/CFS patients and healthy controls were matched for possible confounding factors, such as age, gender, and body mass index. The main objective of this study is to investigate the association between a candidate causal factor (e.g., a genetic factor or the occurrence of a given infection) and ME/CFS. For simplicity, let us assume that this factor has only two possible values, present and absent; the probabilities for that factor being present in healthy controls and suspected cases are represented by θ_0 and θ_1 , respectively. In general, the respective data are given by a two-way contingency table (Table 4.1).

Statistically speaking, we aim to investigate the evidence for an association between

Table 4.1: Two-way contingency table of a typical case-control study where θ_0 and θ_1 are the probabilities of the candidate causal factor being present in healthy controls and ME/CFS-diagnosed cases, respectively.

Causal factor	Controls	ME/CFS-diagnosed cases
Present	θ_0	θ_1
Absent	$1 - \theta_0$	$1 - \theta_1$

ME/CFS and the causal factor. This is translated to the following hypotheses

$$H_0 : \theta_0 = \theta_1 \text{ versus } H_1 : \theta_0 \neq \theta_1 .$$

One can then use the classical Pearson's χ^2 test, where p-values < 0.05 indicate a significant association at the 5% significance level.

In this scenario, our objective is to study the impact of misdiagnosis on the power of the Pearson's χ^2 test to detect an association with the disease. With this objective, we considered seven simplifying assumptions:

- I. ME/CFS-diagnosed cases are a mix of apparent and genuine patients of the disease;
- II. The causal factor is only associated with genuine ME/CFS patients;
- III. Apparent cases are similar to healthy controls as far as the association with the causal factor is concerned;
- IV. The chance of an ME/CFS misdiagnosis is only dependent on the true clinical status of the cases and not on the confounding factors;
- V. The true association is independent of disease duration and disease triggers, among other factors occurring during the disease course;
- VI. Healthy controls were not misdiagnosed as such;
- VII. The value of the candidate causal factor can be determined perfectly in each individual.

The first assumption is simply the invocation of misdiagnosis in a strict sense (i.e., they are actually patients of another disease). The second assumption determines that there is a true association between the causal factor and ME/CFS. In the third assumption, we determine

that the apparent ME/CFS cases share with healthy controls the same probability of the causal factor being present, θ_0 . The fourth and fifth assumptions simplify the determination of what a misdiagnosed case can be, linking it exclusively to the true/apparent category, thus, rejecting other potential disease-related factors that may influence the disease association. The sixth assumption aims at excluding the situation in which healthy controls could include undiagnosed genuine ME/CFS patients.

Note that the above assumptions are for mathematical convenience and represent the minimal set of conditions that enable the derivation of simple formulas for the probability of the causal factor being present in putative cases. As a consequence, the data simulation procedure is simplified. Additional assumptions can be invoked, but they would lead to a more-complex data simulation procedure. This is the case of also assuming that genuine cases are divided into several sub-types with different degrees of association with the causal factor. This situation, although more realistic, is beyond the scope of this paper due to its higher modelling complexity. On the other hand, the apparently different assumption in which misdiagnosis does not depend on the clinic and the clinicians who performed the diagnoses falls under the umbrella of the fourth assumption, where putative confounding factors would be given by a confounding factor referring to the participating clinics if applicable and another one referring to the clinicians.

Based on the above assumptions, the probability of the causal factor being present in ME/CFS-diagnosed cases can be expressed as follows

$$\theta_1 = \gamma\theta_0 + (1 - \gamma)\theta_1^*, \quad (4.1)$$

where γ is the probability of misdiagnosing an apparent case as a genuine one, and θ_1^* is the probability of the candidate causal factor being present in genuine ME/CFS cases. If misdiagnosis could be an observable outcome, the above 2×2 contingency table could be augmented as shown in Supplementary Table A.1.

A more complex situation emerges from the previous scenario where the candidate causal factor cannot be determined perfectly in each individual. As a consequence, there is the pos-

sibility of having misdiagnosis together with misclassification of the causal factor. This is particularly relevant to serological studies that aim at investigating whether the presence of specific antibodies is associated with ME/CFS (Ruiz-Pablos et al. 2021) or whether these antibodies can be used for disease diagnosis (Sepúlveda et al. 2022). Note that the serological evaluation of a suspected case is not mandatory by consensual definitions of ME/CFS (Fukuda et al. 1994).

To model this new situation, the above assumption VII is replaced with two additional assumptions:

- VII. There are only two possible serological outcomes for each individual: seronegative or seropositive;
- VIII. The sensitivity and specificity of the serological classification are identical for all of the individuals.

The revised assumption VII excludes the situation where the serological classification can contemplate an indeterminate status due to the laboratory protocol (Cliff et al. 2019) or the presence of multiple serological populations (Sepúlveda et al. 2015). Similarly to assumption V for misdiagnosis, the new assumption VIII intends to disregard the effect of confounders (i.e., age or gender) and disease-related factors (i.e., disease duration or disease severity) on the performance of the serological classification.

Under the validity of assumptions I–VIII, the probability of the candidate causal factor being present in a ME/CFS-diagnosed patient can be extended to

$$\theta_1 = \pi_{se}\gamma\theta_0 + (1 - \pi_{sp})\gamma(1 - \theta_0) + \pi_{se}(1 - \gamma)\theta_1^* + (1 - \pi_{sp})(1 - \gamma)(1 - \theta_1^*) , \quad (4.2)$$

where π_{se} and π_{sp} are the sensitivity and specificity for the serological classification, respectively; see Supplementary Table A.2 for details. Note that when $\pi_{se} = \pi_{sp} = 1$ (perfect serological testing), the above formula converts to Equation (4.1).

4.2.2 Simulation study

To investigate the impact of the above misdiagnosis scenarios, we performed a comprehensive simulation study using the R statistical software, version 4.1.0 (R Core Team 2020). Individuals from each group were selected in accordance with the study's sampling distribution, as shown in Supplementary Equation (A.1). We assumed the same sample size for ME/CFS patients and healthy controls (i.e., $n_0 = n_1$, with n_0 and n_1 being the sample sizes for healthy controls and ME/CFS-diagnosed patients in each simulated scenario, respectively). We considered the following sample sizes per study group: 100, 250, 500, 1000, 2500, and 5000.

To parameterise the simulation study, we first specified the association between the candidate causal factor and genuine ME/CFS patients by the odds ratio (hereafter denoted as Δ_T) and the probability of the presence of the causal candidate factor in healthy controls and apparent ME/CFS cases (θ_0). We considered the true association (i.e., Δ_T) between genuine ME/CFS cases and the causal factor to vary from weak to strong values (i.e., $\Delta_T \in \{1.25, 1.5, 2, 3, 5, 10\}$). We also specified $\theta_0 \in \{0.05, 0.1, 0.25, 0.5\}$. If data comes from a genetic association study, θ_0 could represent the minor allele frequency of a given single nucleotide variant in the healthy population. Note that, having θ_0 and Δ_T fixed in the respective values, the value of θ_1^* can be estimated, as shown in Supplementary Equation (A.2). The misdiagnosis probability (or rate) γ was varied from 0 to 1 (all diagnosed individuals are genuine and apparent ME/CFS cases, respectively) with a lag of 0.01.

To simulate data from the second misdiagnosis scenario, we considered fixed parameters $\Delta_T = 3$ and $\theta_0 = 0.25$. For parameters π_{se} and π_{sp} , we considered all possible combinations of 0.80, 0.90, 0.925, 0.975, and 1.0, where $\pi_{se} = \pi_{sp} = 1$ corresponded to the first scenario.

For each misdiagnosis scenario, parameter set, and sample size, we simulated 10,000 data sets to estimate the power of detecting an association under the presence of misdiagnosis. A detailed description of the simulation procedure can be found elsewhere (Malato et al. 2021; 2022). In each data set, we rejected the presence of association if the p-value of Pearson's χ^2 test was greater than the usual 5% level of significance. For each parameter combination, the power $(1 - \beta)$ was estimated by the proportion of the simulated data sets in which an

association was detected. To facilitate the understanding of the simulation results, we specified a target power of at least 80%.

4.2.3 Application to two ME/CFS studies

We also studied the impact of misdiagnosis on published data from a candidate gene association study and an immunological evaluation study. The first study recruited 201 healthy controls and 305 ME/CFS patients whose symptoms complied with the Canadian Consensus Criteria (Steiner et al. 2020). Five single-nucleotide polymorphisms (SNPs) were evaluated in all participants. The study found significant associations of rs2476601 and rs3087243 with ME/CFS whose onset was triggered by an acute infection.

The second study refers to serological data on 251 ME/CFS patients and 107 healthy controls from the UK ME/CFS Biobank (Cliff et al. 2019). These serological data referred to antibody positivity to each of six different herpesviruses: human cytomegalovirus (CMV), Epstein-Barr virus (EBV), herpes simplex virus 1 and 2 (HSV1 and HSV2), varicella-zoster virus (VZV), and human herpesvirus (HHV6). Antibody positivity per herpesvirus was previously determined by different lab protocols that did not provide any information about the specificity and sensitivity of the resulting serological classification.

In both studies, we estimated the power of detecting an association as a function of misdiagnosis probability, γ , using simulated data generated from the reported associations, as explained later.

4.3 Results

4.3.1 Simulation study: impact of ME/CFS misdiagnosis

The power to detect an association with ME/CFS decreased with the misdiagnosis probability (Figure 4.1 and Figure 4.2). The maximum power was achieved when the diagnosed individuals were all genuine ME/CFS cases ($\gamma = 0$). When the diagnosed individuals were all apparent ME/CFS cases ($\gamma = 1$), the corresponding power matched the 5% significance level. This result

was a direct consequence of assumption III, in which the misdiagnosed cases were considered identical to healthy controls as far as the association with the candidate causal factor was concerned.

As expected, the most optimistic scenarios were associated with $\Delta_T = 5$ or 10 (i.e., strong associations between the candidate causal factor and ME/CFS). In these scenarios, one could find a maximum misdiagnosis probability for which the power of 80% was achieved (Table 4.2). For $\Delta_T = 10$, a misdiagnosis probability of 0.53 was sufficient to ensure the desired power for sample sizes greater than or equal to 100 individuals per study group ($n_1 \geq 100$), irrespective of θ_0 . This minimum probability was reduced to 0.24 for $\Delta_T = 5$.

Similar optimistic scenarios were observed for sample sizes of 2500 and 5000 individuals per study group with the exception of the case of lowest $\Delta_T = 1.25$. Combining these large sample sizes with strong associations between the candidate causal factor and the true ME/CFS cases, failing to achieve the target power only occurred when almost all the cases were misdiagnosed (with misdiagnosis probability greater than or equal to 0.88).

Unsurprisingly, the most pessimistic situations were related to $\Delta_T = 1.25, 1.5$, $n_0 = n_1 = 100$, or a combination of the two. When $\Delta_T = 1.25$, the sample size had to increase to 2500 or 5000 individuals per group in order to achieve the target power. Therefore, for this weak association, the chance of finding reproducible results was very low, even under the assumption of a perfect diagnosis. As a consequence, testing the “common disease, common variant hypothesis” in ME/CFS is likely to fail in future genetic associations. Finally, the case of $n_0 = n_1 = 100$ was particularly problematic given that it was not possible to find any value misdiagnosis probability in which the desired power could be achieved for $\Delta_T \leq 2$ (Figure 4.1).

4.3.2 Simulation study: impact of ME/CFS misdiagnosis and misclassification on the candidate causal factor

We then simulated the data of a hypothetical association study in which there were both imperfect diagnoses and misclassification of the candidate causal factor (Figure 4.2). This situation underpins any serological association study in ME/CFS, given the estimation of

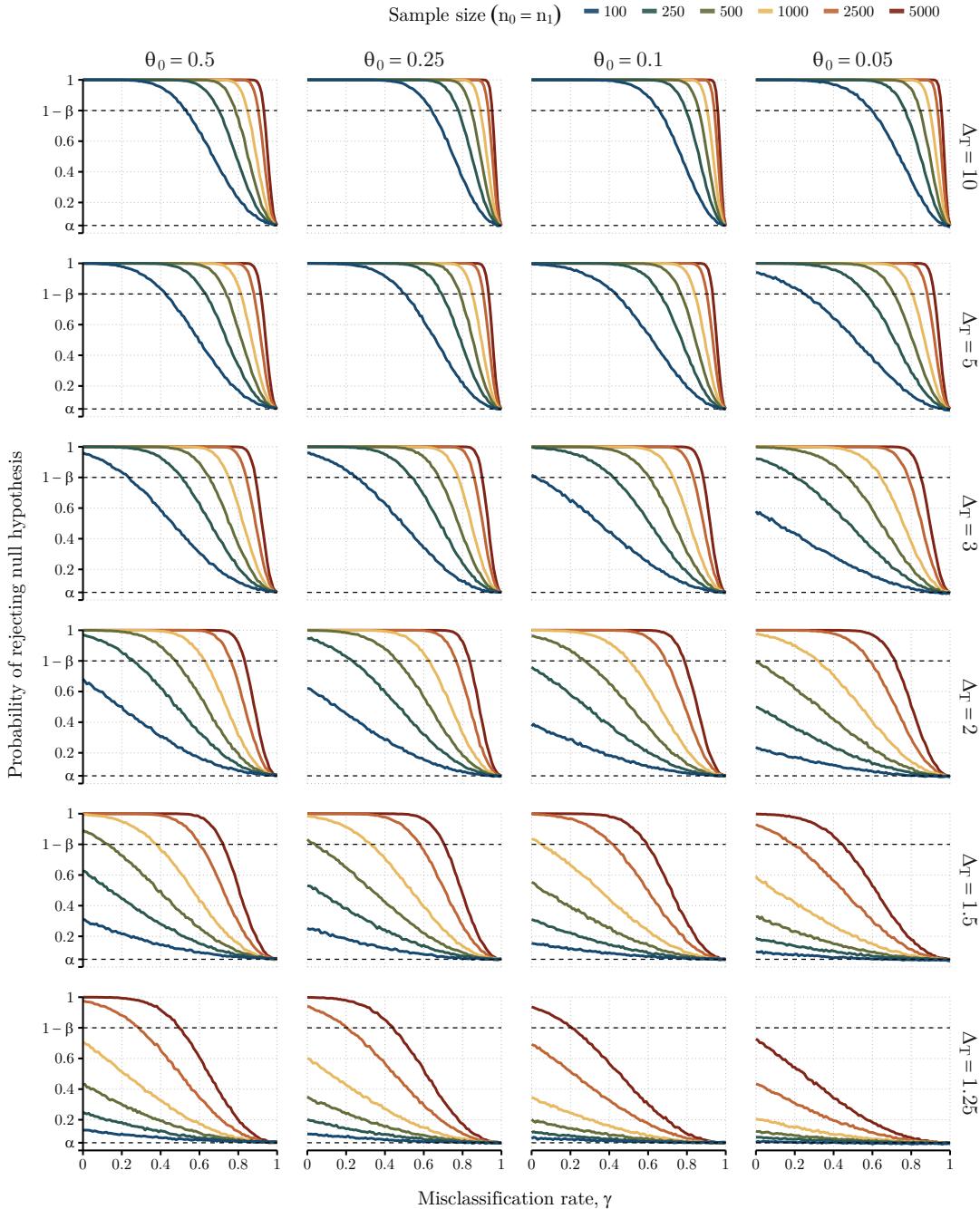


Figure 4.1: Probabilities of detecting an association (i.e., rejecting H_0) as a function of the misdiagnosis rate. Each column represents the values attributed to the risk allele frequency found in matched healthy controls and false positive ME/CFS cases ($\theta_0 \in \{0.05, 0.1, 0.25, 0.5\}$). Each row varies the true odds ratio for the association between risk allele frequency assessed between true positive cases and healthy controls ($\Delta_T \in \{1.25, 1.5, 2, 3, 5, 10\}$). Power was estimated for different sample sizes of 100, 250, 500, 1000, 2500, and 5000 ($n_0 = n_1$), represented by lines with different colours in each scenario. The upper dashed line indicates the target power of 80% (i.e., $1 - \beta = 0.80$). The lower dashed line indicates the 5% significance level.

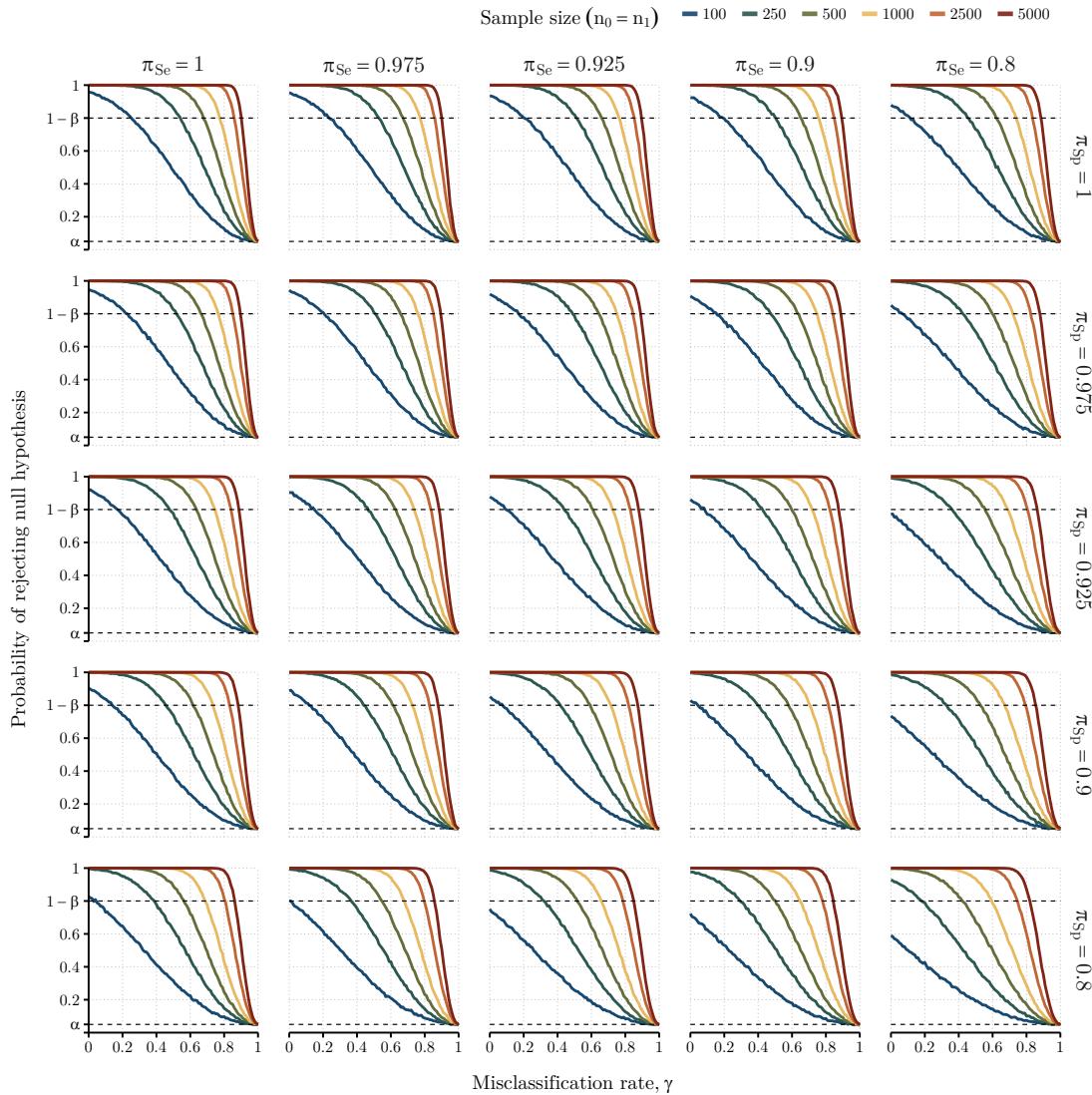


Figure 4.2: Probabilities of detecting an association (i.e., rejecting H_0) as a function of the misdiagnosis rate. Each scenario represents simulated results with a different combination of sensitivity (π_{se}) and specificity (π_{sp}) for the serological test for columns and rows, respectively. Power was estimated for different sample sizes of 100, 250, 500, 1000, 2500, and 5000 ($n_0 = n_1$), represented by lines with different colours in each scenario, with the probability of exposure in healthy controls fixed as $\theta_0 = 0.25$ and true odds ratio $\Delta_T = 3$. The upper dashed line indicates the target power of 80% (i.e., $1 - \beta = 0.80$). The lower dashed line indicates the 5% significance level.

Table 4.2: Maximum values of misdiagnosis probability γ that ensure the minimum power of 80% to detect a genuine association Δ_T as a function of θ_0 and sample size n per group. Cells with no value indicate that the minimum power could not be reached in the respective parameter combination.

$\Delta_T \backslash \theta_0$	0.05	0.1	0.25	0.5	n_i
10	0.59	0.65	0.64	0.53	100
5	0.24	0.43	0.50	0.42	
3	—	0.02	0.25	0.23	
2	—	—	—	—	
1.5	—	—	—	—	
1.25	—	—	—	—	
10	0.77	0.79	0.77	0.70	250
5	0.56	0.66	0.69	0.63	
3	0.20	0.41	0.53	0.50	
2	—	—	0.23	0.26	
1.5	—	—	—	—	
1.25	—	—	—	—	
10	0.84	0.86	0.84	0.78	500
5	0.70	0.76	0.78	0.73	
3	0.47	0.60	0.67	0.65	
2	—	0.27	0.46	0.47	
1.5	—	—	0.04	0.13	
1.25	—	—	—	—	
10	0.89	0.90	0.89	0.84	1000
5	0.80	0.84	0.85	0.81	
3	0.64	0.72	0.77	0.75	
2	0.32	0.50	0.62	0.62	
1.5	—	0.05	0.32	0.38	
1.25	—	—	—	—	
10	0.93	0.94	0.93	0.90	2500
5	0.88	0.90	0.90	0.88	
3	0.78	0.83	0.85	0.84	
2	0.58	0.69	0.76	0.76	
1.5	0.18	0.42	0.58	0.59	
1.25	—	—	0.20	0.28	
10	0.95	0.95	0.95	0.93	5000
5	0.91	0.93	0.93	0.91	
3	0.84	0.88	0.90	0.88	
2	0.71	0.78	0.83	0.83	
1.5	0.44	0.59	0.70	0.72	
1.25	—	0.20	0.44	0.49	

seropositivity of all individuals could be affected by the sensitivity and specificity associated with the classification rule used. At this point, it was clear that for values of $\Delta_T = 1.25, 1.5$, and 2 , the desired power was not often achieved for sample sizes smaller than 500 individuals per group in the case of perfect classification of the causal factor. Therefore, the additional assumption of imperfect classification of the candidate causal factor would make the previously estimated power even worse. Because of that, we only performed our simulation study on the more optimistic scenario in which $\Delta_T = 3$ (Table 4.3).

4.3.3 Application to data from two ME/CFS studies

We illustrated the problem of misdiagnosis in data from two ME/CFS studies (Cliff et al. 2019; Steiner et al. 2020). We started with data from a candidate gene association study (Steiner et al. 2020). In this study, some genetic associations were only found to be significant when comparing healthy controls to ME/CFS patients with an infectious disease trigger onset (Table 4.4). The estimated allele-related odds ratios varied from 0.84 ($95\% \text{ CI} = [0.56, 1.27]$) (rs1799724, TNF) to 1.63 ($95\% \text{ CI} = [1.04, 2.55]$) (rs2476601, PTPN22). In our re-analysis, we investigated the impact of misdiagnosis if a replication study were conducted in a similar population. In line with the original study, no genotyping errors were assumed for the genetic data. The reported odds ratios were assumed to be the true ones for the population, and data were simulated with the same allele frequencies as reported in the original study.

Again, the estimated probability of detecting an association decreased with the misdiagnosis probability (Figure 4.3A). More importantly, when the misdiagnosis probability was low ($\gamma < 0.09$), it was possible to achieve the minimum power of 80% for the allele association reported for rs3087243 in *CTLA4*. Therefore, the target power cannot be ensured for $\gamma > 0.09$. For the remaining SNPs, the target power was never achieved, irrespective of the misdiagnosis probability. This is particularly problematic for rs2476601 in *PTPN22* whose association was reported to be significant at the 5% significance level. For this SNP, the misdiagnosis probability of approximately 0.10 had an estimated power of about 50% . This result implies that the chance of replicating the reported association was no better than flipping a coin.

Table 4.3: Maximum values of misdiagnosis probability γ that still ensures a power of rejecting the null hypotheses of at least 80% for $\Delta_T = 3$ and $\theta_0 = 0.25$, where π_{se} and π_{sp} represent sensitivity and specificity associated with the classification of the candidate, respectively. See Table 4.2 for more information.

π_{sp}	π_{se}	1	0.975	0.925	0.9	0.8	n
1	0.25	0.25	0.23	0.20	0.19	0.11	100
	0.975	0.22	0.20	0.17	0.15	0.06	
	0.925	0.17	0.14	0.09	0.08	—	
	0.9	0.13	0.11	0.07	0.04	—	
	0.8	0.03	—	—	—	—	
0.975	0.53	0.52	0.51	0.50	0.45	—	250
	0.925	0.51	0.50	0.48	0.47	0.42	
	0.9	0.47	0.46	0.43	0.42	0.36	
	0.8	0.45	0.43	0.41	0.39	0.32	
	0.8	0.38	0.36	0.31	0.29	0.18	
0.925	0.67	0.67	0.66	0.65	0.62	—	500
	0.975	0.66	0.65	0.64	0.63	0.59	
	0.9	0.63	0.62	0.60	0.59	0.55	
	0.8	0.61	0.61	0.59	0.57	0.52	
	0.8	0.56	0.54	0.51	0.50	0.42	
0.9	0.77	0.77	0.76	0.75	0.73	—	1000
	0.975	0.76	0.75	0.74	0.74	0.72	
	0.925	0.74	0.73	0.72	0.71	0.68	
	0.8	0.73	0.72	0.71	0.70	0.67	
	0.8	0.68	0.67	0.65	0.64	0.59	
0.8	0.85	0.85	0.85	0.84	0.83	—	2500
	0.975	0.85	0.85	0.84	0.84	0.82	
	0.925	0.84	0.83	0.82	0.82	0.80	
	0.9	0.83	0.82	0.81	0.81	0.79	
	0.8	0.80	0.79	0.78	0.78	0.74	
0.75	0.90	0.90	0.89	0.89	0.88	—	5000
	0.975	0.89	0.89	0.89	0.88	0.87	
	0.925	0.88	0.88	0.87	0.87	0.86	
	0.9	0.88	0.87	0.87	0.87	0.85	
	0.8	0.86	0.85	0.84	0.84	0.81	

Table 4.4: Reported associations of a candidate gene association study (Steiner et al. 2020) where $\hat{\theta}_0$ represents the frequencies of the non-reference allele for healthy controls and $\hat{\Delta}_T$ is the odds ratio of these allele frequencies when comparing ME/CFS patients with an infectious disease trigger to healthy controls. P-values are associated with the Pearson's χ^2 test for 2×2 contingency tables.

SNP	Gene	$\hat{\theta}_0$	$\hat{\Delta}_T$	CI ($\hat{\Delta}_T$)	p-value
rs3087243	<i>CTLA4</i>	0.56	1.54	(1.17, 2.03)	0.002
rs2476601	<i>PTPN22</i>	0.08	1.63	(1.04, 2.55)	0.033
rs1799724	<i>TNF</i>	0.13	0.84	(0.56, 1.27)	0.409
rs1800629	<i>TNF</i>	0.16	0.89	(0.61, 1.30)	0.551
rs3807306	<i>IRF5</i>	0.51	0.94	(0.72, 1.22)	0.637

The second study referred to putative associations of six herpes virus infections with ME/CFS using antibody positivity data (Cliff et al. 2019). In these data, all individuals were classified as seronegative or seropositive for each antibody used. Under the assumption of perfect serological classification and diagnosis, the associations of these serological data with severely affected ME/CFS patients ranged from 0.65 (95% CI = [0.21, 1.97]) to 1.60 (95% CI = [0.83, 3.09]) for EBV and HSV1, respectively (Table 4.5). In this study, no association was deemed significant at the usual significance level of 5%, according to the original study (p-values ≥ 0.16).

Table 4.5: Summary of serological findings from Cliff et al. (2019), where $\hat{\theta}_0$ represents the seroprevalence of healthy controls and $\hat{\Delta}_T$ refers to the odds ratio for being seropositive when comparing severely-affected ME/CFS patients with healthy controls. The 95% CI ($\hat{\Delta}_T$) p-values are associated with the Pearson's χ^2 test for 2×2 contingency tables.

Herpes virus	$\hat{\theta}_0$	$\hat{\Delta}_T$	95% CI ($\hat{\Delta}_T$)	p-value
HSV1	0.42	1.60	(0.83, 3.09)	0.163
HSV2	0.34	1.36	(0.69, 2.66)	0.377
EBV	0.93	0.65	(0.21, 1.97)	0.442
CMV	0.37	0.84	(0.42, 1.67)	0.613
VZV	0.97	0.75	(0.12, 4.63)	0.757
HHV6	0.95	1.27	(0.24, 6.79)	0.776

The original serological classification was based on a cut-off in the antibody levels determined by the 2σ -rule; the cut-off is the mean plus twice the standard deviation of a known or hypothetical seronegative population. Under the assumption of a normal distribution for the

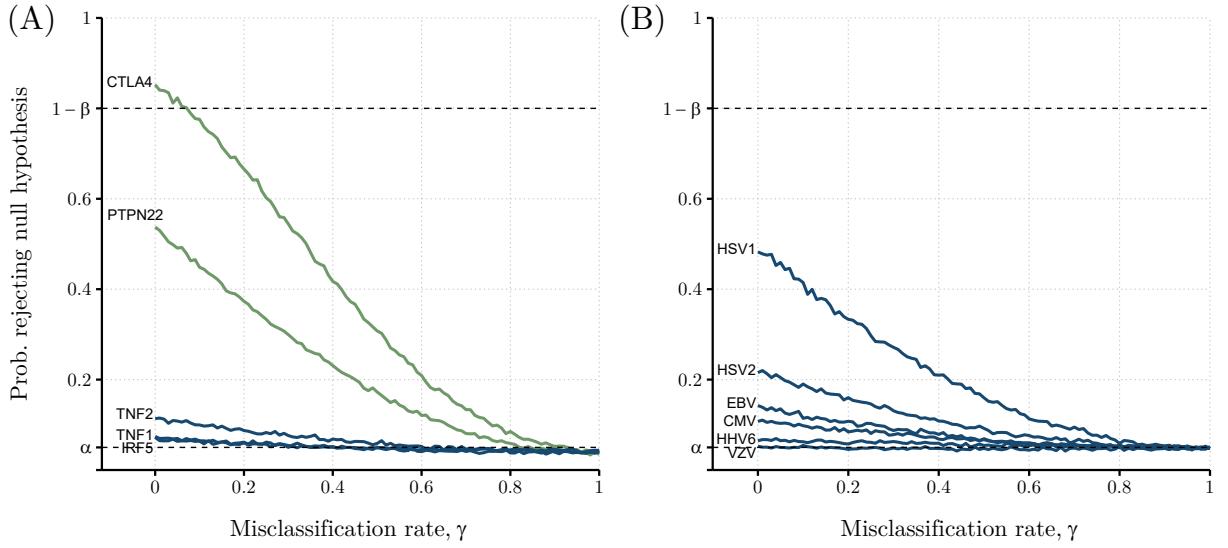


Figure 4.3: The relationship between the misdiagnosis probability (or rate) and the probability of detecting an association (i.e., rejecting the H_0) estimated from simulated data from two previously published studies: (A) Data from five different SNPs (genes *PTPN22*, *CTLA4*, *TNF* (*TNF1*-rs1799724 and *TNF2*-rs1800629), and *IRF5*); (B) Data of antibody positivity related to six human herpesviruses (CMV, EBV, HSV1 and HSV2, VZV, and HHV6). For each study, risk allele frequencies or the probability of exposure and true odds ratio were determined by Steiner et al. (2020) ($n_0 = 201$; $n_1 = 305$) and Cliff et al. (2019) ($n_0 = 107$; $n_1 = 251$; $\pi_{se} = \pi_{sp} = 0.975$), with determined values shown in Table 4.4 and Table 4.5, respectively. Green lines indicate candidate risk factors where a significant association with the disease was found in the original study. Blue lines show non-significant ME/CFS risk factors. The upper dashed line indicates the target power, where the probability of rejecting the null hypothesis is $1 - \beta = 0.80$. The lower dashed line indicates the significance level used, $\alpha = 0.05$.

seronegative population, the expected specificity of the serological specificity is approximately 0.975 (Domingues et al. 2021b). We assumed this value for π_{sp} . For simplicity, we assumed $\pi_{se} = \pi_{sp}$. Again, we simulated data from this scenario as the original study and estimated the probability of detecting an association as a function of the misdiagnosis probability. In this study, the minimum power of 80% could not be reached for any of the antibodies (Figure 4.3B). The best case was the antibody data related to HSV1. In this case, the maximal power was around 0.50 in the absence of misdiagnosis. This power dropped to 0.30 when $\gamma = 0.25$. For the remaining cases, the power was almost less than 0.20. This could partially be explained by the fact that θ_0 is higher than 0.93 for antibody data related to EBV, HHV6, and VZV.

4.4 Discussion

This study investigated the impact of misdiagnosis on the reproducibility of ME/CFS association studies. Our simulation study showed that strong associations with ME/CFS can be detected with reasonable power even under a non-negligible misdiagnosis rate. However, strong associations might not be the case of ME/CFS given the difficulty in finding a disease biomarker (Scheibebogen et al. 2017) and a clear genetic signature of the disease (Herrera et al. 2018; Tanigawa et al. 2019; Dibble et al. 2020; Hajdarevic et al. 2022).

Studies with sample sizes larger than 500 individuals per study group are able to compensate for the reduction in power due to misdiagnosis alone. This minimum sample size increases when, besides misdiagnosis, there is also the possibility of not determining the presence of the causal factor perfectly. In general, large studies are becoming common in well-known and highly-funded diseases, such as cancer, cardiovascular diseases (Giri et al. 2019), and autoimmune disorders (International Multiple Sclerosis Genetics Consortium (IMSGC) et al. 2013; Bjornevik et al. 2022). However, large ME/CFS studies are currently unfeasible due to limited funding and poor societal recognition of the disease (Pheby et al. 2021). This problem can be somehow minimised by using data from the United Kingdom ME/CFS Biobank that includes biological samples of more than 500 individuals (Lacerda et al. 2018). Another solution is to conduct multi-centric studies (Scheibebogen et al. 2017). Increasing sample size via data from self-reported ME/CFS cases (as performed in studies based on the UK Biobank) does not seem a viable solution because the chance of misdiagnosis is too high for obtaining reliable results. This problem is clearly illustrated in a Polish study where 1400 individuals were believed to be suffering from ME/CFS, but only 69 individuals actually complied with a consensual ME/CFS case definition (Słomko et al. 2019).

Current serological association studies of ME/CFS neglect the possibility of misclassifying seropositive individuals. In addition, it is common to leave the sensitivity and specificity of the respective serological classification unreported. This research practice adds to the list of other factors that can contribute to the lack of reproducibility of ME/CFS serological studies (Ariza 2020). Genetic association studies of ME/CFS also neglect the possibility of misclassifying the

genotypes of the individuals. This neglect is reasonable in most studies given that genotype error rates are often below 1%, and rare genetic markers with higher genotype errors are typically excluded from the analysis (Grabowska et al. 2020; Hajdarevic et al. 2021; 2022).

Our results are based on the assumption that disease association is independent of possible confounding factors. This assumption seems appropriate for randomised clinical studies or studies based on the analysis of specific subgroups, such as only focusing on adult women with an infection at the disease onset. However, it is also known that age, gender, and exposure to a given infectious agent can affect the results (Domingues et al. 2021a; Szklarski et al. 2021). Therefore, the assumption might not be true in general.

Our results are also based on the assumption that the controls are indeed healthy. Interestingly, ME/CFS patients and some healthy controls might have the same symptoms profile and similar levels of fatigue (Cella and Chalder 2010; Malato et al. 2021). More importantly, the use of self-reported healthy controls (Loebel et al. 2017; Szklarski et al. 2021) or control samples from existing blood banks (Kaushik et al. 2005; Johnston et al. 2016; Lande et al. 2020) are also common practices in ME/CFS research. According to these research practices, a more realistic assumption is to divide healthy controls into genuine and apparent controls. However, we anticipate that the statistical power to detect a putative disease association is further reduced in this more general scenario. To avoid this scenario, a thorough clinical assessment should also be performed in putative healthy controls.

This study was framed in terms of ME/CFS misdiagnosis in a strict sense. However, from a modelling standpoint, this framing is mathematically equivalent to the situation where ME/CFS-diagnosed cases can be partitioned into two subgroups of genuine patients but with distinct pathological mechanisms and where the association is only present in one of these subgroups. Therefore, our results are directly applicable to this alternative situation but with caution. As alluded to in the introduction, ME/CFS might not be one but several diseases under the same umbrella term, as suggested by genomic data (Kerr et al. 2008; Zhang et al. 2010). Having said that, a more realistic situation is to have multiple subgroups with different degrees of association with the potential causal factor. Therefore, there is a need to extend our simulation study to this situation.

In conclusion, current case-control association studies of ME/CFS seem to have limited power to mitigate the effect of misdiagnosis in the detection of putative disease associations. A sample size of 500 or 1000 individuals per study group is a minimal requirement to detect mild-to-moderate associations with a high power under the assumption of misdiagnosis. These sample sizes are attainable from multi-centric studies; these studies require extensive collaboration among ME/CFS researchers. Under the impossibility of increasing sample size, research efforts should be made towards reducing the rate of strict misdiagnosis. This can be achieved by following existing recommendations for research reports of ME/CFS, such as reporting the screening laboratory tests and the cut-off values for exclusion (Jason et al. 2012). It can also be achieved by the continued search for alternative diagnoses and co-morbidities (Nacul et al. 2021). In the end, a better understanding of multiple disease pathways leading to ME/CFS leads to better diagnoses, and, therefore, one should ultimately aim to study homogeneous cohorts of patients where the chance of strict misdiagnosis is reduced.

Bibliography

- M. E. Ariza. Commentary: Antibodies to Human Herpesviruses in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Patients. *Frontiers in Immunology*, 11, 2020. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2020.01400>.
- K. Bjornevik, M. Cortese, B. C. Healy, J. Kuhle, M. J. Mina, Y. Leng, S. J. Elledge, D. W. Niebuhr, A. I. Scher, K. L. Munger, and A. Ascherio. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science (New York, N.Y.)*, 375(6578):296–301, Jan. 2022. ISSN 1095-9203. doi: 10.1126/science.abj8222.
- D. Brown, C. Birch, J. Younger, and E. Worthey. ME/CFS: whole genome sequencing uncovers a misclassified case of glycogen storage disease type 13 previously diagnosed as ME/CFS. *Molecular Genetics and Metabolism*, 132:S194–S195, Apr. 2021. doi: 10.1016/s1096-7192(21)00388-7.
- K. G. Brurberg, M. S. Fønhus, L. Larun, S. Flottorp, and K. Malterud. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *BMJ Open*, 4(2):e003973, Feb. 2014. doi: 10.1136/bmjopen-2013-003973.
- B. M. Carruthers, A. K. Jain, K. L. D. Meirleir, D. L. Peterson, N. G. Klimas, A. M. Lerner, A. C. Bested, P. Flor-Henry, P. Joshi, A. C. P. Powles, J. A. Sherkey, and M. I. v. d. Sande. Myalgic Encephalomyelitis/Chronic Fatigue

- Syndrome. *Journal of Chronic Fatigue Syndrome*, 11(1):7–115, Jan. 2003. doi: 10.1300/j092v11n01_02.
- J. Castro-Marrero, M. D. Cordero, N. Sáez-Francas, C. Jimenez-Gutierrez, F. J. Aguilar-Montilla, L. Aliste, and J. Alegre-Martin. Could mitochondrial dysfunction be a differentiating marker between chronic fatigue syndrome and fibromyalgia? *Antioxidants & Redox Signaling*, 19(15):1855–1860, Nov. 2013. ISSN 1557-7716. doi: 10.1089/ars.2013.5346.
- M. Celli and T. Chalder. Measuring fatigue in clinical and community settings. *Journal of Psychosomatic Research*, 69(1):17–22, July 2010. ISSN 0022-3999. doi: 10.1016/j.jpsychores.2009.10.007. URL <https://www.sciencedirect.com/science/article/pii/S0022399909004176>.
- J. M. Cliff, E. C. King, J.-S. Lee, N. Sepúlveda, A.-S. Wolf, C. Kingdon, E. Bowman, H. M. Dockrell, L. Nacul, E. Lacerda, and E. M. Riley. Cellular Immune Function in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Frontiers in Immunology*, 10, Apr. 2019. doi: 10.3389/fimmu.2019.00796. URL <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2019.00796/full>.
- K. E. Conroy, M. F. Islam, and L. A. Jason. Evaluating case diagnostic criteria for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS): toward an empirical case definition. *Disability and Rehabilitation*, 45(5):840–847, Mar. 2023. ISSN 1464-5165. doi: 10.1080/09638288.2022.2043462.
- J. J. Dibble, S. J. McGrath, and C. P. Ponting. Genetic risk factors of ME/CFS: a critical review. *Human Molecular Genetics*, 29(R1):R117–R124, Sept. 2020. ISSN 0964-6906. doi: 10.1093/hmg/ddaa169. URL <https://doi.org/10.1093/hmg/ddaa169>.
- T. D. Domingues, A. D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, F. Westermeier, C. Scheibenbogen, J. M. Cliff, L. Nacul, E. M. Lacerda, H. Mourão, and N. Sepúlveda. Herpesviruses Serology Distinguishes Different Subgroups of Patients From the United Kingdom Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Biobank. *Frontiers in Medicine*, 8, July 2021a. doi: 10.3389/fmed.2021.686736.
- T. D. Domingues, H. Mourão, and N. Sepúlveda. Analysis of antibody data using Finite Mixture Models based on Scale Mixtures of Skew-Normal distributions, Mar. 2021b. URL <https://www.medrxiv.org/content/10.1101/2021.03.08.21252807v1>. ISSN: 2125-2807 Pages: 2021.03.08.21252807.
- K. Fukuda, S. E. Straus, I. Hickie, M. C. Sharpe, J. G. Dobbins, and A. Komaroff. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Annals of Internal Medicine*, 121(12):953, Dec. 1994. doi: 10.7326/0003-4819-121-12-199412150-00009.
- A. Giri, J. N. Hellwege, J. M. Keaton, J. Park, C. Qiu, H. R. Warren, E. S. Torstenson, C. P. Kovesdy, Y. V. Sun, O. D. Wilson, C. Robinson-Cohen, C. L. Roumie, C. P. Chung, K. A. Birdwell, S. M. Damrauer, S. L. DuVall, D. Klarin, K. Cho, Y. Wang, E. Evangelou, C. P. Cabrera, L. V. Wain, R. Shrestha, B. S. Mautz, E. A. Akwo, M. Sargurupremraj, S. Debette, M. Boehnke, L. J. Scott, J. Luan, J.-H. Zhao, S. M. Willems, S. Thériault, N. Shah, C. Oldmeadow, P. Almgren, R. Li-Gao, N. Verweij, T. S. Boutin, M. Mangino, I. Ntalla, E. Feofanova, P. Surendran, J. P. Cook, S. Karthikeyan, N. Lahrouchi, C. Liu, N. Sepúlveda, T. G. Richardson, A. Kraja, P. Amouyel, M. Farrall,

- N. R. Poulter, M. Laakso, E. Zeggini, P. Sever, R. A. Scott, C. Langenberg, N. J. Wareham, D. Conen, C. N. A. Palmer, J. Attia, D. I. Chasman, P. M. Ridker, O. Melander, D. O. Mook-Kanamori, P. v. d. Harst, F. Cucca, D. Schlessinger, C. Hayward, T. D. Spector, M.-R. Jarvelin, B. J. Hennig, N. J. Timpson, W.-Q. Wei, J. C. Smith, Y. Xu, M. E. Matheny, E. E. Siew, C. Lindgren, K.-H. Herzig, G. Dedoussis, J. C. Denny, B. M. Psaty, J. M. M. Howson, P. B. Munroe, C. Newton-Cheh, M. J. Caulfield, P. Elliott, J. M. Gaziano, J. Concato, P. W. F. Wilson, P. S. Tsao, D. R. Velez Edwards, K. Susztak, C. J. O'Donnell, A. M. Hung, and T. L. Edwards. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nature Genetics*, 51(1):51–62, Jan. 2019. ISSN 1546-1718. doi: 10.1038/s41588-018-0303-9. URL <https://www.nature.com/articles/s41588-018-0303-9>. Number: 1 Publisher: Nature Publishing Group.
- A. D. Grabowska, E. M. Lacerda, L. Nacul, and N. Sepúlveda. Review of the Quality Control Checks Performed by Current Genome-Wide and Targeted-Genome Association Studies on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Pediatrics*, 8, 2020. ISSN 2296-2360. URL <https://www.frontiersin.org/articles/10.3389/fped.2020.00293>.
- R. Hajdarevic, A. Lande, I. Rekeland, A. Rydland, E. B. Strand, D. D. Sosa, L. E. Creary, O. Mella, T. Egeland, O. D. Saugstad, O. Fluge, B. A. Lie, and M. K. Viken. Fine mapping of the major histocompatibility complex (MHC) in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) suggests involvement of both HLA class I and class II loci. *Brain, Behavior, and Immunity*, 98:101–109, Nov. 2021. ISSN 0889-1591. doi: 10.1016/j.bbi.2021.08.219. URL <https://www.sciencedirect.com/science/article/pii/S0889159121005092>.
- R. Hajdarevic, A. Lande, J. Mehlsen, A. Rydland, D. D. Sosa, E. B. Strand, O. Mella, F. Pociot, O. Fluge, B. A. Lie, and M. K. Viken. Genetic association study in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) identifies several potential risk loci. *Brain, Behavior, and Immunity*, 102:362–369, May 2022. ISSN 0889-1591. doi: 10.1016/j.bbi.2022.03.010. URL <https://www.sciencedirect.com/science/article/pii/S0889159122000782>.
- S. Herrera, W. C. de Vega, D. Ashbrook, S. D. Vernon, and P. O. McGowan. Genome-epigenome interactions associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Epigenetics*, 13(12):1174–1190, Dec. 2018. ISSN 1559-2294. doi: 10.1080/15592294.2018.1549769. URL <https://doi.org/10.1080/15592294.2018.1549769>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/15592294.2018.1549769>.
- International Multiple Sclerosis Genetics Consortium (IMSGC), A. H. Beecham, N. A. Patsopoulos, D. K. Xifara, M. F. Davis, A. Kemppinen, C. Cotsapas, T. S. Shah, C. Spencer, D. Booth, A. Goris, A. Oturai, J. Saarela, B. Fontaine, B. Hemmer, C. Martin, F. Zipp, S. D'Alfonso, F. Martinelli-Boneschi, B. Taylor, H. F. Harbo, I. Kockum, J. Hillert, T. Olsson, M. Ban, J. R. Oksenberg, R. Hintzen, L. F. Barcellos, Wellcome Trust Case Control Consortium 2 (WTCCC2), International IBD Genetics Consortium (IIBDGC), C. Agliardi, L. Alfredsson, M. Alizadeh, C. Anderson, R. Andrews, H. B. Søndergaard, A. Baker, G. Band, S. E. Baranzini, N. Barizzone, J. Barrett, C. Bellenguez, L. Bergamaschi, L. Bernardinelli, A. Berthele, V. Biberacher, T. M. C. Binder, H. Blackburn, I. L.

- Bomfim, P. Brambilla, S. Broadley, B. Brochet, L. Brundin, D. Buck, H. Butzkueven, S. J. Caillier, W. Camu, W. Carpentier, P. Cavalla, E. G. Celius, I. Coman, G. Comi, L. Corrado, L. Cosemans, I. Cournu-Rebeix, B. A. C. Cree, D. Cusi, V. Damotte, G. Defer, S. R. Delgado, P. Deloukas, A. di Sapiro, A. T. Dilthey, P. Donnelly, B. Dubois, M. Duddy, S. Edkins, I. Elovaara, F. Esposito, N. Evangelou, B. Fiddes, J. Field, A. Franke, C. Freeman, I. Y. Frohlich, D. Galimberti, C. Gieger, P-A. Gourraud, C. Graetz, A. Graham, V. Grummel, C. Guaschino, A. Hadjixenofontos, H. Hakonarson, C. Halfpenny, G. Hall, P. Hall, A. Hamsten, J. Harley, T. Harrower, C. Hawkins, G. Hellenthal, C. Hillier, J. Hobart, M. Hoshi, S. E. Hunt, M. Jagodic, I. Jelčić, A. Jochim, B. Kendall, A. Kermode, T. Kilpatrick, K. Koivisto, I. Konidari, T. Korn, H. Kronsbein, C. Langford, M. Larsson, M. Lathrop, C. Lebrun-Frenay, J. Lechner-Scott, M. H. Lee, M. A. Leone, V. Leppä, G. Liberatore, B. A. Lie, C. M. Lill, M. Lindén, J. Link, F. Luessi, J. Lycke, F. Macciardi, S. Männistö, C. P. Manrique, R. Martin, V. Martinelli, D. Mason, G. Mazibrada, C. McCabe, I.-L. Mero, J. Mescheriakova, L. Moutsianas, K.-M. Myhr, G. Nagels, R. Nicholas, P. Nilsson, F. Piehl, M. Pirinen, S. E. Price, H. Quach, M. Reunanen, W. Robberecht, N. P. Robertson, M. Rodegher, D. Rog, M. Salvetti, N. C. Schnetz-Boutaud, F. Sellebjerg, R. C. Selter, C. Schaefer, S. Shaunak, L. Shen, S. Shields, V. Siffrin, M. Slee, P. S. Sorensen, M. Sorosina, M. Sospedra, A. Spurkland, A. Strange, E. Sundqvist, V. Thijs, J. Thorpe, A. Ticca, P. Tienari, C. van Duijn, E. M. Visser, S. Vucic, H. Westerlind, J. S. Wiley, A. Wilkins, J. F. Wilson, J. Winkelmann, J. Zajicek, E. Zindler, J. L. Haines, M. A. Pericak-Vance, A. J. Ivinson, G. Stewart, D. Hafler, S. L. Hauser, A. Compston, G. McVean, P. De Jager, S. J. Sawcer, and J. L. McCauley. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature Genetics*, 45(11):1353–1360, Nov. 2013. ISSN 1546-1718. doi: 10.1038/ng.2770.
- L. A. Jason, E. R. Unger, J. D. Dimitrakoff, A. P. Fagin, M. Houghton, D. B. Cook, G. D. Marshall, N. Klimas, and C. Snell. Minimum data elements for research reports on CFS. *Brain, Behavior, and Immunity*, 26(3):401–406, Mar. 2012. ISSN 0889-1591. doi: 10.1016/j.bbi.2012.01.014. URL <https://www.sciencedirect.com/science/article/pii/S0889159112000165>.
- L. A. Jason, M. Sunnquist, A. Brown, M. Evans, S. D. Vernon, J. Furst, and V. Simonis. Examining case definition criteria for chronic fatigue syndrome and myalgic encephalomyelitis. *Fatigue : biomedicine, health & behavior*, 2(1):40–56, Jan. 2014. ISSN 2164-1846. doi: 10.1080/21641846.2013.862993. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3912876/>.
- L. A. Jason, B. Kot, M. Sunnquist, A. Brown, J. Reed, J. Furst, J. L. Newton, E. B. Strand, and S. D. Vernon. Comparing and contrasting consensus versus empirical domains. *Fatigue: Biomedicine, Health & Behavior*, 3(2):63–74, Apr. 2015. ISSN 2164-1846. doi: 10.1080/21641846.2015.1017344. URL <https://doi.org/10.1080/21641846.2015.1017344>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/21641846.2015.1017344>.
- L. A. Jason, S. Ravichandran, B. Z. Katz, B. H. Natelson, and H. F. Bonilla. Establishing a consensus on ME/CFS exclusionary illnesses. *Fatigue: Biomedicine, Health & Behavior*, 11(1):1–13, Jan. 2023. ISSN 2164-1846,

- 2164-1862. doi: 10.1080/21641846.2022.2150487. URL <https://www.tandfonline.com/doi/full/10.1080/21641846.2022.2150487>.
- S. Johnston, D. Staines, A. Klein, and S. Marshall-Gradisnik. A targeted genome association study examining transient receptor potential ion channels, acetylcholine receptors, and adrenergic receptors in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *BMC Medical Genetics*, 17(1):79, Nov. 2016. ISSN 1471-2350. doi: 10.1186/s12881-016-0342-y. URL <https://doi.org/10.1186/s12881-016-0342-y>.
- N. Kaushik, D. Fear, S. C. M. Richards, C. R. McDermott, E. F. Nuwaysir, P. Kellam, T. J. Harrison, R. J. Wilkinson, D. a. J. Tyrrell, S. T. Holgate, and J. R. Kerr. Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. *Journal of Clinical Pathology*, 58(8):826–832, Aug. 2005. ISSN 0021-9746, 1472-4146. doi: 10.1136/jcp.2005.025718. URL <https://jcp.bmjjournals.org/content/58/8/826>. Publisher: BMJ Publishing Group Section: Original article.
- J. R. Kerr, R. Petty, B. Burke, J. Gough, D. Fear, L. I. Sinclair, D. L. Matthey, S. C. M. Richards, J. Montgomery, D. A. Baldwin, P. Kellam, T. J. Harrison, G. E. Griffin, J. Main, D. Enlander, D. J. Nutt, and S. T. Holgate. Gene Expression Subtypes in Patients with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *The Journal of Infectious Diseases*, 197(8):1171–1184, Apr. 2008. ISSN 0022-1899. doi: 10.1086/533453. URL <https://doi.org/10.1086/533453>.
- R. S. König, W. C. Albrich, C. R. Kahlert, L. S. Bahr, U. Löber, P. Vernazza, C. Scheibenbogen, and S. K. Forslund. The Gut Microbiome in Myalgic Encephalomyelitis (ME)/Chronic Fatigue Syndrome (CFS). *Frontiers in Immunology*, 12, 2022. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2021.628741>.
- E. M. Lacerda, K. Mudie, C. C. Kingdon, J. D. Butterworth, S. O’Boyle, and L. Nacul. The UK ME/CFS Biobank: A Disease-Specific Biobank for Advancing Clinical Research Into Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Neurology*, 9:1026, 2018. ISSN 1664-2295. doi: 10.3389/fneur.2018.01026.
- A. Lande, O. Fluge, E. B. Strand, S. T. Flåm, D. D. Sosa, O. Mella, T. Egeland, O. D. Saugstad, B. A. Lie, and M. K. Viken. Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Scientific Reports*, 10(1):5267, Mar. 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-62157-x. URL <https://www.nature.com/articles/s41598-020-62157-x>. Number: 1 Publisher: Nature Publishing Group.
- E.-J. Lim, Y.-C. Ahn, E.-S. Jang, S.-W. Lee, S.-H. Lee, and C.-G. Son. Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *Journal of Translational Medicine*, 18(1), Feb. 2020. doi: 10.1186/s12967-020-02269-0.
- M. Loebel, M. Eckey, F. Sotzny, E. Hahn, S. Bauer, P. Grabowski, J. Zerweck, P. Holenya, L. G. Hanitsch, K. Wittke, P. Borchmann, J.-U. Rüffer, F. Hiepe, K. Ruprecht, U. Behrends, C. Meindl, H.-D. Volk, U. Reimer, and C. Scheibenbogen. Serological profiling of the EBV immune response in Chronic Fatigue Syndrome using a peptide mi-

- croarray. *PLOS ONE*, 12(6):e0179124, Dec. 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0179124. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0179124>. Publisher: Public Library of Science.
- J. a. Malato, L. Graça, L. Nacul, E. M. Lacerda, , and N. Sepúlveda. Statistical challenges of investigating a disease with a complex diagnosis. In P. Milheiro, A. Pacheco, B. de Sousa, I. F. Alves, I. Pereira, M. J. a. Polidoro, and S. Ramos, editors, *Estatística: Desafios transversais às ciências com dados*, Eds. Sociedade portuguesa de estatística, pages 153–167. Sociedade Portuguesa de Estatística, 2021. ISBN 978-972-8890-47-6. URL <https://www.spestatistica.pt/storage/app/uploads/public/609/28f/6d0/60928f6d08a0c016386627.pdf>.
- J. a. Malato, L. Graça, and N. Sepúlveda. Impact of Misclassification and Imperfect Serological Tests in Association Analyses of ME/CFS Applied to COVID-19 Data. In R. Bispo, L. Henriques-Rodrigues, R. Alpizar-Jara, and M. de Carvalho, editors, *Recent Developments in Statistics and Data Science*, Springer Proceedings in Mathematics & Statistics, pages 215–225, Cham, 2022. Springer International Publishing. ISBN 978-3-031-12766-3. doi: 10.1007/978-3-031-12766-3_15.
- L. Nacul, E. M. Lacerda, C. C. Kingdon, H. Curran, and E. W. Bowman. How have selection bias and disease misclassification undermined the validity of myalgic encephalomyelitis/chronic fatigue syndrome studies? *Journal of Health Psychology*, 24(12):1765–1769, Oct. 2019. ISSN 1461-7277. doi: 10.1177/1359105317695803.
- L. Nacul, F. J. Authier, C. Scheibenbogen, L. Lorusso, I. B. Helland, J. A. Martin, C. A. Sirbu, A. M. Mengshoel, O. Polo, U. Behrends, H. Nielsen, P. Grabowski, S. Sekulic, N. Sepulveda, F. Estévez-López, P. Zalewski, D. F. H. Pheby, J. Castro-Marrero, G. K. Sakkas, E. Capelli, I. Brundsdlund, J. Cullinan, A. Krumina, J. Bergquist, M. Murovska, R. C. W. Vermuelen, and E. M. Lacerda. European Network on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (EUROMENE): Expert Consensus on the Diagnosis, Service Provision, and Care of People with ME/CFS in Europe. *Medicina*, 57(5):510, May 2021. ISSN 1648-9144. doi: 10.3390/medicina57050510. URL <https://www.mdpi.com/1648-9144/57/5/510>.
- D. F. H. Pheby, D. Araja, U. Berkis, E. Brenna, J. Cullinan, J.-D. de Korwin, L. Gitto, D. A. Hughes, R. M. Hunter, D. Trepel, and X. Wang-Steverding. A Literature Review of GP Knowledge and Understanding of ME/CFS: A Report from the Socioeconomic Working Group of the European Network on ME/CFS (EUROMENE). *Medicina*, 57(1):7, Jan. 2021. ISSN 1648-9144. doi: 10.3390/medicina57010007. URL <https://www.mdpi.com/1648-9144/57/1/7>. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.
- S. Rasa, Z. Nora-Krukla, N. Henning, E. Eliassen, E. Shikova, T. Harrer, C. Scheibenbogen, M. Murovska, B. K. Prusty, and the European Network on ME/CFS (EUROMENE). Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 16(1):268, Oct. 2018. ISSN

- 1479-5876. doi: 10.1186/s12967-018-1644-y. URL <https://doi.org/10.1186/s12967-018-1644-y>.
- W. C. Reeves, D. Wagner, R. Nisenbaum, J. F. Jones, B. Gurbaxani, L. Solomon, D. A. Papanicolaou, E. R. Unger, S. D. Vernon, and C. Heim. Chronic Fatigue Syndrome – A clinically empirical approach to its definition and study. *BMC Medicine*, 3(1):19, Dec. 2005. ISSN 1741-7015. doi: 10.1186/1741-7015-3-19. URL <https://doi.org/10.1186/1741-7015-3-19>.
- M. C. Rivera, C. Mastronardi, C. Silva-Aldana, M. Arcos-Burgos, and B. Lidbury. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Comprehensive Review. *Diagnostics*, 9(3):91, Aug. 2019. doi: 10.3390/diagnostics9030091.
- M. Ruiz-Pablos, B. Paiva, R. Montero-Mateo, N. Garcia, and A. Zabaleta. Epstein-Barr Virus and the Origin of Myalgic Encephalomyelitis or Chronic Fatigue Syndrome. *Frontiers in Immunology*, 12, 2021. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2021.656797>.
- C. Scheibenbogen, H. Freitag, J. Blanco, E. Capelli, E. Lacerda, J. Authier, M. Meeus, J. C. Marrero, Z. Nora-Krukla, E. Oltra, and others. The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE. *Journal of translational medicine*, 15(1):162, 2017. doi: <https://doi.org/10.1186/s12967-017-1263-z>.
- N. Sepúlveda, G. Stresman, M. T. White, and C. J. Drakeley. Current Mathematical Models for Analyzing Anti-Malarial Antibody Data with an Eye to Malaria Elimination and Eradication. *Journal of Immunology Research*, 2015:738030, 2015. ISSN 2314-7156. doi: 10.1155/2015/738030.
- N. Sepúlveda, J. Carneiro, E. Lacerda, and L. Nacul. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome as a Hyper-Regulated Immune System Driven by an Interplay Between Regulatory T Cells and Chronic Human Herpesvirus Infections. *Frontiers in Immunology*, 10, 2019. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02684>.
- N. Sepúlveda, J. a. Malato, F. Sotzny, A. D. Grabowska, A. Fonseca, C. Cordeiro, L. Graça, P. Bicek, U. Behrends, J. Mautner, F. Westermeier, E. M. Lacerda, and C. Scheibenbogen. Revisiting IgG Antibody Reactivity to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Its Potential Application to Disease Diagnosis. *Frontiers in Medicine*, 9, 2022. ISSN 2296-858X. doi: 10.3389/fmed.2022.921101. URL <https://www.frontiersin.org/articles/10.3389/fmed.2022.921101>.
- J. Słomko, J. L. Newton, S. Kujawski, M. Tafil-Klawe, J. Klawe, D. Staines, S. Marshall-Gradisnik, and P. Zalewski. Prevalence and characteristics of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) in Poland: a cross-sectional study. *BMJ open*, 9(3):e023955, Mar. 2019. ISSN 2044-6055. doi: 10.1136/bmjopen-2018-023955.
- M. E. B. Smith, H. D. Nelson, E. Haney, M. Pappas, M. Daeges, N. Wasson, and M. McDonagh. Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Evidence report/technology assessment*, (219):1–433, Dec. 2014. ISSN 1530-4396. doi: 10.23970/ahrqcerta219. URL <https://doi.org/10.23970/ahrqcerta219>.

- [23970/AHRQEPCERTA219.](#)
- S. Steiner, S. C. Becker, J. Hartwig, F. Sotzny, S. Lorenz, S. Bauer, M. Löbel, A. B. Stitrich, P. Grabowski, and C. Scheibenbogen. Autoimmunity-Related Risk Variants in PTPN22 and CTLA4 Are Associated With ME/CFS With Infectious Onset. *Frontiers in Immunology*, 11, 2020. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2020.00578>.
- M. Szklarski, H. Freitag, S. Lorenz, S. C. Becker, F. Sotzny, S. Bauer, J. Hartwig, H. Heidecke, K. Wittke, C. Kedor, L. G. Hanitsch, P. Grabowski, N. Sepúlveda, and C. Scheibenbogen. Delineating the Association Between Soluble CD26 and Autoantibodies Against G-Protein Coupled Receptors, Immunological and Cardiovascular Parameters Identifies Distinct Patterns in Post-Infectious vs. Non-Infection-Triggered Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Immunology*, 12, Apr. 2021. doi: 10.3389/fimmu.2021.644548.
- Y. Tanigawa, J. Li, J. M. Justesen, H. Horn, M. Aguirre, C. DeBoever, C. Chang, B. Narasimhan, K. Lage, T. Hastie, C. Y. Park, G. Bejerano, E. Ingelsson, and M. A. Rivas. Components of genetic associations across 2,138 phenotypes in the UK Biobank highlight adipocyte biology. *Nature Communications*, 10(1):4064, Sept. 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-11953-9. URL <https://www.nature.com/articles/s41467-019-11953-9>. Number: 1 Publisher: Nature Publishing Group.
- K. Wirth and C. Scheibenbogen. A Unifying Hypothesis of the Pathophysiology of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS): Recognitions from the finding of autoantibodies against β 2-adrenergic receptors. *Autoimmunity Reviews*, 19(6):102527, June 2020. ISSN 1873-0183. doi: 10.1016/j.autrev.2020.102527.
- E. Wood, K. H. Hall, and W. Tate. Role of mitochondria, oxidative stress and the response to antioxidants in myalgic encephalomyelitis/chronic fatigue syndrome: A possible approach to SARS-CoV-2 'long-haulers'? *Chronic Diseases and Translational Medicine*, 7(1):14–26, Mar. 2021. ISSN 2589-0514. doi: 10.1016/j.cdtm.2020.11.002.
- L. Zhang, J. Gough, D. Christmas, D. L. Matthey, S. C. M. Richards, J. Main, D. Enlander, D. Honeybourne, J. G. Ayres, D. J. Nutt, and J. R. Kerr. Microbial infections in eight genomic subtypes of chronic fatigue syndrome/myalgic encephalomyelitis. *Journal of Clinical Pathology*, 63(2):156–164, Feb. 2010. ISSN 1472-4146. doi: 10.1136/jcp.2009.072561.

Chapter 5

Revisiting IgG antibody reactivity to EBV in ME/CFS and its potential application to disease diagnosis

N. Sepúlveda, J. Malato, F. Sotzny, A.D. Grabowska, A. Fonseca, C. Cordeiro, L. Graça, P. Biecek, U. Behrends, J. Mautner, F. Westermeier, E.M. Lacerda, and C. Scheibenbogen. Revisiting IgG Antibody Reactivity to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Its Potential Application to Disease Diagnosis. *Frontiers in Medicine*. 2022. 9:921101. doi: <https://doi.org/10.3389/fmed.2022.921101>.

Abstract

Infections by the Epstein-Barr virus (EBV) are often at the disease onset of patients suffering from myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). However, serological analyses of these infections remain inconclusive when comparing patients with healthy controls. In particular, it is unclear if certain EBV-derived antigens eliciting antibody responses have a biomarker potential for disease diagnosis. With this purpose, we re-analysed a previously published microarray data on the IgG antibody responses against 3,054 EBV-related antigens in 92 patients with ME/CFS and 50 healthy controls. This re-analysis consisted of constructing different regression models for binary outcomes with the ability to classify patients and healthy controls. In these models, we tested for a possible interaction of different antibodies with age and gender. When analysing the whole data set, there were no antibody responses that could distinguish patients from healthy controls. A similar finding was obtained when comparing patients with non-infectious or unknown disease trigger with healthy controls. However, when data analysis was restricted to the comparison between healthy controls and patients with a putative infection at their disease onset, we could identify stronger antibody responses against two candidate antigens (EBNA4_0529 and EBNA6_0070). Using antibody responses to these two antigens together with age and gender, the final classification model had an estimated sensitivity and specificity of 0.833 and 0.720, respectively. This reliable case-control discrimination suggested the use of the antibody levels related to these candidate

viral epitopes as biomarkers for disease diagnosis in this subgroup of patients. To confirm this finding, a follow-up study will be conducted in a separate cohort of patients.

Keywords: Epstein-Barr virus; Myalgic encephalomyelitis/Chronic fatigue syndrome; antigen mimicry; biomarker discovery; patient stratification

5.1 Introduction

Infections by the ubiquitous Epstein-Barr virus (EBV) are linked to multiple sclerosis, rheumatoid arthritis, systemic erythematosus lupus, lymphomas, among other known diseases (Shannon-Lowe et al. 2017; Houen and Trier 2021; Bjornevik et al. 2022). A less-known disease where EBV infections are also important is myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) (Koo 1989; Rasa et al. 2018; Ruiz-Pablos et al. 2021). The hallmark symptom of this condition is an unexplained but persistent fatigue that cannot be alleviated by rest and that can increase upon minimal physical and emotional effort (Rivera et al. 2019; Bateman et al. 2021). In ME/CFS, acute EBV infections are reported by a subset of patients at the onset of their symptoms (Hickie et al. 2006; Domingues et al. 2021). Reactivation of latent EBV infections has also been described during the disease course (Shikova et al. 2020). However, current evidence remains inconclusive on whether the prevalence of these reactivations is either higher or lower in patients than in healthy controls (Lee et al. 2021). This conflicting evidence notwithstanding, ME/CFS patients show deficient B- and T-cell responses against EBV and altered antibody profiles when compared with healthy controls (Lerner et al. 2012; Loebel et al. 2014; Kerr 2019; Domingues et al. 2021). Finally, CD4+ T cells recognising self-peptides on HLA-DR15, the strongest genetic risk factor for multiple sclerosis, have been shown to cross-react with peptides derived from EBV (Wang et al. 2020). Multiple sclerosis patients share many symptoms with the ones suffering from ME/CFS (Morris and Maes 2013; Gaber et al. 2014; Malato et al. 2021). EBV antigens were also reported to share sequence homology with human peptides derived from the myelin basic protein (Wucherpfennig and Strominger 1995; Holmøy et al. 2004; Lünemann et al. 2008), lactoperoxidase (Loebel et al. 2017), and anoctamin-2 (Tengvall et al. 2019; Sepúlveda 2021). These observations suggest that molecu-

lar mimicry between human and EBV-derived antigens could play a role in the pathogenesis of ME/CFS. This suggestion is in line with our recent hypothesis that links the pathogenesis of ME/CFS to chronically activated immune responses (Sepúlveda et al. 2019). Our assumption raises the possibility that the immune system of some ME/CFS patients is oscillating between an activation state that attempts controlling latent herpesviruses infections and the suppression of deleterious autoimmune responses via the activation of regulatory T cells (Sepúlveda et al. 2019). Thus, considering the growing body of evidence that links EBV infection to the pathogenesis of ME/CFS, studies that aim at elucidating underlying mechanisms are needed.

A major problem in investigating ME/CFS is the non-existence of a robust biomarker that could ascertain the disease diagnosis. In the past, different discovery studies suggested certain cytokines, antibodies against self and non-self-antigens, microRNAs, and methylation markers as potential disease biomarkers (Scheibenberg et al. 2017). Antibodies against EBV antigens are of particular interest as disease biomarkers given the above evidence connecting this virus with the disease and routine application of serological assays in the clinical practice. However, EBV antigens included in commercial kits are mostly markers of exposure to the infection and are unable to distinguish between patients with ME/CFS and healthy controls (Cliff et al. 2019). This distinction can only be made when comparing a subset of clinically diagnosed ME/CFS patients with an EBV infection trigger to healthy controls (Domingues et al. 2021). A serological evaluation of antibodies against less-studied EBV antigens did not identify any that could be used as a specific disease biomarker (Blomberg et al. 2019). However, this antibody evaluation was done using a limited number of EBV-derived antigens and no subgroup analysis was performed. The lack of patient stratification in ME/CFS studies reduces the chance of reproducing the same findings in follow-up studies (Jason et al. 2005; Scheibenberg et al. 2017). Therefore, it is still possible to identify alternative antigens whose antibody responses could be used as disease biomarkers for a subgroup of patients.

Recently, we analysed antibody responses against more than 3,000 overlapping antigens derived from 14 EBV proteins (Loebel et al. 2017). The aim of this study was to extract an antibody signature against EBV in ME/CFS patients when compared to healthy controls. In the present study, we extended the analysis of the obtained data with the specific objective

of optimising biomarker discovery. In particular, we compared patients with or without an infectious trigger at disease onset to healthy controls in order to discover EBV-derived antigens whose antibody responses could be used for ME/CFS diagnosis.

5.2 Materials and methods

5.2.1 Study participants

Ninety-two ME/CFS patients were recruited between 2011 and 2015 at the Charité outpatient clinic for immunodeficiencies at the Institute of Medical Immunology in the Charité Universitätsmedizin Berlin, Germany. Additional fifty individuals were recruited from the employees of the same clinic, who self-reported to be healthy and to not suffer from fatigue. However, neither clinical nor laboratory assessment was performed to confirm the healthy status of those individuals. ME/CFS patients and healthy controls were matched for gender and age (Table 5.1) with 50% of women and an overall average of ~43 years of age. Fifty-four out of 92 patients (58.7%) reported an acute infection at their disease onset, whilst the remaining 38 patients (41.3%) reported either a disease trigger other than an infection, did not know their disease onset or the information about the disease trigger was missing. These two subgroups were also matched for age and gender (Table 5.1).

Table 5.1: Basic characteristics of ME/CFS patients and healthy controls, where p-values refer to the comparison between ME/CFS groups and healthy controls.

Group	Female			Age, years	
	N	%	P-value	Mean (age range)	P-value
Healthy controls	50	50.0	—	42.4 (25–61)	—
ME/CFS (all)	92	51.1	0.901	43.7 (25–66)	0.453
With infectious trigger	54	50.0	~1.000	43.2 (17–66)	0.585
Unknown trigger or without infectious trigger	38	52.6	0.807	44.4 (24–66)	0.679

5.2.2 Peptide array

Data under analyses refer to the signal intensities derived from IgG antibody responses to 3,054 EBV-associated peptides measured by a seroarray described in detail in the original study (Loebel et al. 2017). These peptides consisted of partially overlapping 15 amino acids (15-mer) and covered the full length of the following proteins (Supplementary Table B.1): BALF-2, BALF-5, BFRF-3, BLLF-1, BLLF-3, BLRF-2, BMRF-1, BZLF-1, EBNA-1, EBNA-3, EBNA-4, EBNA-6, LMP-1, and LMP-2. The 15-mer peptides overlapped in 11 amino acids. The amino-acid sequences of these peptides were representative of the following EBV strains: AG876 (West Africa, EBV type 2), B95.8 (USA, EBV type 1), GD1 (China, EBV type 1), Cao (China, EBV type 1), Raji (Nigeria, EBV type 1), and P3HR-1 (Nigeria, EBV type 2). These data are freely available in Supplementary File S1 of the original study (Loebel et al. 2017).

5.2.3 Statistical analysis

We used the Pearson's χ^2 test to compare ME/CFS patients to healthy controls in terms of gender distribution. The non-parametric Mann-Whitney test was used to compare the medians of the respective age distributions. There was evidence for age- and gender-matched distributions if the p-values of these tests were greater than the significance level of 5%.

We first performed a multivariate analysis using (i) the classical principal component analysis (PCA) and (ii) computing different correlation matrices using Spearman's correlation coefficient (which is invariant to monotonic changes in the scale of the data, is robust against the presence of outliers, and does not depend on the normality assumption). We then performed linear discriminant analyses (LDA) to determine the best linear combination of all the antibody responses that could distinguish ME/CFS patients and their subgroups from healthy individuals. A similar analysis was done to compare the two subgroups of ME/CFS patients.

The outcome of each LDA was the estimated classification probability for each individual. These estimated probabilities were then analyzed by the respective receiver operating characteristic (ROC) curve where 1 – specificity and sensitivity are plotted against each other as a function of the cutoff of the underlying classification probability. After computing each

ROC curve, we calculated the respective area under the curve (AUC) and its 95% confidence interval to determine the accuracy of the classification irrespective of the cut-off used. In general, an AUC = 0.50 is indicative of a complete random classification of the individuals, while AUC = 1.00 implies that the constructed classifier perfectly predicts the true class membership of each individual.

We performed further antibody-wide association analyses related to the following comparisons (or classification exercises): (i) healthy controls versus all the ME/CFS patients; (ii) healthy controls versus ME/CFS patients with an infectious trigger; (iii) healthy controls versus ME/CFS patients with a non-infectious or unknown trigger; and (iv) ME/CFS patients with an infectious trigger versus the remaining ME/CFS patients. In each association analysis, we first estimated three regression models: logistic model, probit model, and complementary log-log model. In these models, the disease status was the outcome variable, age and gender were the respective covariates. To determine the best link function for the outcome variable, we selected the model with the lowest Akaike's information criterion (AIC). For the best link function ("the null model"), we estimated the respective ROC and its AUC as described above.

We fitted five different logistic models, including the main effects and all the interaction terms related to age, gender, and the antibody response under analysis: (i) a model with main effects only and no interaction terms; (ii) a model with an interaction term between age and the antibody response; (iii) a model with an interaction term between gender and the antibody response; (iv) a model with two interaction terms between age and the antibody response and between gender and the antibody response; (v) a model with all two-way and three-way interaction terms related to age, gender, and the antibody response. We compared each of these models with the null one using Wilks's likelihood ratio test, where low p-values provide evidence for these models, including effects of an antibody response. We reported the minimum p-value obtained from these model comparisons. Finally, we adjusted the minimum p-values of each analysis. This adjustment was made using the Benjamini-Yekutieli procedure ensuring a global false discovery rate (FDR) of 5% under the assumption of dependent tests (Benjamini and Yekutieli 2001). In this analysis, adjusted p-values < 0.05 indicated statistically significant results.

To filter out redundant antibody responses, we pooled all the significant antibody responses in a single model. The effect and interaction terms of these antibody responses were defined according to the most significant model obtained in the previous stage of analysis. We performed a backward stepwise model selection. The resulting model was finally evaluated in terms of predictive performance using ROC analysis as described above.

The above analysis was primarily done for the whole data set irrespective of the ME/CFS subgroups. We repeated the same analysis to compare each subgroup of ME/CFS patients (with infectious and non-infectious or unknown disease trigger) with the healthy controls. Finally, we repeated the analysis to compare the two subgroups of ME/CFS patients.

5.2.4 Statistical software

The statistical analysis was performed in the R software version 4.0.3 with core functions and the following packages: *MASS* v7.3-56 to perform stepwise model selection (Venables and Ripley 2002), *pROC* v1.18.0 to estimate the ROC curve and the respective AUC (Robin et al. 2011), *OptimalCutpoints* v1.1-5 to estimate the optimal cutoff and the associated sensitivity/specificity (López-Ratón et al. 2014). The full reproducible code is freely available from NS or JMal upon request.

5.3 Results

5.3.1 Principal component and linear discriminant analyses

We first performed a PCA to discriminate patients with ME/CFS and their subgroups from healthy controls (Figures 5.1A–C). A similar analysis was done for discriminating patients with an infectious trigger from the remaining patients (Figure 5.1D).

The proportion of variance explained by the first principal component varied from 35.4% (Figure 5.1D) to 44.6% (Figure 5.1C) referring to the comparisons between the two subgroups of ME/CFS patients, and between healthy controls and patients with non-infectious or unknown disease trigger, respectively. These high estimates suggested that different antibody levels were

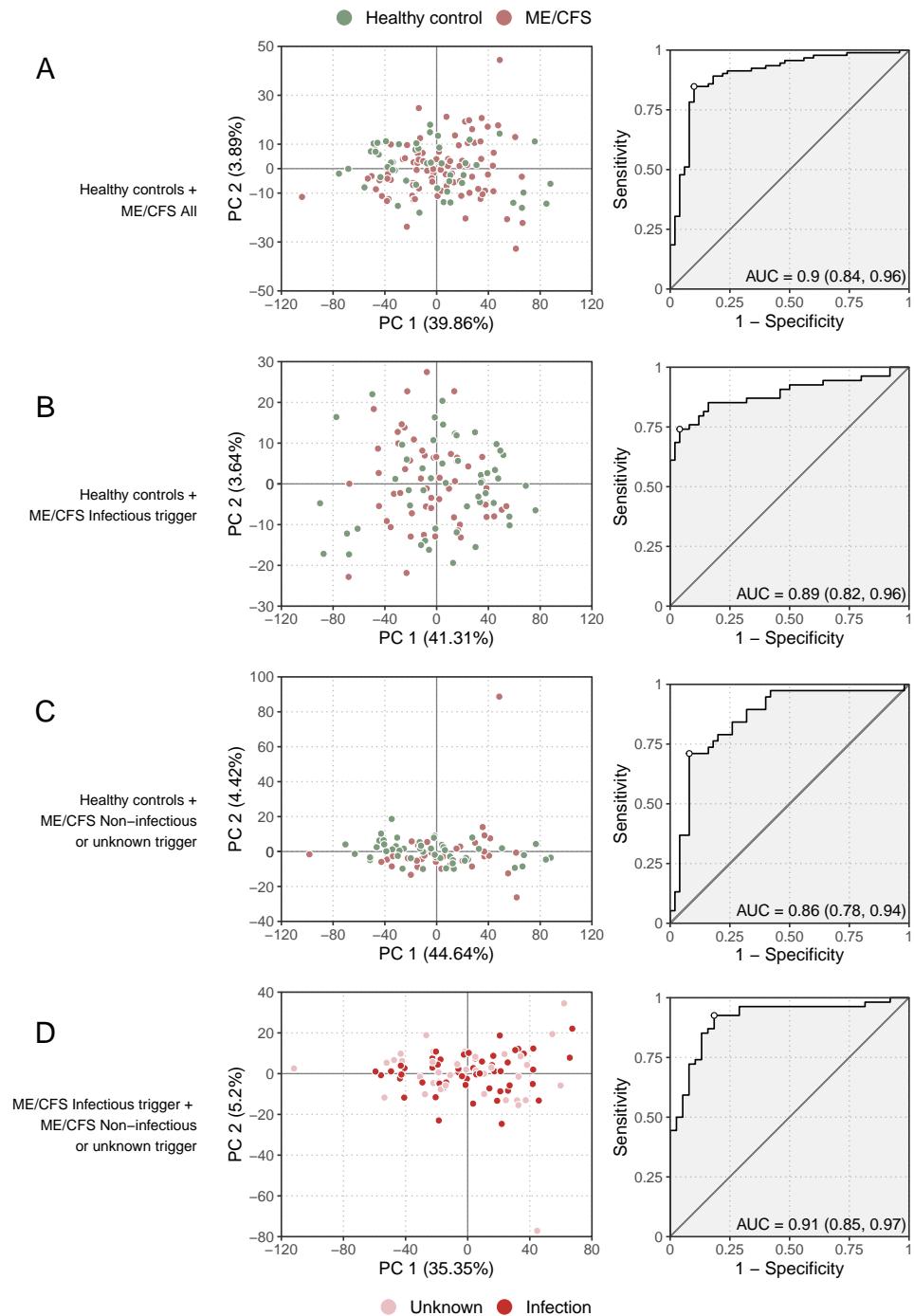


Figure 5.1: Preliminary multivariate analysis of the data. Scatterplots of the first two principal components (left plots) and the ROC curve and its AUC of the respective LDA (right plots) when comparing all the ME/CFS patients to healthy controls (A), ME/CFS patients with an infectious trigger to healthy controls (B), ME/CFS patients with a non-infectious or unknown trigger to healthy controls (C), and ME/CFS patients with an infectious trigger to the remaining patients (D). The percentage of the variance explained by each principal component is shown in each axis within brackets.

correlated with each other. This interpretation was confirmed by determining the distributions of Spearman's correlation coefficient between all possible pairs of antibodies using data from each study group (Supplementary Figure B.1). In particular, the antibody levels were positively correlated with each other with median correlation estimates of 0.56, 0.56, 0.40, and 0.48 for healthy controls, all the ME/CFS patients, ME/CFS patients with an infectious disease trigger, and the remaining ME/CFS patients, respectively. Interestingly, the median correlation estimate was decreased in ME/CFS patients with an infectious trigger when compared to other study groups. This finding suggested that the production of the antibodies against the EBV-derived antigens could be reduced in these patients when compared to healthy controls or patients with non-infectious or unknown disease trigger.

The visualisation of the first two components did not reveal a clear discrimination between healthy controls and ME/CFS patients (or their subgroups). To improve this analysis, we then performed different LDAs in search of a linear combination of the antibody measurements that could be used for disease diagnosis. The performance of the constructed classifiers ranged from 0.86 (Figure 5.1C) to 0.91 (Figure 5.1D) referring to the classification of healthy controls and ME/CFS patients with non-infectious or unknown disease trigger and the classification of the two subgroups of ME/CFS patients, respectively. Therefore, the results of this analysis indicate that the antibody data could discriminate different study groups.

5.3.2 Antibody-wide association analysis

The next step of the analysis was to identify specific antibody responses that could be used to discriminate the different study groups. With this purpose, we first determined the best “null” model among the logistic, probit, and complementary log-log models. All of them included age and gender and their interaction as covariates for each comparison between any two study groups (Supplementary Table B.2). The best “null” models were the following: (i) complementary log-log – comparison between healthy controls and all the ME/CFS patients (AUC = 0.574; 95% CI = [0.475, 0.672]); (ii) probit – comparison between healthy controls and ME/CFS patients with an infectious trigger (AUC = 0.606; 95% CI = [0.496, 0.715]); (iii)

complementary log-log – comparison between healthy controls and ME/CFS patients with a non-infectious or unknown trigger ($AUC = 0.556$; 95% CI = [0.429, 0.683]); and (iv) logit – comparison between the two subgroups of ME/CFS groups ($AUC = 0.596$; 95% CI = [0.471, 0.720]). The 95% confidence interval for the AUC of these null models included 0.50 and therefore, the respective predicted classification was consistent with a random guess. Such a result was in agreement with the age and gender matching between different study groups and healthy controls (Table 5.1).

We performed further antibody-wide association analyses controlling for a global FDR of 5%. The comparison between healthy controls and all the ME/CFS patients did not identify any significant antibody associations with the disease (Figure 5.2A). The top 5 antibodies, although not statistically significant, were EBNA6_0066, BLRF2_0005, EBNA4_0392, EBNA4_0497, and EBNA4_0529 (adjusted p-values = 0.181, 0.326, 0.326, 0.326, and 0.326, respectively).

When the comparison was limited to healthy controls and ME/CFS patients with an infectious trigger, we identified three significant antibodies related to the following antigens (Figure 5.2B): EBNA6_0066, EBNA6_0070, and EBNA4_0529 (adjusted p-values = 0.005, 0.005, and 0.038, respectively). The first two antigens were shared between AG876, B95.8, and GD1 strains, while the third one was derived from the B95.8 strain. We compared ME/CFS patients with non-infectious or unknown disease trigger to healthy controls, and found no significant differences in the antibody responses (Figure 5.2C). The same finding was obtained when we compared the two subgroups of ME/CFS patients (Figure 5.2D). The top 5 antibodies related to these analyses can be found in Supplementary Table B.3.

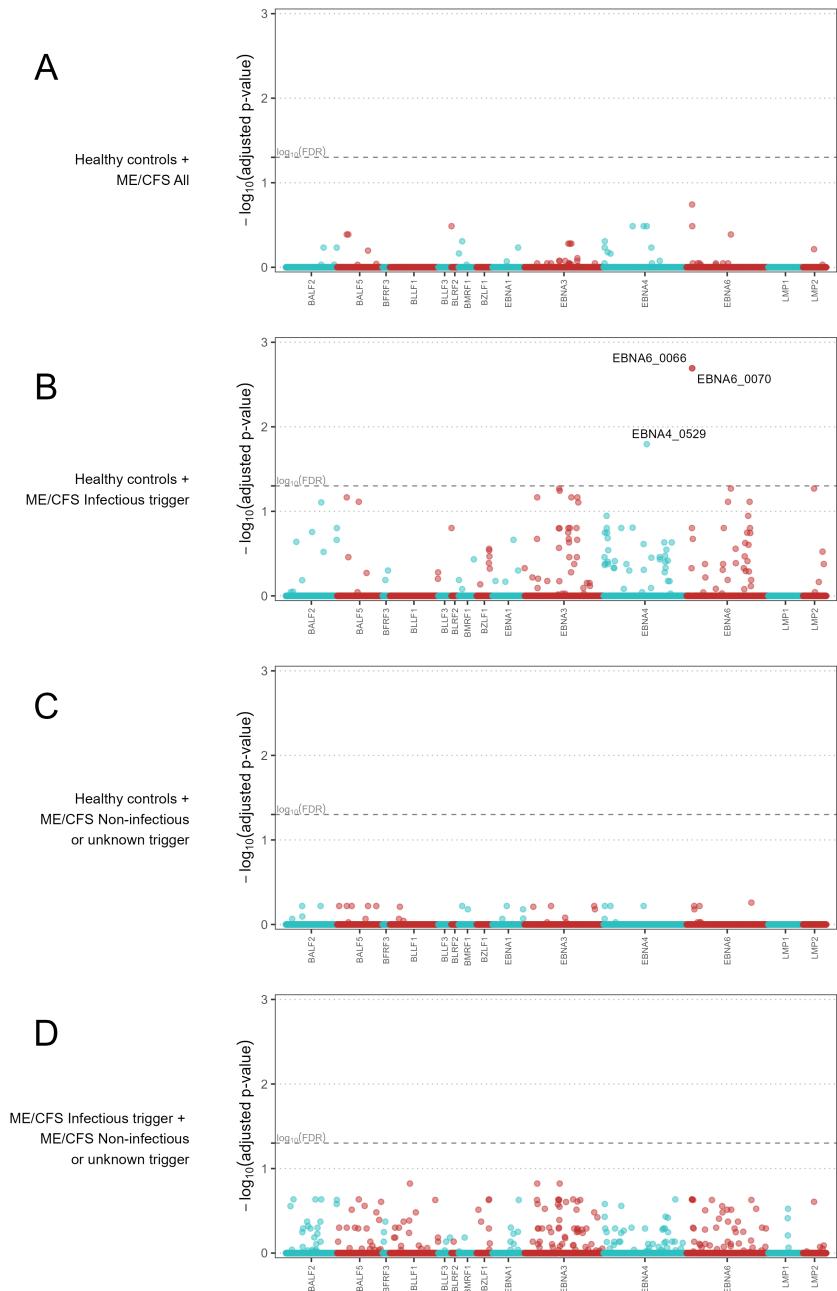


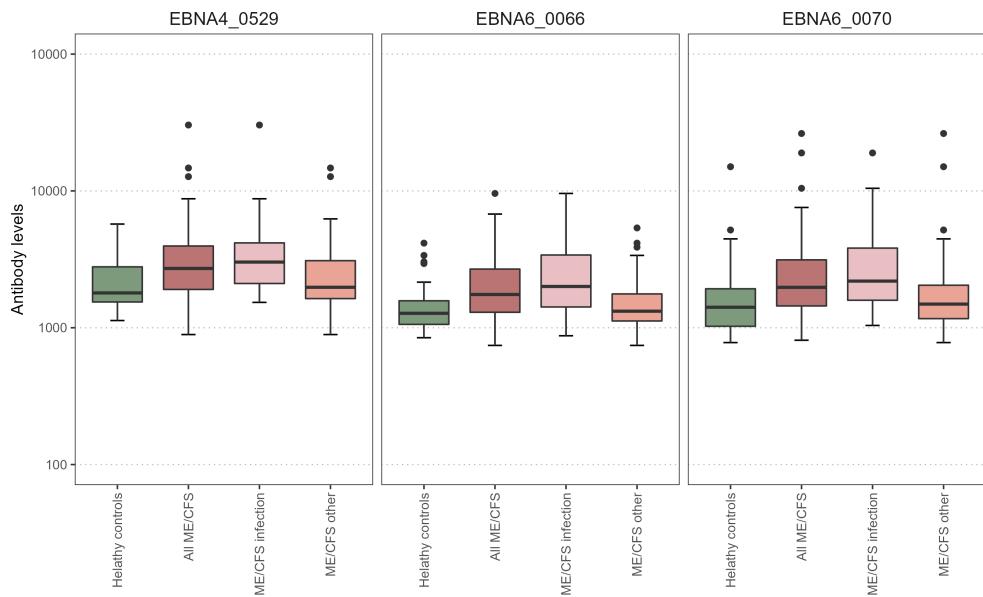
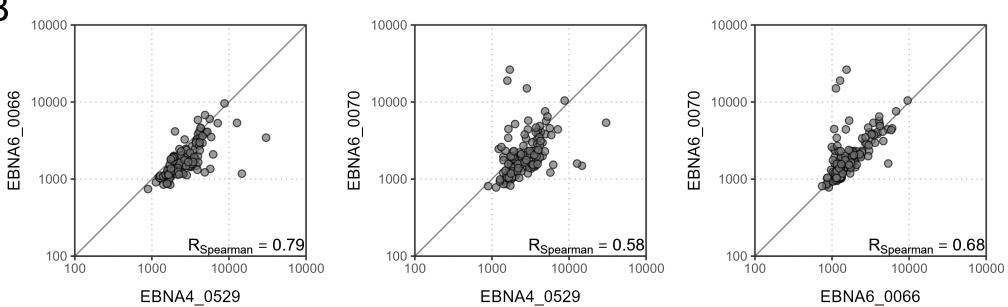
Figure 5.2: Antibody-wide association analyses when comparing all the ME/CFS patients to healthy controls (A), ME/CFS patients with an infectious trigger to healthy controls (B), ME/CFS patients with a noninfectious or unknown trigger to healthy controls (C), and ME/CFS patients with an infectious trigger to the remaining patients (D). The x-axes comprise each antibody while the y-axes represent the $-\log_{10}(\text{adjusted p-value})$ of the respective association. In the x-axes, the antibodies were ordered alphabetically first by the protein name and then by the starting point of the antigen within the protein. Adjusted p-values were calculated according to the Benjamini-Yekutieli procedure for a global FDR of 5% under the assumption of dependent data. Dashed line represents the threshold for statistical significance (i.e., $-\log_{10}(FDR) = 0.05$) and $-\log_{10}(\text{adjusted p-value}) > 1.30$ were considered statistically significant.

5.3.3 Analysis of candidate antigens for classifying ME/CFS patients with infectious trigger

We then analysed in detail the impact of the antibody levels against the three candidate antigens on the classification of ME/CFS patients with an infectious trigger. Antibody levels were increased in this subgroup of ME/CFS patients when compared to healthy controls (Figure 5.3A). The same evidence could not be found when comparing all the ME/CFS patients to healthy controls (Figure 5.3A). Data related to EBNA4_0529, EBNA6_0066 and EBNA6_0070 were significantly correlated with each other (Spearman's correlation coefficients higher than 0.58; Figure 5.3B). The correlation between the levels of antibodies against EBNA6_0066 and EBNA6_0070 could be explained by the fact that these two peptides are 15-mers overlapping 11 amino acids with each other (Loebel et al. 2017). In contrast, it was unclear why the levels of antibodies against EBNA4_0529 and EBNA6_0066 were highly correlated (Spearman's correlation coefficient = 0.79), considering that these antigens did not share a high sequence homology (Figure 5.3C).

Given the high correlation between antibody levels related to these antigens, a statistical redundancy was expected when using their data for patients' classification purpose. This redundancy was confirmed when the three candidate antibodies were included as covariates in the same model. A stepwise variable selection procedure led to the exclusion of the antibody levels related to EBNA6_0066 from the final classification model.

The final model included the main effects of antibodies to EBNA4_0529 and EBNA6_0070 and the two-way interaction of the latter with age and gender (Table 5.2). On the one hand, the \log_{10} -levels of antibodies related to EBNA4 increased the probability of being a patient (coefficient estimate = 2.25, standard error = 1.09). In particular, the odds of being a patient were estimated to increase ~ 9.5 ($e^{2.25}$) times per fold-change in the levels of these antibodies. On the other hand, the effects of antibody levels related to EBNA6_0070 on the probability of an individual being an ME/CFS patients were not so trivial to ascertain (Figure 5.4A). In particular, women with high EBNA6_0070 antibody levels showed an increasing estimated probability of being a patient with increasing age. In contrast, the probability profile of being

A**B****C**

Amino-acid sequences of candidate antigens

EBNA4_0529	P	Q	Q	P	R	A	G	R	R	G	P	C	V	F	F
EBNA6_0066	N	R	G	W	M	Q	R	I	R	R	R	R	R	R	R
EBNA6_0070	M	Q	R	I	R	R	R	R	R	R	R	A	A	L	S

Figure 5.3: Statistical analysis of the antibody levels related to EBNA4_0529, EBNA6_0066, and EBNA6_0070. (A) Boxplots of the data per study group. (B) Scatterplots and the respectively Spearman's correlation coefficients (R) in the whole dataset. (C) Amino acid sequences of EBNA4_0529, EBNA6_0066, and EBNA6_0070.

patient was different in men. In that case, younger men with low EBNA6_0070 antibody levels or older men with high EBNA6 antibody levels had a higher probability of being a patient.

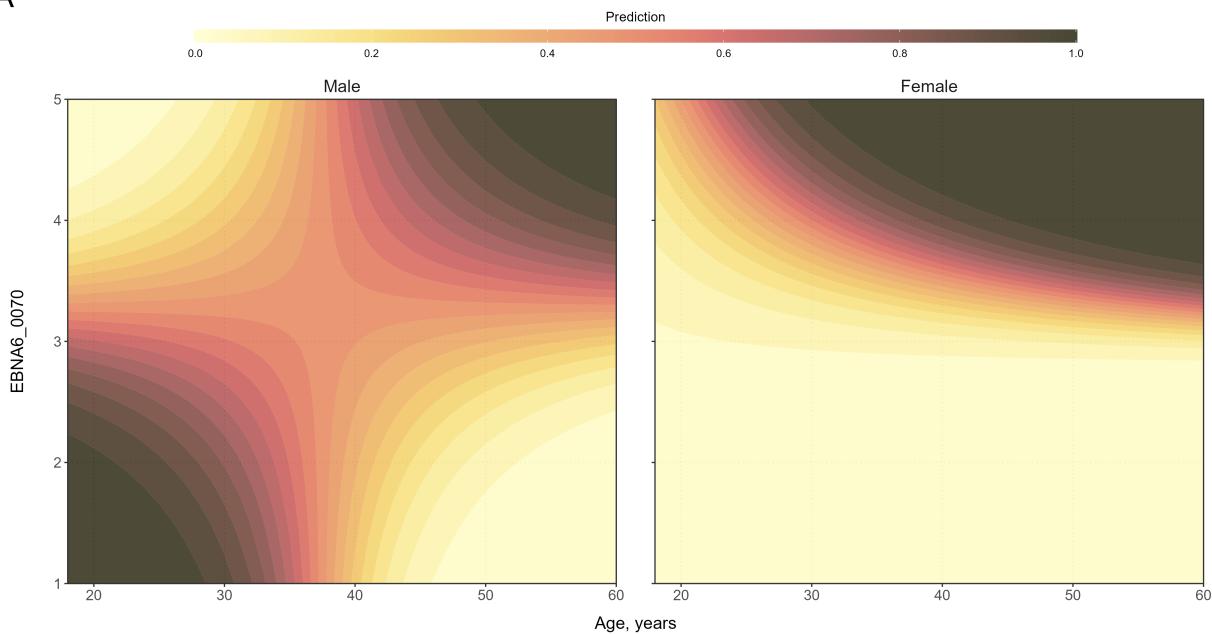
The AUC of the classification predicted by the final model was estimated at 0.835 with a 95% CI = (0.759, 0.911) (Figure 5.4B). This estimate suggested that the combination of these two antibodies together with age and gender could be used for the diagnosis of patients with an infectious trigger. The optimal sensitivity and specificity were estimated at 0.833 and 0.720, respectively. Therefore, ME/CFS patients were better discriminated than healthy controls by this model.

When the same classification model was applied to the whole cohort of ME/CFS patients, the AUC decreased to 0.731 with a 95% CI = (0.648, 0.814). This could be explained by the cohort of patients with a non-infectious or unknown trigger in which the performance of the classification model was close to a random guess (AUC = 0.583; 95% CI = [0.461, 0.705]).

Table 5.2: Estimates of the final complementary log-log model to discriminate ME/CFS patients with an infectious disease trigger from healthy controls.

Model term	Coefficient estimate (SE)	P-value
Intercept	10.67 (10.33)	0.302
Age (in years)	-0.49 (0.26)	0.060
Gender (Woman)	-17.33 (6.85)	0.011
EBNA4_0529	2.25 (1.09)	0.039
EBNA6_0070	-5.62 (3.09)	0.069
Age × Gender	0.07 (0.04)	0.070
Gender × EBNA6_0070	4.05 (1.75)	0.021
Age × EBNA6_0070	0.15 (0.08)	0.062

A



B

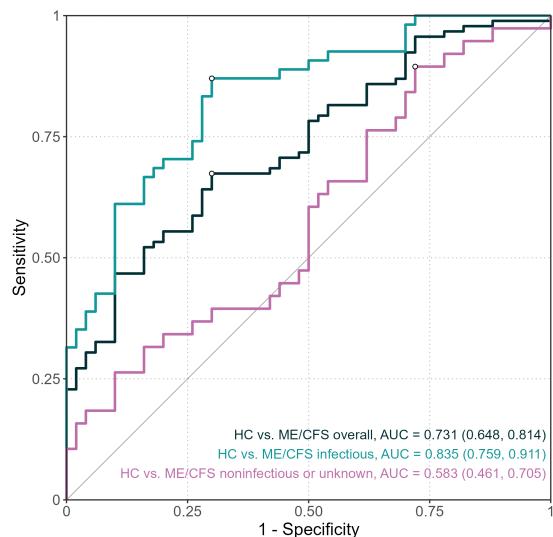


Figure 5.4: Analysis of the final classification model for predicting ME/CFS patients with an infectious trigger when compared to healthy controls. (A) Contour plots of the probability of being a patient as a function of age and EBNA6_0070 antibody levels, for men and women, respectively. The prediction values were calculated by fixing $\log_{10}(\text{EBNA4_0529})$ at the respective mean value. (B) ROC curves and the respective AUC (95% confidence interval shown within brackets) when using the model to compare different groups of ME/CFS patients to healthy controls.

5.4 Discussion

This study, based on previously published data, aimed to discover EBV-derived antigens that could elicit distinct antibody responses in ME/CFS patients when compared to healthy controls. The key finding was the identification of two candidate antigens inducing increased antibody responses in ME/CFS patients with an infectious trigger. The high sensitivity and specificity of our classification model including these antibodies suggest their potential for diagnosis of this subgroup of affected individuals. For ME/CFS patients without an infectious trigger, we could not find any antigens causing antibody responses that could be used for diagnostic purposes. This finding is in agreement with an extensive serological investigation of different herpesviruses in ME/CFS patients (Blomberg et al. 2019). This negative finding supports the hypothesis that EBV plays a role in the group of ME/CFS patients with an infectious trigger. In a subset of patients, infectious mononucleosis caused by primary EBV infection can be documented as a trigger (Domingues et al. 2021). In many others, no infection with a specific pathogen could be associated with the disease onset (Rasa et al. 2018). A tempting hypothesis from our finding is that EBV reactivation which can occur during other infections may play until now an underestimated role in triggering ME/CFS. In line with this concept, a recent study showed that EBV reactivation during Covid-19 is a risk factor for Long Covid which also includes ME/CFS (Su et al. 2022). Alternatively, the responses to the EBNA6 peptides are due to a cross-reactivity to other pathogens, as outlined below.

Other findings of this study pointed to three key challenges associated with the discovery of a biomarker. Firstly, it is difficult to identify a disease-specific biomarker for all the ME/CFS patients. Thus, given the heterogeneous nature of ME/CFS, it is pivotal to stratify patients adequately (Jason et al. 2005), based on age, gender, and disease trigger for biomarker discovery (Scheibenberg et al. 2017). In this regard, the identification of antibody patterns specific to ME/CFS patients with an infectious trigger was in agreement with other studies where significant results could be found for the same subgroup of patients (Steiner et al. 2020; Domingues et al. 2021; Szklarski et al. 2021). However, given the vast number of infectious agents associated with ME/CFS (Rasa et al. 2018; Blomberg et al. 2018), it is worth noting that

this subgroup of patients could be further subdivided according to the nature of the causative infection. In this regard, the data about the infectious agents that could have initiated ME/CFS are either inconclusive or simply based on self-reported history in most patients, as demonstrated by the data from the United Kingdom ME/CFS Biobank, where only a minority of patients had their infection confirmed with the lab test (Domingues et al. 2021).

Secondly, the final classification model included non-trivial statistical interactions of antibodies against EBNA6_0070 with both age and gender. This finding implies that significant interactions between candidate biomarkers and confounding factors might be overlooked by analysts or, even when tested, they are likely to be discarded due to the small sample sizes to detect them. The presence of these interactions might be yet another factor that contributes to the lack of reproducibility between biomarker studies on ME/CFS. A proposed strategy to overcome this limitation is to conduct more advanced statistical analyses including the application of machine learning techniques which intrinsically consider the complexity of a large set of clinical and biological data, as demonstrated in drug discovery (Gupta et al. 2021).

Thirdly, the interaction between the candidate antibodies against EBNA6_0070 and gender implied a remarkable distinct antibody signature between male and female patients. Again, this finding is in line with gender differences in immunity to viral infection (Jacobsen and Klein 2021). In particular, men have typically lower antibody responses when vaccinated and are more susceptible to infections than women (Aaby et al. 2020). In this regard, our study suggested that the higher probability of younger man being an ME/CFS patient is associated with lower levels of antibodies against the antigen EBNA6_0070. In contrast, female and male patients seemed to be at higher risk with higher antibodies at increasing age suggesting that at least a subset develop these antibody responses later in life. An implication of having a different antibody profiling between men and women is that analysis of each gender should be performed separately. At the same time, it is important to note that epidemiological data on ME/CFS suggested approximately a disease ratio of three women to one man (Nacul et al. 2011; Johnston et al. 2016; Chu et al. 2019). Therefore, if gender is an important stratification factor for biomarker discovery, studies should be designed toward a more balanced gender ratio. Similar sample sizes between male and female cohorts ensure comparable statistical

power when analysing data from each sex separately.

Both EBNA4_0529 and EBNA6_0070 antigens are derived from proteins whose genetic expression typically occurs during the EBV type III latency. Therefore, the acquisition of the respective antibodies might have occurred during initial B-cell transformation and immortalization. It could also be acquired slowly over time, given that the type III latency pattern can be detected sporadically in lymphoid follicles where EBV-infected B cells can proliferate and mimic a germinal center reaction program (Thorley-Lawson 2015). We can hypothesize from our data that both male and female patients developing higher antibody responses against this antigen later in life are at an increased risk of developing ME/CFS suggesting that reactivation of EBV plays a role. In male patients a subgroup with lower EBNA6 antibodies early in life is at risk of developing ME/CFS, too. Using the recent analytical framework of ME/CFS natural progression (Nacul et al. 2020), antibodies against these antigens are more likely to be biomarkers of patients suffering from ME/CFS more than 2 years of disease rather than the ones either in prodromal period or at early stages in line with our findings. Based on that assumption, these antibodies seemed more appropriate for diagnosing putative patients with delayed disease diagnosis rather than early suspected cases. However, it is known that the delay of ME/CFS diagnosis is a recurrent problem in the clinic (Bateman et al. 2021; Nacul et al. 2021). As such, we anticipate a higher utility of these antibodies when redeployed to real-world screening. Another practical implication of using these antibodies as biomarkers is the possibility of developing routine ELISA kits that can be standardised across different laboratories and easily scalable for large population screenings. Notwithstanding these promising practical expectations, it is important to emphasise that past studies also suggested potential disease biomarkers (Scheibenbogen et al. 2017) and, therefore, it is imperative to replicate the findings of this study with different cohorts of patients.

An interesting observation is that both EBNA6_0066 and EBNA6_0070 contain an arginine-repeat sequence. Such a sequence has homologies with putative epitopes from several human proteins (Sospedra et al. 2005). Such homologies suggest a potential molecular mimicry between the viral and human antigens. Molecular mimicry can trigger deleterious autoimmune responses as hypothesised for ME/CFS pathogenesis (Blomberg et al. 2018; Phelan et al. 2020).

Molecular mimicry between human and microbial antigens has been also hypothesised for several autoimmune diseases (Rojas et al. 2018), such as multiple sclerosis and rheumatoid arthritis, and Long Covid, whose patients share similar symptoms with ME/CFS ones (Moss-Morris and Chalder 2003; Gaber et al. 2014; Ali et al. 2017; Komaroff and Lipkin 2021). Interestingly, T cell clones recognising such arginine-repeat sequences were isolated from a patient with multiple sclerosis supporting our concept of epitope mimicry (Sospedra et al. 2005). Finally, arginine-repeat sequences are found in various other pathogens including enteroviruses and human papillomavirus which are also triggers of ME/CFS (Rasa et al. 2018).

Further we can hypothesise that peptides highly enriched in arginine residues might be particularly susceptible to citrullination, in which arginine residues are post-translationally converted to citrulline. These post-translational modifications occur during cell death under normal physiological conditions. However, under chronic inflammation, the accumulation of citrullinated (auto)antigens in inflamed sites might lead to deleterious autoimmune responses, thus, promoting the onset of different autoimmune diseases (Alghamdi et al. 2019). A potential cross-reactivity between microbial and citrullinated human antigens could also be a mechanism by which an autoimmune disease can be triggered. In rheumatoid arthritis, antibodies against EBNA-1 peptides were shown to cross-react with denatured collagen and keratin (Birkenfeld et al. 1990). However, in the present study, we could not find any antibodies against EBNA-1-derived peptides to be associated with ME/CFS. Interestingly, the serum levels of citrulline were reported to be elevated in ME/CFS patients when compared to healthy controls (Pall 2002). However, another study could not confirm this finding, but instead provided evidence for increased plasma levels of arginine residues (Naviaux et al. 2016).

Another source of antigen modification is the process of generating new and more immunogenic epitopes from ubiquitous molecules upon oxidative and nitrosative stress. In ME/CFS, IgM antibodies against several of these neoepitopes, including NO-Arginine, were increased in patients (Maes et al. 2006). In all of these possible scenarios, it is imperative to investigate the stability of this candidate biomarker antigen to post-translational modifications that could be occurred and eventually increased during the disease course.

In conclusion, this study identified two candidate antigens whose antibodies could be used

to identify ME/CFS patients with an infectious trigger. To strengthen our findings, two other cohorts of patients are currently studied, including the well-characterised ME/CFS patients with different disease triggers and healthy controls from the United Kingdom ME/CFS biobank (Domingues et al. 2021).

Bibliography

- P Aaby, C. S. Benn, K. L. Flanagan, S. L. Klein, T. R. Kollmann, D. J. Lynn, and F. Shann. The non-specific and sex-differential effects of vaccines. *Nature Reviews Immunology*, 20(8):464–470, Aug. 2020. ISSN 1474-1741. doi: 10.1038/s41577-020-0338-x.
- M. Alghamdi, D. Alasmari, A. Assiri, E. Mattar, A. A. Aljaddawi, S. G. Alattas, and E. M. Redwan. An Overview of the Intrinsic Role of Citrullination in Autoimmune Disorders. *Journal of Immunology Research*, 2019:7592851, 2019. ISSN 2314-7156. doi: 10.1155/2019/7592851.
- S. Ali, F. Matcham, K. Irving, and T. Chalder. Fatigue and psychosocial variables in autoimmune rheumatic disease and chronic fatigue syndrome: A cross-sectional comparison. *Journal of Psychosomatic Research*, 92:1–8, Jan. 2017. ISSN 1879-1360. doi: 10.1016/j.jpsychores.2016.11.002.
- L. Bateman, A. C. Bested, H. F. Bonilla, B. V. Chheda, L. Chu, J. M. Curtin, T. T. Dempsey, M. E. Dimmock, T. G. Dowell, D. Felsenstein, D. L. Kaufman, N. G. Klimas, A. L. Komaroff, C. W. Lapp, S. M. Levine, J. G. Montoya, B. H. Natelson, D. L. Peterson, R. N. Podell, I. R. Rey, I. S. Ruhoy, M. A. Vera-Nunez, and B. P. Yellman. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Essentials of Diagnosis and Management. *Mayo Clinic Proceedings*, 96(11):2861–2878, Nov. 2021. ISSN 1942-5546. doi: 10.1016/j.mayocp.2021.07.004.
- Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, Aug. 2001. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1013699998. URL <https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-4/The-control-of-the-false-discovery-rate-in-multiple-testing/10.1214/aos/1013699998.full>. Publisher: Institute of Mathematical Statistics.
- P Birkenfeld, N. Haratz, G. Klein, and D. Sulitzeanu. Cross-reactivity between the EBNA-1 p107 peptide, collagen, and keratin: implications for the pathogenesis of rheumatoid arthritis. *Clinical Immunology and Immunopathology*, 54(1):14–25, Jan. 1990. ISSN 0090-1229. doi: 10.1016/0090-1229(90)90002-8.
- K. Bjornevik, M. Cortese, B. C. Healy, J. Kuhle, M. J. Mina, Y. Leng, S. J. Elledge, D. W. Niebuhr, A. I. Scher, K. L. Munger, and A. Ascherio. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science (New York, N.Y.)*, 375(6578):296–301, Jan. 2022. ISSN 1095-9203. doi: 10.1126/science.abj8222.

- J. Blomberg, C.-G. Gottfries, A. Elfaitouri, M. Rizwan, and A. Rosén. Infection Elicited Autoimmunity and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: An Explanatory Model. *Frontiers in Immunology*, 9, Feb. 2018. doi: 10.3389/fimmu.2018.00229.
- J. Blomberg, M. Rizwan, A. Böhlin-Wiener, A. Elfaitouri, P. Julin, O. Zachrisson, A. Rosén, and C.-G. Gottfries. Antibodies to Human Herpesviruses in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Patients. *Frontiers in Immunology*, 10, Aug. 2019. doi: 10.3389/fimmu.2019.01946.
- L. Chu, I. J. Valencia, D. W. Garvert, and J. G. Montoya. Onset Patterns and Course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Pediatrics*, 7, Feb. 2019. doi: 10.3389/fped.2019.00012.
- J. M. Cliff, E. C. King, J.-S. Lee, N. Sepúlveda, A.-S. Wolf, C. Kingdon, E. Bowman, H. M. Dockrell, L. Nacul, E. Lacerda, and E. M. Riley. Cellular Immune Function in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Frontiers in Immunology*, 10, Apr. 2019. doi: 10.3389/fimmu.2019.00796. URL <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2019.00796/full>.
- T. D. Domingues, A. D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, F. Westermeier, C. Scheibenbogen, J. M. Cliff, L. Nacul, E. M. Lacerda, H. Mouríño, and N. Sepúlveda. Herpesviruses Serology Distinguishes Different Subgroups of Patients From the United Kingdom Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Biobank. *Frontiers in Medicine*, 8, July 2021. doi: 10.3389/fmed.2021.686736.
- T. A.-Z. K. Gaber, W. W. Oo, and H. Ringrose. Multiple Sclerosis/Chronic Fatigue Syndrome overlap: When two common disorders collide. *NeuroRehabilitation*, 35:529–534, 2014. ISSN 18786448, 10538135. doi: 10.3233/NRE-141146.
- R. Gupta, D. Srivastava, M. Sahu, S. Tiwari, R. K. Ambasta, and P. Kumar. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular Diversity*, 25(3):1315–1360, Aug. 2021. ISSN 1573-501X. doi: 10.1007/s11030-021-10217-3.
- I. Hickie, T. Davenport, D. Wakefield, U. Vollmer-Conna, B. Cameron, S. D. Vernon, W. C. Reeves, and A. Lloyd. Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study. *BMJ (Clinical research ed.)*, 333(7568):575, Sept. 2006. doi: 10.1136/bmj.38933.585764.ae.
- T. Holmøy, E. O. Kvale, and F. Vartdal. Cerebrospinal fluid CD4+ T cells from a multiple sclerosis patient cross-recognize Epstein-Barr virus and myelin basic protein. *Journal of Neurovirology*, 10(5):278–283, Oct. 2004. ISSN 1355-0284. doi: 10.1080/13550280490499524.
- G. Houen and N. H. Trier. Epstein-Barr Virus and Systemic Autoimmune Diseases. *Frontiers in Immunology*, 11, 2021. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2020.587380>.
- H. Jacobsen and S. L. Klein. Sex Differences in Immunity to Viral Infections. *Frontiers in Immunology*, 12:720952, 2021. ISSN 1664-3224. doi: 10.3389/fimmu.2021.720952.
- L. A. Jason, K. Corradi, S. Torres-Harding, R. R. Taylor, and C. King. Chronic Fatigue Syndrome: The Need for

- Subtypes. *Neuropsychology Review*, 15(1):29–58, Mar. 2005. doi: 10.1007/s11065-005-3588-2.
- S. C. Johnston, D. R. Staines, and S. M. Marshall-Gradisnik. Epidemiological characteristics of chronic fatigue syndrome/myalgic encephalomyelitis in Australian patients. *Clinical Epidemiology*, 8:97–107, May 2016. doi: 10.2147/CLEPS96797. URL <https://www.dovepress.com/epidemiological-characteristics-of-chronic-fatigue-nbspsyndromemyalgic-peer-reviewed-full>. Publisher: Dove Press.
- J. R. Kerr. Epstein-Barr Virus Induced Gene-2 Upregulation Identifies a Particular Subtype of Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *Frontiers in Pediatrics*, 7, 2019. ISSN 2296-2360. URL <https://www.frontiersin.org/articles/10.3389/fped.2019.00059>.
- A. L. Komaroff and W. I. Lipkin. Insights from myalgic encephalomyelitis/chronic fatigue syndrome may help unravel the pathogenesis of postacute COVID-19 syndrome. *Trends in Molecular Medicine*, 27(9):895–906, Sept. 2021. ISSN 1471-499X. doi: 10.1016/j.molmed.2021.06.002.
- D. Koo. Chronic fatigue syndrome. A critical appraisal of the role of Epstein-Barr virus. *Western Journal of Medicine*, 150(5):590–596, May 1989. ISSN 0093-0415. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1026689/>.
- J.-S. Lee, E. M. Lacerda, L. Nacul, C. C. Kingdon, J. Norris, S. O’Boyle, C. h. Roberts, L. Palla, E. M. Riley, and J. M. Cliff. Salivary DNA Loads for Human Herpesviruses 6 and 7 Are Correlated With Disease Phenotype in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Medicine*, 8:656692, Aug. 2021. ISSN 2296-858X. doi: 10.3389/fmed.2021.656692. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8378328/>.
- A. M. Lerner, M. E. Ariza, M. Williams, L. Jason, S. Beqaj, J. T. Fitzgerald, S. Lemeshow, and R. Glaser. Antibody to Epstein-Barr Virus Deoxyuridine Triphosphate Nucleotidohydrolase and Deoxyribonucleotide Polymerase in a Chronic Fatigue Syndrome Subset. *PLOS ONE*, 7(11):e47891, Nov. 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0047891. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0047891>. Publisher: Public Library of Science.
- M. Loebel, K. Strohschein, C. Giannini, U. Koelsch, S. Bauer, C. Doeblis, S. Thomas, N. Unterwalder, V. v. Baehr, P. Reinke, M. Knops, L. G. Hanitsch, C. Meisel, H.-D. Volk, and C. Scheibenbogen. Deficient EBV-Specific B- and T-Cell Response in Patients with Chronic Fatigue Syndrome. *PLoS ONE*, 9(1):e85387, Jan. 2014. doi: 10.1371/journal.pone.0085387.
- M. Loebel, M. Eckey, F. Sotzny, E. Hahn, S. Bauer, P. Grabowski, J. Zerweck, P. Holenya, L. G. Hanitsch, K. Wittke, P. Borchmann, J.-U. Rüffer, F. Hiepe, K. Ruprecht, U. Behrends, C. Meindl, H.-D. Volk, U. Reimer, and C. Scheibenbogen. Serological profiling of the EBV immune response in Chronic Fatigue Syndrome using a peptide microarray. *PLOS ONE*, 12(6):e0179124, Dec. 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0179124. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0179124>. Pub-

- lisher: Public Library of Science.
- J. D. Lünemann, I. Jelcić, S. Roberts, A. Lutterotti, B. Tackenberg, R. Martin, and C. Münz. EBNA1-specific T cells from patients with multiple sclerosis cross react with myelin antigens and co-produce IFN-gamma and IL-2. *The Journal of Experimental Medicine*, 205(8):1763–1773, Aug. 2008. ISSN 1540-9538. doi: 10.1084/jem.20072397.
- M. López-Ratón, M. X. Rodríguez-Álvarez, C. C. Suárez, and F. G. Sampedro. OptimalCutpoints: An R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software*, 61(8):1–36, 2014. doi: 10.18637/jss.v061.i08.
- M. Maes, I. Mihaylova, and J.-C. Leunis. Chronic fatigue syndrome is accompanied by an IgM-related immune response directed against neopitopes formed by oxidative or nitrosative damage to lipids and proteins. *Neuro endocrinology letters*, 27(5):615–621, Oct. 2006. ISSN 2354-4716.
- J. a. Malato, L. Graça, L. Nacul, E. M. Lacerda, , and N. Sepúlveda. Statistical challenges of investigating a disease with a complex diagnosis. In P. Milheiro, A. Pacheco, B. de Sousa, I. F. Alves, I. Pereira, M. J. a. Polidoro, and S. Ramos, editors, *Estatística: Desafios transversais às ciências com dados*, Eds. Sociedade portuguesa de estatística, pages 153–167. Sociedade Portuguesa de Estatística, 2021. ISBN 978-972-8890-47-6. URL <https://www.spestatistica.pt/storage/app/uploads/public/609/28f/6d0/60928f6d08a0c016386627.pdf>.
- G. Morris and M. Maes. Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics. *BMC Medicine*, 11(1), Sept. 2013. doi: 10.1186/1741-7015-11-205.
- R. Moss-Morris and T. Chalder. Illness perceptions and levels of disability in patients with chronic fatigue syndrome and rheumatoid arthritis. *Journal of Psychosomatic Research*, 55(4):305–308, Oct. 2003. ISSN 0022-3999. doi: 10.1016/s0022-3999(03)00013-8.
- L. Nacul, S. O’Boyle, L. Palla, F. E. Nacul, K. Mudie, C. C. Kingdon, J. M. Cliff, T. G. Clark, H. M. Dockrell, and E. M. Lacerda. How Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) Progresses: The Natural History of ME/CFS. *Frontiers in Neurology*, 11, 2020. ISSN 1664-2295. doi: <https://doi.org/10.3389/fneur.2020.00826>. URL <https://www.frontiersin.org/articles/10.3389/fneur.2020.00826>.
- L. Nacul, F. J. Authier, C. Scheibenbogen, L. Lorusso, I. B. Helland, J. A. Martin, C. A. Sirbu, A. M. Mengshoel, O. Polo, U. Behrends, H. Nielsen, P. Grabowski, S. Sekulic, N. Sepulveda, F. Estévez-López, P. Zalewski, D. F. H. Pheby, J. Castro-Marrero, G. K. Sakkas, E. Capelli, I. Brundslund, J. Cullinan, A. Krumina, J. Bergquist, M. Murovska, R. C. W. Vermuelen, and E. M. Lacerda. European Network on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (EUROMENE): Expert Consensus on the Diagnosis, Service Provision, and Care of People with ME/CFS in Europe. *Medicina*, 57(5):510, May 2021. ISSN 1648-9144. doi: 10.3390/medicina57050510. URL <https://www.mdpi.com/1648-9144/57/5/510>.

- L. C. Nacul, E. M. Lacerda, D. Pheby, P. Campion, M. Molokhia, S. Fayyaz, J. C. Leite, F. Poland, A. Howe, and M. L. Drachler. Prevalence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in three regions of England: a repeated cross-sectional study in primary care. *BMC Medicine*, 9(1):91, Dec. 2011. ISSN 1741-7015. doi: 10.1186/1741-7015-9-91. URL <http://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-9-91>.
- R. K. Naviaux, J. C. Naviaux, K. Li, A. T. Bright, W. A. Alaynick, L. Wang, A. Baxter, N. Nathan, W. Anderson, and E. Gordon. Metabolic features of chronic fatigue syndrome. *Proceedings of the National Academy of Sciences*, 113(37):E5472–E5480, Sept. 2016. doi: 10.1073/pnas.1607571113. URL <https://www.pnas.org/doi/full/10.1073/pnas.1607571113>. Publisher: Proceedings of the National Academy of Sciences.
- M. L. Pall. Levels of Nitric Oxide Synthase Product Citrulline Are Elevated in Sera of Chronic Fatigue Syndrome Patients. *Journal Of Chronic Fatigue Syndrome*, 10(3-4):37–41, Jan. 2002. ISSN 1057-3321. doi: 10.1300/J092v10n03_04. URL https://doi.org/10.1300/J092v10n03_04. Publisher: Taylor & Francis _eprint: https://doi.org/10.1300/J092v10n03_04.
- J. Phelan, A. D. Grabowska, and N. Sepúlveda. A potential antigenic mimicry between viral and human proteins linking Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) with autoimmunity: The case of HPV immunization. *Autoimmunity Reviews*, 19(4):102487, Apr. 2020. doi: 10.1016/j.autrev.2020.102487.
- S. Rasa, Z. Nora-Kruk, N. Henning, E. Eliassen, E. Shikova, T. Harrer, C. Scheibenbogen, M. Murovska, B. K. Prusty, and the European Network on ME/CFS (EUROMENE). Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 16(1):268, Oct. 2018. ISSN 1479-5876. doi: 10.1186/s12967-018-1644-y. URL <https://doi.org/10.1186/s12967-018-1644-y>.
- M. C. Rivera, C. Mastronardi, C. Silva-Aldana, M. Arcos-Burgos, and B. Lidbury. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Comprehensive Review. *Diagnostics*, 9(3):91, Aug. 2019. doi: 10.3390/diagnostics9030091.
- X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12:77, 2011.
- M. Rojas, P. Restrepo-Jiménez, D. M. Monsalve, Y. Pacheco, Y. Acosta-Ampudia, C. Ramírez-Santana, P. S. C. Leung, A. A. Ansari, M. E. Gershwin, and J.-M. Anaya. Molecular mimicry and autoimmunity. *Journal of Autoimmunity*, 95:100–123, Dec. 2018. ISSN 1095-9157. doi: 10.1016/j.jaut.2018.10.012.
- M. Ruiz-Pablos, B. Paiva, R. Montero-Mateo, N. Garcia, and A. Zabaleta. Epstein-Barr Virus and the Origin of Myalgic Encephalomyelitis or Chronic Fatigue Syndrome. *Frontiers in Immunology*, 12, 2021. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2021.656797>.
- C. Scheibenbogen, H. Freitag, J. Blanco, E. Capelli, E. Lacerda, J. Authier, M. Meeus, J. C. Marrero, Z. Nora-Kruk, E. Oltra, and others. The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE. *Journal of translational medicine*, 15(1):162, 2017. doi: <https://doi.org/10.1186/1475-2867-15-162>.

- s12967-017-1263-z.
- N. Sepúlveda. Impact of genetic variation on the molecular mimicry between Anoctamin-2 and Epstein-Barr virus nuclear antigen 1 in Multiple Sclerosis. *Immunology Letters*, 238:29–31, Oct. 2021. ISSN 1879-0542. doi: 10.1016/j.imlet.2021.07.007.
- N. Sepúlveda, J. Carneiro, E. Lacerda, and L. Nacul. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome as a Hyper-Regulated Immune System Driven by an Interplay Between Regulatory T Cells and Chronic Human Herpesvirus Infections. *Frontiers in Immunology*, 10, 2019. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02684>.
- C. Shannon-Lowe, A. B. Rickinson, and A. I. Bell. Epstein–Barr virus-associated lymphomas. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1732):20160271, Sept. 2017. doi: 10.1098/rstb.2016.0271. URL <https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0271>. Publisher: Royal Society.
- E. Shikova, V. Reshkova, Kumanova, S. Raleva, D. Alexandrova, N. Capo, M. Murovska, and E. N. o. ME/CFS (EU-ROMENE). Cytomegalovirus, Epstein-Barr virus, and human herpesvirus-6 infections in patients with myalgic \cyrchar\cyrencephalomyelitis/chronic fatigue syndrome. *Journal of Medical Virology*, Mar. 2020. doi: 10.1002/jmv.25744.
- M. Sospedra, Y. Zhao, H. zur Hausen, P. A. Muraro, C. Hamashin, E.-M. de Villiers, C. Pinilla, and R. Martin. Recognition of conserved amino acid motifs of common viruses and its role in autoimmunity. *PLoS pathogens*, 1(4):e41, Dec. 2005. ISSN 1553-7366. doi: 10.1371/journal.ppat.0010041.
- S. Steiner, S. C. Becker, J. Hartwig, F. Sotzny, S. Lorenz, S. Bauer, M. Löbel, A. B. Stittrich, P. Grabowski, and C. Scheibenbogen. Autoimmunity-Related Risk Variants in PTPN22 and CTLA4 Are Associated With ME/CFS With Infectious Onset. *Frontiers in Immunology*, 11, 2020. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2020.00578>.
- Y. Su, D. Yuan, D. G. Chen, R. H. Ng, K. Wang, J. Choi, S. Li, S. Hong, R. Zhang, J. Xie, S. A. Kornilov, K. Scherler, A. J. Pavlovitch-Bedzyk, S. Dong, C. Lausted, I. Lee, S. Fallen, C. L. Dai, P. Baloni, B. Smith, V. R. Duvvuri, K. G. Anderson, J. Li, F. Yang, C. J. Duncombe, D. J. McCulloch, C. Rostomily, P. Troisch, J. Zhou, S. Mackay, Q. DeGottardi, D. H. May, R. Taniguchi, R. M. Gittelman, M. Klinger, T. M. Snyder, R. Roper, G. Wojciechowska, K. Murray, R. Edmark, S. Evans, L. Jones, Y. Zhou, L. Rowen, R. Liu, W. Chour, H. A. Algren, W. R. Berrington, J. A. Wallick, R. A. Cochran, M. E. Micikas, ISB-Swedish COVID-19 Biobanking Unit, T. Wrin, C. J. Petropoulos, H. R. Cole, T. D. Fischer, W. Wei, D. S. B. Hoon, N. D. Price, N. Subramanian, J. A. Hill, J. Hadlock, A. T. Magis, A. Ribas, L. L. Lanier, S. D. Boyd, J. A. Bluestone, H. Chu, L. Hood, R. Gottardo, P. D. Greenberg, M. M. Davis, J. D. Goldman, and J. R. Heath. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell*, 185(5): 881–895.e20, Mar. 2022. ISSN 1097-4172. doi: 10.1016/j.cell.2022.01.014.
- M. Szklarski, H. Freitag, S. Lorenz, S. C. Becker, F. Sotzny, S. Bauer, J. Hartwig, H. Heidecke, K. Wittke, C. Kedor,

- L. G. Hanitsch, P. Grabowski, N. Sepúlveda, and C. Scheibenbogen. Delineating the Association Between Soluble CD26 and Autoantibodies Against G-Protein Coupled Receptors, Immunological and Cardiovascular Parameters Identifies Distinct Patterns in Post-Infectious vs. Non-Infection-Triggered Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Immunology*, 12, Apr. 2021. doi: 10.3389/fimmu.2021.644548.
- K. Tengvall, J. Huang, C. Hellström, P. Kammer, M. Biström, B. Ayoglu, I. Lima Bomfim, P. Stridh, J. Butt, N. Brenner, A. Michel, K. Lundberg, L. Padyukov, I. E. Lundberg, E. Svenungsson, I. Ernberg, S. Olafsson, A. T. Dilthey, J. Hillert, L. Alfredsson, P. Sundström, P. Nilsson, T. Waterboer, T. Olsson, and I. Kockum. Molecular mimicry between Anoctamin 2 and Epstein-Barr virus nuclear antigen 1 associates with multiple sclerosis risk. *Proceedings of the National Academy of Sciences of the United States of America*, 116(34):16955–16960, Aug. 2019. ISSN 1091-6490. doi: 10.1073/pnas.1902623116.
- D. A. Thorley-Lawson. EBV Persistence—Introducing the Virus. *Current Topics in Microbiology and Immunology*, 390(Pt 1):151–209, 2015. ISSN 0070-217X. doi: 10.1007/978-3-319-22822-8_8.
- W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Statistics and Computing. Springer, New York, NY, 2002. ISBN 978-1-4419-3008-8 978-0-387-21706-2. doi: 10.1007/978-0-387-21706-2. URL <http://link.springer.com/10.1007/978-0-387-21706-2>.
- J. Wang, I. Jelcic, L. Mühlenbruch, V. Haunerding, N. C. Toussaint, Y. Zhao, C. Cruciani, W. Faigle, R. Naghavian, M. Foege, T. M. C. Binder, T. Eiermann, L. Opitz, L. Fuentes-Font, R. Reynolds, W. W. Kwok, J. T. Nguyen, J.-H. Lee, A. Lutterotti, C. Münz, H.-G. Rammensee, M. Hauri-Hohl, M. Sospedra, S. Stevanovic, and R. Martin. HLA-DR15 Molecules Jointly Shape an Autoreactive T Cell Repertoire in Multiple Sclerosis. *Cell*, 183(5):1264–1281.e20, Nov. 2020. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2020.09.054. URL [https://www.cell.com/cell/abstract/S0092-8674\(20\)31251-4](https://www.cell.com/cell/abstract/S0092-8674(20)31251-4). Publisher: Elsevier.
- K. W. Wucherpfennig and J. L. Strominger. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell*, 80(5):695–705, Mar. 1995. ISSN 0092-8674. doi: 10.1016/0092-8674(95)90348-8.

Chapter 6

Association analysis between symptomology and herpesvirus IgG antibody concentrations in ME/CFS and MS

T.D. Domingues¹, J. Malato¹, A.D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, P. Biecek, L. Graça, H. Mourão, C. Scheibenbogen, F. Westermeier, L. Nacul, J.M. Cliff, E. Lacerda, and N. Sepúlveda. Association analysis between symptomology and herpesvirus IgG antibody concentrations in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and multiple sclerosis. *Heliyon*. 2023; 9(7):E18250. doi: <https://doi.org/10.1016/j.heliyon.2023.e18250>.

Abstract

Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) and multiple sclerosis (MS) are two complex and multifactorial diseases whose patients experience persistent fatigue, cognitive impairment, among other shared symptoms. The onset of these diseases has also been linked to acute herpesvirus infections or their reactivations. In this work, we re-analysed a previously-described dataset related to IgG antibody responses to 6 herpesviruses (CMV–cytomegalovirus; EBV–Epstein-Barr virus; HHV6–human herpesvirus-6; HSV1 and HSV2–herpes simplex virus-1 and -2, respectively; VZV–varicella-zoster virus) from the United Kingdom ME/CFS biobank. The primary goal was to report the underlying symptomology and its association with herpesvirus IgG antibodies using data from 4 disease-trigger-based subgroups of ME/CFS patients ($n = 222$) and patients with MS ($n = 46$). The secondary objective was to assess whether serological data could distinguish ME/CFS and its subgroup from MS using a SuperLearner (SL) algorithm. There was evidence for a significant negative association between temporary eyesight disturbance and CMV antibody concentrations and for a significant positive association between bladder problems and EBV antibody concentrations in the MS group. In the

¹Authors contributed equally to this work.

ME/CFS or its subgroups, the most significant antibody-symptom association was obtained for increasing HSV1 antibody concentration and brain fog, a finding in line with a negative impact of HSV1 exposure on cognitive outcomes in both healthy and disease conditions. There was also evidence for a higher number of significant antibody-symptom associations in the MS group than in the ME/CFS group. When we combined all the serological data in an SL algorithm, we could distinguish three ME/CFS subgroups (unknown disease trigger, non-infection trigger, and an infection disease trigger confirmed in the lab at the time of the event) from the MS group. However, we could not find the same for the remaining ME/CFS group (related to an unconfirmed infection disease). In conclusion, IgG antibody data explains more the symptomatology of MS patients than the one of ME/CFS patients. Given the fluctuating nature of symptoms in ME/CFS patients, the clinical implication of these findings remains to be determined with a longitudinal study. This study is likely to ascertain the robustness of the associations during natural disease course.

Keywords: Enzyme-linked immunosorbent assay; Epstein-barr virus; Cytomegalovirus; Human herpesvirus-6; Varicella-zoster virus; Herpes simplex virus-1 and -2; SuperLearner; United Kingdom ME/CFS biobank

6.1 Introduction

Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) is a widely neglected disease characterised by persistent fatigue, post-exertional malaise (PEM), unrefreshing sleep, among other symptoms related to multiple body systems (Rivera et al. 2019). Despite the research efforts (Scheibenbogen et al. 2017), there are no objective biomarkers for the diagnosis and prognosis of ME/CFS. The absence of this key clinical tool delays disease diagnosis and subsequent treatment (Nacul et al. 2021). It also slows down research progress due to lack of finding's reproducibility across studies (Nacul et al. 2019; Malato et al. 2023). Notwithstanding all these problems, there is a growing body of evidence for an autoimmune component for the origin and chronicity of the disease, especially in patients with an acute infection at their disease onset (Sotzny et al. 2018; Blomberg et al. 2018; Sepúlveda et al. 2019; Steiner et al. 2020). The main candidate proteins for this deleterious autoimmune phenomenon are the adrenergic receptors (Loebel et al. 2016; Bynke et al. 2020; Freitag et al. 2021). However, other human proteins, such as Anoctamin-2 and thyroid peroxidase, have also been suggested (Loebel et al. 2017; Sepúlveda 2021).

The autoimmune hypothesis for the aetiology of ME/CFS directed research efforts towards the identification of key differences between ME/CFS and different autoimmune diseases, including multiple sclerosis (MS) (Ramos et al. 2016; Loebel et al. 2017). In this regard, patients with MS were deemed an important disease control group given that they experience chronic fatigue as a major manifestation of their disease. These patients and the ME/CFS ones also share some neurological symptoms, such as brain fog, memory loss, cognitive impairment, and photosensitivity/photophobia (Morris and Maes 2013). In addition, the root cause of both diseases has been linked to the infections by herpesviruses (Rasa et al. 2018; Ariza 2021; Sedighi et al. 2022; Rasa-Dzelzkaleja et al. 2023; Khalesi et al. 2023), such as Epstein-Barr virus (EBV) (Ruiz-Pablos et al. 2021; Bjornevik et al. 2022; Soldan and Lieberman 2023) and human herpesvirus-6 (HHV6) (Engdahl et al. 2019; Lee et al. 2021; Lundström and Gustafsson 2022; Kasimir et al. 2022). Finally, it was recently hypothesised that ME/CFS and MS share reduced craniospinal compliance and dilated pressured bridging cortical veins (Bateman and Bateman 2024).

The similarity between these two diseases also prompted the creation of the United Kingdom ME/CFS biobank (UKMEB), a large sample resource for the research community, where patients with MS were included as a disease control group (Lacerda et al. 2017; 2018). Until now, most of the UKMEB-based studies compared ME/CFS patients to healthy controls (Rodrigues et al. 2019; Almenar-Pérez et al. 2020; Blauensteiner et al. 2021; Bertinat et al. 2022; Vogl et al. 2022; González-Cebrián et al. 2022) with a few exceptions where the MS group was also included in the analysis (Jain et al. 2017; Lacerda et al. 2019a; Melvin et al. 2019; Cliff et al. 2019). For example, one study reported significant differences in the frequency of several immune-cell populations between MS and ME/CFS patients (Cliff et al. 2019); however, some of these differences were not found when comparing these groups to healthy controls. The same study showed that MS patients had a significantly higher seroprevalence to EBV-derived nuclear antigen-1 (EBNA1) antigen when comparing to both healthy controls and ME/CFS patients divided according to their disease severity. Another study reported evidence for similar average concentration of the stress-related growth/differentiation factor 15 peptide between MS and ME/CFS group (Melvin et al. 2019). Therefore, it is important to increase current

knowledge about the pathological differences between ME/CFS and MS using data from the UKMEB.

In a previous study, we divided the cohort of ME/CFS patients from the UKMEB into 4 subgroups according to their putative disease triggers and compared them to healthy controls in terms of IgG antibody responses to 6 herpesviruses (Domingues et al. 2021). We found reductions in antibody concentrations to EBV in the ME/CFS subgroup without a putative infection trigger and to CMV in the ME/CFS subgroup with a confirmed infection at disease onset. In the present study, we analysed symptoms data from the UKMEB using the MS cohort as the control group and assessed the respective association with the same herpesvirus IgG antibody data mentioned above. We also integrated the IgG antibody data in a SuperLearner (SL) algorithm with the objective of assessing the classification power to distinguish each ME/CFS subgroup from the MS one. Overall, our analyses aimed at providing new data on what differentiates ME/CFS and its subgroups from MS. Given that there are more published studies related to the role of herpesviruses on MS than on ME/CFS, most of the findings reported here are not particularly novel (e.g., high EBV IgG antibody concentrations) for this disease.

6.2 Materials and methods

6.2.1 Study participants

The study participants were 222 adult patients with ME/CFS and 46 adult patients with MS whose serological data were available from the UKMEB. As in our previous study (Domingues et al. 2021), we divided patients with ME/CFS into the following four subgroups according to the disease trigger: S0—an unknown trigger ($n = 42$); S1—a non-infectious trigger ($n = 42$); S2—an infection that was not evaluated by a lab test at the time of the event ($n = 92$); and S3—an infection that was confirmed by a lab test at the time of the event ($n = 46$). The diagnosis of these patients complied with either the 1994 US Center for Disease Control and Prevention criteria (CDC-1994) (Fukuda et al. 1994) or the 2003 Canadian Consensus Criteria

(CCC-2003) (Carruthers et al. 2003). All putative study participants were excluded if they had (i) taken antiviral medication or drugs known to alter immune function in the preceding 3 months; (ii) had any vaccinations in the preceding 3 months; (iii) had a history of acute and chronic infectious diseases, such as hepatitis B and C, tuberculosis, HIV (but not herpes virus or other retrovirus infection); (iv) another chronic disease such as cancer, coronary heart disease, or uncontrolled diabetes; (v) a severe mood disorder; (vi) been pregnant or breastfeeding in the preceding 12 months; or (vii) were morbidly obese ($BMI \geq 40$). Further information on the UKMEB can be found elsewhere (Lacerda et al. 2017; 2018). The severity of ME/CFS patients was divided into mild/moderate and severely affected (home or bed bound). MS patients had a previous diagnosis made by neurologists from the UK National Health System. No assessment of disease severity was made for these patients.

6.2.2 Symptomology assessment

At the recruitment, all the study participants were asked to report their disease duration and disease course. They also answered a symptom's assessment questionnaire on the presence or absence of 58 different symptoms in the previous 7 days (Supplementary Table C.1). The symptoms were related to the following domains: autonomic ($n = 10$); immunological ($n = 9$); neuroendocrine ($n = 6$); neurocognitive ($n = 19$); pain ($n = 6$); post-exertional malaise ($n = 6$); and sleep function ($n = 2$). The frequency of each symptom per study group can be found in Supplementary Table C.2.

6.2.3 Herpesvirus IgG antibodies

As mentioned in the Introduction, we re-analysed the same herpesvirus IgG antibody concentration dataset, as described in the original and follow-up study (Cliff et al. 2019; Domingues et al. 2021). Briefly, antibody quantification of each participant was performed at recruitment using different commercial ELISAs. Data referred to plasma concentrations (in arbitrary units per millilitre, U/ml) of IgG antibodies against human cytomegalovirus (CMV), EBV (EBNA1 and EBV-derived Viral Capsid Antigen, EBV-VCA), HHV6, herpes simplex virus-1 (HSV1), her-

pes simplex virus-2 (HSV2), and varicella-zoster virus (VZV). In the manufacturer's protocols, the suggested cutoff values for seropositivity were 12 U/ml for HSV1, HSV 2, VZV, CMV and EBV, and 12.5 U/ml for HHV6. The maximum of plasma concentrations per antibody was the following: CMV–304 U/ml; EBV-EBNA1–200 U/ml; EBV-VCA–468 U/ml; HHV6–419 U/ml; HSV1–257 U/ml; HSV2–367 U/ml; VZV–301 U/ml. Given that these maximum values translated into a narrow fold change in \log_{10} scale beyond the seropositivity cut-off values, we chose to present some of the results using \log_2 scale.

6.2.4 Statistical analysis

Basic quantitative characteristics (e.g., age and disease duration) were summarised by means, median, min, max, and standard deviation. Absolute frequencies and percentages were used for summarising basic categorical variables (e.g., gender and disease course). To compare different study groups in terms of these data, we used the non-parametric Kruskal-Wallis test and the Pearson's χ^2 test for quantitative and categorical variables, respectively. We used a significance level of 5% in these tests.

Symptomology data from the UKMEB were here analysed for the first time. In this analysis, we removed 9 symptoms, because the frequency of missing data was higher than 25% in each disease group or overall. For each symptom, we calculated an age-adjusted odds ratio (OR) and the respective confidence interval using a logistic regression model where age was included as a confounding factor and MS was used as the reference group. The level of each confidence interval was adjusted by the Bonferroni correction to ensure a global confidence level of at least 95%. This correction was used due to its tendency to be more conservative towards the null hypothesis (i.e., absence of association/homogeneity between groups).

With respect to serological data, we first estimated seroprevalences in each study group by the proportion of individuals above the cutoff points recommended by the respective ELISA protocol. We calculated the respective 95% confidence interval using the Pearson-Clopper exact method. We performed the Pearson's χ^2 test to assess the homogeneity of seroprevalences to a given herpesvirus. For the respective quantitative data, we reported the median and

the interquartile range (IQR) in each study group and herpesvirus antibody. To compare the median antibody concentrations across all study groups, we performed the non-parametric Kruskal-Wallis tests using a significance level of 5%.

Similarly to our previous study (Domingues et al. 2021), we computed the Receiver Operating Characteristic (ROC) curve for each serological data, using all possible cut-off values to discriminate cases (ME/CFS subgroups) from controls (MS patients). We then estimated the respective areas under these curves (AUC) and computed their 95% confidence interval. For this estimation, we checked whether antibody concentrations of cases were either higher or lower than controls and chose the direction that provided the maximum AUC. In this analysis, there was evidence for a random classification if the confidence intervals for AUC included the value 0.5. We finally computed the optimal cut-point for each of these ROC curves by maximising the significance of association provided by the Pearson's χ^2 test. We finally adjusted the p-values associated with a given serological variable for multiple testing using the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995). This adjustment ensured a global false discovery rate (FDR) of 5%.

In contrast with our previous study, we now used an SL algorithm (Laan et al. 2007) to combine the serological data of multiple herpesviruses and determined the power of these data in discriminating patients of each ME/CFS subgroup from MS patients. In this algorithm, we estimated 4 classifiers for binary outcomes: Elastic-Net logistic regression, linear discriminant analysis, quadratic discriminant analysis, and random forests. For each pairwise comparison, these classifiers were trained using 10-fold cross-validation. The estimated probabilities from each individual model were then pooled together by a weighted average estimated by the SL algorithm. We finally performed a similar ROC-based analysis of these results, as described above.

The final step of the study contemplated the association analysis between symptomology and herpesvirus IgG antibodies in each study group. With this purpose, we used the non-parametric Mann-Whitney test to compare the concentration of a given herpesvirus antibody in absence and presence of the symptom under analysis in each study group. When analysing data from a given study group, we adjusted the raw p-values using the Benjamini-Hochberg

procedure in order to ensure a global FDR of 5%.

All the analyses were conducted in the R software version 4.2.2 (R Core Team 2020) using the following packages: *table1* to summarise the baseline characteristics (Rich 2023), *pROC* to perform the ROC analyses (Robin et al. 2011), *OptimalCutpoints* to estimate the optimal cutpoints in the ROC curves (López-Ratón et al. 2014), and *SuperLearner* to perform the analysis based on the SL algorithm (Polley et al. 2023).

6.2.5 Ethical approval

Ethical approval was granted by the London School of Hygiene & Tropical Medicine Ethics Committee (Ref. 6123) and the National Research Ethics Service (NRES) London-Bloomsbury Research Ethics Committee (REC ref. October 11, 1760, IRAS ID: 77765). All participants provided written informed consent for data collection (questionnaire, clinical measurement and laboratory tests), and for allowing their samples to be available to any research receiving ethical approval (33).

6.3 Results

6.3.1 Basic characteristics of the study participants

A summary of the basic characteristics can be found in Table 6.1. The four ME/CFS subgroups did not differ substantially in terms of the age distribution, with average values of 44.5, 41.0, 43.4, and 41.1 years old for S0, S1, S2, and S3, respectively. Patients with MS were significantly older (average of 48.8 years old) than the ME/CFS patients (Kruskal-Wallis test, $p = 0.007$). Hence, we used age-adjusted ORs for the symptomology data. The percentages of female patients ranged from 69.6% (S3) to 82.6% (MS) across the five study groups. This range was not statistically significant (Pearson's χ^2 test, $p = 0.578$).

Most of ME/CFS patients had a diagnosis complying with both CDC-1994/CCC-2003. The differences in proportion of different diagnostic combinations were not statistically significant among the ME/CFS (Pearson's χ^2 test, $p = 0.144$). The percentage of severely-affected ME/CFS

patients was 9.5% in both S0 and S1. This percentage was significantly higher in the remaining subgroups (30.4% in S2 and 26.1% in S3; Pearson's χ^2 test, $p = 0.007$). The disease duration was approximately the same across the ME/CFS subgroups (mean values between 11.8 and 13.2 years; Kruskal-Wallis test, $p = 0.754$). No information was available on disease duration for the MS patients.

Finally, in terms of disease course, there was a large proportion of missing data across the 5 study groups. Notwithstanding this data limitation, the most prevalent disease course was "fluctuating symptoms" in both ME/CFS and MS cohorts. However, differences in proportion of distinct disease course profiles were not statistically significant (Pearson's χ^2 test, $p = 0.317$).

Table 6.1: Basic characteristics of the study participants where p-values refer to the Kruskal-Wallis test and the Pearson's χ^2 test for quantitative and categorical variables, respectively.

	ME/CFS_S0 (n = 42)	ME/CFS_S1 (n = 42)	ME/CFS_S2 (n = 92)	ME/CFS_S3 (n = 46)	MS (n = 46)	p-value
Age						0.007
Mean (SD)	44.5 (10.7)	41.0 (13.0)	43.4 (10.9)	41.1 (10.5)	48.8 (7.10)	
Median [Min, Max]	44.0 [23.0, 60.0]	42.5 [18.0, 60.0]	44.0 [18.0, 59.0]	42.0 [19.0, 57.0]	50.5 [31.0, 58.0]	
Gender						0.578
Female	33 (78.6%)	32 (76.2%)	74 (80.4%)	32 (69.6%)	38 (82.6%)	
Male	9 (21.4%)	10 (23.8%)	18 (19.6%)	14 (30.4%)	8 (17.4%)	
ME/CFS diagnosis						0.143
1994 CDC Only	9 (21.4%)	9 (21.4%)	6 (6.5%)	6 (13.0%)	—	
2003 CCC Only	0 (0%)	1 (2.4%)	1 (1.1%)	1 (2.2%)	—	
1994 CDC/2003 CCC	33 (78.6%)	32 (76.2%)	85 (92.4%)	39 (84.8%)	—	
Disease course						0.317
Constantly getting worse	4 (9.5%)	4 (9.5%)	16 (17.4%)	8 (17.4%)	10 (21.7%)	
Constantly improving	0 (0%)	1 (2.4%)	0 (0%)	1 (2.2%)	0 (0%)	
No change	1 (2.4%)	1 (2.4%)	7 (7.6%)	2 (4.3%)	2 (4.3%)	
Relapsing and remitting	3 (7.1%)	4 (9.5%)	6 (6.5%)	8 (17.4%)	10 (21.7%)	
Fluctuating	16 (38.1%)	15 (35.7%)	30 (32.6%)	11 (23.9%)	14 (30.4%)	
Missing data	18 (42.9%)	17 (40.5%)	33 (35.9%)	16 (34.8%)	10 (21.7%)	
Disease severity						0.007
Mild/Moderate	38 (90.5%)	38 (90.5%)	64 (69.6%)	34 (73.9%)	—	
Severe	4 (9.5%)	4 (9.5%)	28 (30.4%)	12 (26.1%)	—	
Disease onset (infection)						—
Don't know	42 (100%)	0 (0%)	0 (0%)	0 (0%)	14 (30.4%)	
No, did not have an infection at onset	0 (0%)	42 (100%)	0 (0%)	0 (0%)	22 (47.8%)	
Yes, but not confirmed by a lab test	0 (0%)	0 (0%)	92 (100%)	0 (0%)	1 (2.2%)	
Yes, confirmed by a lab test	0 (0%)	0 (0%)	0 (0%)	46 (100%)	0 (0%)	
Missing data	0 (0%)	0 (0%)	0 (0%)	0 (0%)	9 (19.6%)	
Disease duration						0.754
Mean (SD)	12.7 (8.39)	11.8 (8.60)	12.1 (8.52)	13.2 (8.22)	—	
Median [Min, Max]	11.4 [1.40, 37.0]	10.3 [1.60, 38.0]	11.2 [0.200, 39.9]	12.0 [1.10, 29.3]	—	
Missing	1 (2.4%)	2 (4.8%)	2 (2.2%)	0 (0%)	46 (100%)	

6.3.2 Analysis of symptomology

An initial analysis of symptoms' data revealed that the age-adjusted ORs for the presence of each symptom when comparing the overall ME/CFS group to the MS one tended to be significantly higher than 1 (Figure 6.1, Figure 6.2). This observation suggested that ME/CFS cohort as a whole or divided in its 4 subgroups had symptoms with a higher frequency than patients with MS. In all pairwise comparisons, the highest age-adjusted ORs were obtained for PEM lasting more than 24 hours. This finding was expected given that PEM is a key symptom for the ME/CFS diagnosis according to CCC-2003. Thirteen symptoms had similar frequencies in ME/CFS and MS cohorts: disorientation, worsening post-stress, brain fog, muscle weakness, intolerance to heat/cold, muscle stiffness in the morning, intolerance to standing up, decreased in sexual function, temporary eyesight disturbance, poor coordination, bladder problems, tingling or numbness in arms and/or legs, and loss of balance (Figure 6.1).

The number of significant age-adjusted ORs reduced in the subgroup analysis, probably due to a reduction in sample size per ME/CFS subgroup (Figure 6.2A–D). On the one hand, ME/CFS_S0 and ME/CFS_S1 had 18 and 16 symptoms whose frequency was significantly increased in relation to the MS group, respectively. Thirteen symptoms were shared between these two ME/CFS groups. In contrast, this number of significant symptoms was 34 and 29 for ME/CFS_S2 and ME/CFS_S3, respectively. Hence, both ME/CFS subgroups not related to a putative infectious disease trigger were the most similar ones to the MS group. Given that these subgroups had a lower frequency of severely-affected patients than the remaining groups, the MS group seems to be composed of patients with a mild/moderate symptomology.

Finally, there was no evidence for any symptom whose presence was significantly more prevalent in the MS group than in the overall ME/CFS group or its subgroups (Figure 6.1, Figure 6.2).

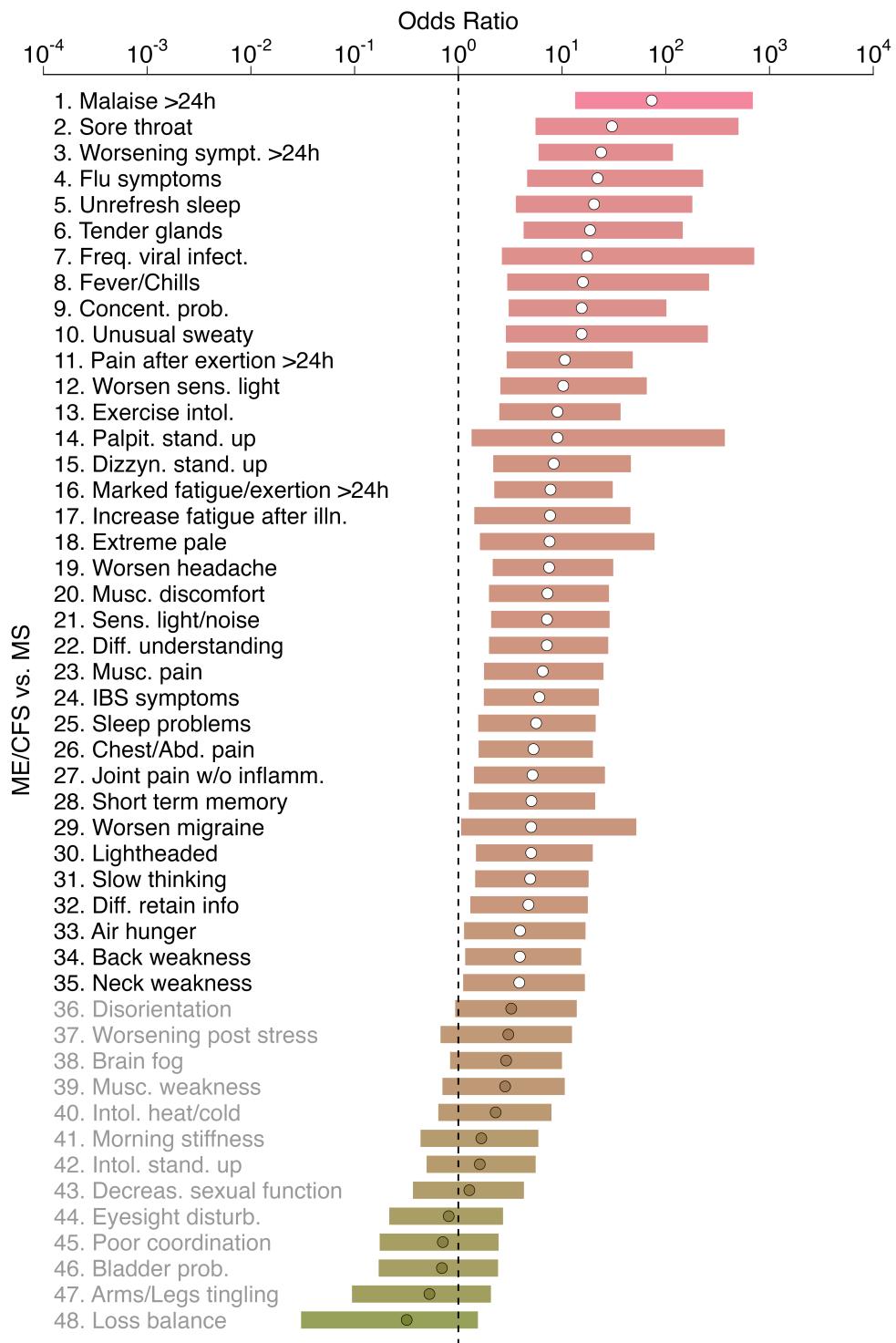


Figure 6.1: Age-adjusted odds ratios (dots) ordered by magnitude and their Bonferroni-adjusted 95% confidence intervals (horizontal bars) for the presence of each of 48 symptoms when comparing the whole ME/CFS group to the MS group (reference). White-filled dots refer to symptoms where there was evidence of a higher frequency in the ME/CFS group than in the MS group. Grey-filled dots refer to symptoms where there was evidence for the same frequency in both cohorts.

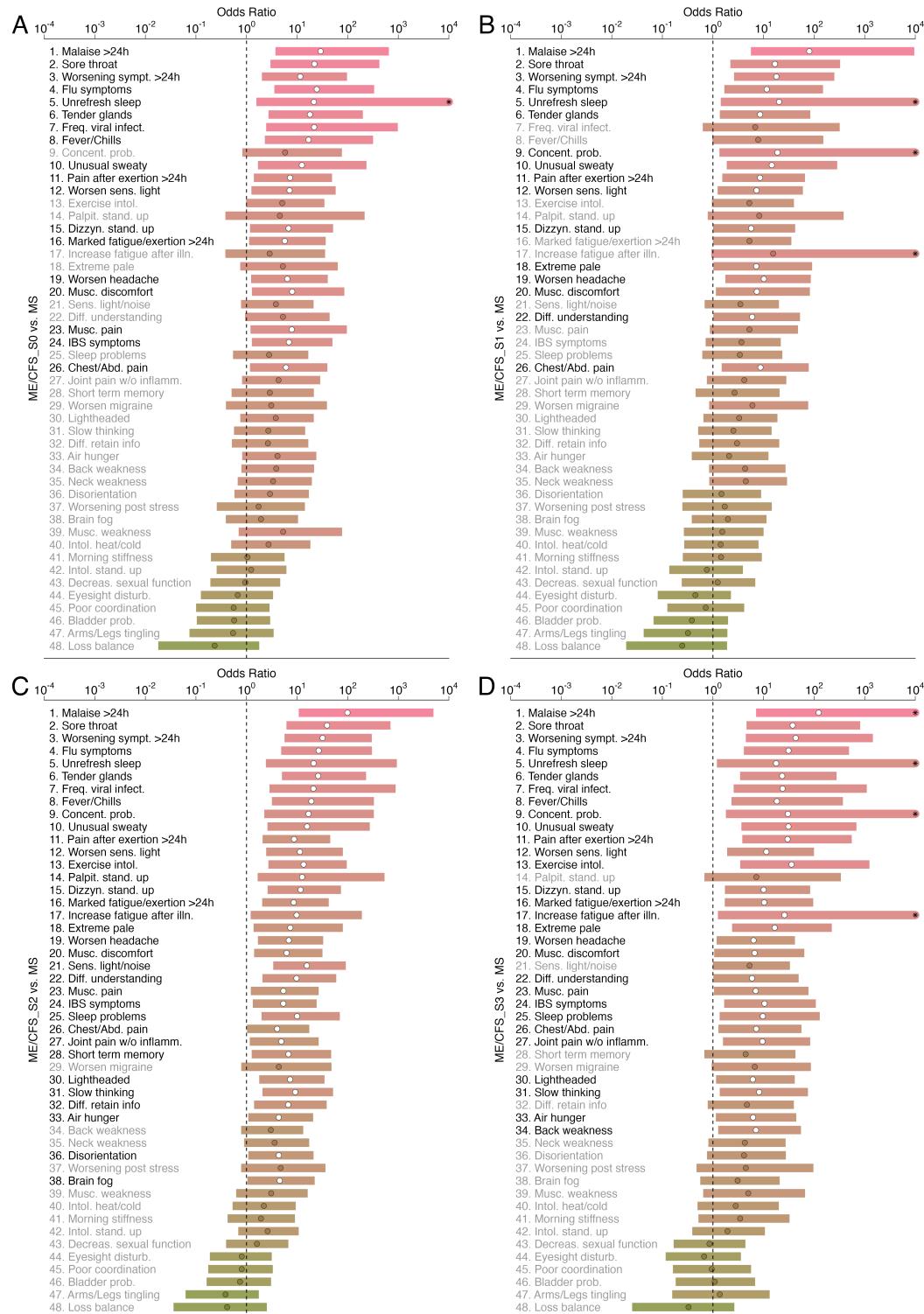


Figure 6.2: Age-adjusted odds ratios (dots) and their Bonferroni-adjusted 95% confidence intervals (horizontal bars) for the presence of each of 48 symptoms when comparing ME/CFS_S0 (A), ME/CFS_S1 (B), ME/CFS_S2 (C), and ME/CFS_S3 (D) to the MS group. The “*” symbol in some of the bars denotes an upper limit beyond the maximum value used for the xx axis. See Figure 6.1 legend for further information.

6.3.3 Univariate analysis of herpesvirus IgG antibody data

As stated in the Introduction, EBV and HHV6 infections are thought to be at the origin and a risk factor for both MS (Lundström and Gustafsson 2022; Soldan and Lieberman 2023) and ME/CFS (Ruiz-Pablos et al. 2021). Given this evidence, one could hypothesise that IgG antibody concentrations related to these viruses could be elevated in both diseases. In this regard, we previously found evidence for similar seroprevalence to each herpesvirus, including EBV and HHV6, when comparing mild/moderate and severely affected ME/CFS patients, MS patients, and healthy controls (Cliff et al. 2019). Similar evidence was found for the respective antibody concentration levels except a higher EBNA1-VCA IgG antibody concentration in the MS cohort.

Here, we repeated a similar analysis using data of seroprevalence and IgG antibody concentration levels for 5 study groups to test whether a disease-trigger stratification of the ME/CFS group could provide additional information (Table 6.2 and Supplementary Figure C.1). Interestingly, the most significant differences remained to be related to both EBV-related IgG antibodies. These differences were in part driven by a higher seroprevalence or a higher median IgG antibody concentration in the MS group, even when compared to the ME/CFS_S3 group where confirmed EBV infections were reported to be the cause of the disease (Domingues et al. 2021). Therefore, our re-analysis using a different ME/CFS patient's stratification was once again against the prior expectation that IgG antibody concentrations related to different herpesviruses are equivalent in ME/CFS and MS.

In agreement with our previous analysis when we compared ME/CFS subgroups to MS (Domingues et al. 2021), differences in CMV seroprevalence groups were in the vicinity of statistical significance. This result was due to lower seroprevalences to this virus in ME/CFS_S1 and ME/CFS_S3. For the remaining herpesvirus antibodies, there were no significant differences among the study groups in terms of seroprevalence or IgG antibody concentrations.

We extended this analysis with the objective of discriminating patients with MS from healthy controls combining these data with a ROC curve approach (Supplementary Table C.3). The most significant AUC result was found for the antibodies against EBV-VCA. In this case, the

Table 6.2: Seroprevalence (SeroP in percentage), median concentration and the respective IQR per study group and herpesvirus IgG antibody, where p-value refers to the Pearson's χ^2 test for seroprevalence-based analysis and the Kruskal-Wallis test for the corresponding quantitative data.

Herpesvirus (Antigen)	Analysis	ME/CFS_S0	ME/CFS_S1	ME/CFS_S2	ME/CFS_S3	MS	p-value
CMV	SeroP (95% CI)	42.9 (27.7; 59.0)	25.6 (13.5; 41.2)	37.0 (27.1; 47.7)	18.8 (8.9; 32.6)	45.0 (29.3; 61.5)	0.035
	Median (IQR)	5.4 (78.5)	4.6 (13.3)	6 (59.8)	5.4 (1.5)	5.8 (64.4)	0.407
EBV (EBNA1)	SeroP (95% CI)	71.4 (55.4; 84.3)	58.1 (42.1; 73.0)	75.0 (64.9; 83.4)	75.0 (60.4; 86.4)	95.0 (83.1; 99.4)	0.004
	Median (IQR)	39.3 (17.8)	24.6 (17.1)	37.2 (23.5)	32.9 (26.4)	46.9 (17.8)	0.006
EBV (VCA)	SeroP (95% CI)	88.1 (74.4; 96.0)	74.4 (58.8; 86.5)	84.8 (75.8; 91.4)	95.8 (85.7; 99.5)	92.5 (79.6; 98.4)	0.030
	Median (IQR)	99.4 (64.4)	88.0 (64.8)	108.9 (49.9)	129.5 (48.6)	155.4 (33.8)	0.003
HHV6	SeroP (95% CI)	85.7 (71.5; 94.6)	90.7 (77.9; 97.4)	92.4 (84.9; 96.9)	100.0 (92.6; 100.0)	90.0 (76.3; 97.2)	0.146
	Median (IQR)	38.3 (21.1)	42.5 (24.3)	47.5 (25.2)	43.0 (29.7)	43.4 (50.8)	0.418
HSV1	SeroP (95% CI)	50.0 (34.2; 65.8)	34.9 (21.0; 50.9)	43.5 (33.2; 54.2)	54.2 (39.2; 68.6)	57.5 (40.9; 73.0)	0.207
	Median (IQR)	12.1 (89.5)	6.6 (56.5)	7.6 (108.8)	16.5 (119.7)	22.6 (109.6)	0.273
HSV2	SeroP (95% CI)	50.0 (34.2; 65.8)	30.2 (17.2; 46.6)	35.9 (26.1; 46.5)	45.8 (31.4; 60.8)	40.0 (24.9; 56.7)	0.309
	Median (IQR)	12.2 (61.4)	5.0 (33.5)	4.1 (45.0)	6.2 (41.8)	8.1 (41.8)	0.650
VZV	SeroP (95% CI)	95.2 (83.8; 99.4)	93.0 (80.9; 98.5)	100.0 (96.1; 100.0)	97.9 (88.9; 99.9)	97.5 (86.8; 99.9)	0.169
	Median (IQR)	129.5 (46.2)	119.7 (20.8)	135.7 (48.4)	142.2 (35.8)	156.2 (36.7)	0.069

respective antibody concentrations were significantly increased in patients with MS. The estimate of AUC was 0.641 (95% CI = [0.538, 0.744]) with an optimal sensitivity and specificity of approximately 0.650. Another significant result was found for antibodies against HSV1 (AUC = 0.626; 95% CI = [0.531, 0.721]). Interestingly, these antibodies were in higher concentrations in healthy controls than in MS patients. The corresponding sensitivity and specificity were estimated at 0.950 and 0.327, respectively. This result indicated that MS patients are well characterised by a low HSV1 antibody concentration. The data from the remaining antibodies did not provide any evidence for a significant result.

We then performed a classification exercise of each subgroup of ME/CFS patients using patients with MS as the control group (Table 6.3). In general, the MS cohort tended to have higher antibody concentrations than the ME/CFS patients, irrespective of the herpesvirus under analysis apart from a single exception, ME/CFS_S0 vs. MS for CMV. However, the classification power of each antibody was not significant as illustrated by the 95% confidence interval for the AUC that contained the value 0.50. The most optimistic scenarios were observed for the pairwise comparisons involving the EBV-related antigens. In particular, the best classification was obtained for ME/CFS_S1 vs. MS for EBNA1 (AUC = 0.741; 95% CI = [0.634, 0.848]). In this case, the sensitivity and specificity were estimated at 0.465 and 0.950, respectively, using

a cutoff of 20.0 U/ml. These estimates suggested that high antibody concentrations to this EBV antigen were able to discriminate MS patients almost perfectly. However, the same could not be said for detecting cases from this ME/CFS subgroup.

ME/CFS_S1 seemed the most different one from the MS group in terms of herpesvirus IgG antibodies (Table 6.3). This interpretation was supported by a significant AUC for antibodies against EBNA1 (as reported above), EBV-VCA (AUC = 0.695; 95% CI = [0.580, 0.811]; Se = 0.442, Sp = 0.900), HSV1 (AUC = 0.635; 95% CI = [0.513, 0.756]; Se = 0.256, Sp = 0.975), and VZV (AUC = 0.667; 95% CI = [0.548, 0.786]; Se = 0.744, Sp = 0.630). This result suggested that combining data from multiple antibodies could help discriminating these two groups of patients.

6.3.4 Combined analysis of IgG antibody data using an SL algorithm

In this step of the analysis, we integrated all IgG antibody data in several classifiers to distinguish patients of each ME/CFS subgroup from patients with MS. For each pairwise comparison, these classifiers were then assembled into a final classifier using an SL algorithm (Figure 6.3).

This combined data analysis provided evidence for an AUC significantly different from 0.5 (random guess) for ME/CFS_S0 (AUC = 0.658; 95% CI = [0.536, 0.779]), ME/CFS_S1 (AUC = 0.731; 95% CI = [0.622, 0.841]), and ME/CFS_S3 (AUC = 0.707; 95% CI = [0.599, 0.816]) subgroups using the MS group as a control. The highest sensitivity was obtained for the ME/CFS_S2 group (0.772), but at the cost of a poor specificity related to the MS group (0.075) (Figure 6.3). The most balanced sensitivity and specificity estimates were observed for ME/CFS_S0 vs. MS (0.619 and 0.725, respectively). In the remaining pairwise comparisons (ME/CFS_S1 and ME/CFS_S3), there was evidence of a high specificity (higher than 0.90) but a modest sensitivity (up to 0.512).

Table 6.3: Area under the Receiver Operating Characteristic curve (AUC) and its 95% confidence interval (CI), optimal cutoff and associated sensitivity (Se) and specificity (Sp) to discriminate ME/CFS_S0, ME/CFS_S1, ME/CFS_S2, ME/CFS_S3 subgroups (cases) from patients with multiple sclerosis (MS) used as controls. In the Direction column, the symbols “>” and “<” represent higher value in MS cases than in healthy controls and vice-versa, respectively. In the Cutoff column, the p-value within parenthesis is associated with the Pearson’s χ^2 test for 2×2 tables after adjusting for an FDR of 5%. In the AUC column, the symbol “*” denote the cases where there was evidence of an AUC different from 0.50 (random guess).

Herpesvirus (Antigen)	Comparison (vs. MS)	Direction	AUC (95% CI)	Cutoff in U/ml (p-value)	Se	Sp
CMV	ME/CFS_S0	controls < cases	0.551 (0.421; 0.680)	2.845 (0.0214)	0.952	0.325
	ME/CFS_S1	controls > cases	0.547 (0.418; 0.677)	29.21 (0.0701)	0.791	0.425
	ME/CFS_S2	controls < cases	0.507 (0.389; 0.624)	194.91 (0.0256)	0.000	0.900
	ME/CFS_S3	controls > cases	0.545 (0.413; 0.678)	58.755 (0.0028)	0.938	0.375
EBV (EBNA1)	ME/CFS_S0	controls > cases	0.590 (0.463; 0.716)	15.415 (0.0257)	0.286	0.950
	ME/CFS_S1	controls > cases	0.741 (0.634; 0.848)*	20.035 (0.0004)	0.465	0.950
	ME/CFS_S2	controls > cases	0.602 (0.505; 0.700)*	20.505 (0.0035)	0.359	0.950
	ME/CFS_S3	controls > cases	0.621 (0.503; 0.739)*	19.985 (0.0028)	0.396	0.950
EBV (VCA)	ME/CFS_S0	controls > cases	0.668 (0.550; 0.787)*	132.96 (0.0257)	0.643	0.675
	ME/CFS_S1	controls > cases	0.695 (0.580; 0.811)*	71.435 (0.0029)	0.442	0.900
	ME/CFS_S2	controls > cases	0.645 (0.543; 0.746)*	150.345 (0.0144)	0.707	0.575
	ME/CFS_S3	controls > cases	0.566 (0.443; 0.690)	139.315 (0.0687)	0.583	0.650
HHV6	ME/CFS_S0	controls > cases	0.590 (0.465; 0.714)	78.105 (0.0374)	0.881	0.350
	ME/CFS_S1	controls > cases	0.558 (0.429; 0.687)	103.94 (0.0057)	1.000	0.225
	ME/CFS_S2	controls > cases	0.518 (0.403; 0.632)	130.625 (0.0256)	0.978	0.150
	ME/CFS_S3	controls > cases	0.520 (0.391; 0.649)	91.82 (0.0271)	0.938	0.275
HSV1	ME/CFS_S0	controls > cases	0.579 (0.454; 0.704)	2.14 (0.0371)	0.167	1.000
	ME/CFS_S1	controls > cases	0.635 (0.513; 0.756)*	2.84 (0.0104)	0.256	0.975
	ME/CFS_S2	controls > cases	0.596 (0.497; 0.696)	2.755 (0.0303)	0.196	0.975
	ME/CFS_S3	controls > cases	0.550 (0.429; 0.671)	2.8 (0.0182)	0.250	0.975
HSV2	ME/CFS_S0	controls < cases	0.557 (0.432; 0.683)	231.395 (0.0737)	0.119	1.000
	ME/CFS_S1	controls > cases	0.536 (0.409; 0.663)	6.81 (0.1600)	0.628	0.55
	ME/CFS_S2	controls > cases	0.520 (0.410; 0.629)	7.03 (0.1343)	0.609	0.550
	ME/CFS_S3	controls > cases	0.527 (0.402; 0.651)	2.895 (0.1341)	0.771	0.400
VZV	ME/CFS_S0	controls > cases	0.596 (0.472; 0.721)	141.005 (0.0737)	0.571	0.650
	ME/CFS_S1	controls > cases	0.667 (0.548; 0.786)*	139.495 (0.0025)	0.744	0.650
	ME/CFS_S2	controls > cases	0.532 (0.418; 0.646)	148.24 (0.5054)	0.554	0.550
	ME/CFS_S3	controls > cases	0.546 (0.421; 0.672)	180.995 (0.0981)	0.792	0.400

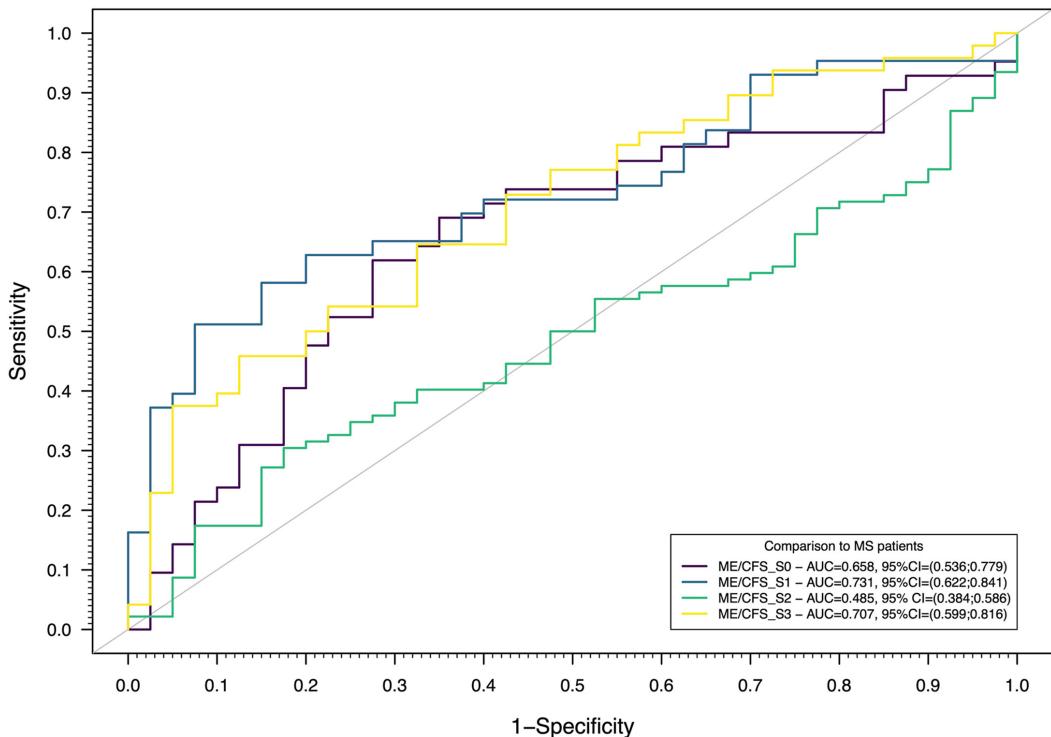


Figure 6.3: ROC curves for the predictions based on an SL algorithm trained with 4 different classifiers (Elastic-Net Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Random Forest) and 10-fold cross-validation using antibody data and patients with MS as the controls. Optimal sensitivities and specificities were estimated at 0.619 and 0.725 for ME/CFS_S0, 0.512 and 0.925 for ME/CFS_S1, 0.772 and 0.075 for ME/CFS_S2, 0.375 and 0.950 for ME/CFS_S3, when compared to the MS group.

As expected from the analysis based on a single antibody, the best AUC was obtained for the comparison between ME/CFS_S1 and MS. When compared to the MS group, the probability of having a patient from this ME/CFS subgroup decreased with the antibody concentrations related to different herpesviruses including EBV (Figure 6.4A–G); see the respective probability profiles for the remaining pairwise comparison in Supplementary Figure C.2, Supplementary Figure C.3, and Supplementary Figure C.4 (ME/CFS_S0 vs. MS, ME/CFS_S2 vs. MS, ME/CFS_S3 vs. MS, respectively). Note that, given the poor discrimination between ME/CFS_S2 and MS, the probability profile of having a patient from this ME/CFS subgroup was almost constant across all the herpesviruses antibodies (Supplementary Figure C.3).

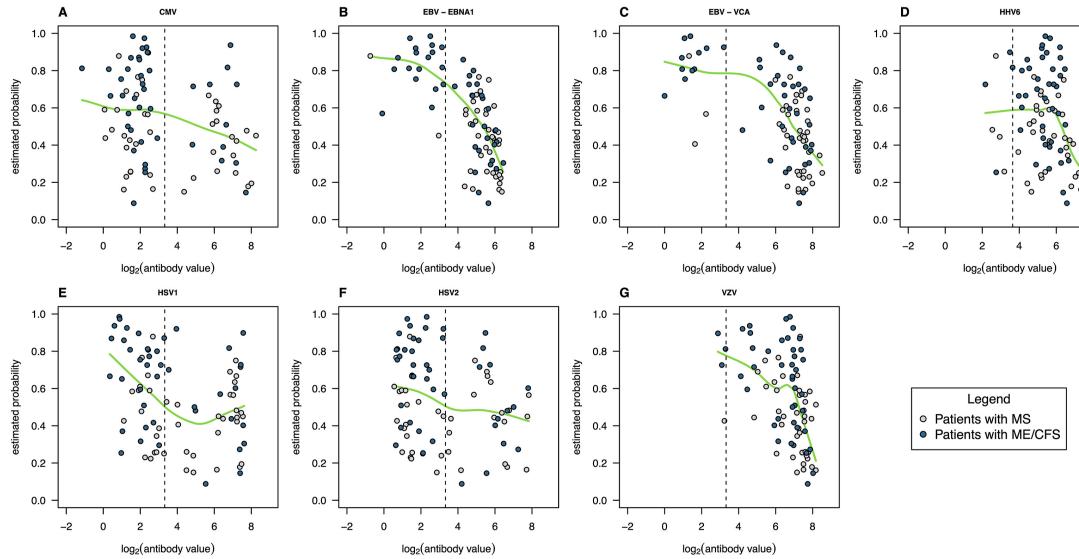


Figure 6.4: Smooth-line approximations (green lines) of the relationship between $\log_2(\text{antibody concentrations})$ and SL-estimated probability of ME/CFS_S1 patient when compared to patients with MS (A–CMV, B–EBV-EBNA1, C–EBV-VCA, D–HHV-6, E–HSV1, F–HSV2, G–VZV). In the plots, each dot represents a patient and the vertical dashed line represents the cut-off value for seropositivity according to the respective lab protocol.

6.3.5 Association analysis between symptomology and herpesvirus IgG antibodies

Lastly, we correlated the data of presence/absence of each symptom with data of each herpesvirus IgG antibody (Figure 6.5A–F). In the case of the MS cohort, HHV6 antibody concentration was significantly and negatively associated with difficulties in understanding, and worsening of symptoms after stress. Similar significant negative association was found between CMV IgG antibody concentrations and eyesight disturbances. The only significant positive association was found for EBV-EBNA1 antibody concentrations and bladder problems.

For the overall ME/CFS group, a positive and a negative association reached statistical significance between brain fog and HSV1 IgG antibody concentrations and between chest/abdominal pain and HHV6 antibody concentrations, respectively. These significant associations could not be confirmed by the subsequent subgroup analysis. For the ME/CFS_S0 subgroup, two significant negative associations were found (neck weakness/HSV1 antibody concentrations and fever chills/EBV-VCA antibody concentrations). In the ME/CFS_S1 subgroup, there was a sig-

nificant negative association between difficulties in retaining/recalling information and VZV IgG antibody concentrations. In the ME/CFS_S2 subgroup, negative and positive associations were found between EBV-VCA antibody concentration and worsening of symptoms after exertion lasting more than 24 hours and between EBNA1 and short-term memory problems, respectively. Finally, feeling lightheaded was significantly and negatively associated with CMV IgG antibody concentrations in the ME/CFS_S3 subgroup. In opposition, bladder problems were significantly and positively associated with VZV antibody concentrations in the same ME/CFS subgroup.

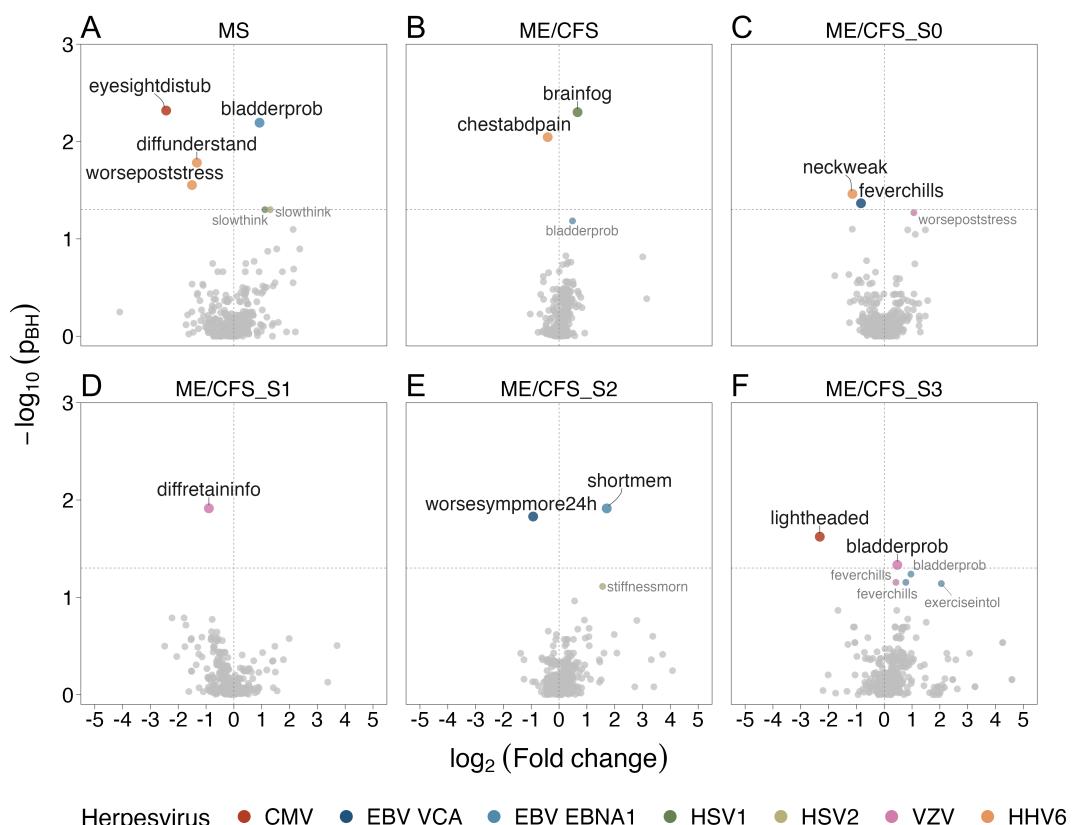


Figure 6.5: Association analysis between symptoms and IgG antibody data for the MS group (A), the overall ME/CFS (B), the ME/CFS_S0 subgroup (C), ME/CFS_S1 subgroup (D), ME/CFS_S2 subgroup (E), and ME/CFS_S3 subgroup (F), where the xx axis refers to $\log_2(\text{mean fold-change})$ between individuals with and without a given symptom, respectively, and the yy axis refers to the logarithm in base 10 of the p-values derived from the Mann-Whitney test and adjusted for an FDR of 5% using the Benjamini-Hochberg procedure ($-\log_{10}(p_{BH})$). See Supplementary Table C.1 for linking the symptom codes presented in each plot to the respective symptom descriptions.

6.4 Discussion

Our study showed that the overall ME/CFS cohort had more symptoms whose frequency was higher than the one observed for the MS patients. This finding was in agreement with a higher frequency of functional symptoms reported by an online self-report survey with the aim at finding the key differentiating symptoms of these two diseases (Jason 2017). Interestingly, sore throat, tender glands, and flu-like symptoms from the immunological domain were also at the top differentiators of both ME/CFS_S2 and ME/CFS_S3 ME/CFS (i.e., subgroups related to a putative infectious trigger) and MS. Again, this finding is in line with another study where tender lymph nodes and flu-like symptoms could correctly distinguish MS from ME/CFS 81% of the time (Ohanian et al. 2016). Therefore, our study provided additional evidence that, besides the presence of hallmark persistent fatigue and PEM, symptoms from the immunological domain are crucial to diagnose ME/CFS and differentiate from MS. This is reflected in the CDC-1994 criterion and CCC-2003, two recommended case definitions for ME/CFS diagnosis, especially in the research setting (Nacul et al. 2021). However, we cannot rule out that difference in symptoms' frequency might be simply due to mild MS cases who did not undergo any treatment together with the presence of severely affected patients in the different ME/CFS subgroups. In this regard, it is important to emphasise that severe MS cases (most likely undergoing any immune-therapy) were excluded from participation in the UKMEB.

We found several positive and negative associations between herpesvirus antibody concentrations and different symptoms in each study group, even after adjusting for multiple testing. This finding is remarkable given the dozens of symptoms evaluated, the modest sample size of each study group, and the use of a non-parametric test that usually reduces the statistical power to detect group differences. However, we cannot rule out that these associations might be affected by patient's subjective perception of each symptom. At the same time, casting doubts on this perception is indirectly supporting theories that ME/CFS is a psychosomatic rather than an organic or a physical condition. A more plausible impact on the robustness of these associations is that they could have resulted from random fluctuations in symptomology occurring over the natural disease course. To confirm or refute this last explanation, one could

conduct a longitudinal study with multiple timepoints and check whether the associations remain valid during follow-up.

The MS group had the highest number of significant antibody-symptom associations (or close to statistical significance). Therefore, infections by these herpesviruses seem to have a higher impact on this group than on the ME/CFS groups. The strongest association was observed for CMV antibody concentrations and the presence of eyesight disturbances. This is an interesting association given that photophobia and photosensitivity were commonly reported by MS patients (Cortese et al. 2018) and CMV is also known to cause retinitis in immunocompromised individuals (Port et al. 2017). In this scenario, a reduction in antibodies against CMV might result in a low-grade ocular infection in MS patients. The second strongest association was found for EBNA1 antibody concentrations and bladder problems. This association is according to the view that bladder dysfunction, also a common symptom observed in MS, is linked to autoantibodies (for example, against muscarinic receptor-3) (McCombe et al. 2009). Our finding suggests that the origin of these autoantibodies could be a cross-reactive antibody response to an EBNA1 peptide mimicking a human protein, as demonstrated for MS (Ayoglu et al. 2016; Tengvall et al. 2019; Lanz et al. 2022). Interestingly, a positive association was also found for bladder problems in the ME/CFS_S3 group but related to VZV antibody concentrations. Given that infections by VZV can also cause bladder dysfunction (Sakakibara et al. 2022), we speculate that patients from this ME/CFS subgroup experienced a recent reactivation of the virus. Notwithstanding the recent observation of VZV reactivation after SARS-CoV-2 infections (Martinez-Reviejo et al. 2022) or under stressful conditions (Rooney et al. 2019), there is weak evidence for VZV reactivation in ME/CFS patients (Koelle et al. 2002; Lee et al. 2021). The remaining significant antibody-symptoms associations in the MS group were both negative and related to HHV6 IgG antibody concentrations. These associations are in opposition to the evidence that antibody concentrations to HHV6 are positively associated with the risk of developing MS (Engdahl et al. 2019). A possible explanation for this contradicting finding is that a lower HHV6 antibody concentration triggers the reactivation of this virus, as observed in individuals after cord blood transplantation (Nakayama et al. 2021). Unfortunately, the quantification of viral DNA was not performed in this study and, therefore, this explanation

could not be tested with data.

As mentioned above, we found a lesser number of significant antibody-symptom associations in ME/CFS. Interestingly, the strongest association was found for increasing HSV1 antibody concentrations and brain fog. It is worth noting that brain fog is a colloquial term used by patients that seems to gather a constellation of symptoms mostly concerning deficits in cognition, such as reduced speed and efficiency of information processing, attention, concentration, and working memory (Ross et al. 2013; Ocon 2013; Krishnan et al. 2022). If brain fog is taken in this sense, then the positive association with HSV1 antibody levels seems interesting given that this virus is known to be neurotropic during latency (Marcocci et al. 2020). The presence and concentration levels of HSV1 IgG antibodies, as measured here, have been positively associated with cognitive deficits in both healthy individuals (Jonker et al. 2014; Tarter et al. 2014; Fruchter et al. 2015) and patients suffering from Alzheimer's disease (Murphy et al. 2021), schizophrenia (Dickerson et al. 2020), and bipolar disorder (Tucker and Bertke 2019). The same association could be significantly found for ME/CFS_S1 and ME/CFS_S2 before adjusting for multiple testing. Hence, a reduction in the statistical power by dividing patients in different strata is a plausible explanation for not finding this association in the ME/CFS subgroup analysis. At this point, it is reasonable to raise the question of why similar association with brain fog was not detected for antibody concentrations related to HHV6, another neurotropic virus whose viral miRNA was found in post-mortem brain biopsies of ME/CFS patients (Kasimir et al. 2022). A possible answer for this question is that, given that exposure to HHV6 is more prevalent in human populations than to HSV1, the respective antibody quantification has shorter dynamic range due to a high percentage of seropositive individuals (Table 6.2 and Supplementary Figure C.1). In theory, this shorter dynamic range implies a higher sample size to detect smaller differences between patients reporting or not brain fog.

In the ME/CFS subgroup analysis, the detected antibody-symptom associations were not consistent across the 4 ME/CFS subgroups. This result suggested a large heterogeneity of different ME/CFS groups in terms of symptoms and their relationship with the underlying herpesviruses IgG antibody concentrations. An analysis based on disease severity could provide a clearer insight into the relationship between these two sets of variables. This possibility

is motivated by a longitudinal study from the UKMEB where disease scores and different symptoms were associated with herpesvirus reactivation (Lee et al. 2021). This alternative analysis by disease severity was beyond the scope of this study, but it will be done in a near future.

Our study also showed that the IgG antibody data were statistically significant to discriminate three ME/CFS subgroups from MS. Such a discrimination capability could be attributed to increased IgG antibody concentrations to multiple herpesviruses in the MS group, as previously found for EBV antigens (Loebel et al. 2017). This capability, although reaching statistical significance, is likely to have a diminished impact in the clinic, because it neither have a high sensitivity to classify these ME/CFS subgroups nor a high specificity to classify MS patients. This observation suggests that one should look for alternative disease-specific biomarkers or to combine these basic antibody measurements with more specific herpesvirus-related antibodies, such as those against EBNA4_0529 and EBNA6_0070 peptides derived from EBV (Sepúlveda et al. 2022). Notwithstanding this observation, it is worth mentioning that the reporting of a less promising finding helps the research community to focus on more promising biomarkers, thus, avoiding the waste of valuable research efforts and resources. It is also in line with the expectations of the UKMEB participants who are less concerned about breakthroughs in medical science and more interested in incremental steps and collaborative efforts which might one day lead to effective diagnosis, treatment, or cure (Lacerda et al. 2019b).

We found that serological data including the EBV antigens were unable to discriminate MS patients from the ME/CFS subgroup of patients who self-reported an infection (not evaluated by a lab test) at their disease onset. This subgroup is the largest in size and the one with the highest frequency of patients who self-reported a flu-like infection (22%) or an unspecified viral infection (35%) at their disease onset (Domingues et al. 2021). Therefore, this subgroup seems very heterogeneous and non-specific. Given that many factors can contribute to MS pathogenesis (Marcocci et al. 2020), the lack of evidence for a serological discrimination between these two groups might be attributed to the presence of multiple factors shared between this heterogeneous subgroup of ME/CFS patients and ME patients. In this line of thought, one could make the case for MS-related drugs such as cyclophosphamide (Patti and

Lo Fermo 2011) and rituximab (Brancati et al. 2021) to be deployed to treat ME/CFS patients, as made elsewhere (Fluge et al. 2019; Rekeland et al. 2020). However, this deployment is likely to be more successful if targeting not all but only patients from this ME/CFS subgroup. However, future studies should be conducted to ascertain further similarities between this subgroup of ME/CFS patients and patients with MS.

There are four major limitations of this study. As already highlighted, the main limitation is that our re-analysis focused on data portraying a single snapshot of these patients. Given the fluctuating clinical course of ME/CFS (Table 6.1), it is likely that the reported antibody-symptom associations might render irreproducible. It is also possible that non-significant associations would become statistically (and clinically) significant in latter time points. A longitudinal study including the measurements at multiple time points would resolve this limitation. In practice, such a study is expensive and affected by other problems, such as patients' compliance or the presence of missing data due to drop-outs. Patient's exhaustion (often triggered by data collection) is also another limiting factor in the context of ME/CFS research. As such, the common and pragmatic approach is to conduct association analyses between different biomarkers and symptoms (or, alternatively, symptom scores) using cross-sectional data, as illustrated not only for ME/CFS but also for Long Covid (Freitag et al. 2021; Blauensteiner et al. 2021; Szklarski et al. 2021; Sotzny et al. 2022).

The second limitation is that IgG antibodies are often used as biomarkers of a past active infection. As such, these antibodies have limited power to judge when the last active infection occurred and whether these herpesviruses are currently active or simply latent. In this regard, our study would have benefited from a quantification of IgM antibodies in the same biological samples. However, additional serological testing was beyond the scope of this study due to lack of funding and inexistence of biological material for many of the study participants. In this scenario, our study should be viewed as an opportunistic re-analysis of already existing data with its own limitations.

The third limitation is that the MS cohort was conveniently enrolled in the UKMEB due to the criterion of excluding any patient following any immune therapy. In this perspective, the recruited MS patients can be seen as "pure patients" or simply "mild cases". On the one

hand, the recruitment of patients without any treatment suggests a selective bias towards sampling patients with less severe symptoms, thus, explaining the higher number of symptoms that are more prevalent in the ME/CFS group than in the MS one. This putative bias was also suggested by a higher number of symptoms whose frequency was similar in MS and ME/CFS_S0/ME/CFS_S1 in which the frequency of severely affected patients was less than 10%. However, low adherence to modifying disease therapy by MS patients is considered a major concern in the management of the disease (Cerghet et al. 2010; Rieckmann et al. 2015) and, therefore, the recruited MS patients are in part representative of this clinical population who do not adhere to any treatment.

The fourth limitation is that disease course profiles do not comply with the standard classification used in MS research; this standard classification is reviewed elsewhere (Klineova and Lublin 2018). These profiles, if following such a standard, would have helped to understand whether the proportions of sampled MS patients at different disease courses would match those found across epidemiological studies of MS, thus, increasing the interpretation on the representativeness of the MS cohort. Notwithstanding this limitation, 22% of the MS patients reported constantly getting worse. This frequency anecdotally coincides with the prevalence of a primary progressive disease course (15%–22%) in MS patients from different parts of the UK (Pugliatti et al. 2006). In addition, all study groups were well matched for the alternative disease course categorisation used here and, therefore, there is evidence that our findings are less likely to be driven by differences in putative disease courses of the patients.

In conclusion, some candidate antibody-symptom associations were identified for ME/CFS patients. However, the clinical impact of these associations remains to be determined given the fluctuating nature of symptoms in ME/CFS patients and the nature of IgG antibodies in simply detecting exposure. To understand the true impact of our findings in the clinic, we propose a future longitudinal study with at least three timepoints. This study is likely to help discerning the robustness of the reported associations during disease course and whether they could be targeted by any existing drug treatment. This study, although theoretical appealing, might be difficult to design with an appropriate statistical power and to execute in the clinic given the lack of funding affecting ME/CFS research, poor societal recognition of the disease,

and possible patient's exhaustion during follow-up.

Bibliography

- E. Almenar-Pérez, L. Sarría, L. Nathanson, and E. Oltra. Assessing diagnostic value of microRNAs from peripheral blood mononuclear cells and extracellular vesicles in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Scientific Reports*, 10(1):2064, Feb. 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-58506-5. URL <https://www.nature.com/articles/s41598-020-58506-5>. Number: 1 Publisher: Nature Publishing Group.
- M. E. Ariza. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: The Human Herpesviruses Are Back! *Biomolecules*, 11(2):185, Feb. 2021. ISSN 2218-273X. doi: 10.3390/biom11020185. URL <https://www.mdpi.com/2218-273X/11/2/185>. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.
- B. Ayoglu, N. Mitsios, I. Kockum, M. Khademi, A. Zandian, R. Sjöberg, B. Forsström, J. Bredenberg, I. Lima Bomfim, E. Holmgren, H. Grönlund, A. O. Guerreiro-Cacais, N. Abdelmagid, M. Uhlén, T. Waterboer, L. Alfredsson, J. Mulder, J. M. Schwenk, T. Olsson, and P. Nilsson. Anoctamin 2 identified as an autoimmune target in multiple sclerosis. *Proceedings of the National Academy of Sciences*, 113(8):2188–2193, Feb. 2016. doi: 10.1073/pnas.1518553113. URL <https://www.pnas.org/doi/full/10.1073/pnas.1518553113>. Publisher: Proceedings of the National Academy of Sciences.
- G. A. Bateman and A. R. Bateman. Chronic fatigue syndrome and multiple sclerosis have reduced craniospinal compliance and dilated pressurized bridging cortical veins. *Medical Hypotheses*, 182:111243, Jan. 2024. ISSN 0306-9877. doi: 10.1016/j.mehy.2023.111243. URL <https://www.sciencedirect.com/science/article/pii/S0306987723002396>.
- Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, Jan. 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.
- R. Bertinat, R. Villalobos-Labra, L. Hofmann, J. Blauensteiner, N. Sepúlveda, and F. Westermeier. Decreased NO production in endothelial cells exposed to plasma from ME/CFS patients. *Vascular Pharmacology*, 143:106953, Apr. 2022. ISSN 1537-1891. doi: 10.1016/j.vph.2022.106953. URL <https://www.sciencedirect.com/science/article/pii/S1537189122000027>.
- K. Bjornevik, M. Cortese, B. C. Healy, J. Kuhle, M. J. Mina, Y. Leng, S. J. Elledge, D. W. Niebuhr, A. I. Scher, K. L. Munger, and A. Ascherio. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science (New York, N.Y.)*, 375(6578):296–301, Jan. 2022. ISSN 1095-9203. doi: 10.1126/science.abj8222.
- J. Blauensteiner, R. Bertinat, L. E. León, M. Riederer, N. Sepúlveda, and F. Westermeier. Altered endothelial dysfunction-related miRs in plasma from ME/CFS patients. *Scientific Reports*, 11(1):10604, May 2021.

- ISSN 2045-2322. doi: 10.1038/s41598-021-89834-9. URL <https://www.nature.com/articles/s41598-021-89834-9>. Number: 1 Publisher: Nature Publishing Group.
- J. Blomberg, C.-G. Gottfries, A. Elfaitouri, M. Rizwan, and A. Rosén. Infection Elicited Autoimmunity and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: An Explanatory Model. *Frontiers in Immunology*, 9, Feb. 2018. doi: 10.3389/fimmu.2018.00229.
- S. Brancati, L. Gozzo, L. Longo, D. C. Vitale, and F. Drago. Rituximab in Multiple Sclerosis: Are We Ready for Regulatory Approval? *Frontiers in Immunology*, 12, 2021. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2021.661882>.
- A. Bynke, P. Julin, C.-G. Gottfries, H. Heidecke, C. Scheibenbogen, and J. Bergquist. Autoantibodies to beta-adrenergic and muscarinic cholinergic receptors in Myalgic Encephalomyelitis (ME) patients – A validation study in plasma and cerebrospinal fluid from two Swedish cohorts. *Brain, Behavior, & Immunity - Health*, 7:100107, Aug. 2020. ISSN 2666-3546. doi: 10.1016/j.bbih.2020.100107. URL <https://www.sciencedirect.com/science/article/pii/S2666354620300727>.
- B. M. Carruthers, A. K. Jain, K. L. D. Meirleir, D. L. Peterson, N. G. Klimas, A. M. Lerner, A. C. Bested, P. Flor-Henry, P. Joshi, A. C. P. Powles, J. A. Sherkey, and M. I. v. d. Sande. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Chronic Fatigue Syndrome*, 11(1):7–115, Jan. 2003. doi: 10.1300/j092v11n01_02.
- M. Cerghet, E. Dobie, J. E. Lafata, L. Schultz, S. Elias, M. Pladenvall, and J. Reuther. Adherence to Disease-Modifying Agents and Association with Quality of Life Among Patients with Relapsing-Remitting Multiple Sclerosis. *International Journal of MS Care*, 12(2):51–58, Jan. 2010. ISSN 1537-2073. doi: 10.7224/1537-2073-12.2.51. URL <https://doi.org/10.7224/1537-2073-12.2.51>.
- J. M. Cliff, E. C. King, J.-S. Lee, N. Sepúlveda, A.-S. Wolf, C. Kingdon, E. Bowman, H. M. Dockrell, L. Nacul, E. Lacerda, and E. M. Riley. Cellular Immune Function in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Frontiers in Immunology*, 10, Apr. 2019. doi: 10.3389/fimmu.2019.00796. URL <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2019.00796/full>.
- A. Cortese, A. Conte, G. Ferrazzano, E. Sgarlata, E. Millefiorini, M. Frontoni, and A. Berardelli. Photophobia in multiple sclerosis. *Multiple Sclerosis and Related Disorders*, 26:55–57, Nov. 2018. ISSN 2211-0348, 2211-0356. doi: 10.1016/j.msard.2018.09.005. URL [https://www.msard-journal.com/article/S2211-0348\(18\)30317-1/fulltext](https://www.msard-journal.com/article/S2211-0348(18)30317-1/fulltext). Publisher: Elsevier.
- F. Dickerson, J. R. Schroeder, V. Nimagaonkar, J. Gold, and R. Yolken. The association between exposure to herpes simplex virus type 1 (HSV-1) and cognitive functioning in schizophrenia: A meta-analysis. *Psychiatry Research*, 291:113157, Sept. 2020. ISSN 0165-1781. doi: 10.1016/j.psychres.2020.113157. URL <https://www.sciencedirect.com/science/article/pii/S0165178119325338>.
- T. D. Domingues, A. D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, F. Westermeier, C. Scheibenbogen, J. M. Cliff, L. Nacul, E. M. Lacerda, H. Mouríño, and N. Sepúlveda. Herpesviruses Serology Distinguishes Different Sub-

- groups of Patients From the United Kingdom Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Biobank. *Frontiers in Medicine*, 8, July 2021. doi: 10.3389/fmed.2021.686736.
- E. Engdahl, R. Gustafsson, J. Huang, M. Biström, I. Lima Bomfim, P. Stridh, M. Khademi, N. Brenner, J. Butt, A. Michel, D. Jons, M. Hortlund, L. Alonso-Magdalena, A. K. Hedström, L. Flamand, M. Ihira, T. Yoshikawa, O. Andersen, J. Hillert, L. Alfredsson, T. Waterboer, P. Sundström, T. Olsson, I. Kockum, and A. Fogdell-Hahn. Increased Serological Response Against Human Herpesvirus 6A Is Associated With Risk for Multiple Sclerosis. *Frontiers in Immunology*, 10, 2019. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02715>.
- O. Fluge, I. G. Rekeland, K. Lien, H. Thürmer, P. C. Borchgrevink, C. Schäfer, K. Sørland, J. Aßmus, I. Ktoridou-Valen, I. Herder, M. E. Gotaas, O. Kvammen, K. A. Baranowska, L. M. L. J. Bohnen, S. S. Martinsen, A. E. Lonar, A.-E. H. Solvang, A. E. S. Gya, O. Bruland, K. Risa, K. Alme, O. Dahl, and O. Mella. B-Lymphocyte Depletion in Patients With Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. 170(9):585, Apr. 2019. doi: 10.7326/m18-1451.
- H. Freitag, M. Szklarski, S. Lorenz, F. Sotzny, S. Bauer, A. Philippe, C. Kedor, P. Grabowski, T. Lange, G. Riemarkasten, H. Heidecke, and C. Scheibenbogen. Autoantibodies to Vasoregulatory G-Protein-Coupled Receptors Correlate with Symptom Severity, Autonomic Dysfunction and Disability in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Clinical Medicine*, 10(16):3675, Aug. 2021. ISSN 2077-0383. doi: 10.3390/jcm10163675.
- E. Fruchter, S. Goldberg, D. Fenchel, I. Grotto, K. Ginat, and M. Weiser. The impact of Herpes simplex virus type 1 on cognitive impairments in young, healthy individuals — A historical prospective study. *Schizophrenia Research*, 168(1):292–296, Oct. 2015. ISSN 0920-9964. doi: 10.1016/j.schres.2015.08.036. URL <https://www.sciencedirect.com/science/article/pii/S092099641500465X>.
- K. Fukuda, S. E. Straus, I. Hickie, M. C. Sharpe, J. G. Dobbins, and A. Komaroff. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Annals of Internal Medicine*, 121(12):953, Dec. 1994. doi: 10.7326/0003-4819-121-12-199412150-00009.
- A. González-Cebrián, E. Almenar-Pérez, J. Xu, T. Yu, W. E. Huang, K. Giménez-Orenga, S. Hutchinson, T. Lodge, L. Nathanson, K. J. Morten, A. Ferrer, and E. Oltra. Diagnosis of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome With Partial Least Squares Discriminant Analysis: Relevance of Blood Extracellular Vesicles. *Frontiers in Medicine*, 9, 2022. ISSN 2296-858X. URL <https://www.frontiersin.org/articles/10.3389/fmed.2022.842991>.
- V. Jain, A. Arunkumar, C. Kingdon, E. Lacerda, and L. Nacul. Prevalence of and risk factors for severe cognitive and sleep symptoms in ME/CFS and MS. *BMC Neurology*, 17(1):117, June 2017. ISSN 1471-2377. doi: 10.1186/s12883-017-0896-0. URL <https://doi.org/10.1186/s12883-017-0896-0>.
- L. A. Jason. Differentiating Multiple Sclerosis from Myalgic Encephalomyelitis and Chronic Fatigue Syndrome.

- Insights in Biomedicine*, 02(02), 2017. doi: 10.21767/2572-5610.10027.
- I. Jonker, H. C. Klein, H. E. Duivis, R. H. Yolken, J. G. M. Rosmalen, and R. A. Schoevers. Association between Exposure to HSV1 and Cognitive Functioning in a General Population of Adolescents. The TRAILS Study. *PLOS ONE*, 9(7):e101549, Jan. 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0101549. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0101549>. Publisher: Public Library of Science.
- F. Kasimir, D. Toomey, Z. Liu, A. C. Kaiping, M. E. Ariza, and B. K. Prusty. Tissue specific signature of HHV-6 infection in ME/CFS. *Frontiers in Molecular Biosciences*, 9, 2022. ISSN 2296-889X. URL <https://www.frontiersin.org/articles/10.3389/fmolb.2022.1044964>.
- Z. Khalesi, V. Tamrchi, M. H. Razizadeh, A. Letafati, P. Moradi, A. Habibi, N. Habibi, J. Heidari, M. Noori, M. Nahid Samiei, Z. Azarash, M. Hoseini, H. Saadati, A. Bahavar, M. Farajzade, S. Saeb, M. Hadadi, M. Sorouri Majd, S. Mothlaghzadeh, P. Fazli, K. Asgari, S. J. Kiani, and S. Ghorbani. Association between human herpesviruses and multiple sclerosis: A systematic review and meta-analysis. *Microbial Pathogenesis*, 177:106031, Apr. 2023. ISSN 0882-4010. doi: 10.1016/j.micpath.2023.106031. URL <https://www.sciencedirect.com/science/article/pii/S0882401023000645>.
- S. Klineova and F. D. Lublin. Clinical Course of Multiple Sclerosis. *Cold Spring Harbor Perspectives in Medicine*, 8(9):a028928, Sept. 2018. ISSN , 2157-1422. doi: 10.1101/cshperspect.a028928. URL <http://perspectivesinmedicine.cshlp.org/content/8/9/a028928>. Publisher: Cold Spring Harbor Laboratory Press.
- D. M. Koelle, S. Barcy, M.-L. Huang, R. L. Ashley, L. Corey, J. Zeh, S. Ashton, and D. Buchwald. Markers of Viral Infection in Monozygotic Twins Discordant for Chronic Fatigue Syndrome. *Clinical Infectious Diseases*, 35(5): 518–525, Sept. 2002. ISSN 1058-4838. doi: 10.1086/341774. URL <https://doi.org/10.1086/341774>.
- K. Krishnan, Y. Lin, K.-R. M. Prewitt, and D. A. Potter. Multidisciplinary Approach to Brain Fog and Related Persisting Symptoms Post COVID-19. *Journal of Health Service Psychology*, 48(1):31–38, Feb. 2022. ISSN 2662-2653. doi: 10.1007/s42843-022-00056-7. URL <https://doi.org/10.1007/s42843-022-00056-7>.
- M. J. v. d. Laan, E. C. Polley, and A. E. Hubbard. Super Learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1), Sept. 2007. ISSN 1544-6115. doi: 10.2202/1544-6115.1309. URL <https://www.degruyter.com/document/doi/10.2202/1544-6115.1309.html>. Publisher: De Gruyter.
- E. M. Lacerda, E. W. Bowman, J. M. Cliff, C. C. Kingdon, E. C. King, J.-S. Lee, T. G. Clark, H. M. Dockrell, E. M. Riley, H. Curran, and L. Nacul. The UK ME/CFS Biobank for biomedical research on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) and Multiple Sclerosis. *Open Journal of Bioresources*, 4(1):4, Feb. 2017. ISSN 2056-5542. doi: 10.5334/ojb.28. URL <https://openbioresources.metajnl.com/articles/10.5334/ojb.28>. Number: 1 Publisher: Ubiquity Press.
- E. M. Lacerda, K. Mudie, C. C. Kingdon, J. D. Butterworth, S. O'Boyle, and L. Nacul. The UK ME/CFS Biobank:

- A Disease-Specific Biobank for Advancing Clinical Research Into Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Neurology*, 9:1026, 2018. ISSN 1664-2295. doi: 10.3389/fneur.2018.01026.
- E. M. Lacerda, K. Geraghty, C. C. Kingdon, L. Palla, and L. Nacul. A logistic regression analysis of risk factors in ME/CFS pathogenesis. *BMC Neurology*, 19(1), Nov. 2019a. doi: 10.1186/s12883-019-1468-2.
- E. M. Lacerda, C. McDermott, C. C. Kingdon, J. Butterworth, J. M. Cliff, and L. Nacul. Hope, disappointment and perseverance: Reflections of people with Myalgic encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) and Multiple Sclerosis participating in biomedical research. A qualitative focus group study. *Health Expectations*, 22(3):373–384, 2019b. ISSN 1369-7625. doi: 10.1111/hex.12857. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/hex.12857>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/hex.12857>.
- T. V. Lanz, R. C. Brewer, P. P. Ho, J.-S. Moon, K. M. Jude, D. Fernandez, R. A. Fernandes, A. M. Gomez, G.-S. Nadj, C. M. Bartley, R. D. Schubert, I. A. Hawes, S. E. Vazquez, M. Iyer, J. B. Zuchero, B. Teegen, J. E. Dunn, C. B. Lock, L. B. Kipp, V. C. Cotham, B. M. Ueberheide, B. T. Aftab, M. S. Anderson, J. L. DeRisi, M. R. Wilson, R. J. M. Bashford-Rogers, M. Platten, K. C. Garcia, L. Steinman, and W. H. Robinson. Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. *Nature*, 603(7900):321–327, Mar. 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04432-7. URL <https://www.nature.com/articles/s41586-022-04432-7>. Number: 7900 Publisher: Nature Publishing Group.
- J.-S. Lee, E. M. Lacerda, L. Nacul, C. C. Kingdon, J. Norris, S. O’Boyle, C. h. Roberts, L. Palla, E. M. Riley, and J. M. Cliff. Salivary DNA Loads for Human Herpesviruses 6 and 7 Are Correlated With Disease Phenotype in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Medicine*, 8:656692, Aug. 2021. ISSN 2296-858X. doi: 10.3389/fmed.2021.656692. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8378328/>.
- M. Loebel, P. Grabowski, H. Heidecke, S. Bauer, L. G. Hanitsch, K. Wittke, C. Meisel, P. Reinke, H.-D. Volk, O. Fluge, O. Mella, and C. Scheibenbogen. Antibodies to adrenergic and muscarinic cholinergic receptors in patients with Chronic Fatigue Syndrome. *Brain, Behavior, and Immunity*, 52:32–39, Feb. 2016. ISSN 0889-1591. doi: 10.1016/j.bbi.2015.09.013. URL <https://www.sciencedirect.com/science/article/pii/S0889159115300209>.
- M. Loebel, M. Eckey, F. Sotzny, E. Hahn, S. Bauer, P. Grabowski, J. Zerweck, P. Holenya, L. G. Hanitsch, K. Wittke, P. Borchmann, J.-U. Rüffer, F. Hiepe, K. Ruprecht, U. Behrends, C. Meindl, H.-D. Volk, U. Reimer, and C. Scheibenbogen. Serological profiling of the EBV immune response in Chronic Fatigue Syndrome using a peptide microarray. *PLOS ONE*, 12(6):e0179124, Dec. 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0179124. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0179124>. Publisher: Public Library of Science.
- W. Lundström and R. Gustafsson. Human Herpesvirus 6A Is a Risk Factor for Multiple Sclerosis. *Frontiers in*

- Immunology, 13, 2022. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2022.840753>.
- M. López-Ratón, M. X. Rodríguez-Álvarez, C. C. Suárez, and F. G. Sampedro. OptimalCutpoints: An R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software*, 61(8):1–36, 2014. doi: 10.18637/jss.v061.i08.
- J. a. Malato, L. Graça, and N. Sepúlveda. Impact of Misdiagnosis in Case-Control Studies of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Diagnostics*, 13(3):531, Jan. 2023. ISSN 2075-4418. doi: 10.3390/diagnostics13030531. URL <https://www.mdpi.com/2075-4418/13/3/531>. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
- M. E. Marcocci, G. Napoletani, V. Protto, O. Kolesova, R. Piacentini, D. D. L. Puma, P. Lomonte, C. Grassi, A. T. Palamara, and G. D. Chiara. Herpes Simplex Virus-1 in the Brain: The Dark Side of a Sneaky Infection. *Trends in Microbiology*, 28(10):808–820, Oct. 2020. ISSN 0966-842X, 1878-4380. doi: 10.1016/j.tim.2020.03.003. URL [https://www.cell.com/trends/microbiology/abstract/S0966-842X\(20\)30074-3](https://www.cell.com/trends/microbiology/abstract/S0966-842X(20)30074-3). Publisher: Elsevier.
- R. Martinez-Reviejo, S. Tejada, G. A. R. Adebanjo, C. Chello, M. C. Machado, F. R. Parisella, M. Campins, A. Tammaro, and J. Rello. Varicella-Zoster virus reactivation following severe acute respiratory syndrome coronavirus 2 vaccination or infection: New insights. *European Journal of Internal Medicine*, 104:73–79, Oct. 2022. ISSN 0953-6205, 1879-0828. doi: 10.1016/j.ejinme.2022.07.022. URL [https://www.ejinme.com/article/S0953-6205\(22\)00270-9/fulltext](https://www.ejinme.com/article/S0953-6205(22)00270-9/fulltext). Publisher: Elsevier.
- P. A. McCombe, T. P. Gordon, and M. W. Jackson. Bladder dysfunction in multiple sclerosis. *Expert Review of Neurotherapeutics*, 9(3):331–340, Mar. 2009. ISSN 1473-7175. doi: 10.1586/14737175.9.3.331. URL <https://doi.org/10.1586/14737175.9.3.331>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1586/14737175.9.3.331>.
- A. Melvin, E. Lacerda, H. M. Dockrell, S. O’Rahilly, and L. Nacul. Circulating levels of GDF15 in patients with myalgic encephalomyelitis/chronic fatigue syndrome. *Journal of Translational Medicine*, 17(1):409, Dec. 2019. ISSN 1479-5876. doi: 10.1186/s12967-019-02153-6. URL <https://doi.org/10.1186/s12967-019-02153-6>.
- G. Morris and M. Maes. Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics. *BMC Medicine*, 11(1), Sept. 2013. doi: 10.1186/1741-7015-11-205.
- M. J. Murphy, L. Fani, M. K. Ikram, M. Ghanbari, and M. A. Ikram. Herpes simplex virus 1 and the risk of dementia: a population-based study. *Scientific Reports*, 11(1):8691, Apr. 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-87963-9. URL <https://www.nature.com/articles/s41598-021-87963-9>. Number: 1 Publisher: Nature Publishing Group.

- L. Nacul, E. M. Lacerda, C. C. Kingdon, H. Curran, and E. W. Bowman. How have selection bias and disease misclassification undermined the validity of myalgic encephalomyelitis/chronic fatigue syndrome studies? *Journal of Health Psychology*, 24(12):1765–1769, Oct. 2019. ISSN 1461-7277. doi: 10.1177/1359105317695803.
- L. Nacul, F. J. Authier, C. Scheibenbogen, L. Lorusso, I. B. Helland, J. A. Martin, C. A. Sirbu, A. M. Mengshoel, O. Polo, U. Behrends, H. Nielsen, P. Grabowski, S. Sekulic, N. Sepulveda, F. Estévez-López, P. Zalewski, D. F. H. Pheby, J. Castro-Marrero, G. K. Sakkas, E. Capelli, I. Brundslund, J. Cullinan, A. Krumina, J. Bergquist, M. Murovska, R. C. W. Vermuelen, and E. M. Lacerda. European Network on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (EUROMENE): Expert Consensus on the Diagnosis, Service Provision, and Care of People with ME/CFS in Europe. *Medicina*, 57(5):510, May 2021. ISSN 1648-9144. doi: 10.3390/medicina57050510. URL <https://www.mdpi.com/1648-9144/57/5/510>.
- H. Nakayama, R. Yamazaki, J. Kato, Y. Koda, M. Sakurai, and T. Mori. Impact of Specific Antibody Level on Human Herpesvirus 6 Reactivation after Allogeneic Stem Cell Transplantation. *Transplantation and Cellular Therapy*, 27(2):174.e1–174.e5, Feb. 2021. ISSN 2666-6367. doi: 10.1016/j.jtct.2020.10.011. URL <https://www.sciencedirect.com/science/article/pii/S2666636720300191>.
- A. Ocon. Caught in the thickness of brain fog: exploring the cognitive symptoms of Chronic Fatigue Syndrome. *Frontiers in Physiology*, 4, 2013. ISSN 1664-042X. URL <https://www.frontiersin.org/articles/10.3389/fphys.2013.00063>.
- D. Ohanian, A. Brown, M. Sunnquist, J. Furst, L. Nicholson, L. Klebek, and L. A. Jason. Identifying Key Symptoms Differentiating Myalgic Encephalomyelitis and Chronic Fatigue Syndrome from Multiple Sclerosis. *Neurology (E-Cronicon)*, 4(2):41–45, 2016.
- F. Patti and S. Lo Fermo. Lights and Shadows of Cyclophosphamide in the Treatment of Multiple Sclerosis. *Autoimmune Diseases*, 2011:961702, Mar. 2011. ISSN 2090-0430. doi: 10.4061/2011/961702. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3087413/>.
- E. Polley, E. LeDell, C. Kennedy, S. Lendle, and M. v. d. Laan. SuperLearner: Super Learner Prediction, July 2023. URL <https://cran.r-project.org/web/packages/SuperLearner/index.html>.
- A. D. Port, A. Orlin, S. Kiss, S. Patel, D. J. D'Amico, and M. P. Gupta. Cytomegalovirus Retinitis: A Review. *Journal of Ocular Pharmacology and Therapeutics*, 33(4):224–234, May 2017. ISSN 1080-7683. doi: 10.1089/jop.2016.0140. URL <https://www.liebertpub.com/doi/10.1089/jop.2016.0140>. Publisher: Mary Ann Liebert, Inc., publishers.
- M. Pugliatti, G. Rosati, H. Carton, T. Riise, J. Drulovic, L. Vécsei, and I. Milanov. The epidemiology of multiple sclerosis in Europe. *European Journal of Neurology*, 13(7):700–722, 2006. ISSN 1468-1331. doi: 10.1111/j.1468-1331.2006.01342.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-1331.2006.01342.x>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-1331.2006.01342.x>.

- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.
- S. Ramos, E. Brenu, S. Broadley, R. Kwiatek, J. Ng, T. Nguyen, S. Freeman, D. Staines, and S. Marshall-Gradisnik. Regulatory T, natural killer T and \upgamma\updelta T cells in multiple sclerosis and chronic fatigue syndrome/myalgic encephalomyelitis: a comparison. *Asian Pacific Journal of Allergy and Immunology*, 2016. doi: 10.12932/ap0733.
- S. Rasa, Z. Nora-Krukla, N. Henning, E. Eliassen, E. Shikova, T. Harrer, C. Scheibenbogen, M. Murovska, B. K. Prusty, and the European Network on ME/CFS (EUROMENE). Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 16(1):268, Oct. 2018. ISSN 1479-5876. doi: 10.1186/s12967-018-1644-y. URL <https://doi.org/10.1186/s12967-018-1644-y>.
- S. Rasa-Dzelzkaleja, A. Krumina, S. Capenko, Z. Nora-Krukla, S. Gravelsina, A. Vilmane, L. Ievina, Y. Shoenfeld, M. Murovska, and the VirA project. The persistent viral infections in the development and severity of myalgic encephalomyelitis/chronic fatigue syndrome. *Journal of Translational Medicine*, 21(1):33, Jan. 2023. ISSN 1479-5876. doi: 10.1186/s12967-023-03887-0. URL <https://doi.org/10.1186/s12967-023-03887-0>.
- I. G. Rekeland, A. Fosså, A. Lande, I. Ktoridou-Valen, K. Sørland, M. Holsen, K. J. Tronstad, K. Risa, K. Alme, M. K. Viken, B. A. Lie, O. Dahl, O. Mella, and O. Fluge. Intravenous Cyclophosphamide in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. An Open-Label Phase II Study. *Frontiers in Medicine*, 7, 2020. ISSN 2296-858X. URL <https://www.frontiersin.org/articles/10.3389/fmed.2020.00162>.
- B. Rich. table1: Tables of Descriptive Statistics in HTML, Jan. 2023. URL <https://cran.r-project.org/web/packages/table1/index.html>.
- P. Rieckmann, A. Boyko, D. Centonze, I. Elovaara, G. Giovannoni, E. Havrdová, O. Hommes, J. Kesselring, G. Kobelt, D. Langdon, J. LeLorier, S. A. Morrow, C. Oreja-Guevara, S. Schippling, C. Thalheim, H. Thompson, and P. Vermersch. Achieving patient engagement in multiple sclerosis: A perspective from the multiple sclerosis in the 21st Century Steering Group. *Multiple Sclerosis and Related Disorders*, 4(3):202–218, May 2015. ISSN 2211-0348, 2211-0356. doi: 10.1016/j.msard.2015.02.005. URL [https://www.msard-journal.com/article/S2211-0348\(15\)00024-3/fulltext](https://www.msard-journal.com/article/S2211-0348(15)00024-3/fulltext). Publisher: Elsevier.
- M. C. Rivera, C. Mastronardi, C. Silva-Aldana, M. Arcos-Burgos, and B. Lidbury. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Comprehensive Review. *Diagnostics*, 9(3):91, Aug. 2019. doi: 10.3390/diagnostics9030091.
- X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12:77, 2011.
- L. S. Rodrigues, L. H. da Silva Nali, C. O. D. Leal, E. C. Sabino, E. M. Lacerda, C. C. Kingdon, L. Nacul, and C. M. Romano. HERV-K and HERV-W transcriptional activity in myalgic encephalomyelitis/chronic fatigue syndrome. *Autoimmunity Highlights*, 10(1):12, Nov. 2019. ISSN 2038-3274. doi: 10.1186/s13317-019-0122-8. URL

- <https://doi.org/10.1186/s13317-019-0122-8>.
- B. V. Rooney, B. E. Crucian, D. L. Pierson, M. L. Laudenslager, and S. K. Mehta. Herpes Virus Reactivation in Astronauts During Spaceflight and Its Application on Earth. *Frontiers in Microbiology*, 10, 2019. ISSN 1664-302X. URL <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00016>.
- A. J. Ross, M. S. Medow, P. C. Rowe, and J. M. Stewart. What is brain fog? An evaluation of the symptom in postural tachycardia syndrome. *Clinical Autonomic Research*, 23(6):305–311, Dec. 2013. ISSN 1619-1560. doi: 10.1007/s10286-013-0212-z. URL <https://doi.org/10.1007/s10286-013-0212-z>.
- M. Ruiz-Pablos, B. Paiva, R. Montero-Mateo, N. Garcia, and A. Zabaleta. Epstein-Barr Virus and the Origin of Myalgic Encephalomyelitis or Chronic Fatigue Syndrome. *Frontiers in Immunology*, 12, 2021. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2021.656797>.
- R. Sakakibara, S. Sawai, and T. Ogata. Varicella-zoster virus infection and autonomic dysfunction. *Autonomic Neuroscience: Basic and Clinical*, 242, Nov. 2022. ISSN 1566-0702. doi: 10.1016/j.autneu.2022.103018. URL [https://www.autonomicneuroscience.com/article/S1566-0702\(22\)00077-7/fulltext](https://www.autonomicneuroscience.com/article/S1566-0702(22)00077-7/fulltext). Publisher: Elsevier.
- C. Scheibenbogen, H. Freitag, J. Blanco, E. Capelli, E. Lacerda, J. Authier, M. Meeus, J. C. Marrero, Z. Nora-Krukle, E. Oltra, and others. The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE. *Journal of translational medicine*, 15(1):162, 2017. doi: <https://doi.org/10.1186/s12967-017-1263-z>.
- S. Sedighi, O. Gholizadeh, S. Yasamineh, S. Akbarzadeh, P. Amini, P. Favakehi, H. Afkhami, A. Firouzi-Amandi, D. Pahlevan, M. Eslami, B. Yousefi, V. Poortahmasebi, and M. Dadashpour. Comprehensive Investigations Relationship Between Viral Infections and Multiple Sclerosis Pathogenesis. *Current Microbiology*, 80(1):15, Dec. 2022. ISSN 1432-0991. doi: 10.1007/s00284-022-03112-z.
- N. Sepúlveda. Impact of genetic variation on the molecular mimicry between Anoctamin-2 and Epstein-Barr virus nuclear antigen 1 in Multiple Sclerosis. *Immunology Letters*, 238:29–31, Oct. 2021. ISSN 1879-0542. doi: 10.1016/j.imlet.2021.07.007.
- N. Sepúlveda, J. Carneiro, E. Lacerda, and L. Nacul. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome as a Hyper-Regulated Immune System Driven by an Interplay Between Regulatory T Cells and Chronic Human Herpesvirus Infections. *Frontiers in Immunology*, 10, 2019. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02684>.
- N. Sepúlveda, J. a. Malato, F. Sotzny, A. D. Grabowska, A. Fonseca, C. Cordeiro, L. Graça, P. Biecek, U. Behrends, J. Mautner, F. Westermeier, E. M. Lacerda, and C. Scheibenbogen. Revisiting IgG Antibody Reactivity to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Its Potential Application to Disease Diagnosis. *Frontiers in Medicine*, 9, 2022. ISSN 2296-858X. doi: 10.3389/fmed.2022.921101. URL <https://www.frontiersin.org/articles/10.3389/fmed.2022.921101>.

- S. S. Soldan and P. M. Lieberman. Epstein–Barr virus and multiple sclerosis. *Nature Reviews Microbiology*, 21(1): 51–64, Jan. 2023. ISSN 1740-1534. doi: 10.1038/s41579-022-00770-5. URL <https://www.nature.com/articles/s41579-022-00770-5>. Number: 1 Publisher: Nature Publishing Group.
- F. Sotzny, J. Blanco, E. Capelli, J. Castro-Marrero, S. Steiner, M. Murovska, and C. Scheibenbogen. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome – Evidence for an autoimmune disease. *Autoimmunity Reviews*, 17(6):601–609, June 2018. doi: 10.1016/j.autrev.2018.01.009.
- F. Sotzny, I. S. Filgueiras, C. Kedor, H. Freitag, K. Wittke, S. Bauer, N. Sepúlveda, D. L. Mathias da Fonseca, G. C. Baiocchi, A. H. C. Marques, M. Kim, T. Lange, D. R. Plaça, F. Luebber, F. M. Paulus, R. De Vito, I. Jurisica, K. Schulze-Forster, F. Paul, J. Bellmann-Strobl, R. Rust, U. Hoppmann, Y. Shoenfeld, G. Riemarkasten, H. Heidecke, O. Cabral-Marques, and C. Scheibenbogen. Dysregulated autoantibodies targeting vaso- and immunoregulatory receptors in Post COVID Syndrome correlate with symptom severity. *Frontiers in Immunology*, 13, 2022. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2022.981532>.
- S. Steiner, S. C. Becker, J. Hartwig, F. Sotzny, S. Lorenz, S. Bauer, M. Löbel, A. B. Stitrich, P. Grabowski, and C. Scheibenbogen. Autoimmunity-Related Risk Variants in PTPN22 and CTLA4 Are Associated With ME/CFS With Infectious Onset. *Frontiers in Immunology*, 11, 2020. ISSN 1664-3224. URL <https://www.frontiersin.org/articles/10.3389/fimmu.2020.00578>.
- M. Szklarski, H. Freitag, S. Lorenz, S. C. Becker, F. Sotzny, S. Bauer, J. Hartwig, H. Heidecke, K. Wittke, C. Kedor, L. G. Hanitsch, P. Grabowski, N. Sepúlveda, and C. Scheibenbogen. Delineating the Association Between Soluble CD26 and Autoantibodies Against G-Protein Coupled Receptors, Immunological and Cardiovascular Parameters Identifies Distinct Patterns in Post-Infectious vs. Non-Infection-Triggered Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Immunology*, 12, Apr. 2021. doi: 10.3389/fimmu.2021.644548.
- K. D. Tarter, A. M. Simanek, J. B. Dowd, and A. E. Aiello. Persistent Viral Pathogens and Cognitive Impairment Across the Life Course in the Third National Health and Nutrition Examination Survey. *The Journal of Infectious Diseases*, 209(6):837–844, Mar. 2014. ISSN 0022-1899. doi: 10.1093/infdis/jit616. URL <https://doi.org/10.1093/infdis/jit616>.
- K. Tengvall, J. Huang, C. Hellström, P. Kammer, M. Biström, B. Ayoglu, I. Lima Bomfim, P. Stridh, J. Butt, N. Brenner, A. Michel, K. Lundberg, L. Padyukov, I. E. Lundberg, E. Svenungsson, I. Ernberg, S. Olafsson, A. T. Dilthey, J. Hillert, L. Alfredsson, P. Sundström, P. Nilsson, T. Waterboer, T. Olsson, and I. Kockum. Molecular mimicry between Anoctamin 2 and Epstein-Barr virus nuclear antigen 1 associates with multiple sclerosis risk. *Proceedings of the National Academy of Sciences of the United States of America*, 116(34):16955–16960, Aug. 2019. ISSN 1091-6490. doi: 10.1073/pnas.1902623116.
- J. D. Tucker and A. S. Bertke. Assessment of cognitive impairment in HSV-1 positive schizophrenia and bipolar patients: Systematic review and meta-analysis. *Schizophrenia Research*, 209:40–47, July 2019. ISSN 0920-9964. doi: 10.1016/j.schres.2019.01.001. URL <https://www.sciencedirect.com/science/article/>

[pii/S0920996419300015](#).

- T. Vogl, I. N. Kalka, S. Klompus, S. Leviatan, A. Weinberger, and E. Segal. Systemic antibody responses against human microbiota flagellins are overrepresented in chronic fatigue syndrome patients. *Science Advances*, 8(38): eabq2422, Sept. 2022. doi: 10.1126/sciadv.abq2422. URL <https://www.science.org/doi/10.1126/sciadv.abq2422>. Publisher: American Association for the Advancement of Science.

Chapter 7

Relation between domain-specific severity profiles and IgG antibody responses in ME/CFS

J. Malato, L. Graça, J.-S. Lee, J.M. Cliff, L. Nacul, E.M. Lacerda, and N. Sepúlveda. Relation between domain-specific severity profiles and IgG antibody responses in ME/CFS. *In preparation.* 2024.

7.1 Introduction

Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) is a disease with unknown aetiology and pathogenesis. Patients afflicted with this disease experience post-exertional malaise (PEM) after what would be considered normal levels of activity (Carruthers et al. 2003) and long-lasting unexplained fatigue that is not alleviated by rest (Fukuda et al. 1994), together with a wide range of incapacitating symptoms from various domains, reflecting the effect of specific systems and pathophysiological mechanisms in the body (see Table 1.1). The lack of established biomarkers leaves ME/CFS diagnosis to be made on the clinical assessment of symptoms and the exclusion of other known diseases that could justify the state of fatigue. This has led to multiple case definitions for the disease to be proposed over the years (Brurberg et al. 2014; Lim and Son 2020), which in practice means that an individual suspected for ME/CFS could be diagnosed by one criterion and excluded by another (Malato et al. 2021). This lack of diagnostic agreement results in research studies that can present discrepant or even competing results (Nacul et al. 2019).

Additionally, the natural symptom heterogeneity observed across ME/CFS individuals suggests that ME/CFS is likely an umbrella term, encompassing a spectrum of phenotypically

similar illnesses that are included together (Malato et al. 2023). As such, dealing with the patients as a whole may be one of the reasons leading to a lack of consistent and reproducible results, possibly due to misdiagnosis of patients. To tackle this, the stratification of suspected cases into similar subgroups could help to identify more specific cluster profiles for research or treatment purposes (Jason et al. 2005; Scheibenbogen et al. 2017). Corroborating this strategy, research studies have shown results on specific subsets of the disease when splitting diagnosed individuals by disease severity or cause of disease onset. For instance, Montoya et al. (2017) correlated proinflammatory cytokines with disease severity in patients, and Cliff et al. (2019) split ME/CFS into subgroups for mild–moderate and severely affected individuals and found increased mucosal-associated invariant T (MAIT) lymphocytes in the latter subgroup, suggesting the effect of the immune system in the exacerbation of the observed symptoms. Other studies have demonstrated differences in subsets reporting acute response to a herpesviruses infection prior to the development of the disease (Domingues et al. 2021; Sepúlveda et al. 2022; Domingues et al. 2023). These results are in line with the hypothesis that at least a subset of ME/CFS patients has an autoimmune aetiology, which could characterise the significant portion of cases reporting a previous infection-like event before the development of the disease (Sotzny et al. 2018). However, there is still no unified strategy to subtype ME/CFS (Jason et al. 2005).

The aim of this preliminary study is to increase our understanding of how related symptoms could explain the response towards six different herpesviruses. We first studied the symptomatic heterogeneity across ME/CFS patients and healthy controls to look for possible discernible clusters. Then, we categorised the symptoms by seven domains (immunological, neuroendocrine, PEM, autonomic, neurocognitive, neurophysiological, and pain) and identified subgroups based on similar profiles of severity. As such, instead of analysing the relationship between all symptoms screened at diagnosis, we considered the seven domains as specific phenotypes associated with particular mechanisms and stratified them instead. Afterwards, antibody IgG concentrations and population seroprevalence towards distinct herpesviruses were measured across the domain subgroups and compared with healthy controls.

The results here presented are part of an undergoing study. Future work will look to study the combinations of domain severity from each individual to ascertain more complete patterns of classification of the disease.

7.2 Materials and methods

7.2.1 Study participants

Throughout the initial stages of the study, we worked with a total sample size of 347 individuals split into two cohorts of 241 ME/CFS patients and 106 healthy controls matched for sex and age. All participants were adults and part of the UK ME/CFS Biobank (UKMEB) (Lacerda et al. 2017). ME/CFS patients were ascertained for compliance by the 1994 US Centre for Disease Control and Prevention Criteria (CDC-1994, Fukuda et al. 1994) or the 2003 Canadian Consensus Criteria (CCC-2003, Carruthers et al. 2003). According to the biobank, aside from the diagnosis requirements, patients would be excluded if they (i) used drugs known to alter immune function in the preceding three months or antiviral medication; (ii) had any vaccinations in the preceding three months; (iii) had a history of acute and chronic infectious diseases (but not herpes viruses); (iv) were diagnosed with another severe illness such as cancer, coronary heart disease, or uncontrolled diabetes; (v) had a severe mood disorder; (vi) had been pregnant or breastfeeding in the preceding 12 months; or (vii) were morbidly obese ($BMI \geq 40 \text{ kg/m}^2$). Additional information related to the recruitment and inclusion of individuals and management of samples by the UKMEB can be found elsewhere (Lacerda et al. 2017; 2018).

Similar to previous studies (Cliff et al. 2019; Domingues et al. 2023; Chapter 6), age, sex, and disease duration were summarised and compared across healthy controls and patients (Supplementary Table D.1). Pearson's χ^2 test was applied to qualitative measures, and the non-parametric two-sided Kruskal-Wallis rank sum test was applied to quantitative values. All statistical tests were performed using a significance level of 5%.

7.2.2 Symptomatology assessment

At enrolment, all participants were asked to answer a symptom assessment form related to the severity range of 57 specific symptoms experienced over the previous seven days (Supplementary Table D.2). Each symptom was rated within an ordinal scale with four possible degrees of severity: absent, mild, moderate, or severe. This symptom assessment was included as part of a questionnaire for ME/CFS diagnosis.

During data preprocessing, the relative frequency of each symptom in the ME/CFS study population was calculated (Figure 7.1). Symptoms with a frequency of unreported answers higher than 20% were identified and removed. Ultimately, 10 symptoms surpassed the missing data threshold and were removed from the study, leaving the analyses to be performed with the remaining 47. The symptoms removed were intolerance to alcohol (alcoholintoler), pain in chest and/or abdomen (chestabdpain), unusually cold hands and/or feet (coldhandsfeet), difficulty retaining information (diffdecisions), difficulty finding or saying words (diffwords), fatigue or exhaustion lasting for more than 24 hours after what would be considered normal levels of activity (fatiguelast24h), palpitations and irregular heartbeats (palpitations), feeling sick or nauseated (sick_nausea), muscle twitching (twitching), and abnormal appetite and/or significant changes in weight (weightchange).

7.2.3 Herpesviruses serological data

Plasma samples from eligible patients were collected for quantification of immunoglobulin G (IgG) antibodies against human cytomegalovirus (CMV), Epstein-Barr virus (EBV) nuclear antigen-1 (EBV EBNA1) and EBV viral capsid antigen (EBV VCA), human herpesvirus-6 (HHV6), herpes simplex virus 1 (HSV-1), herpes simplex virus 2 (HSV-2), and varicella-zoster virus (VZV). Concentration assays were measured by commercial quantitative ELISA and expressed in arbitrary units per millilitre (U/ml). Seropositivity cutoff of each herpesvirus was determined according to the manufacturer's instructions. Under those assumptions, individuals were classified as seronegative or seropositive, with some individuals being considered equivocal. In this study, we treated equivocal individuals as seronegative. More information related

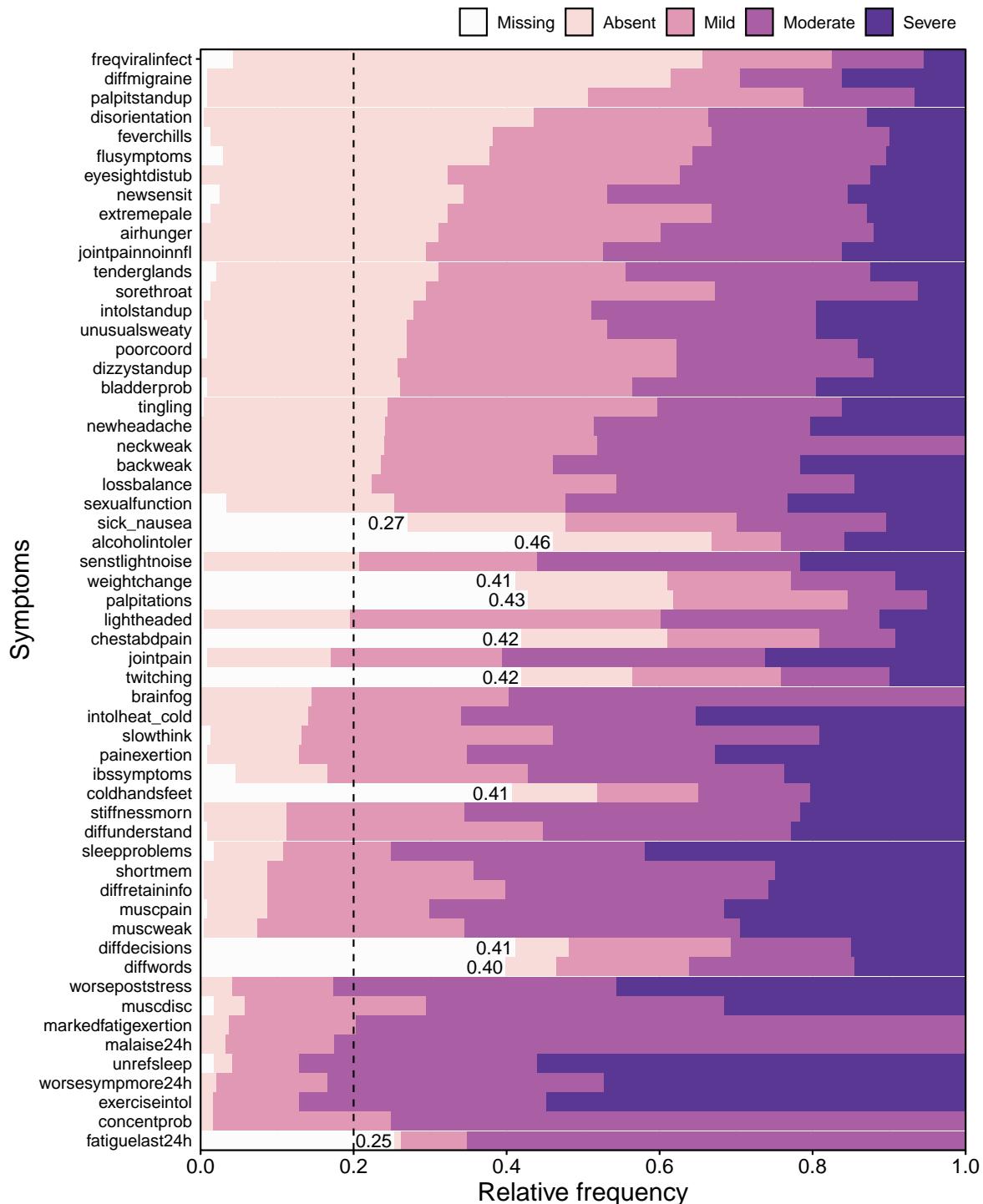


Figure 7.1: Relative frequency of ordinal degrees of severity on each one of the original 57 symptoms available across the population of ME/CFS patients ($n = 241$). Symptoms are ordered by decreasing level of severity and relative proportion. Vertical dashed line at the 0.2 mark designates the missing values cutoff. Symptoms with a proportion of unreported (missing) severity across all individuals higher than that value were identified (relative proportions that surpass the cutoff are indicated in the missing columns) and removed from the analysis. A more detailed description of each symptom can be found elsewhere (Supplementary Table D.2).

to the laboratory procedures can be found in previous studies (Cliff et al. 2019; Domingues et al. 2021).

The use of both data on antibody titrations and seroprevalence serves to complement inferences when comparing cohorts. Since antibodies against different herpesviruses can vary, the analysis of seroprevalence is useful in a context where IgG concentrations tend to be lower, with the population more homogeneously split between seronegative and seropositive.

7.2.4 Symptoms intra- and inter-rater agreement

We used the 47 selected symptoms, each characterised by four possible degrees of severity, to quantify the similarity of symptoms for each one of the healthy controls and ME/CFS-diagnosed individuals in the data set (intra-rater agreement). We computed the overall severity proportions and estimated the entropy using the usual (Shannon) formula (Shannon 1948),

$$H(X) = - \sum_{k=1}^4 p_{ik} \log_2 p_{ik},$$

where, for an individual i , p_{ik} represents the proportion of each one of the severity levels k (in the equation, $k = \{1, 2, 3, 4\}$ represents the four categories of severity, with values 1 for absent, 2 for mild, 3 for moderate, and 4 for severe). On each individual, this proportion for a severity category k varies between 0 (non-existent across symptoms) and 1 (all symptoms with the same degree of severity), and the total sum of all four proportions equals 1. As a result, in cases where an individual experiences the exact same severity level across all symptoms, the estimated entropy would be 0. Conversely, in a scenario where a patient demonstrates an extremely heterogeneous profile, with a high degree of fluctuation from one symptom to the other, we can assume that the severity level follows a Uniform distribution on each symptom with the resulting proportions for the four levels being equally split at 25% ($\frac{1}{4} = 0.25$). Under this scenario, the estimated entropy would reach its maximum possible value of 2 (maximum value of entropy estimated as $\log_2 k$; in this case, we have $\log_2 4 = 2$), signifying maximal uncertainty or diversity in the symptomatological severity pattern of an individual.

To compare the similarity between the severity profile of all 347 participants (inter-rater agreement), we iteratively calculated a pairwise similarity matrix between each two individuals by estimating the Cohen's κ coefficient (Cohen 1960). Afterwards, the resulting similarity matrix was analysed by classical multidimensional scaling (MDS). More information on this method can be seen in a previous application to similar data from the UKMEB (Malato et al. 2021).

7.2.5 Construction of patient subgroups from symptomatological data

Originally, the symptoms recorded were grouped into seven distinct domains: autonomic ($n = 8$), immunological ($n = 7$), neurological cognitive (neurocognitive) ($n = 16$), neuroendocrine ($n = 4$), pain ($n = 5$), post-exertional malaise (PEM, $n = 5$), and sleep function ($n = 2$). This grouping scheme arises from the detailed description of symptoms in the CCC-2003 criteria and from conversations with specialised clinicians at the UKMEB. It serves as a way to organise the potentially afflicted systems and trace back from the combination of expressed symptoms. For this study, the two symptoms related to sleep dysfunction (sleepproblems and unrefsleep) were included in the neurophysiological domain ($n = 7$), which is a larger group that also included other symptoms from both autonomic and neurocognitive domains (Supplementary Table D.2). Henceforth, analysis and stratification of ME/CFS patients were done considering the entirety of symptoms and considering symptoms split into the seven domains.

We implemented the latent class analysis (LCA) modelling method on the severity profiles of ME/CFS patients. Considering all symptoms or each domain, this probabilistic algorithm identifies subgroups (or latent classes) in the symptoms' multivariate categorical data that have a similar profile of severity. Simply put, it defines "clusters" of similar patients, estimating the posterior probability that relates each individual to each one of the proposed subgroups.

Since the number of latent classes in the model is a predetermined parameter required before performing the LCA, we tested different models by sequentially increasing the parameter value between 2 and 10. This means that, at the bare minimum, we performed LCA considering

the ME/CFS population to be stratified into two subgroups; and at the maximum, and admitting a highly heterogeneous symptomatological profile from the patients, the model clustered the patients into ten groups.

The optimal number of latent classes in the models was initially chosen based on the Akaike and the Bayesian information criteria (AIC and BIC, respectively). Both information criteria evaluate and reward goodness of fit while at the same time penalising for the number of parameters considered. This is a way to adjust for the model's complexity and avoid overfitting, with the goal being to choose a more parsimonious model. Looking at the formulas for AIC ($AIC = -2 \ln(\hat{L}) + 2k$), and BIC ($BIC = -2 \ln(\hat{L}) + k \ln(n)$), respectively, $\ln(\hat{L})$ is the maximised log-likelihood function for the model, k is the number of parameters, and n is the sample size used. Therefore, under the principle of parsimony, for a set of candidate models, the "best" model will be the one presenting the smallest score value for AIC or BIC. Because the penalty term varies between the two information criteria, it is likely that the selected models for the same domain have a different number of classes. AIC adds a direct penalty related to the number of parameters in the model (term $2k$). BIC penalty increases logarithmically with sample size (term $k \ln(n)$). This makes BIC the more strict criteria of the two and one should expect it to go for models with fewer parameters.

Once the model with optimal number of latent groups was chosen in each domain, the estimated subgroups were ordered based on the increasing symptom severity profile according to the LCA class membership probabilities. For the subgroups of each domain, a sequential number was attributed, meaning that any group denominated 'g1' classifies individuals within a pattern of reduced severity (mostly absent to mild). Oppositely, subgroups of higher numbers classify the population with a more severe combination of symptoms.

7.2.6 Analysis of serological data

With ME/CFS patients stratified by multiple subgroups, we compared both antibody concentration and seropositivity values against healthy controls with the intent of identifying differences that would relate the profile of symptom severity and past exposure to the assessed

herpesviruses. For each virus, we performed two complementary analyses. First, a two-sided Kruskal-Wallis sum rank test was used to compare the median IgG antibody concentration values (\log_{10} -transformed) on healthy controls and ME/CFS subgroups. We then proceeded with the test's extension, the non-parametric pairwise Mann-Whitney U test, to perform comparisons within population pairs. Second, we created contingency tables for the number of seropositive and seronegative individuals across the cohorts for healthy controls, ME/CFS patients as a whole, and each one of the domains' subgroups, and tested whether there were differences in population frequencies. We used Pearson's χ^2 for independence on comparisons across multiple cohorts, followed by a pairwise Fisher's exact test to iteratively compare between two groups. On both analyses, the level of significance was set at 5%. All estimated (raw) p-values were adjusted for multiple testing post-hoc, using the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995), ensuring a global false discovery rate of 5%.

7.2.7 Statistical software

All statistical analyses were conducted using the R software, version 4.2.3 (2023-03-15) (R Core Team 2020). For the LCA on polytomous outcomes variables we used the package *poLCA* (Linzer and Lewis 2011) and to estimate inter-rater agreement between pairs of individuals we used the package *irr* (Gamer et al. 2012).

7.3 Results

7.3.1 Characterisation of study participants

A summary of the basic characteristics of the study population and considered subpopulations, based on severity and infection trigger, can be found in Supplementary Table D.1. Both study cohorts of healthy controls and ME/CFS had more female patients than males, almost at a ratio of 3 to 1 (Pearson's χ^2 test, $p = 0.819$). The average age was 41.7 and 42.2 years old, respectively, for healthy individuals and patients (Kruskal-Wallis test, $p = 0.827$). ME/CFS patients had an average of 12.4 years of disease duration and the large majority was diagnosed

by both the CDC-1994 and CCC-2003 ($n = 205$, 85.1%), with fewer individuals diagnosed by only the former ($n = 33$, 13.7%), or the latter ($n = 3$, 1.2%).

The ME/CFS subgroups for mild to moderate ($n = 188$, 78.0%) and severely affected patients ($n = 53$, 22.0%) did not differ substantially in sex or age distribution. However, severe patients had significantly longer duration of the disease by an average of 4.4 years (median difference = 5.9 years, Kruskal-Wallis test, $p = 0.001$), and had all but one individual diagnosed by both diagnoses (Pearson's χ^2 test, $p = 0.004$).

The four infection trigger-based subgroups also had similar sex ($p = 0.642$), disease duration ($p = 0.495$), and age distributions ($p = 0.338$). Mean ages were 43.5, 40.2, 43.1, and 40.5 years old for patients who did not know their disease trigger (S0), patients who reported a non-infection trigger (S1), patients who reported an infection trigger but the infection was not confirmed by a lab test (S2), and patients who reported an infection trigger with that infection confirmed by a lab test (S3), respectively. Differently from the binary severity subgroups, there were no significant differences in the diagnosis used on the participants of each subgroup ($p = 0.095$).

7.3.2 Symptom description and similarity profiles

In the ME/CFS population, the distribution of severity levels varied among symptoms (Figure 7.1). Certain symptoms, such as frequent viral infections (freqviralinfect, immunological domain) and different or worse migraines (diffmigraine, pain domain), were absent in the majority of patients, marked as such in 61.4% and 60.6% of the population, respectively. Notably, absent values only appeared in an average of 19.6% of cases across all symptoms. Moderate was the predominant severity level overall, manifesting on average in 31.2% of cases. Conversely, only two symptoms appeared as severe in the majority of patients: unrefreshing sleep (unrefsleep, neurophysiological domain) and intolerance to exercise (exerciseintol, PEM domain and a requisite for the CCC-2003 diagnosis), with 56.0% and 54.8% of the patients study population marking them as severe, respectively.

Among the 47 selected symptoms, five did not exhibit the maximum value of severity.

These were neck weakness (neckweak), brain fog (brainfo), and difficulties in concentrating (concentprob), from the neurocognitive domain, and long-lasting marked physical or mental fatigue and malaise after what would be considered minimal/normal levels of exertion (markedfatigexertion and malaise24h), belonging to the PEM domain. Interestingly, despite not including severe as a severity level, with the exception of brainfog, these symptoms had close to no absent values, being predominantly marked as moderate or mild.

The individual estimated entropy values in the population varied between 0.00 and 1.99, the minimal and almost-maximal possible values, respectively (Figure 7.2A). There were clear differences between healthy controls and ME/CFS patients. Individuals from the healthy population were the only ones with high levels of homogeneity in all symptoms. Of the 106 healthy participants, 13 (12.26%) consistently reported the absence of any symptoms, with the large majority ($n = 91$, 85%) displaying entropy below the midrange value of 1.00. Conversely, the lowest reached entropy in the ME/CFS population was 0.66, with 235 (97.51%) patients reporting estimates above the midrange value. This showed that at the individual level, no ME/CFS patient experienced the same severity level across all recorded symptoms, which, similarly to some healthy controls, would result in an entropy value of zero. The more consistent ME/CFS individuals identified most of their symptoms as either moderate with some mild symptoms or severe with some moderate symptoms. These results indicated that a majority of patients exhibited a heterogeneous combination of severity levels across the assessed symptoms, which contrasted with the cohort of healthy controls that exhibited lower entropy due to the absence of symptoms.

Comparing the inter-rater profile similarity, classical MDS results for the first two components explained 32.42% of the total variance (Figure 7.2B). The first component discriminated between healthy controls and ME/CFS patients. Healthy individuals appeared as the more homogeneous group, with individuals being more closely related than the patients' cohort. The second component seemed to correlate well with entropy, identifying clusters in both cohorts where estimated entropy was lower. Interestingly, the cluster of patients sharing lower values of entropy was the farthest away from healthy controls and identified the ME/CFS group of individuals exhibiting elevated levels of severity on a majority of symptoms (i.e., those con-

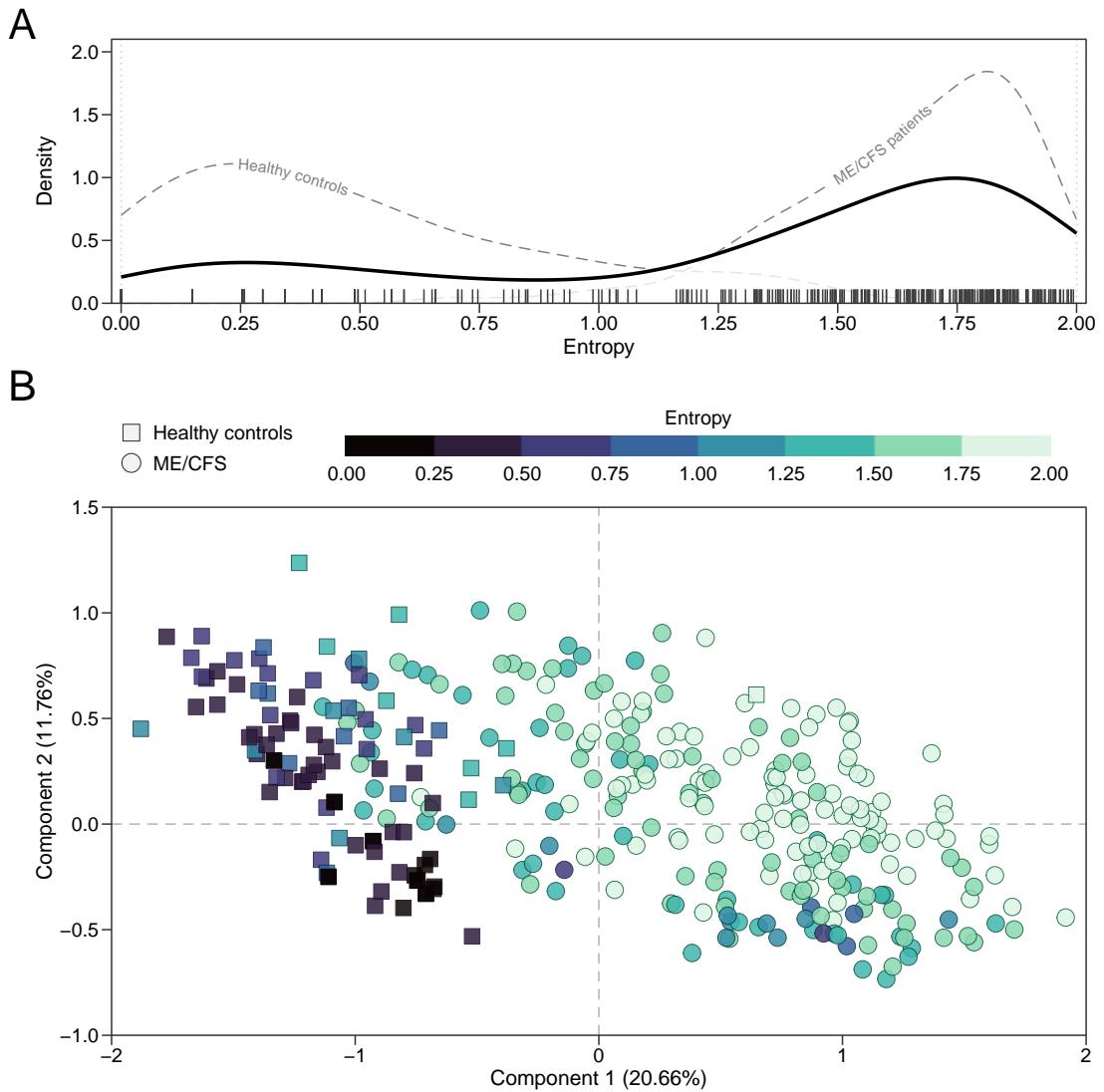


Figure 7.2: Estimation of intra- and inter-rater agreement to assess heterogeneity on the combination of severity levels across the study population. (A) Kernel density estimation of entropy values of the study population (dark line) and individualised cohorts (light-gray dashed lines), with entropy values of each participant identified (rug plot/vertical dashes displayed along the x -axis); (B) Classical multidimensional scaling (MDS) based on Cohen's κ coefficient, assessing the similarity of symptoms between participants.

sidered by clinicians to be severely affected). From the results it is evident that there is a high degree of heterogeneity among the severity profile of assessed symptoms from patients. Only patients at the two opposite extremes showed reduced estimates for entropy: those with reduced severity levels, closer to healthy controls (but still with higher entropy than healthy controls), and those with severe ME/CFS. Patients in between displayed high variability in entropy and symptom profile.

7.3.3 Latent class analysis across symptomatological domains

Latent class models were applied considering 47 symptoms combined into seven domains: immunological ($n = 7$), neuroendocrine ($n = 4$), PEM ($n = 5$), autonomic ($n = 8$), neurocognitive ($n = 16$), neurophysiological ($n = 7$), and pain ($n = 5$) (Supplementary Table D.2). According to AIC, the optimal number of subgroups (or latent classes) when studying the combination of all symptoms was five (domain Total in Figure 7.3A). When considering different domains at a time, the best LCA models identified 3, 4, 4, 5, 6, 6, and 6, respectively, for the immunological, neuroendocrine, PEM, autonomic, neurocognitive, neurophysiological, and pain domains. Conversely, BIC selected the minimum possible number of classes for the immunological and neuroendocrine domains, stratifying the rest into three classes, including the stratification done considering all symptoms (Figure 7.3B).

As detailed in the methodological section, the chosen number of subgroups in each LCA model depends on the applied information criteria. When comparing the selection outcomes, the information curves resulting from AIC exhibited an overall less pronounced variation, particularly around the optimal number of subgroups. In other words, the information value varied less while assessing the addition or removal of subgroups. This variability led to a more distinct number of classes, depending on the considered combination of symptoms, which can be indicative of the patients' large heterogeneity within each domain. On the other hand, given the more pronounced penalty imposed by BIC for the addition of parameters in the model, the value steadily increased as more than 3 subgroups were taken into consideration. Consequently, the number of parameters is more homogeneous for the majority of domains.

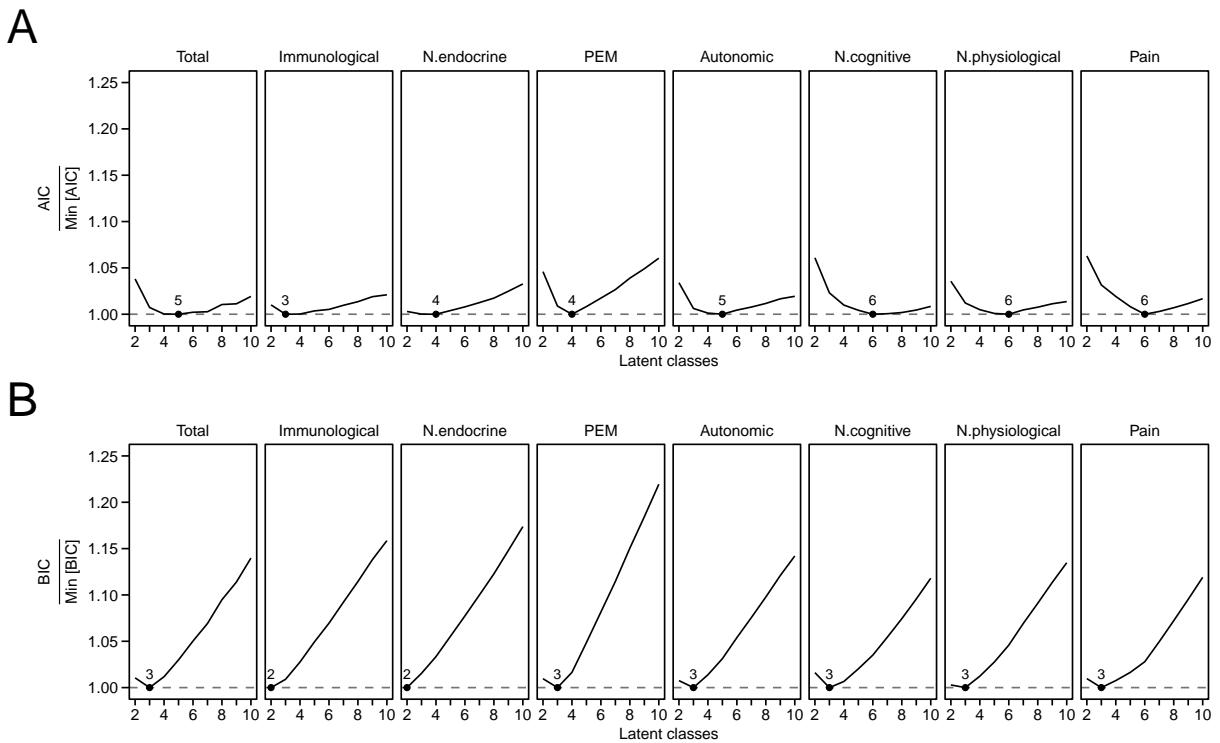


Figure 7.3: Selection of the optimal number of subgroups (latent classes) on each symptomatological domain based on the more parsimonious model, considering (A) Akaike information criterion (AIC) or (B) Bayesian information criterion (BIC). For each panel, latent class models were created with the number of subgroups varying between 2 and 10. The results show a standardisation of the estimated AIC and BIC values for each model as a ratio between each original estimated value and with the identified minimum information value as the denominator (black dot at the horizontal dashed line). As a ratio, results closer to 1.00 (the minimal ratio reference) indicate an overall improvement in the estimates for log-likelihood and penalty from the information criteria.

In the analysis encompassing all symptoms, the five AIC-identified subgroups were ordered from g1 to g5, based on the increasing severity of symptoms, according to the LCA class membership probabilities (Figure 7.4). These probability values indicate which combination of symptom severities is more likely to happen in individuals classified into each subgroup. As such, subgroup g1 classified individuals with higher combinations of absent and mild symptoms, while subgroup g5, at the other extreme, signifies those with class membership probabilities favouring more severe symptoms. The sample sizes in the subgroups for this domain varied (Table 7.1). This imbalance was particularly evident in the middle subgroup, g3, which included only 11.62% of the ME/CFS population (Pearson's χ^2 test, $p = 0.007$).

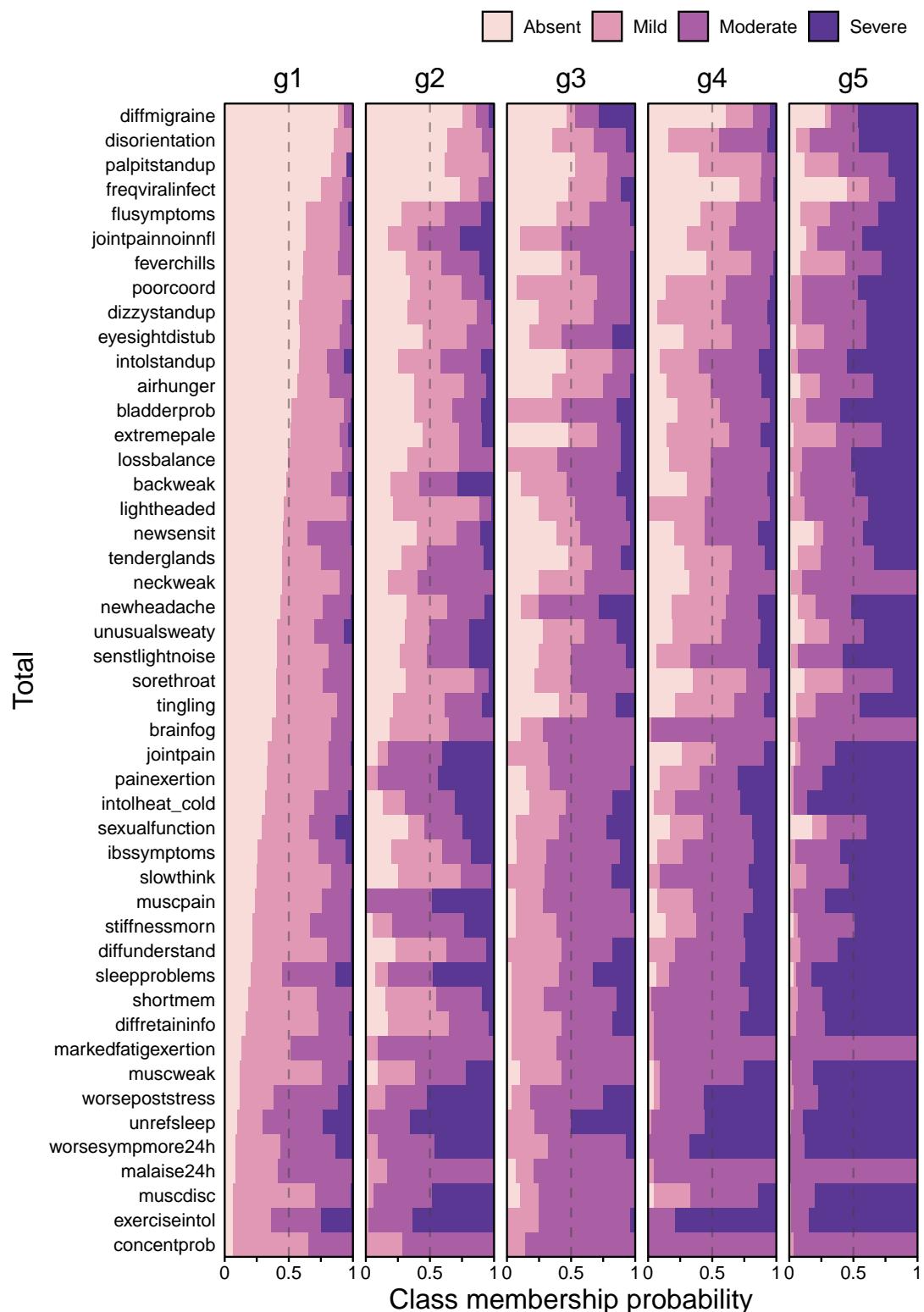


Figure 7.4: Latent class analysis estimated class membership probabilities of symptom severities on different AIC-based subgroups (latent classes), when profiling ME/CFS patients with the totality of available symptoms. Subgroups were ordered by increasing severity in the response probabilities of symptoms. A more detailed description of each symptom can be found elsewhere (Supplementary Table D.2).

Table 7.1: Samples sizes (and rounded percentages, row-wise) of the 241 ME/CFS patients in each AIC-based latent class estimated on each domain. For each domain, the latent classes are arranged by increasing level according to the severity profile from the response probabilities (Figure 7.5). Each domain is independent. The optimal number of classes on each domain was chosen based on the Akaike information criterion (AIC). P-values refer to Pearson's χ^2 test.

Domain	Latent classes, n (%)						P-value
	g1	g2	g3	g4	g5	g6	
Total	60 (24.90)	52 (21.58)	28 (11.62)	43 (17.84)	58 (24.07)	—	0.007
Immunological	124 (51.45)	88 (36.51)	29 (12.03)	—	—	—	<0.001
Neuroendocrine	70 (29.05)	29 (12.03)	62 (25.73)	80 (33.20)	—	—	<0.001
PEM	6 (2.49)	43 (17.84)	71 (29.46)	121 (50.21)	—	—	<0.001
Autonomic	68 (28.22)	34 (14.11)	69 (28.63)	39 (16.18)	31 (12.86)	—	<0.001
Neurocognitive	38 (15.77)	50 (20.75)	34 (14.11)	31 (12.86)	44 (18.26)	44 (18.26)	0.279
Neurophysiological	37 (15.35)	19 (7.88)	53 (21.99)	55 (22.82)	46 (19.09)	31 (12.86)	<0.001
Pain	38 (15.77)	69 (28.63)	44 (18.26)	42 (17.43)	27 (11.20)	21 (8.71)	<0.001

Similar imbalances were observed when performing LCA within specific symptom domains. The immunological ($p < 0.001$), neuroendocrine ($p < 0.001$), and autonomic ($p < 0.001$) domains had one subgroup close to 12% of the population, while the PEM ($p < 0.001$), neurophysiological ($p < 0.001$), and pain ($p < 0.001$) domains had one subgroup encompassing less than 10% of the ME/CFS population. Only the neurocognitive domain showed a balanced number of individuals across all subgroups ($p = 0.279$). In the immunological domain, the majority of patients belonged to the less severe subgroup, g1 ($n = 124$, 51.45%), whereas the more severe subgroup, g3, comprised the smaller proportion ($n = 29$, 12.03%). The neuroendocrine domain had all but subgroup g2 ($n = 29$, 12.03%) with a sample size above 25% of the population. The autonomic and pain domains showed similar results, classifying a small number of individuals into the more severe subgroup, g5 ($n = 31$, 12.86%), and g6 ($n = 21$, 8.71%), respectively. In contrast, the PEM domain displayed an inverse result, with only six individuals (2.49%) classified for the g1 subgroup and more than half the population attributed to the more severe cluster g4 ($n = 121$, 50.21%). This last outcome could be justified by the requirement of symptoms from this domain for the CCC-2003 clinical diagnosis to be considered valid, explaining the reduced number of individuals classified within this cluster, as the class membership probabilities were mostly for the absent severity (Figure 7.5).

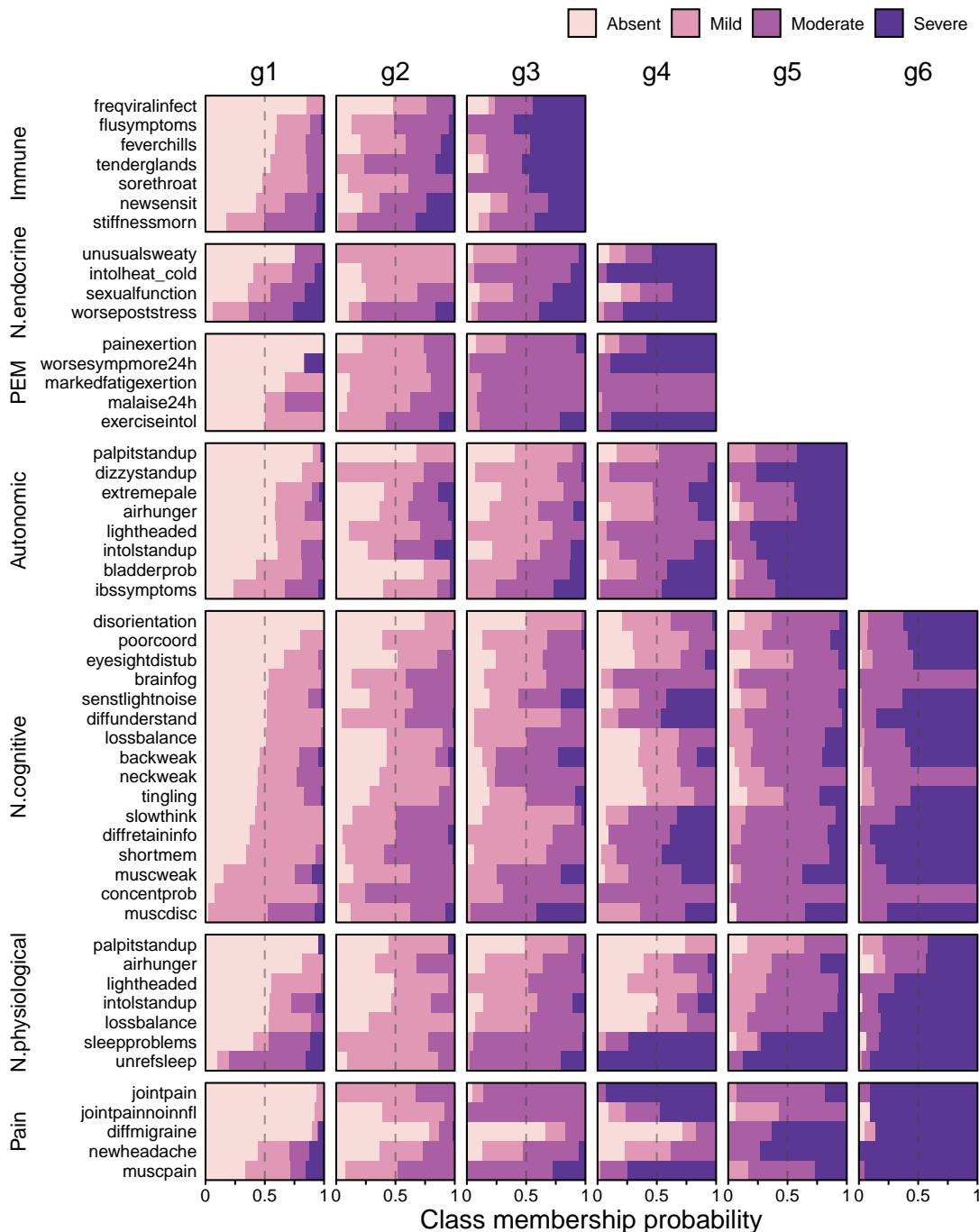


Figure 7.5: Latent class analysis estimated class membership probabilities of symptom severities on different AIC-based subgroups (latent classes), across the different immunological, neuroendocrine, post-exertional malaise (PEM), autonomic, neurocognitive, neurophysiological, and pain domains. Within each domain, subgroups were ordered by increasing severity in the response probabilities of symptoms that make up each domain. A more detailed description of each symptom can be found elsewhere (Supplementary Table D.2).

For the more stringent BIC, the optimal number of latent classes was reduced (Figure 7.3B). This clarified the association between class membership probabilities and the initially recorded severity level of each symptom (Supplementary Figure D.1 and Supplementary Figure D.2). In contrast to the LCA based on AIC, which identified clusters for a combination of symptoms that gradually increased the severity level over the course of three to six classes, BIC created subgroups based on the profile of absence and mild symptoms in a single group, and mild to severe symptoms in one or more subgroups. Neuroendocrine, PEM, neurocognitive, and pain domains are examples of this stratification. Alternatively, this approach allocated an initial subgroup g1, characterised by mostly absent symptoms, while assigning to other groups the presence of symptoms (e.g., g2 in the immunological domain). Additionally, the latter subgroups could be further stratified into an intermediate group with mild to moderate symptoms and a third one with mostly severe symptoms. The stratification with all symptoms (Supplementary Figure D.2), and the autonomic and neurophysiological domains are examples of this pattern of stratification. Regarding sample size distribution, there was a significant class imbalance across all domains except the pain domain (Pearson's χ^2 test, $p = 0.110$) (Supplementary Table D.3). The distributions were in accordance with the AIC-based stratification, most notably in the autonomic domain ($p < 0.001$), with a smaller number of individuals in the more severe subgroup, g3 ($n = 44$, 18.26%), and the PEM domain ($p < 0.001$), with fewer individuals in the initial subgroup, g1 ($n = 45$, 18.67%).

Further analyses were performed considering the AIC-based latent classes since BIC only identified two to three classes across all domains. This implies that BIC was essentially distinguishing between an almost binary stratification (absence of symptoms vs. presence of symptoms, regardless of severity, or absence and mild symptoms vs. moderate and severe symptoms), or a stratification which was close to the original one, creating categories for the more extreme situations, absence and more severe cases, with one other category in-between.

7.3.4 Herpesvirus IgG antibody data

The values of plasma IgG antibodies against the six herpesviruses varied amongst themselves (Figure 7.6A). There were no significant differences observed when comparing the median antibody values of healthy controls and ME/CFS patients. However, upon further stratification of patients into distinct domain subgroups, discernible differences emerged (Table 7.2). The (AIC) LCA subgroups identified significant differences in HSV-1 antibody levels within the immunological (Kruskal-Wallis test, $p = 0.041$, $p_{BH} = 0.285$), PEM ($p = 0.019$, $p_{BH} = 0.068$), autonomic ($p = 0.002$, $p_{BH} = 0.011$), and neurocognitive ($p = 0.022$, $p_{BH} = 0.157$) domains. Other differences were also found in antibody values of herpesvirus HSV-2 in the PEM domain ($p = 0.004$, $p_{BH} = 0.030$), and for herpesvirus EBV VCA within the neurophysiological group of symptoms ($p = 0.004$, $p_{BH} = 0.030$). Following correction for multiple testing, only the autonomic domain in data from HSV-1, the PEM domain at HSV-2, and neurophysiological in EBV VCA remained.

Table 7.2: Selection of significant p-values, both unadjusted and BH-adjusted, from Kruskal-Wallis sum rank test on antibody concentration values across each herpesvirus, comparing different healthy controls and ME/CFS patients under stratification based on symptomatological domains. p, p-values; p_{BH} , BH-adjusted p-values.

Information	Domain	Herpesvirus	Statistic	p	p_{BH}	p, new groups	$p_{BH},$ new groups
AIC	Immunological	HSV-1	8.27	0.041	0.285	0.041	0.285
	PEM	HSV-1	11.75	0.019	0.068	0.009	0.030
	PEM	HSV-2	15.18	0.004	0.030	0.003	0.019
	Autonomic	HSV-1	19.54	0.002	0.011	<0.001	0.002
	Neurocognitive	HSV-1	14.74	0.022	0.157	0.014	0.098
	Neurophysiological	EBV VCA	18.96	0.004	0.030	0.004	0.030
BIC	Immune	HSV-1	7.73	0.021	0.146	—	—

Pairwise tests on the identified herpesvirus-domain associations revealed differences not only between healthy controls and ME/CFS subgroups but also between the subgroups themselves (Figure 7.7A). Disconsidering subgroup g1 in the PEM domain due to its small sample size, there was a significant difference between groups g2 and g3 in the HSV-2 data (Mann-Whitney U test, $p = 0.004$, $p_{BH} = 0.021$). IgG concentrations for g2 were reduced and g3 (along

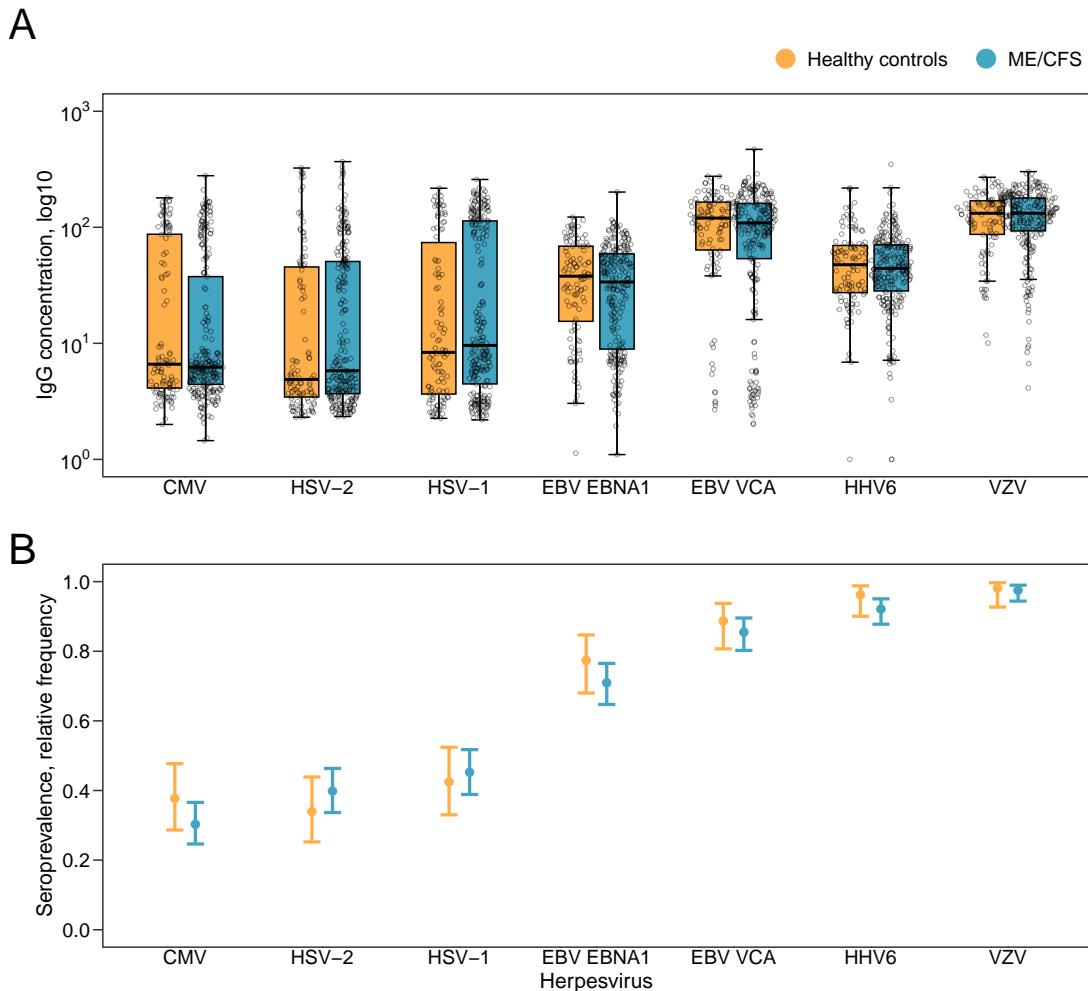


Figure 7.6: Plasma antibody concentrations against herpesviruses and relative seroprevalence in healthy controls and ME/CFS patients. (A) \log_{10} -transformed concentration of IgG immunoglobulins reactive against members of the herpesvirus family; (B) Proportion of seropositive individuals in each cohort, with the upper and lower bounds indicating the 95% CI. Herpesviruses were ordered by the increasing number of seropositive individuals in the data across both cohorts. From a total population of 347 individuals, the number of seropositives were 113 (32.56%), 132 (38.04%), 154 (44.38%), 253 (72.91%), 300 (86.46%), 324 (93.37%), and 339 (97.69%), respectively for CMV, HSV-2, HSV-1, EBV EBNA1, EBV VCA, HHV6, and VZV.

with g4, albeit not significant post-adjustment) were increased. The key difference between the two subgroups is that g2 classified individuals with mild PEM symptoms and, given the importance of this domain for the diagnosis of ME/CFS, could be considered as the lowest possible severity values in ME/CFS (with the small group g1 likely being outliers), and g3 mostly classified individuals with a moderate pattern of PEM symptoms (Figure 7.5). This pattern was also observed in data related to HSV-1, although it was not statistically significant after adjusting for multiple testing.

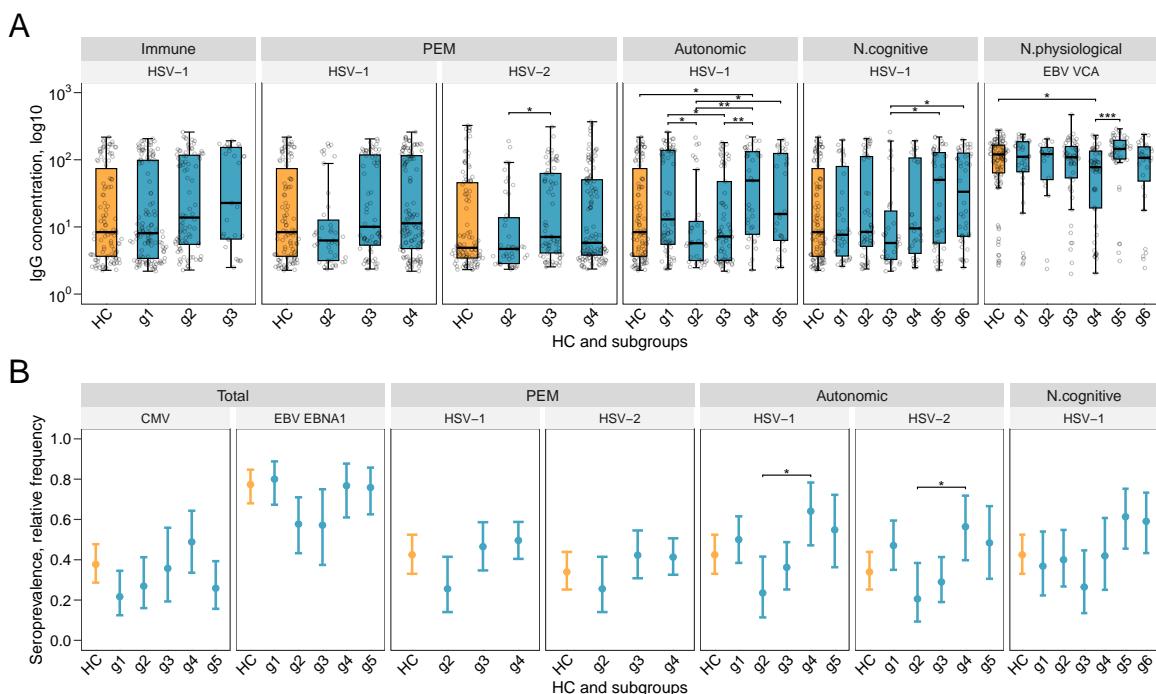


Figure 7.7: Plasma antibody concentrations against herpesviruses and relative seroprevalence in healthy controls and subgroups of ME/CFS patients based on distinct domains. (A) log₁₀-transformed concentration of IgG immunoglobulins reactive against members of the herpesvirus family; (B) Proportion of seropositive individuals in each cohort, with the upper and lower bounds indicating the 95% CI. Significant results from the BH-adjusted pairwise comparisons (Mann-Whitney U test in (A) and Fisher's exact test in (B)) were given an annotation, with *** for $p < 0.001$, ** when $p < 0.01$, and * when $p < 0.05$.

Various differences were observed in the subgroups of the autonomic domain on HSV-1 data. Most notably, reduced concentrations for subgroups g2 and g3, and increased values for subgroup g4. Subgroups g2 and g3 had similar median concentration levels and differed significantly when compared to g4 (g2, $p = 0.001$, $p_{BH} = 0.008$; g3, $p = 0.001$, $p_{BH} = 0.008$).

In the neurocognitive domain in HSV-1, the pairwise analyses identified a reduction in subgroup g3 when compared to the two more severe subgroups, g5 ($p = 0.003$, $p_{BH} = 0.036$) and g6 ($p < 0.002$, $p_{BH} = 0.036$).

Lastly, subgroup g4 from the neurophysiological domain showed reduced antibody concentrations when compared to both healthy controls ($p = 0.004$, $p_{BH} = 0.037$) and the group g5 ($p < 0.001$, $p_{BH} = 0.001$). This particular group classified few individuals with moderate symptoms, presenting a distinct profile characterised by almost exclusively severe levels in the two symptoms related to sleep dysfunction (sleepproblems and unrefsleep) and absent to mild severity levels for the remaining five symptoms (palpitstandup, airhunger, lightheaded, intelstandup, and lossbalance) (Figure 7.5). The observed differences towards healthy controls could be attributed to increased problems with sleep dysfunction, while the differences towards g5 could be linked to reduced severity in the other symptoms, as patients from this group also exhibit severe levels of sleep-related symptoms.

We then merged neighbouring pairs of domain subgroups that were closely related in terms of antibody concentration levels. With this, our intent was to reduce the uncertainty that could arise from working with groups of smaller sample sizes. The pairs formed were subgroups g3–g4 from the PEM domain, subgroups g2–g3 and g4–g5 from the autonomic domain, and subgroups g3–g4 and g4–g5 from the neurocognitive domain (Table 7.2, Figure 7.8A).

This grouping strategy kept relevant the strong differences initially noted in the HSV-1 antibody concentration levels for the autonomic domain (Kruskal-Wallis test, $p < 0.001$, $p_{BH} = 0.002$) and the PEM domain at the HSV-2 levels ($p = 0.003$, $p_{BH} = 0.019$) (Table 7.2). Interestingly this approach also evidenced differences in the PEM domain for the HSV-1 ($p = 0.009$, $p_{BH} = 0.019$). In this domain, the subgroup g3–g4 of moderate to severe ME/CFS patients consistently reported higher concentrations of antibodies, differing from subgroup g2, with noticeably lower concentrations (Mann-Whitney U test, HSV-1, $p = 0.010$, $p_{BH} = 0.046$; HSV-2, $p = 0.008$, $p_{BH} = 0.023$) (Figure 7.8A).

In the autonomic domain, HSV-1 concentrations exhibited similar results, more clearly relating the increase of symptom severity (subgroup g4–g5) with an increase in antibody

concentrations, comparatively to healthy individuals ($p = 0.005$, $p_{BH} = 0.011$) and to subgroup g2–g3 ($p < 0.001$, $p_{BH} < 0.001$).

In the neurocognitive domain, the more severe subgroup g5–g6 also showed a significant increase in antibody concentrations for the same herpesvirus when compared to healthy controls ($p = 0.005$, $p_{BH} = 0.021$).

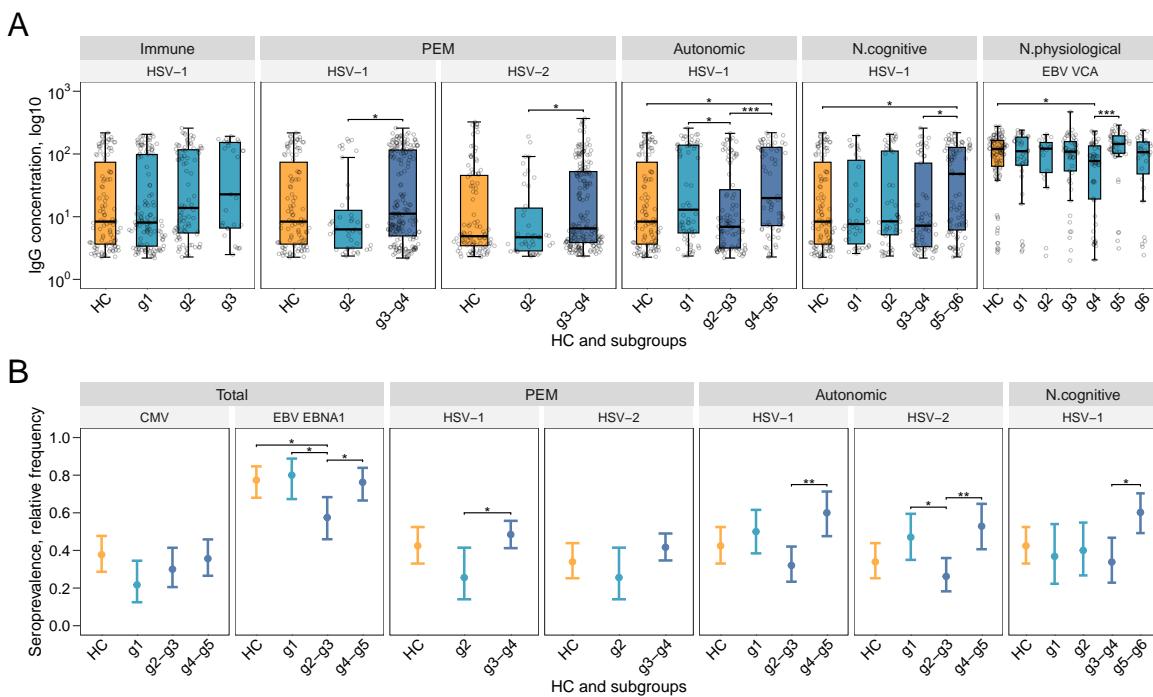


Figure 7.8: Plasma antibody concentrations against herpesviruses and relative seroprevalence in healthy controls and homogenised ME/CFS subgroups based on distinct domains. (A) log₁₀-transformed concentration of IgG immunoglobulins reactive against members of the herpesvirus family; (B) Proportion of seropositive individuals in each cohort, with the upper and lower bounds indicating the 95% CI. Significant results from the BH-adjusted pairwise comparisons (Mann-Whitney U test in (A) and Fisher's exact test in (B)) were given an annotation, with *** for $p < 0.001$, ** when $p < 0.01$, and * when $p < 0.05$.

7.3.5 Herpesvirus seropositivity

The differences observed in the herpesviruses' serology influenced the seroprevalence in the study populations (Figure 7.6B, HC vs. ME/CFS across herpesviruses). Similarly to the previous section, the comparison between healthy controls and ME/CFS patients showed consistent

results across all herpesviruses. However, following the stratification of ME/CFS into the seven specific domains, differences became more apparent (Table 7.3). In line with the results obtained when studying antibody titers, the AIC-based subgroups showed significant differences in HSV-1 seroprevalence in the PEM (Pearson's χ^2 test, $p = 0.022$, $p_{BH} = 0.148$), autonomic ($p = 0.006$, $p_{BH} = 0.020$), and neurocognitive in HSV-1 ($p = 0.021$, $p_{BH} = 0.146$) domains. Differences in HSV-2 seroprevalence in the PEM domain ($p = 0.042$, $p_{BH} = 0.148$) were also observed. Other significant differences appeared, namely in the stratification considering all symptoms across CMV ($p = 0.038$, $p_{BH} = 0.132$) and EBV EBNA1 ($p = 0.025$, $p_{BH} = 0.132$), and autonomic domain at the HSV-2 herpesvirus ($p = 0.005$, $p_{BH} = 0.020$). After correction for multiple testing, all previously significant results were disregarded with the exception of the autonomic relations within prevalence values for the two herpes simplex viruses.

Table 7.3: Selection of significant p-values, both unadjusted and BH-adjusted, from Pearson's χ^2 test on seropositivity values across each herpesvirus, comparing different healthy controls and ME/CFS patients under stratification based on symptomatological domains. p, p-values; p_{BH} , BH-adjusted p-values.

Information	Domain	Herpesvirus	Statistic	p	p_{BH}	p, new groups	$p_{BH},$ new groups
AIC	Total	CMV	11.79	0.038	0.132	—	—
	Total	EBV EBNA1	12.79	0.025	0.132	0.005	0.036
	PEM	HSV-1	11.46	0.022	0.148	0.010	0.069
	PEM	HSV-2	9.89	0.042	0.148	0.020	0.069
	Autonomic	HSV-1	16.39	0.006	0.020	0.003	0.009
	Autonomic	HSV-2	16.88	0.005	0.020	0.001	0.009
	Neurocognitive	HSV-1	14.91	0.021	0.146	0.010	0.069
BIC	Neuroendocrine	HHV6	8.22	0.016	0.115	—	—

Pairwise comparisons on the autonomic results were in line with the ones seen in the previous section (see Figure 7.7A), showing reduced seroprevalence values on subgroups g2 and g3 overall (Figure 7.7B). Other results on both HSV-1 and HSV-2 were the noticeable seroprevalence similarities between the two more severe groups, g4 and g5. This could explain a relation between past exposure to HSV-1 (and/or HSV-2) and the development of more pronounced autonomic symptoms. In the neurocognitive domain, the more severe subgroups, g5 and g6, also showed higher seroprevalence, albeit not significant.

As previously, we merged groups with similar seroprevalence estimates. In line with the serological section, the subgroup pairs were g3–g4 from the PEM domain, pairs g2–g3 and g4–g5 from the autonomic domain, and pairs g3–g4 and g5–g6 from the neurocognitive domain (Figure 7.8B). Two additional pairs were formed for the stratification with all symptoms, g2–g3, and g4–g5. From this, only the seroprevalence values for EBV EBNA1 when clustering for all symptoms yielded a new relation ($p = 0.005$, $p_{BH} = 0.036$) (Table 7.3), which showed a seroprevalence reduction in the intermediate group g2–g3 when compared to healthy controls and remaining ME/CFS subgroups. On par with the HSV-1 IgG concentration values, seroprevalence was also higher in the more severe subgroup g3–g4 from the PEM domain, subgroup g4–g5 from the autonomic domain, and subgroup g5–g6 from the neurocognitive domain.

7.4 Discussion

This preliminary study demonstrates how the wide range of symptoms assessed during ME/CFS diagnosis could be used to stratify patients by mechanistic domains.

The findings from the intra- and inter-rater agreement analyses indicated that patients diagnosed with ME/CFS are quite diverse in terms of symptom severity, both at the individual and population levels (Figure 7.2). Similar to the symptom-based case definitions, the severity profile from all individuals distinguished healthy controls from patients (absence vs. presence of disease) and identified clusters of patients with consistently increased levels of symptom severity (mild to moderate vs. severely affected). Stratification based on this binary phenotype for mild vs. severe ME/CFS has been previously studied and even used to research the link between the disease and herpesviruses, using data from the UKMEB (Cliff et al. 2019; Lee et al. 2021). For instance, Lee et al. (2021) found a positive correlation between the HHV6 viral load and higher severity scores from symptoms in the autonomic, neurocognitive, and pain domains.

The LCA method allowed us to test an unbiased approach to symptom-based clustering based on pre-specified domains. When dividing all symptoms into subgroups, the two more

extreme groups, g1 and g5, showed clearer profiles of more absent and mild, and more severe symptoms, respectively (Figure 7.4). The remaining three subgroups showed a combination of gradually increasing severity in most symptoms. Interestingly, g1 and g5 included a similar number of patients and together accounted for almost half the ME/CFS population.

It is worth noting that PEM and sleep-related symptoms (exerciseintol, malaise24h, marked-fatigexertion, painexertion, and worsesympmore24h, and sleepproblems and unrefsleep, respectively) were the only ones where the sum of class membership probability estimates from mild and severe surpassed the estimated class membership probabilities midpoint value of 0.5 on all five subgroups, being particularly evident in subgroup g1. This highlights the importance of both PEM and sleep disturbances in order to comply with the case diagnosis—the presence of symptoms from the two domains are requirements for the CCC-2003 case definition and are listed as necessary in the CDC-1994 (Fukuda et al. 1994; Carruthers et al. 2003).

The analysis of serological data using patients split by these five subgroups did not yield conclusive results. Similar to the comparisons using the entire ME/CFS population, we did not find any significant differences in either antibody concentrations or seroprevalence comparisons. After combining two subgroups (moderate subgroup g2–g3 and severe subgroup g4–g5), we found a significant decrease in seroprevalence in the moderate group for EBV EBNA1 (Figure 7.8B). However, this decrease in the intermediate subgroup g2–g3 does not provide much information. Nonetheless, it is still noteworthy that g2–g3 consistently showed a far lower percentage of seropositive individuals (57.50%, 95% CI = [45.95%, 68.32%]) when the overall 70.95% (95% CI = [64.71%, 76.51%]) of the ME/CFS population is seropositive towards EBV EBNA1 (Figure 7.6B).

When implementing the LCA method over specific symptom domains, the optimal number of subgroups varied (Figure 7.3A). This suggests that there are different degrees of symptom heterogeneity within each domain. In some, patients are more similar and show fewer subgroups, such as the immunological, neuroendocrine, and PEM domains. Conversely, in the pain, neurophysiological, and neurocognitive domains, ME/CFS patients exhibit a more heterogeneous profile of symptoms, resulting in a larger number of optimal latent classes (Figure 7.5). One example of the latter is the pain domain, where distinct subgroups conveyed clear profiles

of symptom severity. Subgroup g4 classified individuals with severe muscle and multi-joint pain, and subgroup g5 identified individuals reporting severe headaches and migraines together with a moderate level of physical joint and muscle pain. Lastly, subgroup g6 almost exclusively included patients experiencing severe symptoms. Since severe chronic pain greatly reduces the overall well-being and quality of life, there are pharmacological therapies listed as recommendations to relieve pain in ME/CFS (NICE Guideline [NG206] 2021). As such, it makes sense that this was the smaller subgroup from this domain (Table 7.1).

Similar to the previous analysis, IgG concentrations in conjunction with domain subgroups did not reveal a clear relationship between severity and antibody titers. Overall, the most significant differences were observed for HSV-1, particularly in the autonomic and neurocognitive domains (Figure 7.7A). In the autonomic domain, individuals in subgroup g4 had higher antibody levels compared to those in intermediate subgroups, g2 and g3, and healthy controls. Similarly, within the neurocognitive domain, high antibody levels against HSV-1 were observed in the severe groups, g5 and g6, in comparison to the intermediate g3. When combining the subgroups, the differences between the more severe (g4–g5 in the autonomic and g5–g6 in the neurocognitive domain) and healthy controls became more pronounced (Figure 7.8A). Interestingly, the seroprevalence results also showed an increased number of HSV-1 seropositive individuals in the more severe groups of autonomic and neurocognitive domains (Figure 7.7B and Figure 7.8B). This observation could be attributed to the neurotropic nature of the virus (Olsson et al. 2016; Marcocci et al. 2020), as there is evidence that repeated HSV-1 brain infections can lead to neuronal damage (Marcocci et al. 2020; Murphy et al. 2021). However, the inferences regarding positive associations between autonomic and neurocognitive subgroup severity and serological measures for HSV-1 are not so linear. In both domains, subgroups classified as mostly absent to mild (g1 in autonomic, and g1 and g2 in neurocognitive) did not show minimal levels of antibody concentrations or seroprevalence. One speculative explanation could be that these subgroups are closer to healthy controls in terms of symptoms and HSV-1 serological values. However, to confirm this hypothesis would require the individuals from both domains to be the same, and looking at the profile of each individual across all domains was not in the scope of this preliminary study. Nonetheless, despite the lack of concrete

interpretations, this intriguing result should be further explored in subsequent works.

The seroprevalence results for the autonomic domains in HSV-1 were also observed in HSV-2. This correlated result could be influenced by the close relation between the two herpes simplex viruses, as they share similarities in terms of their structure, transmission methods, and symptoms caused (Bradley et al. 2014; Kawaguchi et al. 2018).

The motivation behind this approach stems from the need to simplify the myriad of potential symptom combinations across the ME/CFS population. Conceptually, this task could be viewed as a combinatorial problem. To begin, we consider an extreme scenario under two assumptions: (1) there are no constraints from the usual case definitions used to diagnose ME/CFS, implying that patients can experience any symptom at any range (absent, mild, moderate, or severe); and (2) the 47 assessed symptoms are independent of each other, meaning that the manifestation of one symptom does not influence the occurrence of another potentially related one.

Under these assumptions, the number of distinct and possible symptom profiles would be $4^{47} = 1.98 \times 10^{28}$. Certainly, many of these profiles consist of combinations predominantly featuring absence to mild symptoms, or only a small number of exacerbated and isolated symptoms, insufficient to comply with a comprehensive ME/CFS diagnosis. However, even after removing those less relevant combinations for absent and sporadic symptoms, and reconsidering the previous assumptions of symptom independence and case criteria restrictions, the number of possible combinations would still be immeasurable.

The act of grouping symptoms into specific domains offers a means to mitigate this complexity, simplifying the vast number of potential combinations into a more perceivable number. Under the more extreme scenario, the AIC-based LCA stratification has $3 \times 4^2 \times 5 \times 6^3 = 51,840$ possible combinations. While still a substantial number of possible profiles for ME/CFS patients, it is already a tremendous reduction from what the original consideration was. By removing combinations incompatible with the diagnosis and identifying profiles associated with the domain-specific responses or dysregulation, we could gradually reduce this number into a more manageable subset.

This study presents some limitations that can influence the interpretation of the results. Firstly,

the study is based on a single symptomatic timepoint, which may not fully capture the fluctuating nature of some symptoms experienced by ME/CFS patients. Symptoms can vary over time and depend on the individual's degree of activity or rest. Additionally, personalised treatment can be done to target and reduce the exacerbation of specific symptoms—this could be the cause of the six individuals with a profile of mostly absent symptoms in the PEM domain. Thus, it is not possible to interpret the severity of the reported symptoms in terms of how they are progressing at the time of diagnosis.

Secondly, accurately reporting the severity of a symptom is more challenging than simply indicating its presence or absence. Severity depends on the aforementioned progression at the time of diagnosis, but it can also vary depending on the tolerance of each individual. In this sense, factors such as age, sex, or disease duration could influence the severity profiles and reported in certain domains and overall prognosis (Ghali et al. 2022). While our study did not find differences in overall symptom severity based on age or sex (Supplementary Table D.1), previous research suggests an association between age and increased autonomic dysfunction and more severe symptoms in ME/CFS (Lewis et al. 2013). Additionally, studies in rheumatoid arthritis and fibromyalgia indicate that women can have slightly more pain-related symptoms than men, reporting higher sensitivity (Wolfe et al. 2018; Vogel et al. 2023). Regarding disease duration, individuals can learn to cope with the disease and mitigate the exacerbation of certain symptoms over time. Contrarily to this idea, the more severe group of patients reported experiencing the disease for an average of 4.4 years longer than those with mild to moderate severity (Supplementary Table D.1). Future work may need to consider these possible differences when performing the analysis.

Lastly, the LCA models assume independence of the symptoms within each domain. The assessed symptoms are likely related and some may not even be strictly confined to a single domain (as seen by the inclusion of repeated symptoms into the neurophysiological domain). In this sense, linking each domain to its most common symptoms could be an oversimplification and other less stringent domains could be looked at. Alternatively, other unbiased approaches for stratification have been proposed to capture the diverse profiles observed in ME/CFS patients. For example, a study by Asprusten et al. (2021) used hierarchical clustering methods to

split adolescents diagnosed with ME/CFS into six clusters related to distinct pathophysiological profiles that had not been defined before the analysis.

7.5 Conclusions

The identification of more homogeneous subgroups of patients would certainly improve the consistency and consequent reproducibility of the research done in ME/CFS. The stratification of symptoms into specific domains could be one consideration, allowing for interpreting particular triggers and mechanisms related to the viral onset and expression of the disease.

Our study shows that it is possible to stratify ME/CFS patients by focusing on the profile of particular domains of symptoms related to mechanistic activation. Although the results were limited, there was some evidence for associations between increased severity in autonomic and neurocognitive subgroups and higher antibody levels against HSV-1. To confirm these findings, a subsequent work must involve (1) the simplification or grouping of some of the proposed domains in order to improve the relation between symptoms; (2) the usage of the already identified domain subgroups in other data sets (e.g. profiles based on immune cellular compartments or functional assays) to further corroborate the relation between herpesviruses (namely HSV-1) and autonomic or neurocognitive exacerbation, or to assess if it is possible to identify other distinguishable patterns under other mechanistic assumptions; and (3) look more thoroughly at the combination of subgroups from each individual across all seven domains. Ultimately, advances in this particular topic would grant greater control over what is causing the different symptoms, allowing for more personalised care and treatment to be implemented.

Bibliography

T. T. Asprusten, L. Sletner, and V. B. B. Wyller. Are there subgroups of chronic fatigue syndrome? An exploratory cluster analysis of biological markers. *Journal of Translational Medicine*, 19(1), Jan. 2021. doi: 10.1186/s12967-021-02713-9.

Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to

- Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, Jan. 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.
- H. Bradley, L. E. Markowitz, T. Gibson, and G. M. McQuillan. Seroprevalence of Herpes Simplex Virus Types 1 and 2—United States, 1999–2010. *The Journal of Infectious Diseases*, 209(3):325–333, Feb. 2014. ISSN 0022-1899. doi: 10.1093/infdis/jit458. URL <https://doi.org/10.1093/infdis/jit458>.
- K. G. Brurberg, M. S. Fønhus, L. Larun, S. Flottorp, and K. Malterud. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *BMJ Open*, 4(2):e003973, Feb. 2014. doi: 10.1136/bmjopen-2013-003973.
- B. M. Carruthers, A. K. Jain, K. L. D. Meirleir, D. L. Peterson, N. G. Klimas, A. M. Lerner, A. C. Bested, P. Flor-Henry, P. Joshi, A. C. P. Powles, J. A. Sherkey, and M. I. v. d. Sande. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Chronic Fatigue Syndrome*, 11(1):7–115, Jan. 2003. doi: 10.1300/j092v11n01_02.
- J. M. Cliff, E. C. King, J.-S. Lee, N. Sepúlveda, A.-S. Wolf, C. Kingdon, E. Bowman, H. M. Dockrell, L. Nacul, E. Lacerda, and E. M. Riley. Cellular Immune Function in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Frontiers in Immunology*, 10, Apr. 2019. doi: 10.3389/fimmu.2019.00796. URL <https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2019.00796/full>.
- J. Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1): 37–46, Apr. 1960. ISSN 0013-1644. doi: 10.1177/001316446002000104. URL <https://doi.org/10.1177/001316446002000104>. Publisher: SAGE Publications Inc.
- T. D. Domingues, A. D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, F. Westermeier, C. Scheibenbogen, J. M. Cliff, L. Nacul, E. M. Lacerda, H. Mourão, and N. Sepúlveda. Herpesviruses Serology Distinguishes Different Subgroups of Patients From the United Kingdom Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Biobank. *Frontiers in Medicine*, 8, July 2021. doi: 10.3389/fmed.2021.686736.
- T. D. Domingues, J. a. Malato, A. D. Grabowska, J.-S. Lee, J. Ameijeiras-Alonso, P. Biecek, L. Graça, H. Mourão, C. Scheibenbogen, F. Westermeier, L. Nacul, J. M. Cliff, E. Lacerda, and N. Sepúlveda. Association analysis between symptomatology and herpesvirus IgG antibody concentrations in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and multiple sclerosis. *Heliyon*, 9(7):e18250, July 2023. ISSN 2405-8440. doi: 10.1016/j.heliyon.2023.e18250. URL <https://www.sciencedirect.com/science/article/pii/S2405844023054580>.
- K. Fukuda, S. E. Straus, I. Hickie, M. C. Sharpe, J. G. Dobbins, and A. Komaroff. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Annals of Internal Medicine*, 121(12):953, Dec. 1994. doi: 10.7326/0003-4819-121-12-199412150-00009.
- M. Gamer, J. Lemon, I. Fellows, and P. Singh. irr: Various coefficients of interrater reliability and agreement. R package version 0.84, 2012. URL <https://cran.r-project.org/web/packages/irr/irr.pdf>.
- A. Ghali, C. Lacout, J.-O. Fortrat, K. Depres, M. Ghali, and C. Lavigne. Factors Influencing the Prognosis of Patients

- with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Diagnostics*, 12(10):2540, Oct. 2022. ISSN 2075-4418. doi: 10.3390/diagnostics12102540. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9600584/>.
- L. A. Jason, K. Corradi, S. Torres-Harding, R. R. Taylor, and C. King. Chronic Fatigue Syndrome: The Need for Subtypes. *Neuropsychology Review*, 15(1):29–58, Mar. 2005. doi: 10.1007/s11065-005-3588-2.
- Y. Kawaguchi, Y. Mori, and H. Kimura, editors. *Human Herpesviruses*, volume 1045 of *Advances in Experimental Medicine and Biology*. Springer, Singapore, 2018. ISBN 978-981-10-7229-1 978-981-10-7230-7. doi: 10.1007/978-981-10-7230-7. URL <http://link.springer.com/10.1007/978-981-10-7230-7>.
- E. M. Lacerda, E. W. Bowman, J. M. Cliff, C. C. Kingdon, E. C. King, J.-S. Lee, T. G. Clark, H. M. Dockrell, E. M. Riley, H. Curran, and L. Nacul. The UK ME/CFS Biobank for biomedical research on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) and Multiple Sclerosis. *Open Journal of Bioresources*, 4(1):4, Feb. 2017. ISSN 2056-5542. doi: 10.5334/ojb.28. URL <https://openbioresources.metajnl.com/articles/10.5334/ojb.28>. Number: 1 Publisher: Ubiquity Press.
- E. M. Lacerda, K. Mudie, C. C. Kingdon, J. D. Butterworth, S. O’Boyle, and L. Nacul. The UK ME/CFS Biobank: A Disease-Specific Biobank for Advancing Clinical Research Into Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Neurology*, 9:1026, 2018. ISSN 1664-2295. doi: 10.3389/fneur.2018.01026.
- J.-S. Lee, E. M. Lacerda, L. Nacul, C. C. Kingdon, J. Norris, S. O’Boyle, C. h. Roberts, L. Palla, E. M. Riley, and J. M. Cliff. Salivary DNA Loads for Human Herpesviruses 6 and 7 Are Correlated With Disease Phenotype in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Medicine*, 8:656692, Aug. 2021. ISSN 2296-858X. doi: 10.3389/fmed.2021.656692. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8378328/>.
- I. Lewis, J. Pairman, G. Spickett, and J. L. Newton. Is Chronic Fatigue Syndrome in Older Patients a different disease? – a clinical cohort study. *European Journal of Clinical Investigation*, 43(3):302–308, 2013. ISSN 1365-2362. doi: 10.1111/eci.12046. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/eci.12046>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/eci.12046>.
- E.-J. Lim and C.-G. Son. Review of case definitions for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 18(1):289, Dec. 2020. ISSN 1479-5876. doi: 10.1186/s12967-020-02455-0. URL <https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-020-02455-0>.
- D. A. Linzer and J. B. Lewis. poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, 42:1–29, June 2011. ISSN 1548-7660. doi: 10.18637/jss.v042.i10. URL <https://doi.org/10.18637/jss.v042.i10>.
- J. a. Malato, L. Graça, L. Nacul, E. M. Lacerda, , and N. Sepúlveda. Statistical challenges of investigating a disease with a complex diagnosis. In P. Milheiro, A. Pacheco, B. de Sousa, I. F. Alves, I. Pereira,

- M. J. a. Polidoro, and S. Ramos, editors, *Estatística: Desafios transversais às ciências com dados*, Eds. Sociedade portuguesa de estatística, pages 153–167. Sociedade Portuguesa de Estatística, 2021. ISBN 978-972-8890-47-6. URL <https://www.spestatistica.pt/storage/app/uploads/public/609/28f/6d0/60928f6d08a0c016386627.pdf>.
- J. a. Malato, L. Graça, and N. Sepúlveda. Impact of Misdiagnosis in Case-Control Studies of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Diagnostics*, 13(3):531, Jan. 2023. ISSN 2075-4418. doi: 10.3390/diagnostics13030531. URL <https://www.mdpi.com/2075-4418/13/3/531>. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
- M. E. Marcocci, G. Napoletani, V. Protto, O. Kolesova, R. Piacentini, D. D. L. Puma, P. Lomonte, C. Grassi, A. T. Palamara, and G. D. Chiara. Herpes Simplex Virus-1 in the Brain: The Dark Side of a Sneaky Infection. *Trends in Microbiology*, 28(10):808–820, Oct. 2020. ISSN 0966-842X, 1878-4380. doi: 10.1016/j.tim.2020.03.003. URL [https://www.cell.com/trends/microbiology/abstract/S0966-842X\(20\)30074-3](https://www.cell.com/trends/microbiology/abstract/S0966-842X(20)30074-3). Publisher: Elsevier.
- J. G. Montoya, T. H. Holmes, J. N. Anderson, H. T. Maecker, Y. Rosenberg-Hasson, I. J. Valencia, L. Chu, J. W. Younger, C. M. Tato, and M. M. Davis. Cytokine signature associated with disease severity in chronic fatigue syndrome patients. *Proceedings of the National Academy of Sciences*, 114(34):E7150–E7158, July 2017. doi: 10.1073/pnas.1710519114.
- M. J. Murphy, L. Fani, M. K. Ikram, M. Ghanbari, and M. A. Ikram. Herpes simplex virus 1 and the risk of dementia: a population-based study. *Scientific Reports*, 11(1):8691, Apr. 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-87963-9. URL <https://www.nature.com/articles/s41598-021-87963-9>. Number: 1 Publisher: Nature Publishing Group.
- L. Nacul, E. M. Lacerda, C. C. Kingdon, H. Curran, and E. W. Bowman. How have selection bias and disease misclassification undermined the validity of myalgic encephalomyelitis/chronic fatigue syndrome studies? *Journal of Health Psychology*, 24(12):1765–1769, Oct. 2019. ISSN 1461-7277. doi: 10.1177/1359105317695803.
- NICE Guideline [NG206]. Myalgic encephalomyelitis (or encephalopathy)/chronic fatigue syndrome: diagnosis and management | Guidance, Oct. 2021. URL <https://www.nice.org.uk/guidance/ng206>. Publisher: NICE.
- J. Olsson, H. Lövheim, E. Honkala, P. J. Karhunen, F. Elgh, and E. H. Kok. HSV presence in brains of individuals without dementia: the TASTY brain series. *Disease Models & Mechanisms*, 9(11):1349–1355, Nov. 2016. ISSN 1754-8411. doi: 10.1242/dmm.026674.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL <https://www.R-project.org/>.
- C. Scheibenberg, H. Freitag, J. Blanco, E. Capelli, E. Lacerda, J. Authier, M. Meeus, J. C. Marrero, Z. Nora-Kruk, E. Oltra, and others. The European ME/CFS Biomarker Landscape project: an initiative of the European

- network EUROMENE. *Journal of translational medicine*, 15(1):162, 2017. doi: <https://doi.org/10.1186/s12967-017-1263-z>.
- N. Sepúlveda, J. a. Malato, F. Sotzny, A. D. Grabowska, A. Fonseca, C. Cordeiro, L. Graça, P. Biecek, U. Behrends, J. Mautner, F. Westermeier, E. M. Lacerda, and C. Scheibenbogen. Revisiting IgG Antibody Reactivity to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Its Potential Application to Disease Diagnosis. *Frontiers in Medicine*, 9, 2022. ISSN 2296-858X. doi: 10.3389/fmed.2022.921101. URL <https://www.frontiersin.org/articles/10.3389/fmed.2022.921101>.
- C. E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.
- F. Sotzny, J. Blanco, E. Capelli, J. Castro-Marrero, S. Steiner, M. Murovska, and C. Scheibenbogen. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome – Evidence for an autoimmune disease. *Autoimmunity Reviews*, 17(6):601–609, June 2018. doi: 10.1016/j.autrev.2018.01.009.
- K. Vogel, L. N. Muhammad, J. Song, T. Neogi, C. O. Bingham, M. B. Bolster, W. Marder, A. Wohlfahrt, D. J. Clauw, D. Dunlop, and Y. C. Lee. Sex Differences in Pain and Quantitative Sensory Testing in Patients With Rheumatoid Arthritis. *Arthritis Care & Research*, 75(12):2472–2480, Dec. 2023. ISSN 2151-4658. doi: 10.1002/acr.25178.
- F. Wolfe, B. Walitt, S. Perrot, J. J. Rasker, and W. Häuser. Fibromyalgia diagnosis and biased assessment: Sex, prevalence and bias. *PLOS ONE*, 13(9):e0203755, Sept. 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0203755. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0203755>. Publisher: Public Library of Science.

Chapter 8

The SARS-CoV-2 receptor ACE2 in ME/CFS: A meta-analysis of public DNA methylation and gene expression data

J. Malato, F. Sotzny, S. Bauer, H. Freitag, A. Fonseca, A.D. Grabowska, L. Graça, C. Cordeiro, L. Nacul, E.M. Lacerda, J. Castro-Marrero, C. Scheibenbogen, F. Westermeier, and N. Sepúlveda. The SARS-CoV-2 receptor angiotensin-converting enzyme 2 (ACE2) in myalgic encephalomyelitis/chronic fatigue syndrome: A meta-analysis of public DNA methylation and gene expression data. *Heliyon*. 2021; 7(8). doi: <https://doi.org/10.1016/j.heliyon.2021.e07665>.

Abstract

People with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) often report a high frequency of viral infections and flu-like symptoms during their disease course. Given that this reporting agrees with different immunological abnormalities and altered gene expression profiles observed in the disease, we aimed at answering whether the expression of the human angiotensin-converting enzyme 2 (ACE2), the major cell entry receptor for SARS-CoV-2, is also altered in these patients. In particular, a low expression of ACE2 could be indicative of a high risk of developing Covid-19. We then performed a meta-analysis of public data on CpG DNA methylation and gene expression of this enzyme and its homologous ACE protein in peripheral blood mononuclear cells and related subsets. We found that patients with ME/CFS have decreased methylation levels of four CpG probes in the ACE locus (cg09920557, cg19802564, cg21094739, and cg10468385) and of another probe in the promoter region of the ACE2 gene (cg08559914). We also found a decreased expression of ACE2 but not of ACE in patients when compared to healthy controls. Accordingly, in newly collected data, there was evidence for a significant higher proportion of samples with an ACE2 expression below the limit of detection in patients than healthy controls. Altogether, patients with ME/CFS can be at a higher Covid-19 risk and, if so, they should be considered a priority group for vaccination by public health authorities. To further support this conclusion, similar research is recommended for other human cell entry receptors and cell types, namely, those cells targeted by the virus.

Keywords: Myalgic encephalomyelitis/Chronic fatigue syndrome; SARS-CoV-2; ACE2; Gene expression; DNA methylation

8.1 Introduction

Myalgic encephalomyelitis/Chronic fatigue syndrome (ME/CFS) is a multifactorial and complex disease characterised by two key symptoms: (1) persistent but unexplained fatigue that is not alleviated by rest; and (2) post-exertional malaise upon minimal physical or even mental effort (Fukuda et al. 1994; Carruthers et al. 2003). Although its cause remains unknown, a growing body of evidence strongly associates ME/CFS with several microbial and viral infections, as potential triggering factors (Chu et al. 2019; Johnston et al. 2016). In addition, it is currently hypothesised that reactivations of dormant viral infections also play a role (Rasa et al. 2018; Ariza 2021) due to several immunological abnormalities (Klimas et al. 1990; Lorusso et al. 2009; Brenu et al. 2011). On the molecular basis of the disease, peripheral blood mononuclear cells (PBMCs) have altered gene expression profiles (Kerr et al. 2008), including a decreased abundance of the human angiotensin-converting enzyme 2 (ACE2) (Smith et al. 2011), the main receptor of the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) for cell invasion (Li et al. 2003; Ge et al. 2013; Hoffmann et al. 2020). Altogether, this evidence raises the question about the Covid-19 risk in patients with ME/CFS.

As basic information, ACE2 is encoded by the X-linked *ACE2* gene whose expression is predominant in the lungs, heart, skin, and kidneys (Hamming et al. 2004; To and Lo 2004; Li et al. 2020a; Radzikowska et al. 2020). Its expression can also be detected in monocytes (Rutkowska-Zapała et al. 2015) and activated macrophages (Song et al. 2023). However, the percentage of *ACE2*-expressing cells is below 5% in the main immune-cell populations (Song et al. 2023). Accordingly, current RNA-Seq studies suggest a residual *ACE2* expression in PBMCs from healthy controls (Radzikowska et al. 2020). ACE2 has an amino-acid sequence identity of 41% with its homologous angiotensin-converting enzyme (ACE) (Tipnis et al. 2000). This sequence similarity increases to 61% at the nucleotide level (Tipnis et al. 2000). The enzymes ACE and ACE2 are members of the renin-angiotensin-aldosterone system (RAAS), which reg-

ulates blood pressure and vascular resistance (Westermeier et al. 2015). In particular, ACE and ACE2 have vasoconstriction and vasodilation effects, respectively. Given this counteracting effect, high ACE:ACE2 ratios are possible indicators of severe Covid-19 outcomes, linked to increased reactive oxygen species (ROS) production, vasoconstriction, and inflammation (Pagliaro and Penna 2020).

To answer our research question, we performed a meta-analysis of public DNA methylation and gene expression data of *ACE2* and *ACE* in PBMCs. Similar study was conducted on the DNA methylation pattern of *ACE2* in the same cell type from patients with systemic lupus erythematosus (Sawalha et al. 2020), an autoimmune disease whose symptoms overlap with the ones from ME/CFS (Blomberg et al. 2018). To complement our findings, we also compared the mRNA levels of these two genes in PBMCs from a new cohort of female patients with ME/CFS and healthy women.

8.2 Materials and methods

8.2.1 Eligible diagnostic criteria of ME/CFS

In our meta-analysis, we selected public data from studies using either the 1994 US Center for Disease Control and Prevention criteria (CDC-1994) (Fukuda et al. 1994) or the 2003 Canadian Consensus Criteria (CCC-2003) (Carruthers et al. 2003) for the disease diagnosis. These criteria are defined by the presence of several key symptoms while excluding known medical conditions (e.g., multiple sclerosis or lupus) that can also explain fatigue. The choice of using these two criteria for study selection complies with the research standards set by the European Network on ME/CFS (Pheby et al. 2020).

8.2.2 Analysis of published DNA methylation association studies

Our meta-analysis was based on six genome-wide DNA methylation association studies (Table 8.1), four of which (Brenu et al. 2014; Vega et al. 2014; 2017; Trivedi et al. 2018) were previously reviewed (Almenar-Pérez et al. 2019), and other two published after this review

(Herrera et al. 2018; Helliwell et al. 2020). Briefly, these studies aimed at identifying differentially methylated CpG dinucleotide sites between patients and healthy controls. Illumina methylation arrays were used to measure the respective DNA methylation levels with the exception of a single study (Table 8.1). In this study, the measurements were made by the reduced representation bisulfite sequencing (Helliwell et al. 2020).

Table 8.1: Summary of the six DNA methylation studies under analysis.

Reference	Sample type	ME/CFS patients			Healthy controls, n	Technology (manufacturer)	NCBI GEO Accession number
		n	Sample characteristics	Case definition			
Brenu et al. (2014)	CD4 ⁺ T cells	25	Female/male adults Mean age: 50 years old Mean BMI: not reported	CDC-1994	18	Infinium HumanMethylation450K Array (Illumina)	NA
Vega et al. (2014)	PBMC	12	Female adults Mean age: 41 years old Mean BMI: 23 kg/m ²	CDC-1994 & CCC-2003	12	Infinium HumanMethylation450K Array (Illumina)	GSE59489
Vega et al. (2017)	PBMC	49	Female adults Mean age: 50 years old Mean BMI: 23 kg/m ²	CDC-1994 & CCC-2003	25	Infinium HumanMethylation450K Array (Illumina)	GSE93266
Trivedi et al. (2018)	PBMC	13	Female adults Mean age: 50 years old Mean BMI: 26 kg/m ²	CDC-1994 & CCC-2003	12	Methylation EPIC Array (Illumina)	GSE111183
Herrera et al. (2018)	T lymphocytes	61	Female/male adults Mean age: 32 years old Mean BMI: 27 kg/m ²	CDC-1994 & CCC-2003	48	Infinium HumanMethylation450K Array (Illumina)	GSE156792
Helliwell et al. (2020)	PBMC	10	Female/male adults Mean age: not reported Mean BMI: not reported	CCC-2003	10	Reduced representation Bisulfite sequencing	GSE153667

With respect to the exclusion criteria, one study excluded individuals who were taking beta-blockers or ACE inhibitors (Trivedi et al. 2018). Three studies excluded participants who were treated with immunomodulatory effects or affecting the underlying DNA methylation levels at the time of data collection (Vega et al. 2014; 2017; Herrera et al. 2018).

In four of the published DNA methylation studies, patients and healthy controls were matched for age, gender, and body mass index (Table 8.1) (Vega et al. 2014; 2017; Trivedi et al. 2018). In two other studies, the matching was only based on age and gender (Brenu et al. 2014; Helliwell et al. 2020). Ethnicity was also used for further matching (Trivedi et al. 2018; Herrera et al. 2018) or the same matching could be assumed in studies that only recruited white females (Vega et al. 2014; 2017). The DNA methylation levels were quantified in CD4⁺ T cells (Brenu et al. 2014), PBMCs (Vega et al. 2014; 2017; Trivedi et al. 2018; Helliwell et al. 2020), and T lymphocytes (Herrera et al. 2018).

We conducted a joint analysis of the four array-based studies which made the data available

(Vega et al. 2014; 2017; Trivedi et al. 2018; Herrera et al. 2018). We first retrieved the data from all the CpG probes located in the coding regions and the transcription starting sites (TSS) of *ACE* and *ACE2*, respectively. We then restricted our data analysis to the 27 probes shared between the Infinium HumanMethylation450K and the Infinium HumanMethylationEPIC arrays (Supplementary Table E.1).

Before conducting the statistical analysis itself, we checked whether (1) the selected probes showed a high probability of detection, (2) they were not cross-reactive with other genomic regions, and (3) they were not affected by single nucleotide polymorphisms (SNPs) with high minor allele frequencies (Dedeurwaerder et al. 2014). In the latter criterion, the SNPs included in the selected probes had a minor allele frequency less than 5% in Europeans and North Americans (Figure 8.1A; Supplementary Table E.2) referring to the sampled populations of the studies. All probes passed the remaining basic quality control checks.

We analysed the M-values of a given probe instead of the respective β -values to ensure a good approximation of the Normal distribution to the data (Du et al. 2010). Briefly, the β -values were calculated as the proportion of the methylation signal relative to the total signal for a given probe. The M-values were finally obtained by applying a logit transformation to the β -values.

To analyse the M-values of each probe, we initially estimated a linear regression model where the respective covariates were the study indicator and the disease status of the participants. In this model, we included the main effects of the covariates and the interaction. The model parameters were then estimated by the maximum likelihood method. Note that the main effect of the disease status is usually seen as the pooled effect of this covariate across all studies, as done in meta-analysis.

We then simplified the model using a backward stepwise procedure based on Akaike's information criteria. Since the effect of the study indicator was significant for the data of each probe, we tested the association between ME/CFS and a given probe using a likelihood ratio test. In this test, we compared the model including the study indicator only with the best model including that covariate and the one associated with disease status (i.e., either the model only including the main effects or the model including both main effects and the interaction term).

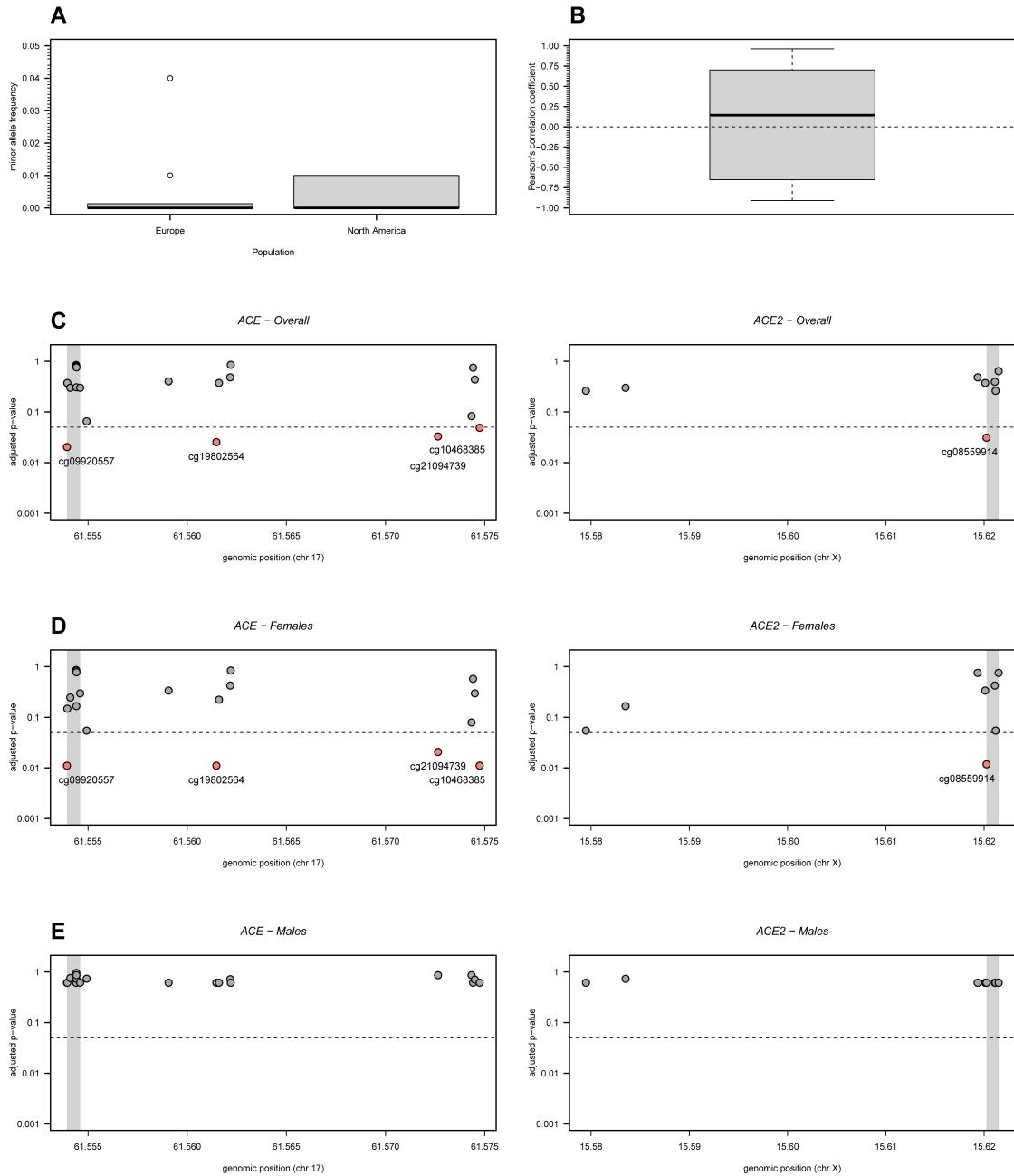


Figure 8.1: DNA methylation analysis of 19 and 8 CpG probes located in the *ACE* and *ACE2* genes, respectively. (A) Minor allele frequency in European and North American populations of SNPs located in the probes under analysis (see the respective data in Supplementary Table E.2). (B) Boxplot of all possible Pearson's correlation coefficients (y axis) between the M-values of the probes under analysis. Horizontal dashed line represents the situation of lack of correlation. (C) Adjusted p-values for the overall association between each probe and ME/CFS. Adjusted p-values were calculated according to the Benjamini-Hochberg procedure with a false discovery rate of 5% (dashed line). Grey areas in the plots represent the TSS of the genes. (D) and (E) The same analyses as shown in C but for women and men separately.

To control for multiple testing, we adjusted the raw p-values using the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995). This adjustment ensured a false discovery rate of 5% under the assumption of independent tests. Pearson's correlation coefficient was used to check the validity of this assumption (Figure 8.1B).

We also repeated the same association analysis for women and men separately. Note that three studies only recruited women (Vega et al. 2014; 2017; Trivedi et al. 2018) while the remaining study recruited both men and women (Herrera et al. 2018). In the latter study, there was no information available about the gender of each participant. In this case, we estimated this missing information using the function *getSex* of the R package *minfi* applied to the genome-wide DNA methylation data (Aryee et al. 2014). The resulting frequencies of men and women matched with those reported in the original study.

In the women-specific analysis, we performed the same association analysis as described above. In the men-specific analysis, we compared a linear regression model with the disease status as the single covariate against another model without that covariate, when analysing data from each probe. The comparison was done by the likelihood ratio test whose p-values were then adjusted for multiple testing in the same way as described above.

Finally, for the study which did not share the respective data (Brenu et al. 2014), we checked whether the reported differentially methylated CpG probes were located in either *ACE* or *ACE2* (see Table 1 from this study). We did the same for the study based on the reduced representation bisulfite sequencing technology (Helliwell et al. 2020) (see Additional File 1 from this study).

8.2.3 Analysis of gene expression studies

Our meta-analysis of gene expression studies was focused on eight reports using microarray technology (Table 8.2) (Whistler et al. 2005; Kaushik et al. 2005; Kerr et al. 2008; Saiki et al. 2008; Pietrangelo et al. 2009; Gow et al. 2009; Smith et al. 2011; Jeffrey et al. 2019). These studies complied with the Minimum Information about a Microarray Experiment (MIAME) standard (Brazma et al. 2001) and, therefore, they were considered to have sufficient quality

for their inclusion in the meta-analysis. In particular, these studies normalised the data which ensured comparability between different samples and between different measurements of the same genes.

Table 8.2: Summary of the 8 microarray-based gene expression studies under analysis, ordered by the year of publication.

Reference	Sample type	ME/CFS patients			Healthy controls, n	Technology (manufacturer)	ACE/ACE2 available	Data availability (NCBI GEO Assessment number)
		n	Sample characteristics	Case definition				
Whistler et al. (2005)	PBMC	5	Female adults Mean age: 42 years old Mean BMI: not reported	CDC-1994	5	Atlas Glass Human 3.8 I Microarray (BD Biosciences Clontech)	No/No	No (NA)
Kaushik et al. (2005)	PBMC	25	Female/male adults Mean age: 41 years old Mean BMI: not reported	CDC-1994	25	Custom microarray (Nimblegen)	Unclear	No (NA)
Kerr et al. (2008)	Whole blood	25	Female/male adults Mean age: 43 years old Mean BMI: not reported	CDC-1994	50	GeneChip Human Genome U133 Plus 2.0 (Affymetrix)	Yes/Yes	No (NA)
Saiki et al. (2008)	Whole blood	11	Female/male adults Mean age: 34 years old Mean BMI: 20.3 kg/m ²	CDC-1994	11	Custom microarray (NA)	Yes/No	Yes (NA) ¹
Pietrangelo et al. (2009)	Muscle biopsies	4	Female/male adults Mean age: 45/37 years old Mean BMI: not reported	CDC-1994	5	Operon V2.0 (CRIBI University of Padova)	Yes/Yes	No (NA)
Gow et al. (2009)	PBMC	8	Male adults Median age: 36 years old Mean BMI: not reported	CDC-1994	7	GeneChip Human Genome U133 (Affymetrix)	Yes/Yes	Yes (GSE14577)
Smith et al. (2011)	PBMC	37	Female/male adults Mean age: 51 years old Mean BMI: 29.4 kg/m ²	CDC-1994	25	MWG 20K human Array (Biotech MWG)	Yes/Yes	No (NA)
Jeffrey et al. (2019)	PBMC	33	Female/male adults Mean age: not reported Mean BMI: not reported	CDC-1994	21	GeneChip Human Gene ST (Affymetrix)	Yes/No	No (NA)

¹Data shared as a supplementary file in the online version of the study.

Gene expression of these studies was performed in PBMCs (5 studies), whole blood (2 studies) and muscle biopsies (one study). One study excluded participants who were taking any regular medication (Gow et al. 2009). Another study reviewed the medications taken by the participants (Smith et al. 2011). However, it was unclear which medications were considered as a part of the exclusion criteria. A third study reported that healthy controls were free from any medication at the time of sampling (Saiki et al. 2008).

Three additional studies using microarray technology (Vernon et al. 2002; Galbraith et al. 2011; Nguyen et al. 2017) were excluded from our meta-analysis due to unclear or ineligible case definitions of ME/CFS. We also excluded four RNA-seq studies (Bouquet et al. 2017; 2019; Sweetman et al. 2019; Raijmakers et al. 2019), because of insufficient reporting on the basic quality control checks. In particular, these studies did not report the percentage of reads

that could be mapped onto the reference transcriptome, the percentage of the transcriptome covered, the average number of mapped reads per transcript, the relationship between the GC content and the mapped read distribution, as recommended elsewhere (Conesa et al. 2016). More importantly, given the high sequence homology between ACE and ACE2, these studies did not explain how their mapping algorithms dealt with reads that could be ambiguously mapped onto different locations in the transcriptome.

The selected studies were conducted in small cohorts of patients with ME/CFS (mean sample size = 18.5; range = 4–37) and healthy controls (mean sample size = 18.6; range = 5–50 individuals) (Table 8.2). In these studies, the patients and healthy controls were matched for age and gender. Different commercial and custom microarray technologies were used for the respective gene expression quantification. There was only one study in which the microarray did not include any probe in the genes of interest (Whistler et al. 2005). Another study used a custom array based on 9,522 genes from the RefSeq database, as available in August 2002 (Kaushik et al. 2005). However, this study did not provide the list of genes included in the respective microarray. In terms of data sharing, one study made the data available in the GEO database (Gow et al. 2009) and another one within the respective publication (Saiki et al. 2008). The latter study used a custom microarray that measured the expression of stress-related genes including *ACE* but excluding *ACE2*.

Before conducting a meta-analysis of the available data, we first re-analysed two studies where the normalized data were available (Saiki et al. 2008; Gow et al. 2009). In the first study (Saiki et al. 2008), we calculated the mean of the \log_2 (fold-change) for *ACE* and the respective standard error. Note that the microarray used in this study did not include any probe in *ACE2*. In the second study (Gow et al. 2009), we initially calculated the mean and the respective standard error of the \log_2 (fold-change) for each probe located in *ACE* and *ACE2*. We then pooled each pair of means for the same gene using the inverse-variance weighting method (Hartung et al. 2008). A third study reported the mean of the \log_2 (fold-change) for *ACE2* and the respective p-value using a two-tailed Student's test (Smith et al. 2011). In this case, we determine the quantile of the t-distribution associated with half of the reported p-value, equated it to the test statistic, and solved the resulting equation as a function of the standard

error. No information was available from this study concerning the expression levels of *ACE*.

Finally, we pooled the different estimates for the same gene from different studies using the inverse-variance weighting method (Hartung et al. 2008).

8.2.4 Analysis of new RNA data on the ACE/ACE2 gene expression in ME/CFS

Study participants

Thirty-seven women with ME/CFS were recruited in 2020 from the outpatient clinic for immunodeficiencies at the Institute for Medical Immunology at the Charité-Universitätsmedizin Berlin, Germany. These patients were diagnosed according to the CCC-2003 while excluding other medical or neurological diseases which could explain fatigue (Carruthers et al. 2003). Thirty-four women with self-reported healthy status were recruited from staff.

Experimental procedure for RNA isolation and expression

Consistently with previous studies of ME/CFS, the gene expression quantification was performed in PBMCs. These cells were isolated from heparinized whole blood by density gradient centrifugation using Biocoll Separating Solution (Merck Millipore). Total RNA was isolated and extracted from 2×10^6 PBMCs according to the manufacturer's instructions (NucleoSpin RNA Kit, Macherey-Nagel, cat. nr. 740955.50). Afterwards cDNA was prepared by reverse transcription (High-Capacity cDNA Reverse Transcription Kit, Applied Biosystems, cat. nr. 4368814) and real-time PCR was performed using TaqMan™ Universal PCR Master Mix (cat. nr. 4305719) and TaqMan™ Gene Expression Assays (cat. nr. 4331182) for *ACE* (Hs00174179_m1), *ACE2* (Hs01085333_m1) and the housekeeping gene *HPRT1* (Hs02800695_m1) (Applied Biosystems). The amplification of *ACE* and *HPRT1* was based on 20 ng template cDNA. For the amplification of *ACE2*, this quantity was increased to 100 ng. All measurements were performed with the ABI7200 and software Step One Plus as absolute quantification according to manufacturer's instruction. Relative gene expression was analysed using the ΔCT method.

Statistical analysis

We first tested whether patients and healthy controls were matched for age using the Kolgomorov-Smirnov test for two independent samples. For statistical convenience, gene expression values were independently transformed for ACE and ACE2 using a Box-Cox transformation (Asar et al. 2017). The parameter estimates of this transformation were 0.303 and 0.225 for ACE and ACE2, respectively. The transformed values for each gene were then analysed as the outcome variable of a linear regression model specifying age and disease status of the participants as the respective covariates. The linear regression model was estimated using the maximum likelihood method. After estimating the models, we tested the Normal distribution in the resulting residuals using the Shapiro-Wilk test. We also visually inspected the assumption of constant variance of the same residuals as a function of the covariates.

Note that we were unable to quantify the *ACE2* expression in 11 patients due to cDNA material below the limit of detection. These problematic samples could be due to a lower expression of *ACE2* in ME/CFS patients than in healthy controls. To test this hypothesis, we compared the respective proportion of samples below the limit of detection using the Pearson's χ^2 test for two-way frequency tables. The significance level of the statistical analysis was set at 5%.

Ethical approval

The protocol of this study was approved by the Ethics Committee of Charité-Universitätsmedizin Berlin in accordance with the 1964 Declaration of Helsinki and its later amendments (reference number EA2/067/20). All patients and healthy controls gave written informed consent to participate in the study.

8.2.5 Statistical software

We performed our statistical analysis in the R software version 4.0.3. In this analysis, we used the following Bioconductor packages: *hgu133a.db*, *hgu133plus2.db*, *IlluminaHumanMethylation450kanno.ilmn12.hg19*, and *IlluminaHumanMethylationEPICannoilm10b2.hg19* to re-

trieve the annotation of the GeneChip HG-U133A, GeneChip U133+2, Infinium HumanMethylation450K Array and HumanMethylationEPIC arrays, respectively; *minfi* to estimate the sex of each individual from DNA methylation data (Aryee et al. 2014). The R scripts are freely available from the first and last authors upon request.

8.3 Results

8.3.1 Meta-analysis of ACE/ACE2 DNA methylation in ME/CFS patients

The oldest DNA methylation study (Brenu et al. 2014) did not make the data available and hence, we screened the list of 120 differentially methylated probes (see Table 1 from this study). Although located in 70 genes, these probes were neither located in *ACE* nor *ACE2*. We also screened the list of differentially methylated probes reported by the study based on the reduced representation bisulfite sequencing technology (see Additional File 1 from Helliwell et al. (2020)). Again, none of these probes was in the *ACE* or *ACE2* loci.

For the four array-based studies (Vega et al. 2014; 2017; Trivedi et al. 2018; Herrera et al. 2018), we conducted a joint analysis of the respective data in accordance with a meta-analysis. We first observed that the M-values of the 27 probes under investigation tended to be uncorrelated with each other (Figure 8.1B). This observation supported the use of the Benjamini-Hochberg procedure to adjust the raw p-values under a multiple testing scenario.

The subsequent analysis suggested four CpG probes in *ACE* to be associated with ME/CFS (Figure 8.1C). The probe cg09920557 belongs to the TSS region of the gene while the remaining probes (cg19802564, cg21094739, and cg10468385) are located in the gene body. The best linear regression models for each probe included both the main effects of the study indicator and of the disease status and the respective interaction term (Supplementary Table E.3). The statistical interaction between these two covariates could be seen when plotting the whole data set (Figure 8.2A). Although not significant, the estimated main effect of the disease status was negative for each of the significantly associated probes.

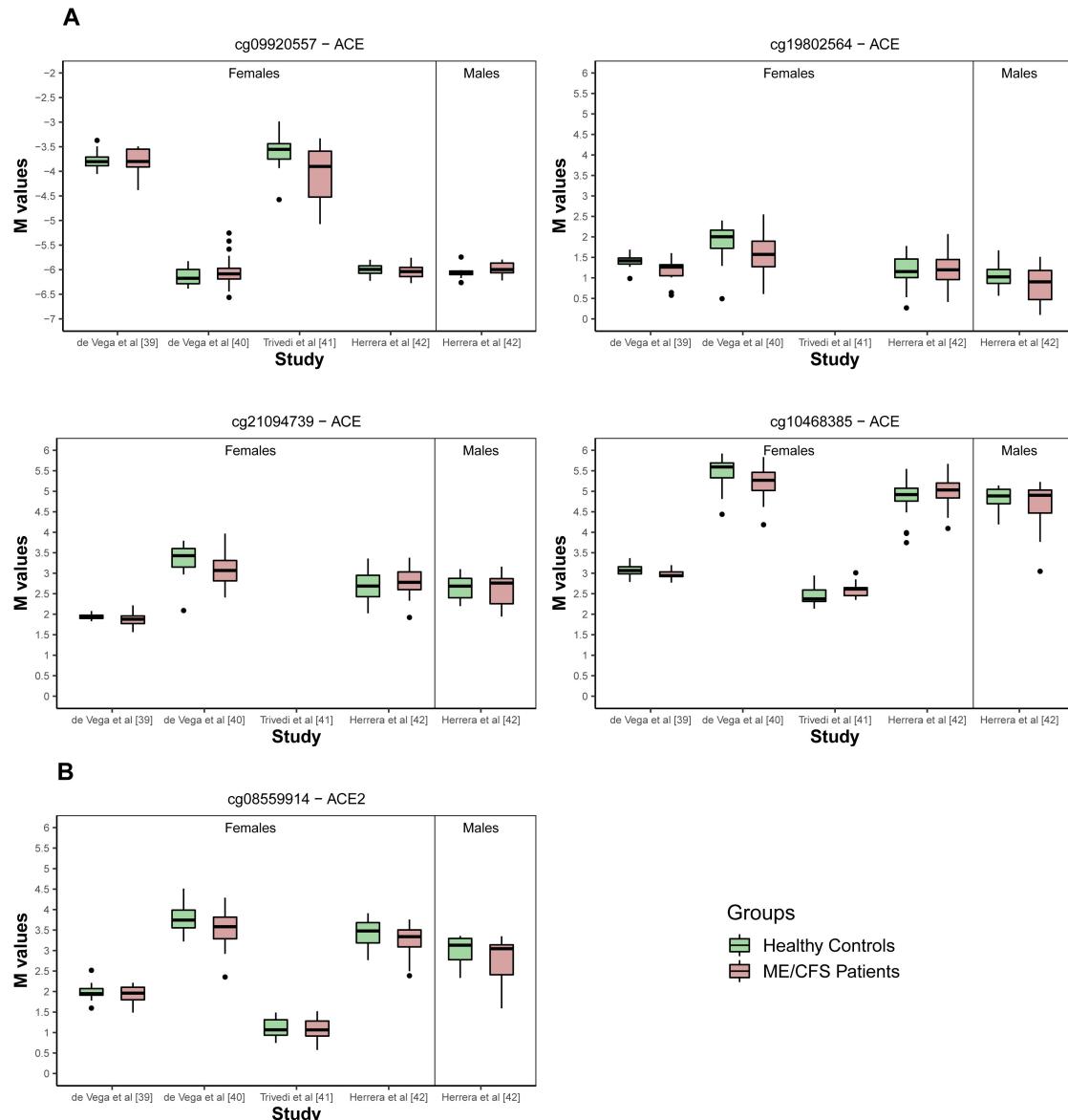


Figure 8.2: Boxplots per study, group and gender of the M-values referring to probes identified in Figure 8.1C and Figure 8.1D. (A) Significant probes located in ACE. (B) Significant probe located in ACE2.

Concerning the probes in *ACE2*, the only significant association with ME/CFS was obtained for cg08559914 located in the TSS region of the gene (Figure 8.1C). According to the best linear regression model for this probe, there was a negative association between the respective M-values and ME/CFS (coefficient estimate = -0.141 with a standard error of 0.048; Figure 8.2B and Supplementary Table E.3). Given that a hypomethylated promoter region is typically indicative of an increased expression of the respective gene, this finding suggested an increased *ACE2* expression in patients with ME/CFS.

We then repeated the same analysis for women and men separately. For women, we obtained the same disease associations, as described above (Figure 8.1D and Supplementary Table E.3). For men, we did not find any significant associations, probably due to data from a single study (Herrera et al. 2018) (Figure 8.1E).

8.3.2 Meta-analysis of ACE/ACE2 gene expression in ME/CFS patients

We first conducted a re-analysis of the two studies in which the expression levels of *ACE* or *ACE2* were available for each participant (Figure 8.3A) (Saiki et al. 2008; Gow et al. 2009). In the first study (Saiki et al. 2008), there was evidence for an increased expression of *ACE* in patients with ME/CFS (mean of the \log_2 (fold-change) = 0.265; 95% CI = [0.089, 0.441]). In the second study (Gow et al. 2009), the means of the \log_2 (fold-change) were estimated at 0.012 (95% CI = [-0.012, 0.036]) and 0.004 (95% CI = [-0.014, 0.022]) for the two probes in *ACE*. The corresponding estimates for the two probes in *ACE2* were -0.038 (95% CI = [-0.085, 0.009]) and -0.037 (95% CI = [-0.083, 0.008]) (Figure 8.3A). The pooled estimates for this study were 0.007 (95% CI = [-0.006, 0.020]) and -0.038 (95% CI = [-0.067, -0.008]) for *ACE* and *ACE2*, respectively.

Although not sharing the data, there was a study (Smith et al. 2011) that reported a significant negative association between ME/CFS and *ACE2* expression (see online Supplementary Table 2 of this study). In this case, we obtained the following mean of the \log_2 (fold-change) = -2.396 and 95% CI = (-4.518, -0.273).

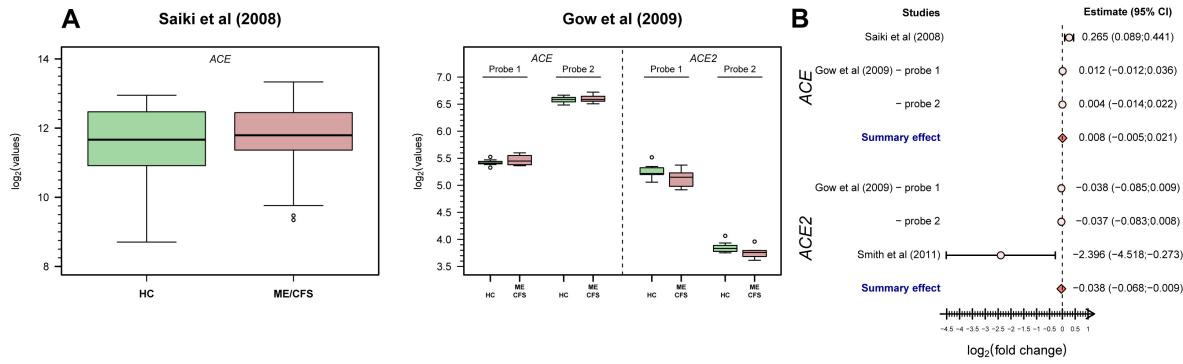


Figure 8.3: Analysis of *ACE*/*ACE2*-related data from eligible microarray-based gene expression studies. (A) Boxplots of the data from these studies (Saiki et al. 2008; Gow et al. 2009). (B) Forest plot for the study-specific and pooled estimate of the mean of the \log_2 (fold-change) between patients with ME/CFS and healthy controls using data shown in A.

We then pooled the estimates from different studies for the same gene: 0.008 (95% CI = [−0.005, 0.021]) and −0.038 (95% CI = [−0.068, −0.009]) for *ACE* and *ACE2*, respectively (Figure 8.3B). Therefore, our meta-analysis suggested a reduced expression of *ACE2* but not of *ACE* in patients with ME/CFS when comparing to healthy controls.

Finally, the remaining gene expression studies neither shared the respective data nor reported any differential *ACE*/*ACE2* expression between patients and healthy controls.

8.3.3 Analysis of *ACE*/*ACE2* gene expression from a new female cohort

To complement our findings from the above meta-analysis, we measured the *ACE* and *ACE2* mRNA levels in PBMCs from 37 women with ME/CFS (mean age = 41.1 years old) and 34 healthy women (mean age = 37.4 years old) (Table 8.3). Patients and healthy participants were matched for age (Kolmogorov-Smirnov test, $p = 0.38$). There was no information about the disease duration for 4 patients. The average disease duration for the remaining patients was 5.4 months in relation to the time of diagnosis (range = 0–24 months).

We observed higher mRNA levels of *ACE* than of *ACE2* (Table 8.4, Figure 8.4A). There was no evidence for a significant correlation between *ACE* and *ACE2* expression levels (Spearman's correlation coefficient = −0.120) (Figure 8.4B). In contrast to the above meta-analysis, we

Table 8.3: Summary statistics for the gene expression of *ACE* and *ACE2* from the German female study participants where data of *ACE2* were only available for 26 affected patients.

Summary statistic	Healthy controls	ME/CFS patients
N	34	37
Mean age (range), years	37.4 (23, 65)	41.1 (19, 60)
Mean disease duration since diagnostic (range), months	—	5.4 (0, 24)
ACE		
Geometric mean	0.153	0.144
Interquartile range	0.087	0.073
ACE2		
Geometric mean	0.002	0.001
Interquartile range	0.005	0.004

could not find a reduced expression of *ACE2* in patients with ME/CFS using the complete case scenario (Table 8.4). However, there were 11 (29.7%) of the 37 samples from patients in which the expression level of *ACE2* was below the limit of detection. This proportion of samples was significantly higher than that for healthy controls given that the expression of *ACE2* could be quantified in all the samples (29.7% versus 0%; Pearson's χ^2 test, $p = 0.002$). Consequently, we could not rule out that the patients with ME/CFS from this cohort have a decreased expression of *ACE2* when compared to healthy controls. Finally, in accordance with our meta-analysis, there was no evidence of differential expression of *ACE* between patients and healthy controls from this cohort.

8.4 Discussion

In this work, we investigated potential differences in *ACE*/*ACE2* DNA methylation and expression levels between patients with ME/CFS and healthy controls. With the identification of these differences, we expected to determine the health risk of patients with ME/CFS if infected by SARS-CoV-2. However, we stumbled upon hurdles related to (i) data unavailability for a possible re-analysis, (ii) availability of data derived from PBCMs and related subsets in which

Table 8.4: Analysis of the linear regression models for the Box-Cox-transformed *ACE* and *ACE2* mRNA levels where data were only available for 26 ME/CFS patients.

Analyses	Estimate (SE)	P-value
Box-Cox transformed <i>ACE</i>		
Intercept	0.541 (0.032)	≤ 0.001
Age	0.001 (0.001)	0.328
Disease status (ME/CFS)	-0.013 (0.018)	0.481
Box-Cox transformed <i>ACE2</i>		
Intercept	0.307 (0.038)	≤ 0.001
Age	-0.001 (0.001)	0.137
Disease status (ME/CFS)	-0.006 (0.021)	0.789

ACE2 is not particularly expressed, (iii) studies with unclear data quality, and (iv) studies using disease case definitions that are not recommended for research. As a consequence, we could not provide a more definite answer to our main research question.

Notwithstanding these difficulties, we could identify four CpG probes on *ACE* and another one on *ACE2* with decreased DNA methylation levels in patients with ME/CFS. This finding suggested an increased expression of the respective genes. However, our meta-analysis of public data suggested the opposite. Such decrease in *ACE2* expression was partially confirmed by new data in which there was a significant higher proportion of samples below the limit of detection in patients with ME/CFS than in healthy controls. Nonetheless, it was clear that *ACE2* is not particularly expressed in PBMCs from both patients with ME/CFS and healthy controls, as mentioned in the introduction.

In general, *ACE2* downregulation is known to occur after host-cell entry by SARS-CoV-2 (Datta et al. 2020). This downregulation is particularly problematic in individuals affected by cardiovascular diseases, diabetes, and other medical conditions, due to their low *ACE2* levels before the infection (Verdecchia et al. 2020). SARS-CoV-2 infection is then expected to further increase the *ACE*:*ACE2* ratio, thus, promoting vasoconstriction, increased production of ROS and inflammation in patients with these co-morbidities (Pagliaro and Penna 2020). In this scenario, a putative reduction of the *ACE2* expression makes patients with ME/CFS similar to

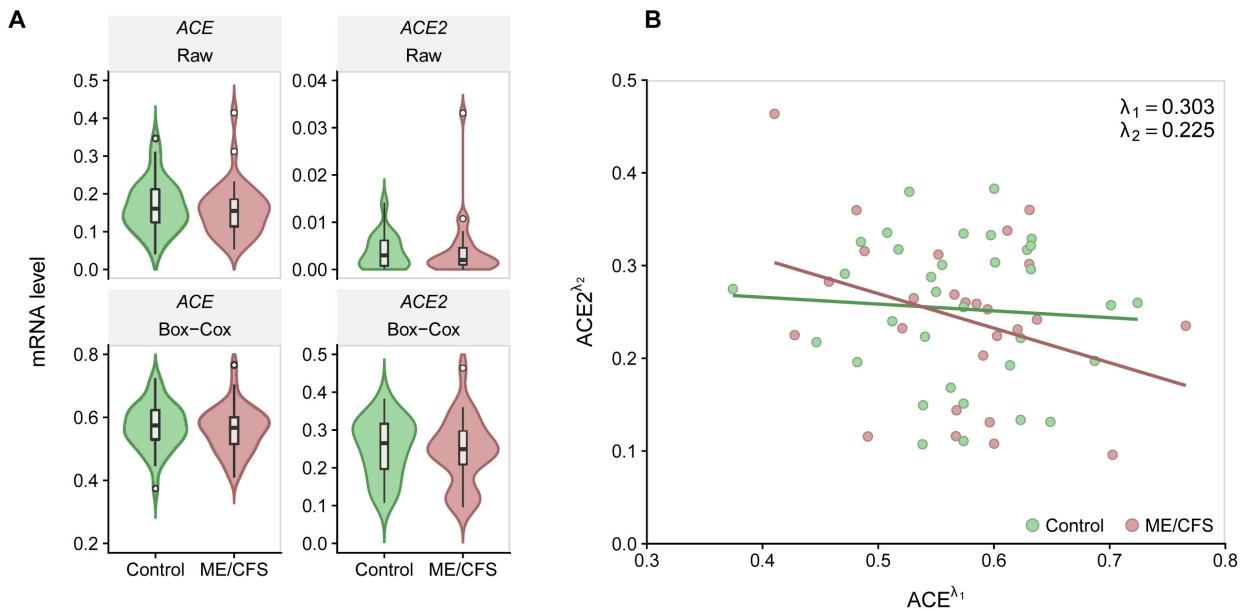


Figure 8.4: Analysis of ACE and ACE2 expression levels from the German study. (A) Violin plots of ACE (left side) and ACE2 (right side) mRNA raw data (upper row) and transformed data using a Box-Cox transformation (lower row). (B) Scatterplot between the transformed ACE and ACE2 expression levels (Spearman's correlation coefficient = -0.120).

these patients with a high risk for Covid-19. As a consequence, patients with ME/CFS could be considered a priority group for vaccination by public health authorities. The fundamental question is then to know whether our findings based on PBMCs could recreate what occurs in pulmonary epithelial and endothelial cells, the main targets of SARS-CoV-2. Future research should be conducted to answer this question, as similarly done in past studies aiming at understanding how the gene expression profiles from PBMCs could mimic those present in other tissues affected by a given disease (Takamura et al. 2007; Manoel-Caetano et al. 2012; Gerling et al. 2013).

Given the residual ACE2 expression in PBMCs under normal conditions, one is tempted to say that SARS-CoV-2 does not infect these cells. However, earlier studies on SARS-CoV-1 found this virus within T lymphocytes, macrophages, and dendritic cells (Tay et al. 2020). More recently, an *in vitro* study was able to infect PBMCs with SARS-CoV-2 (Codo et al. 2020). Monocytes are particularly susceptible to such infections. In this context, one cannot rule out that SARS-CoV-2 might use alternative receptors when infecting PBMCs.

Among the alternative receptors for SARS-CoV-2, the human transmembrane protease serine 2 (TMPRSS2) was suggested as a strong candidate (Sungnak et al. 2020) due to its role on SARS-CoV-1 infection (Matsuyama et al. 2010; Glowacka et al. 2011). This protease seems to induce SARS-CoV-2 cell entry through endocytosis via a mechanism of ACE2 cleavage (Hoffmann et al. 2020). Another candidate receptor is the A disintegrin and metallopeptidase domain 17 protein (ADAM17) recognised by the immune system as a stress-response signal (Düsterhöft et al. 2019). Like TMPRSS2, ADAM17 can also cleave ACE2 but with a reduced viral invasion efficiency (Heurich et al. 2014).

With respect to the role of these proteases in ME/CFS, a targeted gene expression study analysed ADAM17 and other stress-response proteins (Saiki et al. 2008). This study did not report any differential expression of this protease between patients with ME/CFS and healthy controls. However, this study is likely to be affected by a low statistical power due to small sample sizes for both groups. In addition, one of the selected DNA methylation studies suggested a decrease in the DNA methylation levels of one ADAM17-related CpG probe in patients with ME/CFS (Trivedi et al. 2018).

Dipeptidyl peptidase-4 (DPP4), also known as the lymphocyte cell surface protein CD26, was found to be the main receptor for the Middle East respiratory syndrome–related coronavirus (van Doremalen et al. 2014; Widagdo et al. 2019). In contrast to ACE2, this surface protein is highly abundant in PBMCs including CD4⁺ and CD8⁺ T cells (Radzikowska et al. 2020). Bioinformatic analysis also suggested a strong interaction potential between this protein and SARS-CoV-2 (Li et al. 2020b; Vankadari and Wilce 2020). Finally, DPP4 inhibitors were found to be protective against severe Covid-19 in patients with diabetes mellitus when compared to RAAS blockers (Rhee et al. 2021). After initial concerns, this finding combined with others suggested an interesting therapeutic avenue against Covid-19 using DPP4 blockers (Scheen 2021).

Interestingly, there is evidence for an increased proportion of natural killer cells and T cells expressing DPP4/CD26⁺ in patients with ME/CFS (Klimas et al. 1990; Fletcher et al. 2010). However, the number of DPP4/CD26 molecules was significantly reduced in T lymphocytes and natural killer cells of these patients (Fletcher et al. 2010). If DPP4 is indeed a relevant receptor

for immune-cell invasion by SARS-CoV-2, research about this receptor should be prioritised when analysing PBMCs from patients with ME/CFS.

Sialic acids were also hypothesised as binding receptors used by SARS-CoV-2, as reported for other human coronaviruses (Sun 2021). These acids are highly expressed in the epithelium cells of the lungs and oral cavity (Cross and Ruhl 2018). *In vitro* and *in silico* studies demonstrated the same binding potential for SARS-CoV-2 (Awasthi et al. 2020; Baker et al. 2020; Milanetti et al. 2021). However, the ACE2 glycosylation inhibition studies suggested that sialic acids on ACE2 receptor prevent ACE2-virus interaction (Chu et al. 2021; Yang et al. 2020). Again, detailed research on these putative receptors could help to determine the health risk of patients with ME/CFS when infected by SARS-CoV-2.

It was suggested that the arousal state experienced by patients with ME/CFS protects them against microbial infections (Sulheim et al. 2014). This suggestion came from a clinical trial where patients were treated with clonidine to decrease such a state. Treated patients got their symptoms worsened and had their inflammation markers increased during the trial. In contrast, basic epidemiological studies reported many patients with frequent viral infections and flu-like symptoms (Johnston et al. 2016; Chu et al. 2019; Słomko et al. 2019). The question is how an infection by SARS-CoV-2 lies in this contrasting evidence. A possible answer can be given with the assistance of the so-called sustained arousal model of ME/CFS (Wyller et al. 2009). According to this model, a sustained arousal state promotes in the long-run deleterious alterations of different body systems, including the immune system. Similar prediction was made by a recent study discussing the natural history of ME/CFS (Nacul et al. 2020). If so, patients with longer disease durations are more likely to show these immunological alterations than patients at the early stages of the disease. However, we could not analyse the effect of disease duration on our results, because this variable was not available in the public data sets included in our meta-analyses.

Finally, our original idea was also to include a meta-analysis of ACE/ACE2 data from published genome-wide association studies on ME/CFS (Smith et al. 2011; Schlauch et al. 2016; Herrera et al. 2018; Perez et al. 2019; Dibble et al. 2020). However, we could not materialise this idea, because such studies did not make their data publicly available. Nevertheless,

evidence is scarce for a putative role of *ACE*/*ACE2* polymorphisms on ME/CFS. Two studies reported many candidate SNPs for such association, but none was located in *ACE* or *ACE2* (Smith et al. 2011; Schlauch et al. 2016). Two other studies did not find any significant SNPs associated with ME/CFS (Herrera et al. 2018; Dibble et al. 2020). The most optimistic study reported thousands of SNPs related to the disease (Perez et al. 2019). However, this study did not perform all the basic quality control checks (Grabowska et al. 2020).

8.5 Conclusions

Notwithstanding the low expression of *ACE2* in PBMCs in general, there is evidence for a decreased expression of the gene in these cells from patients with ME/CFS. If PBMCs can qualitatively recreate what is occurring in the main cellular targets of SARS-CoV-2, then patients with this disease could be at a higher Covid-19 risk. In this regard, a recent preliminary report suggested that patients with ME/CFS got their symptoms worsened upon SARS-CoV-2 infection (Action for M.E. 2021). Altogether, these patients could be considered a priority group for vaccination against Covid-19, even though vaccines could trigger ME/CFS (Gherardi et al. 2019; Phelan et al. 2020) or even exacerbate ME/CFS symptoms as the case of the natural immunisation by SARS-CoV-2. To further consolidate the existing evidence, future research should prioritise the collection of data from the main cellular targets in patients with ME/CFS. Further investigation should be also conducted on alternative SARS-CoV-2 receptors (i.e., DPP4 and sialic acids). At last, future research should also consider investigating putative sex differences in patients with ME/CFS given that, in general, men are more affected by Covid-19 than women (Gadi et al. 2020).

Bibliography

- Action for M.E. Report on the impact of Covid-19 on ME. Apr. 2021. URL <https://www.meaction.net/wp-content/uploads/2021/04/Report-on-the-impact-of-Covid-19-on-ME.pdf>.
E. Almenar-Pérez, T. Ovejero, T. Sánchez-Fito, J. A. Espejo, L. Nathanson, and E. Oltra. Epigenetic Compo-

- nents of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Uncover Potential Transposable Element Activation. *Clinical Therapeutics*, 41(4):675–698, Apr. 2019. ISSN 0149-2918, 1879-114X. doi: 10.1016/j.clinthera.2019.02.012. URL [https://www.clinicaltherapeutics.com/article/S0149-2918\(19\)30072-4/fulltext](https://www.clinicaltherapeutics.com/article/S0149-2918(19)30072-4/fulltext). Publisher: Elsevier.
- M. E. Ariza. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: The Human Herpesviruses Are Back! *Biomolecules*, 11(2):185, Feb. 2021. ISSN 2218-273X. doi: 10.3390/biom11020185. URL <https://www.mdpi.com/2218-273X/11/2/185>. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.
- M. J. Aryee, A. E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A. P. Feinberg, K. D. Hansen, and R. A. Irizarry. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, 30(10):1363–1369, May 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btu049. URL <https://doi.org/10.1093/bioinformatics/btu049>.
- O. Asar, O. Ilk, and O. Dag. Estimating Box-Cox power transformation parameter via goodness-of-fit tests. *Communications in Statistics - Simulation and Computation*, 46(1):91–105, Jan. 2017. ISSN 0361-0918. doi: 10.1080/03610918.2014.957839. URL <https://doi.org/10.1080/03610918.2014.957839>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/03610918.2014.957839>.
- M. Awasthi, S. Gulati, D. P. Sarkar, S. Tiwari, S. Kateriya, P. Ranjan, and S. K. Verma. The Sialoside-Binding Pocket of SARS-CoV-2 Spike Glycoprotein Structurally Resembles MERS-CoV. *Viruses*, 12(9):909, Aug. 2020. ISSN 1999-4915. doi: 10.3390/v12090909.
- A. N. Baker, S.-J. Richards, C. S. Guy, T. R. Congdon, M. Hasan, A. J. Zwetsloot, A. Gallo, J. R. Lewandowski, P. J. Stansfeld, A. Straube, M. Walker, S. Chessa, G. Pergolizzi, S. Dedola, R. A. Field, and M. I. Gibson. The SARS-CoV-2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device. *ACS Central Science*, 6(11):2046–2052, Nov. 2020. ISSN 2374-7943. doi: 10.1021/acscentsci.0c00855. URL <https://doi.org/10.1021/acscentsci.0c00855>. Publisher: American Chemical Society.
- Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, Jan. 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.
- J. Blomberg, C.-G. Gottfries, A. Elfaitouri, M. Rizwan, and A. Rosén. Infection Elicited Autoimmunity and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: An Explanatory Model. *Frontiers in Immunology*, 9, Feb. 2018. doi: 10.3389/fimmu.2018.00229.
- J. Bouquet, J. L. Gardy, S. Brown, J. Pfeil, R. R. Miller, M. Morshed, A. Avina-Zubieta, K. Shojania, M. McCabe, S. Parker, M. Uyaguari, S. Federman, P. Tang, T. Steiner, M. Otterstater, R. Holt, R. Moore, C. Y. Chiu, and D. M. P. and. RNA-Seq Analysis of Gene Expression, Viral Pathogen, and B-Cell/T-Cell Receptor Signatures in Complex Chronic Disease. *Clinical Infectious Diseases*, 64(4):476–481, Jan. 2017. doi: 10.1093/cid/ciw767.
- J. Bouquet, T. Li, J. L. Gardy, X. Kang, S. Stevens, J. Stevens, M. VanNess, C. Snell, J. Potts, R. R. Miller, M. Morshed,

- M. McCabe, S. Parker, M. Uyaguari, P. Tang, T. Steiner, W.-S. Chan, A.-M. D. Souza, A. Mattman, D. M. Patrick, and C. Y. Chiu. Whole blood human transcriptome and virome analysis of ME/CFS patients experiencing post-exertional malaise following cardiopulmonary exercise testing. *PLOS ONE*, 14(3):e0212193, Mar. 2019. doi: 10.1371/journal.pone.0212193.
- A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genetics*, 29(4):365–371, Dec. 2001. ISSN 1061-4036. doi: 10.1038/ng1201-365.
- E. Brenu, D. Staines, and S. Marshall-Gradisnik. Methylation Profile of CD4+ T Cells in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *Journal of Clinical & Cellular Immunology*, 5(3):228–14, 2014. ISSN 2155-9899. doi: 10.4172/2155-9899.1000228. URL <https://research-repository.griffith.edu.au/handle/10072/68935>. Accepted: 2017-05-03T16:10:59Z Publisher: Omics Publishing Group.
- E. W. Brenu, M. L. v. Driel, D. R. Staines, K. J. Ashton, S. B. Ramos, J. Keane, N. G. Klimas, and S. M. Marshall-Gradisnik. Immunological abnormalities as potential biomarkers in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *Journal of Translational Medicine*, 9(1):81, 2011. doi: 10.1186/1479-5876-9-81.
- B. M. Carruthers, A. K. Jain, K. L. D. Meirleir, D. L. Peterson, N. G. Klimas, A. M. Lerner, A. C. Bested, P. Flor-Henry, P. Joshi, A. C. P. Powles, J. A. Sherkey, and M. I. v. d. Sande. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Chronic Fatigue Syndrome*, 11(1):7–115, Jan. 2003. doi: 10.1300/j092v11n01_02.
- H. Chu, B. Hu, X. Huang, Y. Chai, D. Zhou, Y. Wang, H. Shuai, D. Yang, Y. Hou, X. Zhang, T. T.-T. Yuen, J.-P. Cai, A. J. Zhang, J. Zhou, S. Yuan, K. K.-W. To, I. H.-Y. Chan, K.-Y. Sit, D. C.-C. Foo, I. Y.-H. Wong, A. T.-L. Ng, T. T. Cheung, S. Y.-K. Law, W.-K. Au, M. A. Brindley, Z. Chen, K.-H. Kok, J. F.-W. Chan, and K.-Y. Yuen. Host and viral determinants for efficient SARS-CoV-2 infection of the human lung. *Nature Communications*, 12(1):134, Jan. 2021. ISSN 2041-1723. doi: 10.1038/s41467-020-20457-w. URL <https://www.nature.com/articles/s41467-020-20457-w>. Number: 1 Publisher: Nature Publishing Group.
- L. Chu, I. J. Valencia, D. W. Garvert, and J. G. Montoya. Onset Patterns and Course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Pediatrics*, 7, Feb. 2019. doi: 10.3389/fped.2019.00012.
- A. C. Codo, G. G. a. Davanzo, L. d. B. Monteiro, G. F. de Souza, S. P. Muraro, J. a. V. Virgilio-da Silva, J. S. Prodonoff, V. C. Carregari, C. A. O. de Biagi Junior, F. Crunfli, J. L. Jimenez Restrepo, P. H. Vendramini, G. Reis-de Oliveira, K. Bispo Dos Santos, D. A. Toledo-Teixeira, P. L. Parise, M. C. Martini, R. E. Marques, H. R. Carmo, A. Borin, L. D. Coimbra, V. O. Boldrini, N. S. Brunetti, A. S. Vieira, E. Mansour, R. G. Ulaf, A. F. Bernardes, T. A. Nunes, L. C. Ribeiro, A. C. Palma, M. V. Agrela, M. L. Moretti, A. C. Sposito, F. B. Pereira, L. A. Velloso, M. A. R. Vinolo, A. Damasio, J. L. Proença-Módena, R. F. Carvalho, M. A. Mori, D. Martins-de Souza, H. I. Nakaya, A. S. Farias, and P. M. Moraes-Vieira. Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through

- a HIF-1/Glycolysis-Dependent Axis. *Cell Metabolism*, 32(3):437–446.e5, Sept. 2020. ISSN 1932-7420. doi: 10.1016/j.cmet.2020.07.007.
- A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szczęśniak, D. J. Gaffney, L. L. Elo, X. Zhang, and A. Mortazavi. A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1):13, Jan. 2016. ISSN 1474-760X. doi: 10.1186/s13059-016-0881-8. URL <https://doi.org/10.1186/s13059-016-0881-8>.
- B. W. Cross and S. Ruhl. Glycan recognition at the saliva - oral microbiome interface. *Cellular Immunology*, 333: 19–33, Nov. 2018. ISSN 1090-2163. doi: 10.1016/j.cellimm.2018.08.008.
- P. K. Datta, F. Liu, T. Fischer, J. Rappaport, and X. Qin. SARS-CoV-2 pandemic and research gaps: Understanding SARS-CoV-2 interaction with the ACE2 receptor and implications for therapy. *Theranostics*, 10(16):7448–7464, 2020. ISSN 1838-7640. doi: 10.7150/thno.48076.
- S. Dedeurwaerder, M. Defrance, M. Bizet, E. Calonne, G. Bontempi, and F. Fuks. A comprehensive overview of Infinium HumanMethylation450 data processing. *Briefings in Bioinformatics*, 15(6):929–941, Nov. 2014. ISSN 1477-4054. doi: 10.1093/bib/bbt054.
- J. J. Dibble, S. J. McGrath, and C. P. Ponting. Genetic risk factors of ME/CFS: a critical review. *Human Molecular Genetics*, 29(R1):R117–R124, Sept. 2020. ISSN 0964-6906. doi: 10.1093/hmg/ddaa169. URL <https://doi.org/10.1093/hmg/ddaa169>.
- P. Du, X. Zhang, C.-C. Huang, N. Jafari, W. A. Kibbe, L. Hou, and S. M. Lin. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, 11(1): 587, Nov. 2010. ISSN 1471-2105. doi: 10.1186/1471-2105-11-587. URL <https://doi.org/10.1186/1471-2105-11-587>.
- S. Düsterhöft, J. Lokau, and C. Garbers. The metalloprotease ADAM17 in inflammation and cancer. *Pathology, Research and Practice*, 215(6):152410, June 2019. ISSN 1618-0631. doi: 10.1016/j.prp.2019.04.002.
- M. A. Fletcher, X. R. Zeng, K. Maher, S. Levis, B. Hurwitz, M. Antoni, G. Broderick, and N. G. Klimas. Biomarkers in chronic fatigue syndrome: evaluation of natural killer cell function and dipeptidyl peptidase IV/CD26. *Plos One*, 5(5):e10817, May 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0010817.
- K. Fukuda, S. E. Straus, I. Hickie, M. C. Sharpe, J. G. Dobbins, and A. Komaroff. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Annals of Internal Medicine*, 121(12):953, Dec. 1994. doi: 10.7326/0003-4819-121-12-199412150-00009.
- N. Gadi, S. C. Wu, A. P. Spihlman, and V. R. Moulton. What's Sex Got to Do With COVID-19? Gender-Based Differences in the Host Immune Response to Coronaviruses. *Frontiers in Immunology*, 11:2147, 2020. ISSN 1664-3224. doi: 10.3389/fimmu.2020.02147.
- S. Galbraith, B. Cameron, H. Li, D. Lau, U. Vollmer-Conna, and A. R. Lloyd. Peripheral blood gene expression in postinfective fatigue syndrome following from three different triggering infections. *The Journal of Infectious Diseases*

- Diseases*, 204(10):1632–1640, Nov. 2011. ISSN 1537-6613. doi: 10.1093/infdis/jir612.
- X.-Y. Ge, J.-L. Li, X.-L. Yang, A. A. Chmura, G. Zhu, J. H. Epstein, J. K. Mazet, B. Hu, W. Zhang, C. Peng, Y.-J. Zhang, C.-M. Luo, B. Tan, N. Wang, Y. Zhu, G. Crameri, S.-Y. Zhang, L.-F. Wang, P. Daszak, and Z.-L. Shi. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*, 503(7477):535–538, Nov. 2013. ISSN 1476-4687. doi: 10.1038/nature12711. URL <https://www.nature.com/articles/nature12711>. Number: 7477 Publisher: Nature Publishing Group.
- I. C. Gerling, R. A. Ahokas, G. Kamalov, W. Zhao, S. K. Bhattacharya, Y. Sun, and K. T. Weber. Gene Expression Profiles of Peripheral Blood Mononuclear Cells Reveal Transcriptional Signatures as Novel Biomarkers for Cardiac Remodeling in Rats with Aldosteronism and Hypertensive Heart Disease. *JACC. Heart failure*, 1(6): 10.1016/j.jchf.2013.09.003, Dec. 2013. ISSN 2213-1779. doi: 10.1016/j.jchf.2013.09.003. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3885325/>.
- R. K. Gherardi, G. Crépeaux, and F.-J. Authier. Myalgia and chronic fatigue syndrome following immunization: macrophagic myofasciitis and animal studies support linkage to aluminum adjuvant persistency and diffusion in the immune system. *Autoimmunity Reviews*, 18(7):691–705, July 2019. ISSN 1873-0183. doi: 10.1016/j.autrev.2019.05.006.
- I. Glowacka, S. Bertram, M. A. Müller, P. Allen, E. Soilleux, S. Pfefferle, I. Steffen, T. S. Tsegaye, Y. He, K. Gnirss, D. Niemeyer, H. Schneider, C. Drosten, and S. Pöhlmann. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *Journal of Virology*, 85(9):4122–4134, May 2011. ISSN 1098-5514. doi: 10.1128/JVI.02232-10.
- J. W. Gow, S. Hagan, P. Herzyk, C. Cannon, P. O. Behan, and A. Chaudhuri. A gene signature for post-infectious chronic fatigue syndrome. *BMC Medical Genomics*, 2(1), June 2009. doi: 10.1186/1755-8794-2-38.
- A. D. Grabowska, E. M. Lacerda, L. Nacul, and N. Sepúlveda. Review of the Quality Control Checks Performed by Current Genome-Wide and Targeted-Genome Association Studies on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Pediatrics*, 8, 2020. ISSN 2296-2360. URL <https://www.frontiersin.org/articles/10.3389/fped.2020.00293>.
- I. Hamming, W. Timens, M. Bulthuis, A. Lely, G. Navis, and H. van Goor. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *The Journal of Pathology*, 203(2):631–637, 2004. ISSN 1096-9896. doi: 10.1002/path.1570. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/path.1570>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/path.1570>.
- J. Hartung, G. Knapp, and B. K. Sinha. *Statistical Meta-Analysis with Applications*. Wiley Series in Probability and Statistics. Wiley, 1 edition, July 2008. ISBN 978-0-470-29089-7 978-0-470-38634-7. doi: 10.1002/9780470386347. URL <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470386347>.

- A. M. Helliwell, E. C. Sweetman, P. A. Stockwell, C. D. Edgar, A. Chatterjee, and W. P. Tate. Changes in DNA methylation profiles of myalgic encephalomyelitis/chronic fatigue syndrome patients reflect systemic dysfunctions. *Clinical Epigenetics*, 12(1):167, Nov. 2020. ISSN 1868-7083. doi: 10.1186/s13148-020-00960-z. URL <https://doi.org/10.1186/s13148-020-00960-z>.
- S. Herrera, W. C. de Vega, D. Ashbrook, S. D. Vernon, and P. O. McGowan. Genome-epigenome interactions associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Epigenetics*, 13(12):1174–1190, Dec. 2018. ISSN 1559-2294. doi: 10.1080/15592294.2018.1549769. URL <https://doi.org/10.1080/15592294.2018.1549769>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/15592294.2018.1549769>.
- A. Heurich, H. Hofmann-Winkler, S. Gierer, T. Liepold, O. Jahn, and S. Pöhlmann. TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. *Journal of Virology*, 88(2):1293–1307, Jan. 2014. ISSN 1098-5514. doi: 10.1128/JVI.02202-13.
- M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M. A. Müller, C. Drosten, and S. Pöhlmann. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2):271–280.e8, Apr. 2020. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2020.02.052. URL [https://www.cell.com/cell/abstract/S0092-8674\(20\)30229-4](https://www.cell.com/cell/abstract/S0092-8674(20)30229-4). Publisher: Elsevier.
- M. G. Jeffrey, L. Nathanson, K. Aenlle, Z. M. Barnes, M. Baig, G. Broderick, N. G. Klimas, M. A. Fletcher, and T. J. A. Craddock. Treatment Avenues in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Split-gender Pharmacogenomic Study of Gene-expression Modules. *Clinical Therapeutics*, 41(5):815–835.e6, May 2019. doi: 10.1016/j.clinthera.2019.01.011.
- S. C. Johnston, D. R. Staines, and S. M. Marshall-Gradisnik. Epidemiological characteristics of chronic fatigue syndrome/myalgic encephalomyelitis in Australian patients. *Clinical Epidemiology*, 8:97–107, May 2016. doi: 10.2147/CLEPS96797. URL <https://www.dovepress.com/epidemiological-characteristics-of-chronic-fatigue-nbspsyndromemyalgic-peer-reviewed-full>. Publisher: Dove Press.
- N. Kaushik, D. Fear, S. C. M. Richards, C. R. McDermott, E. F. Nuwaysir, P. Kellam, T. J. Harrison, R. J. Wilkinson, D. a. J. Tyrrell, S. T. Holgate, and J. R. Kerr. Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. *Journal of Clinical Pathology*, 58(8):826–832, Aug. 2005. ISSN 0021-9746, 1472-4146. doi: 10.1136/jcp.2005.025718. URL <https://jcp.bmjjournals.com/content/58/8/826>. Publisher: BMJ Publishing Group Section: Original article.
- J. R. Kerr, R. Petty, B. Burke, J. Gough, D. Fear, L. I. Sinclair, D. L. Matthey, S. C. M. Richards, J. Montgomery, D. A. Baldwin, P. Kellam, T. J. Harrison, G. E. Griffin, J. Main, D. Enlander, D. J. Nutt, and S. T. Holgate. Gene Expression Subtypes in Patients with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *The Journal of*

- Infectious Diseases*, 197(8):1171–1184, Apr. 2008. ISSN 0022-1899. doi: 10.1086/533453. URL <https://doi.org/10.1086/533453>.
- N. G. Klimas, F. R. Salvato, R. Morgan, and M. A. Fletcher. Immunologic abnormalities in chronic fatigue syndrome. *Journal of Clinical Microbiology*, 28(6):1403–1410, 1990. ISSN 0095-1137. URL <https://jcm.asm.org/content/28/6/1403>.
- M.-Y. Li, L. Li, Y. Zhang, and X.-S. Wang. Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infectious Diseases of Poverty*, 9(1):45, Apr. 2020a. ISSN 2049-9957. doi: 10.1186/s40249-020-00662-x. URL <https://doi.org/10.1186/s40249-020-00662-x>.
- W. Li, M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, and M. Farzan. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426(6965):450–454, Nov. 2003. ISSN 1476-4687. doi: 10.1038/nature02145. URL <https://www.nature.com/articles/nature02145>. Number: 6965 Publisher: Nature Publishing Group.
- Y. Li, Z. Zhang, L. Yang, X. Lian, Y. Xie, S. Li, S. Xin, P. Cao, and J. Lu. The MERS-CoV Receptor DPP4 as a Candidate Binding Target of the SARS-CoV-2 Spike. *iScience*, 23(8):101400, Aug. 2020b. ISSN 2589-0042. doi: 10.1016/j.isci.2020.101400.
- L. Lorusso, S. V. Mikhaylova, E. Capelli, D. Ferrari, G. K. Ngonga, and G. Ricevuti. Immunological aspects of chronic fatigue syndrome. *Autoimmunity Reviews*, 8(4):287–291, Feb. 2009. doi: 10.1016/j.autrev.2008.08.003.
- F. S. Manoel-Caetano, D. J. Xavier, A. F. Evangelista, P. Takahashi, C. V. Collares, D. Puthier, M. C. Foss-Freitas, M. C. Foss, E. A. Donadi, G. A. Passos, and E. T. Sakamoto-Hojo. Gene expression profiles displayed by peripheral blood mononuclear cells from patients with type 2 diabetes mellitus focusing on biological processes implicated on the pathogenesis of the disease. *Gene*, 511(2):151–160, Dec. 2012. ISSN 1879-0038. doi: 10.1016/j.gene.2012.09.090.
- S. Matsuyama, N. Nagata, K. Shirato, M. Kawase, M. Takeda, and F. Taguchi. Efficient Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by the Transmembrane Protease TMPRSS2. *Journal of Virology*, 84(24):12658–12664, Dec. 2010. doi: 10.1128/jvi.01542-10. URL <https://journals.asm.org/doi/10.1128/jvi.01542-10>. Publisher: American Society for Microbiology.
- E. Milanetti, M. Miotto, L. Di Rienzo, M. Nagaraj, M. Monti, T. W. Golbek, G. Gosti, S. J. Roeters, T. Weidner, D. E. Otzen, and G. Ruocco. In-Silico Evidence for a Two Receptor Based Strategy of SARS-CoV-2. *Frontiers in Molecular Biosciences*, 8:690655, 2021. ISSN 2296-889X. doi: 10.3389/fmolb.2021.690655.
- L. Nacul, S. O’Boyle, L. Palla, F. E. Nacul, K. Mudie, C. C. Kingdon, J. M. Cliff, T. G. Clark, H. M. Dockrell, and E. M. Lacerda. How Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) Progresses: The Natural History of ME/CFS. *Frontiers in Neurology*, 11, 2020. ISSN 1664-2295. doi: <https://doi.org/10.3389/fneur.2020.00826>. URL <https://www.frontiersin.org/articles/10.3389/fneur.2020.00826>.

- C. B. Nguyen, L. Alsøe, J. M. Lindvall, D. Sulheim, E. Fagermoen, A. Winger, M. Kaarbø, H. Nilsen, and V. B. Wyller. Whole blood gene expression in adolescent chronic fatigue syndrome: an exploratory cross-sectional study suggesting altered B cell differentiation and survival. *Journal of Translational Medicine*, 15(1):102, May 2017. ISSN 1479-5876. doi: 10.1186/s12967-017-1201-0.
- P. Pagliaro and C. Penna. ACE/ACE2 Ratio: A Key Also in 2019 Coronavirus Disease (Covid-19)? *Frontiers in Medicine*, 7, 2020. ISSN 2296-858X. URL <https://www.frontiersin.org/articles/10.3389/fmed.2020.00335>.
- M. Perez, R. Jaundoo, K. Hilton, A. D. Alamo, K. Gemayel, N. G. Klimas, T. J. A. Craddock, and L. Nathanson. Genetic Predisposition for Immune System, Hormone, and Metabolic Dysfunction in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Pilot Study. *Frontiers in Pediatrics*, 7, May 2019. doi: 10.3389/fped.2019.00206.
- D. F. H. Pheby, D. Araja, U. Berkis, E. Brenna, J. Cullinan, J.-D. d. Korwin, L. Gitto, D. A. Hughes, R. M. Hunter, D. Trepel, and X. Wang-Steverding. The Development of a Consistent Europe-Wide Approach to Investigating the Economic Impact of Myalgic Encephalomyelitis (ME/CFS): A Report from the European Network on ME/CFS (EUROMENE). *Healthcare*, 8(2):88, Apr. 2020. doi: 10.3390/healthcare8020088.
- J. Phelan, A. D. Grabowska, and N. Sepúlveda. A potential antigenic mimicry between viral and human proteins linking Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) with autoimmunity: The case of HPV immunization. *Autoimmunity Reviews*, 19(4):102487, Apr. 2020. doi: 10.1016/j.autrev.2020.102487.
- T. Pietrangelo, R. Mancinelli, L. Toniolo, G. Montanari, J. Vecchiet, G. Fanò, and S. Fulle. Transcription profile analysis of vastus lateralis muscle from patients with chronic fatigue syndrome. *International Journal of Immunopathology and Pharmacology*, 22(3):795–807, 2009. ISSN 0394-6320. doi: 10.1177/039463200902200326.
- U. Radzikowska, M. Ding, G. Tan, D. Zhakparov, Y. Peng, P. Wawrzyniak, M. Wang, S. Li, H. Morita, C. Altunbulakli, M. Reiger, A. U. Neumann, N. Lunjani, C. Traidl-Hoffmann, K. C. Nadeau, L. O'Mahony, C. Akdis, and M. Sokolowska. Distribution of ACE2, CD147, CD26, and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and COVID-19 risk factors. *Allergy*, 75(11): 2829–2845, 2020. ISSN 1398-9995. doi: 10.1111/all.14429. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/all.14429>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/all.14429>.
- R. P. H. Raijmakers, A. F. M. Jansen, S. P. Keijmel, R. t. Horst, M. E. Roerink, B. Novakovic, L. A. B. Joosten, J. W. M. v. d. Meer, M. G. Netea, and C. P. Bleeker-Rovers. A possible role for mitochondrial-derived peptides humanin and MOTS-c in patients with Q fever fatigue syndrome and chronic fatigue syndrome. *Journal of Translational Medicine*, 17(1), May 2019. doi: 10.1186/s12967-019-1906-3.
- S. Rasa, Z. Nora-Krukla, N. Henning, E. Eliassen, E. Shikova, T. Harrer, C. Scheibenbogen, M. Murovska, B. K. Prusty, and the European Network on ME/CFS (EUROMENE). Chronic viral infections in myalgic encephalomyeli-

- tis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 16(1):268, Oct. 2018. ISSN 1479-5876. doi: 10.1186/s12967-018-1644-y. URL <https://doi.org/10.1186/s12967-018-1644-y>.
- S. Y. Rhee, J. Lee, H. Nam, D.-S. Kyoung, D. W. Shin, and D. J. Kim. Effects of a DPP-4 Inhibitor and RAS Blockade on Clinical Outcomes of Patients with Diabetes and COVID-19. *Diabetes & Metabolism Journal*, 45(2):251–259, Mar. 2021. ISSN 2233-6087. doi: 10.4093/dmj.2020.0206.
- M. Rutkowska-Zapała, M. Suski, R. Szatanek, M. Lenart, K. Węglarczyk, R. Olszanecki, T. Grodzicki, M. Strach, J. Gąsowski, and M. Siedlar. Human monocyte subsets exhibit divergent angiotensin I-converting activity. *Clinical and Experimental Immunology*, 181(1):126–132, July 2015. ISSN 0009-9104. doi: 10.1111/cei.12612. URL <https://doi.org/10.1111/cei.12612>.
- T. Saiki, T. Kawai, K. Morita, M. Ohta, T. Saito, K. Rokutan, and N. Ban. Identification of Marker Genes for Differential Diagnosis of Chronic Fatigue Syndrome. *Molecular Medicine*, 14(9-10):599–607, June 2008. doi: 10.2119/2007-00059.saiki.
- A. H. Sawalha, M. Zhao, P. Coit, and Q. Lu. Epigenetic dysregulation of ACE2 and interferon-regulated genes might suggest increased COVID-19 susceptibility and severity in lupus patients. *Clinical Immunology*, 215:108410, June 2020. ISSN 1521-6616. doi: 10.1016/j.clim.2020.108410. URL <https://www.sciencedirect.com/science/article/pii/S1521661620302394>.
- A. J. Scheen. DPP-4 inhibition and COVID-19: From initial concerns to recent expectations. *Diabetes & Metabolism*, 47(2):101213, Mar. 2021. ISSN 1878-1780. doi: 10.1016/j.diabet.2020.11.005.
- K. A. Schlauch, S. F. Khaiboullina, K. L. De Meirleir, S. Rawat, J. Petereit, A. A. Rizvanov, N. Blatt, T. Mijatovic, D. Kulick, A. Palotás, and V. C. Lombardi. Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome. *Translational Psychiatry*, 6(2):e730–e730, Feb. 2016. ISSN 2158-3188. doi: 10.1038/tp.2015.208. URL <https://www.nature.com/articles/tp2015208>. Number: 2 Publisher: Nature Publishing Group.
- J. Słomko, J. L. Newton, S. Kujawski, M. Tafil-Klawe, J. Klawe, D. Staines, S. Marshall-Gradisnik, and P. Zalewski. Prevalence and characteristics of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) in Poland: a cross-sectional study. *BMJ open*, 9(3):e023955, Mar. 2019. ISSN 2044-6055. doi: 10.1136/bmjopen-2018-023955.
- A. K. Smith, H. Fang, T. Whistler, E. R. Unger, and M. S. Rajeevan. Convergent Genomic Studies Identify Association of GRIK2 and NPAS2 with Chronic Fatigue Syndrome. *Neuropsychobiology*, 64(4):183–194, 2011. doi: 10.1159/000326692.
- X. Song, W. Hu, H. Yu, L. Zhao, Y. Zhao, X. Zhao, H.-H. Xue, and Y. Zhao. Little to no expression of angiotensin-converting enzyme-2 on most human peripheral blood immune cells but highly expressed on tissue macrophages. *Cytometry Part A*, 103(2):136–145, 2023. ISSN 1552-4930. doi: 10.1002/cyto.a.24285. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cyto.a.24285>. _eprint:

- <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cyto.a.24285>.
- D. Sulheim, E. Fagermoen, A. Winger, A. M. Andersen, K. Godang, F. Müller, P. C. Rowe, J. P. Saul, E. Skovlund, M. G. Øie, and V. B. Wyller. Disease mechanisms and clonidine treatment in adolescent chronic fatigue syndrome: a combined cross-sectional and randomized clinical trial. *JAMA pediatrics*, 168(4):351–360, Apr. 2014. ISSN 2168-6211. doi: 10.1001/jamapediatrics.2013.4647.
- X.-L. Sun. The role of cell surface sialic acids for SARS-CoV-2 infection. *Glycobiology*, 31(10):1245–1253, Nov. 2021. ISSN 1460-2423. doi: 10.1093/glycob/cwab032.
- W. Sungnak, N. Huang, C. Bécavin, M. Berg, R. Queen, M. Litvinukova, C. Talavera-López, H. Maatz, D. Reichart, F. Sampaziotis, K. B. Worlock, M. Yoshida, J. L. Barnes, and HCA Lung Biological Network. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature Medicine*, 26(5):681–687, May 2020. ISSN 1546-170X. doi: 10.1038/s41591-020-0868-6.
- E. Sweetman, M. Ryan, C. Edgar, A. MacKay, R. Vallings, and W. Tate. Changes in the transcriptome of circulating immune cells of a New Zealand cohort with myalgic encephalomyelitis/chronic fatigue syndrome. *International Journal of Immunopathology and Pharmacology*, 33:2058738418820402, 2019. ISSN 2058-7384. doi: 10.1177/2058738418820402.
- T. Takamura, M. Honda, Y. Sakai, H. Ando, A. Shimizu, T. Ota, M. Sakurai, H. Misu, S. Kurita, N. Matsuzawa-Nagata, M. Uchikata, S. Nakamura, R. Matoba, M. Tanino, K.-i. Matsubara, and S. Kaneko. Gene expression profiles in peripheral blood mononuclear cells reflect the pathophysiology of type 2 diabetes. *Biochemical and Biophysical Research Communications*, 361(2):379–384, Sept. 2007. ISSN 0006-291X. doi: 10.1016/j.bbrc.2007.07.006.
- M. Z. Tay, C. M. Poh, L. Rénia, P. A. MacAry, and L. F. P. Ng. The trinity of COVID-19: immunity, inflammation and intervention. *Nature Reviews Immunology*, 20(6):363–374, June 2020. ISSN 1474-1741. doi: 10.1038/s41577-020-0311-8. URL <https://www.nature.com/articles/s41577-020-0311-8>. Number: 6 Publisher: Nature Publishing Group.
- S. R. Tipnis, N. M. Hooper, R. Hyde, E. Karran, G. Christie, and A. J. Turner. A Human Homolog of Angiotensin-converting Enzyme: CLONING AND FUNCTIONAL EXPRESSION AS A CAPTOPRIL-INSENSITIVE CARBOYPEPTIDASE *. *Journal of Biological Chemistry*, 275(43):33238–33243, Oct. 2000. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M002615200. URL [https://www.jbc.org/article/S0021-9258\(20\)89003-6/abstract](https://www.jbc.org/article/S0021-9258(20)89003-6/abstract). Publisher: Elsevier.
- K. To and A. W. Lo. Exploring the pathogenesis of severe acute respiratory syndrome (SARS): the tissue distribution of the coronavirus (SARS-CoV) and its putative receptor, angiotensin-converting enzyme 2 (ACE2). *The Journal of Pathology*, 203(3):740–743, 2004. ISSN 1096-9896. doi: 10.1002/path.1597. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/path.1597>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/path.1597>.
- M. S. Trivedi, E. Oltra, L. Sarria, N. Rose, V. Beljanski, M. A. Fletcher, N. G. Klimas, and L. Nathanson. Identifi-

- cation of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome-associated DNA methylation patterns. *PLOS ONE*, 13(7):e0201066, July 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0201066. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201066>. Publisher: Public Library of Science.
- N. van Doremalen, K. L. Miazgowicz, S. Milne-Price, T. Bushmaker, S. Robertson, D. Scott, J. Kinne, J. S. McLellan, J. Zhu, and V. J. Munster. Host species restriction of Middle East respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4. *Journal of Virology*, 88(16):9220–9232, Aug. 2014. ISSN 1098-5514. doi: 10.1128/JVI.00676-14.
- N. Vankadari and J. A. Wilce. Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26. *Emerging Microbes & Infections*, 9(1):601–604, 2020. ISSN 2222-1751. doi: 10.1080/22221751.2020.1739565.
- W. C. d. Vega, S. D. Vernon, and P. O. McGowan. DNA Methylation Modifications Associated with Chronic Fatigue Syndrome. *PLoS ONE*, 9(8):e104757, Aug. 2014. doi: 10.1371/journal.pone.0104757.
- W. C. d. Vega, S. Herrera, S. D. Vernon, and P. O. McGowan. Epigenetic modifications and glucocorticoid sensitivity in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *BMC Medical Genomics*, 10(1), Feb. 2017. doi: 10.1186/s12920-017-0248-3.
- P. Verdecchia, C. Cavallini, A. Spanevello, and F. Angeli. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. *European Journal of Internal Medicine*, 76:14–20, June 2020. ISSN 0953-6205. doi: 10.1016/j.ejim.2020.04.037. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7167588/>.
- S. D. Vernon, E. R. Unger, I. M. Dimulescu, M. Rajeevan, and W. C. Reeves. Utility of the Blood for Gene Expression Profiling and Biomarker Discovery in Chronic Fatigue Syndrome. *Disease Markers*, 18(4):193–199, 2002. ISSN 0278-0240. doi: 10.1155/2002/892374. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3851413/>.
- F. Westermeier, M. Bustamante, M. Pavez, L. García, M. Chiong, M. P. Ocaranza, and S. Lavandero. Novel players in cardioprotection: Insulin like growth factor-1, angiotensin-(1–7) and angiotensin-(1–9). *Pharmacological Research*, 101:41–55, Nov. 2015. ISSN 1043-6618. doi: 10.1016/j.phrs.2015.06.018. URL <https://www.sciencedirect.com/science/article/pii/S1043661815001322>.
- T. Whistler, J. F. Jones, E. R. Unger, and S. D. Vernon. Exercise responsive genes measured in peripheral blood of women with Chronic Fatigue Syndrome and matched control subjects. *BMC Physiology*, 5(1):5, 2005. doi: 10.1186/1472-6793-5-5.
- W. Widagdo, N. M. A. Okba, W. Li, A. de Jong, R. L. de Swart, L. Begeman, J. M. A. van den Brand, B.-J. Bosch, and B. L. Haagmans. Species-Specific Colocalization of Middle East Respiratory Syndrome Coronavirus Attachment and Entry Receptors. *Journal of Virology*, 93(16):e00107–19, Aug. 2019. ISSN 1098-5514. doi: 10.1128/JVI.00107-19.

- V. B. Wyller, H. R. Eriksen, and K. Malterud. Can sustained arousal explain the Chronic Fatigue Syndrome? *Behavioral and brain functions: BBF*, 5:10, Feb. 2009. ISSN 1744-9081. doi: 10.1186/1744-9081-5-10.
- Q. Yang, T. A. Hughes, A. Kelkar, X. Yu, K. Cheng, S. Park, W.-C. Huang, J. F. Lovell, and S. Neelamegham. Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-glycan elaboration. *eLife*, 9:e61552, Oct. 2020. ISSN 2050-084X. doi: 10.7554/eLife.61552.

Part III

URGENCY AND

THE CASE OF SARS-COV-2

OMICRON BA.5 IN PORTUGAL

Chapter 9

Risk of BA.5 infection among persons exposed to previous SARS-CoV-2 variants

J. Malato, R.M. Ribeiro, P.P. Leite, P. Casaca, E. Fernandes, C. Antunes, V.R. Fonseca, M.C. Gomes, and L. Graca. Risk of BA.5 infection among persons exposed to previous SARS-CoV-2 variants. *New England Journal of Medicine*. 2022; 387(10):953–954. doi: <https://doi.org/10.1056/NEJMc2209479>.

9.1 Introduction

In recent months, omicron (B.1.1.529) became the dominant variant of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), displaying some degree of immune evasion (Qu et al. 2022). The initial omicron subvariants, BA.1 and BA.2, are being progressively displaced by BA.5 in many countries, possibly owing to greater transmissibility and partial evasion of BA.1- and BA.2-induced immunity (Yu et al. 2022; Cao et al. 2022). The protection afforded by BA.1 against infection by the BA.5 subvariant is critical because adapted vaccines under clinical trials are based on BA.1.

Portugal was one of the first countries affected by a BA.5 predominance. We used the national coronavirus disease 2019 (Covid-19) registry (SINAVE) to calculate the risk of BA.5 infection among persons with documented infection with past variants, including BA.1 and BA.2. The registry includes all reported cases in the country, regardless of clinical presentation.

The national SARS-CoV-2 genetic surveillance identified periods when different variants represented more than 90% of the isolates (Instituto Nacional de Saúde Doutor Ricardo Jorge 2022). We identified all persons who had a first infection in periods of dominance of each

variant, to calculate their infection risk during the period of BA.5 dominance (Figure 9.1A). We pooled BA.1 and BA.2 because of the slow transition between the two subvariants in the population. Finally, we calculated the risk of BA.5 infection for the population that did not have any documented infection before BA.5 dominance (June 1, 2022).

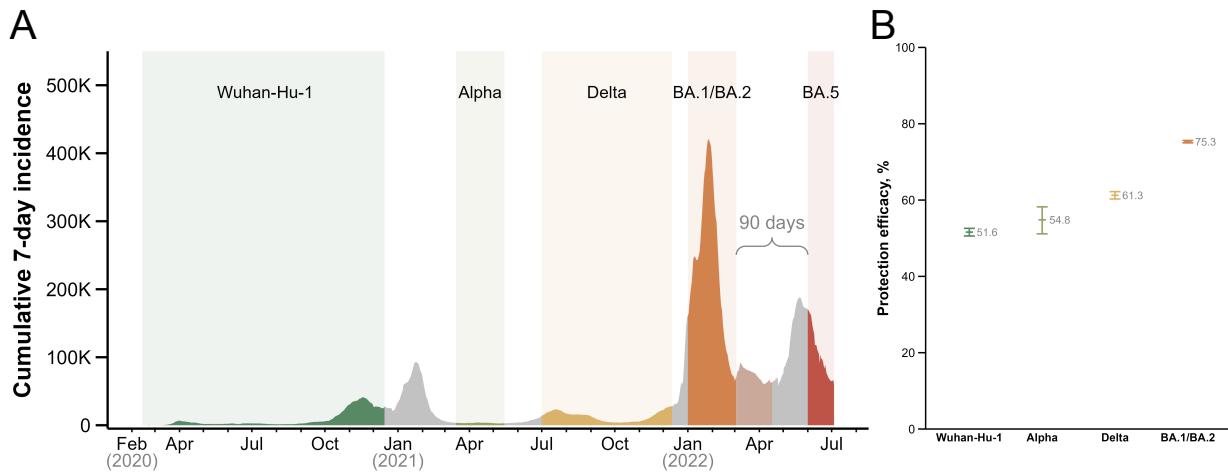


Figure 9.1: Protective effect of previous SARS-CoV-2 infection on infection with the Omicron BA.5 subvariant. As shown in Panel A, we identified the periods (in different colors) when one variant was represented in more than 90% of sample isolates (data from the national SARS-CoV-2 genetic diversity surveillance (Instituto Nacional de Saúde Doutor Ricardo Jorge 2022)). The periods in gray represent times when more than one variant was in circulation. Given the relatively slow transition between dominance by the omicron BA.1 subvariant and dominance by the omicron BA.2 subvariant, we pooled BA.1 and BA.2 in the analysis. We did not include anyone infected in the 90 days before dominance by the omicron BA.5 subvariant. Panel B shows protection efficacy against infection during the period of BA.5 dominance (from June 1, 2022) among persons with one infection in the periods of dominance of different variants, as represented in Panel A, as compared with persons without any documented infection until June 1. Persons with two infections before June 1 were not included in the study. Vertical bars represent 95% confidence intervals.

We found that previous SARS-CoV-2 infection had a protective effect against BA.5 infection (Figure 9.1B and Table 9.1), and this protection was maximal for previous infection with BA.1 or BA.2. These data should be considered in the context of breakthrough infections in a highly vaccinated population, given that in Portugal more than 98% of the study population completed the primary vaccination series before 2022.

The study design cannot eliminate all confounders (see discussion in Section 9.3). In addition, one limitation is the putative effect of immune waning in a population with hybrid immunity (previous infection and vaccination). We found that BA.1 or BA.2 infection in vaccinated

persons provided higher protection against BA.5 than infection with pre-omicron variants, in line with a recent report with a test-negative design (Altarawneh et al. 2022). However, BA.1 or BA.2 infections occurred closer to the period of BA.5 dominance than infections with previous variants. There is a perception that the protection afforded by previous BA.1 or BA.2 infection is very low, given the high number of BA.5 infections among persons with previous BA.1 or BA.2 infection. Our data indicate that this perception is probably a consequence of the larger pool of persons with BA.1 or BA.2 infection than with infection by other subvariants, and it is not supported by the data.

Overall, we found that breakthrough infections with the BA.5 subvariant were less likely among persons with a previous SARS-CoV-2 infection history in a highly vaccinated population, especially for previous BA.1 or BA.2 infection, than among uninfected persons.

Table 9.1: Risk of omicron BA.5 infection according to previous infection history. We included in the study all the population 12 years and older. Under “1st infection” is the number of individuals at risk for a second infection by BA.5 (i.e., all individuals with a second infection before June 1st were excluded). Reinfections were defined as two positive tests by the same individual more than 90 days apart. Note that the risk is dependent on the epidemic situation in Portugal from June 1st to the end of the study (July 4th), affecting all groups equally. RR, relative risk; CI, confidence interval.

	Uninfected on June 1st	1st infection	BA.5 infection	Absolute Risk	RR (95% CI)	Protection Efficacy, % (95% CI)
Uninfected	5 328 287	—	367 783	0.069	—	—
Wuhan-Hu-1	—	267 448	9 031	0.034	0.484 (0.474, 0.494)	51.6 (50.6, 52.6)
Alpha	—	20 004	647	0.032	0.452 (0.418, 0.489)	54.8 (51.1, 58.2)
Delta	—	232 831	6 329	0.027	0.387 (0.378, 0.397)	61.3 (60.3, 62.2)
BA.1/BA.2	—	1 557 635	22 793	0.015	0.247 (0.244, 0.250)	75.3 (75.0, 75.6)

9.2 Methods

9.2.1 Participant selection

The population included in the study was all Portuguese residents aged 12 years and older, obtained from the National Census 2021 (Instituto Nacional de Estatística 2022).

We used the national Covid-19 registry (SINAVE) to obtain information on all notified cases of infection, irrespective of clinical presentation. The “uninfected” population was defined as the population over 12 years of age without a documented infection in the registry. The number of uninfected people in June 1st 2022 (the start of the study period) was 5,328,287, representing 57% of the Portuguese population over 12.

The data available in the national Covid-19 registry (SINAVE) only include cases of tests (PCR tests and rapid antigen tests) performed by healthcare workers in accredited diagnostic facilities. Testing by an accredited facility is a requisite for access to social security compensation for days of isolation—this is a reason for the comprehensiveness of the registry and the exclusive inclusion of validated tests. Only tests performing above the EU-defined minimum for test sensitivity and specificity are used in Portugal. Furthermore, until recently, Portugal had a wide and mandatory testing policy, requiring the presentation of tests for access to several locations, even for vaccinated people (namely, access to entertainment, sports, or healthcare venues).

It is anticipated that the population we classified as “uninfected” contained individuals with a prior unnoticed infection. This issue is discussed below.

We used the national SARS-CoV-2 genetic surveillance database (Instituto Nacional de Saúde Doutor Ricardo Jorge 2022) to identify periods when different variants represented >90% of the sample isolates, as also used in other studies (Altarawneh et al. 2022). With this information, we identified the individuals who were infected in the period of dominance of each variant (Wuhan-Hu-1, Alpha, Delta, BA.1/BA.2, BA.5; Table 9.2). We pooled the BA.1 and BA.2 infections, given the slow transition between the period of dominance of these two subvariants.

Table 9.2: Periods of dominance of the different SARS-CoV-2 variants and omicron subvariants in Portugal. In each of the periods, the variant or subvariant was represented in >90% of sample isolates of the Portuguese residents (data from the national SARS-CoV-2 genetic diversity surveillance (Instituto Nacional de Saúde Doutor Ricardo Jorge 2022)).

Variant or Omicron subvariant	Start	End
Wuhan-Hu-1	14 Feb 20	15 Dec 20
Alpha	15 Mar 21	15 May 21
Delta	01 Jul 21	12 Dec 21
BA.1	01 Jan 22	06 Feb 22
BA.2	27 Mar 22	17 Apr 22
BA.5	01 Jun 22	Present ¹

¹We used a dataset until July 4th 2022.

We excluded from the analyses all individuals who had more than one infection before June 1st (flowchart in Figure 9.2).

Reinfection was defined as two positive tests in the same individual, at least 90 days apart (World Health Organization 2022). Consequently, all cases of infection in the 90 days before the start of BA.5 dominance were not included, as these would not classify as “in risk of reinfection” for the entire duration of the test period under the definition above.

In summary, the population included in the study comprises: (1) All individuals resident in Portugal aged 12 years and older without a documented infection until June 1st 2022; (2) All individuals resident in Portugal aged 12 years and older with a single documented infection before June 1st, when this infection occurred during periods of clear dominance (>90% of cases) of the different variants, but not in the 90 days before June 1st.

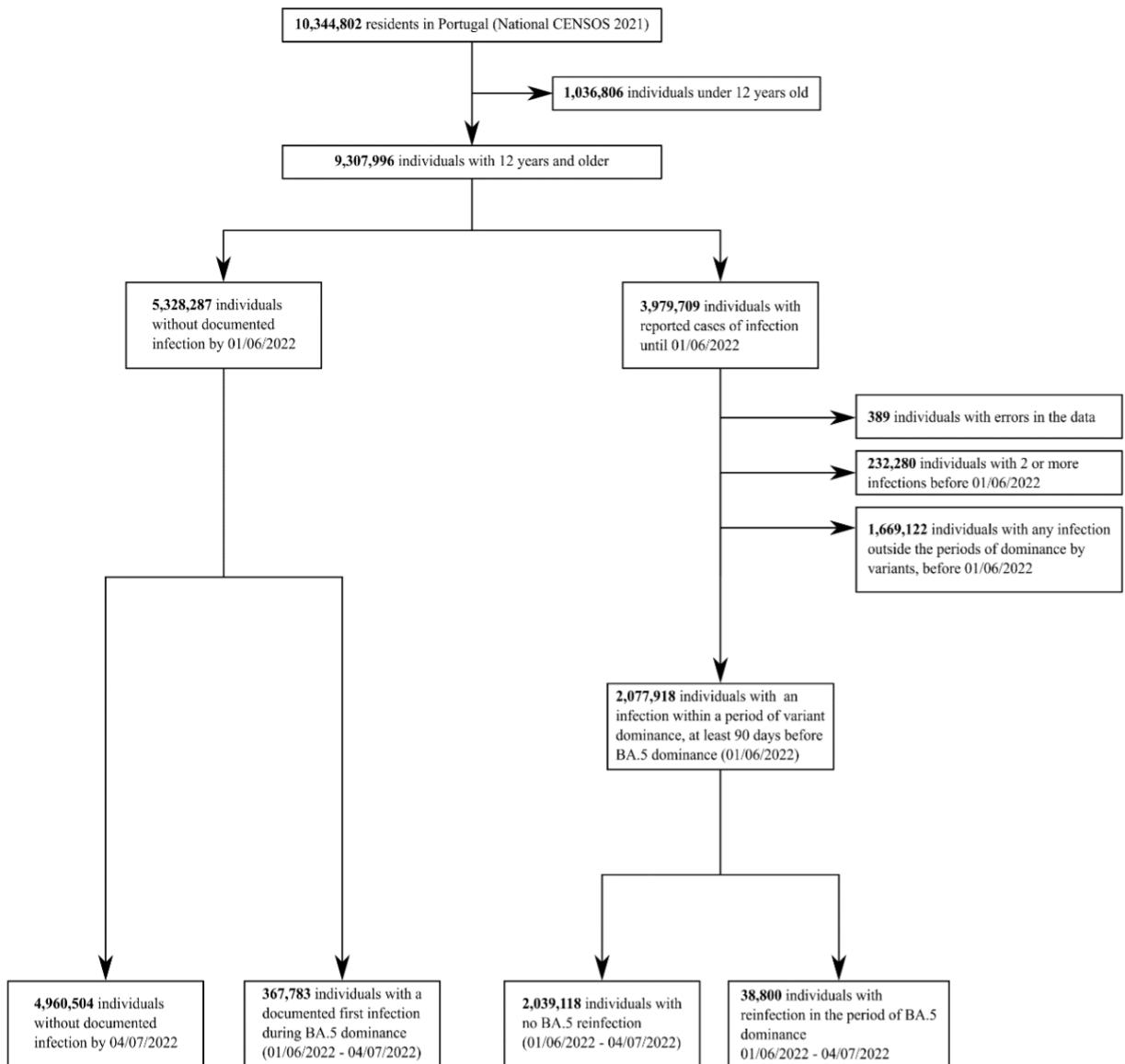


Figure 9.2: Flowchart describing the population selection.

9.2.2 Vaccination coverage

The vaccine coverage with the primary vaccination series in the Portuguese residents over 12 years was >98% by the end of 2021. The primary series of the vaccination campaign used EU/EMA-authorised vaccines: Comirnaty (Pfizer/BioNTech), 69%; Spikevax (Moderna), 12%; Vaxzevria (AstraZeneca), 13%; and Janssen 6%.

At the start of the BA.5 period of dominance (June 1st), the coverage with the first booster was 82%, exclusively using mRNA vaccines (77% Comirnaty and 23% Spikevax). A second booster was not yet in use except for a highly specific (and small) population of patients with severe immunosuppression.

9.2.3 Statistical analysis

We calculated, for all people in the study, the absolute risk of BA.5 infection between June 1st and July 4th, as the number of cases of BA.5 infection in the period under consideration (June 1st to July 4th) over the number of people at risk in each group (uninfected, and single previous infection with Wuhan-Hu-1, Alpha, Delta or BA.1/BA.2). We also used these numbers to estimate the relative risk (RR), that is, the ratio of the probability of BA.5 reinfection in those with a single previous infection (for different variants) and the probability of BA.5 infection within the previously uninfected. Protection efficacy was estimated, in percentage, as $(1 - RR) \times 100\%$. Confidence intervals for the RR were calculated using the normal approximation method.

Further, we performed a sensitivity analysis for the possibility of undiagnosed cases of SARS-CoV-2 in the uninfected population. We used data from the National Serological Panel (from November 2021, assaying the presence of antibodies against the N protein to exclude vaccination seropositivity) to infer that, at that time, there were 29.2% more people who had been infected with SARS-CoV-2 than officially reported (Instituto Nacional de Saúde Doutor Ricardo Jorge 2021). We calculated how many more infections this would correspond to on June 1st, 2022, and removed these infected unreported cases from our uninfected population. We estimated that 2% of these would have also been reinfected with BA.5 during the period

Table 9.3: Risk of omicron BA.5 infection according to previous infection history, considering an estimate of unreported cases of infection. We used data from the national Covid-19 serologic survey that estimated that an additional 29.2% of cases of infection which were not reported (Instituto Nacional de Saúde Doutor Ricardo Jorge 2021). We calculated the overall absolute risk of reinfection during the period of BA.5 dominance for all the infections as 0.020. The values under “Estimated true uninfected” were obtained by subtracting the number of unreported infections and BA.5 reinfections (assuming a 0.020 absolute risk). RR, relative risk; CI, confidence interval.

	Uninfected June 1st 2022	1st infection	BA.5 infection	Absolute risk	RR (95% CI)	Protection Efficacy, % (95% CI)
Infection non-reported	—	(605 944)	(12 208)	0.020	—	—
Estimated true uninfected	4 722 343	—	355 575	0.075	—	—
Wuhan-Hu-1	—	267 448	9 031	0.034	0.443 (0.434, 0.453)	55.7 (54.7, 56.6)
Alpha	—	20 004	647	0.032	0.412 (0.381, 0.445)	58.8 (55.5, 61.9)
Delta	—	232 831	6 329	0.027	0.355 (0.346, 0.364)	64.5 (63.6, 65.4)
BA.1/BA.2	—	1 557 635	22 793	0.015	0.232 (0.229, 0.235)	76.8 (76.5, 77.1)

of interest (being 0.020 the absolute risk for reinfection for the global population with a prior infection). We thus removed that number from the BA.5 infected cases (Table 9.3). With these new values, we recalculated the absolute risk of BA.5 in the previously uninfected and the RR. To further ascertain the sensitivity of our results, we also tested the assumptions that maybe there were only 20% or up to 40% more people who had been infected than officially reported and repeated the calculations described above.

9.3 Supplementary discussion

In this study, we divided the Portuguese epidemic curve into time strata, each characterised by the dominance of one of the SARS-CoV-2 variants or subvariants. We then estimated the risk of infection during the Omicron/BA.5 period, for naive individuals and for those infected in every stratum at least 90 days before the BA.5 period.

We found that, among a highly vaccinated population, the risk of infection by BA.5 was

greater for individuals who had no documented infection, as compared to those who acquired hybrid immunity following a SARS-CoV-2 infection. We also found those infected with the Omicron BA.1/BA.2 subvariants were less at risk of a BA.5 infection than individuals infected with a previous variant. This heightened protection following BA.1/BA.2 infection can be due to the induction of a more effective immune protection towards BA.5 and/or to the shorter time elapsed between infection and exposure to the Omicron BA.5 subvariant.

We used a registry-based study that lacks the precision of a test-negative design (Ayoub et al. 2022). However, the large number of cases that we used, resorting to the entire resident population of Portugal over 12 years old, led to a risk estimate for individuals with prior BA.1/BA.2 infection extremely close to an estimate from Qatar based on a test-negative design (Altarawneh et al. 2022). The effectiveness of protection of BA.1/BA.2 infection against BA.4/BA.5 was calculated in the study from Qatar as 79.7% (95% CI = [74.3, 83.9]), in striking agreement with our data: 75.3% (95% CI = [75.0, 75.6]). Our estimate for effectiveness of protection for individuals infected with SARS-CoV-2 pre-omicron variants was higher than in the study from Qatar, but not too dissimilar from a recent systematic review and meta-analysis that evaluated protection against omicron (Forecasting Team and Lim 2022).

The risk of infection is influenced by the immune status following vaccination and infection and by the time elapsed between the first infection and the BA.5 period due to immune waning. Both issues may be interlocked: in situations where some degree of immune evasion is present, the waning of protection appears to be more rapid. Omicron subvariants are known to differ from pre-omicron variants in having a higher capacity to evade humoral immunity (Arora et al. 2022; Qu et al. 2022; Yu et al. 2022; Cao et al. 2022). It was demonstrated for immune responses directed towards SARS-CoV-2 that the decline of neutralising antibodies over time appears to follow a similar slope (Levin et al. 2021). Hence, partial immune evasion can lead to a lower titre of neutralising antibodies that, with a similar slope of decline, will tend to be faster in reaching a level of neutralising antibodies unable to sustain protection from infection. The relationship between immune evasion and more rapid waning was well illustrated for omicron subvariants (Chemaitelly et al. 2022b). Still, the protection that we are calculating is the protection relevant for public health since infection with previous variants occurred at

different times in the past. In any case, for public health, the critical information is related to the protection against BA.5 observed with individuals infected with previous variants at the time they were in circulation, as it represents the current situation.

In addition, for the interval that we studied (3–5 months after BA.1/BA.2 infection), it is important that the protection conferred by this hybrid immunity remains very significant compared with people vaccinated and without documented infection, or with infection with pre-omicron variants. Indeed, our results are consistent with other studies showing that omicron subvariants are superior to pre-omicron variants in leading to greater protection against other omicron subvariants (Altarawneh et al. 2022; Ayoub et al. 2022; Forecasting Team and Lim 2022; Chemaitley et al. 2022a). Subsequent studies, with longer follow-up, will be necessary to establish the extent of a putative (but likely) effect of immune waning in this hybrid immunity.

A possible source of bias in our results is the presence of undocumented infections among the “uninfected” group of individuals (without a positive test in the registry). A national serologic survey was performed by the National Health Institute Ricardo Jorge before vaccination and at different times following vaccination (using SARS-CoV-2 specific anti-N IgG) (Instituto Nacional de Saúde Doutor Ricardo Jorge 2021). With data from the last national serologic survey, we estimated the population with serologic evidence of SARS-CoV-2 infection that was not detected by testing (i.e., that was not present in the Covid-19 registry) as 29.2% of the notified infections (Instituto Nacional de Saúde Doutor Ricardo Jorge 2021).

We modelled the impact of the presence within the “uninfected” group of unreported infected individuals (in a frequency coherent with the national serologic survey). For this, we calculated the number of all SARS-CoV-2 infections before June 1st and multiplied this number by 0.292 to obtain the number of undetected cases (605,944; Table 9.3). This number of undetected cases was subtracted from the “uninfected”. Next, we removed the reinfections with BA.5 assigned to “uninfected” for the 605,944 individuals that were, in fact, infected. For this, we used the absolute risk calculated for the entire population of individuals that had an infection until 90 days before June 1st 2022—this strategy considers the relative contribution of the number of infections with the distinct variants (and even periods when no variant was dominant) and their risk. The absolute risk was calculated as 0.020. Hence, the number of BA.5

infections that was excluded from the previously uninfected was 12,208. It is not necessary to distribute the new cases of infection for the different groups of variants as these extra cases of infection, and their respective reinfections would not change the relative risks (since this distribution would be proportional across all groups).

Table 9.3 shows that including the unreported positive cases led to a slight increase in the effectiveness of protection (for BA.1/BA.2 from 75% to 77%) (Figure 9.3A). We further explored this sensitivity by considering that unreported cases comprised 20% or 40% of infected cases and redoing the protection calculations (Figure 9.3B). In all situations, consideration of unreported cases lead to an increase in the risk of primary BA.5 infection in the previously uninfected group, and thus an increase in the relative protection of infection with a prior SARS-CoV-2 variant.

We note that we could also consider unreported cases in each of the groups with previous infection with different variants. This would increase by some percentage the number of first infections with Wuhan-Hu-1, Alpha, Delta, and BA.1/BA.2, leading to a lower absolute risk of secondary infection with BA.5 in these groups. Relative to no previous infection, this would increase the protection of having had an infection before. Finally, the number of cases of BA.5 could also be larger, because of unreported cases. However, any increase in these numbers would affect all groups (uninfected and previously infected) in a similar proportion, because we are considering BA.5 infections over the same time frame in all cases. Thus, putative unreported cases in BA.5 would have minimal impact on the relative protection of the previous infections. We assumed that the frequency of unreported cases of infection remained relatively constant over the periods of dominance with different variants. It is very likely that the frequency of unreported cases increased during the period of BA.5 dominance, given the change in the testing policy in Portugal. However, for the reasons given above, the change in this period is expected to affect all groups similarly. It is anticipated that more subtle changes in the frequency of unreported cases occurred over the previous periods, given the maintenance of public health policies, namely regarding testing. This is also confirmed by the relatively constant frequency of estimated unreported cases in the three periods evaluated by the national Covid-19 serologic survey (Instituto Nacional de Saúde Doutor Ricardo Jorge 2021).

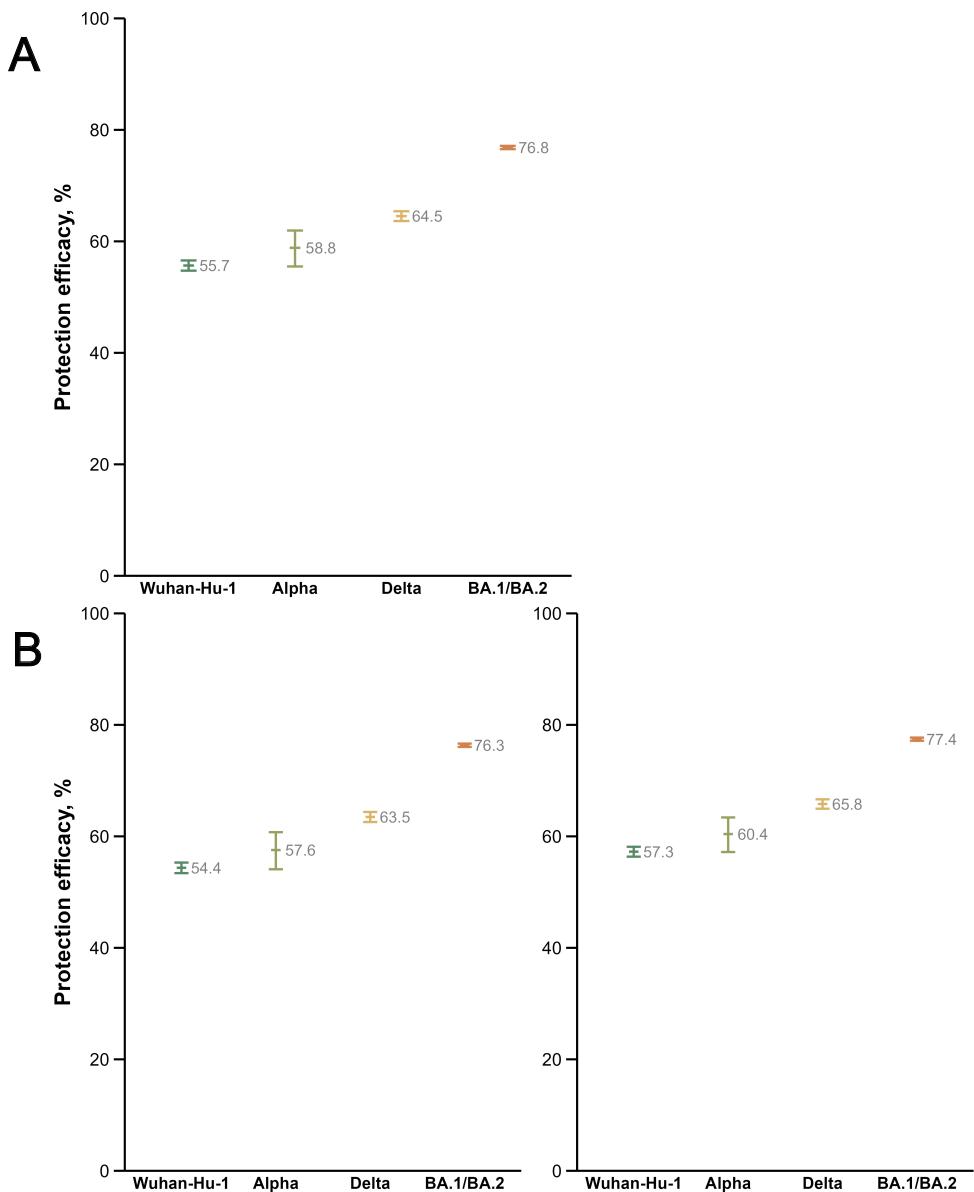


Figure 9.3: Estimates of the impact of unreported cases of infection among the population absent from the national Covid-19 registry. The most recent national Covid-19 serologic survey estimated that an additional 29.2% of cases of infection were not notified (Instituto Nacional de Saúde Doutor Ricardo Jorge 2021). This figure was calculated based on the seroprevalence of SARS-CoV-2 anti-N IgG in the population. (A) We calculated the protection efficacy for prior infection with different SARS-CoV-2 variants in relation to the uninfected population, following correction for the estimated unreported cases of infection (removed from the “uninfected” group). (B) To further explore the method’s sensitivity, we calculated protection efficacy for scenarios where the unreported cases were 20% (left) and 40% (right) of all notified infections.

The sensitivity analysis with a variable percentage of false negatives confirmed that the main conclusions are similar despite this issue. The experimental design does not account, however, for characteristics that may differ between infected and uninfected: some high-risk occupations, lifestyles, and living environments may place individuals consistently at higher risk, that were likely to be infected earlier in the pandemic. Nevertheless, our data is coherent with several other reports consistently showing greater protection of omicron subvariants towards reinfection with other omicron subvariants (Altarawneh et al. 2022; Chemaitlely et al. 2022a).

The popular perception that BA.1/BA.2 infection cannot protect against BA.5 reinfection, a consequence of the observation that many people infected in the first two months of 2022 (BA.1/BA.2 dominance) were reinfected by BA.5 is a perception error. Probably explained by the very high number of infections caused by BA.1/BA.2. In Portugal, there were 1.97 million documented infections in January-February/2022 alone, which compares with 1.42 million during the entire epidemic before.

Our results, together with a similar report (Altarawneh et al. 2022), show that infection with BA.1/BA.2 of a population mostly vaccinated provided significant protection against BA.5 reinfection. At a time when BA.1-based adapted vaccines are under clinical development, this information is of great importance.

Bibliography

- H. N. Altarawneh, H. Chemaitlely, H. H. Ayoub, M. R. Hasan, P. Coyle, H. M. Yassine, H. A. Al-Khatib, F. M. Benslimane, Z. Al-Kanaani, E. Al-Kuwari, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, H. F. Abdul-Rahim, G. K. Nasrallah, M. G. Al-Kuwari, A. A. Butt, H. E. Al-Romaihi, M. H. Al-Thani, A. Al-Khal, R. Bertolini, P. Tang, and L. J. Abu-Raddad. Protection of SARS-CoV-2 natural infection against reinfection with the Omicron BA.4 or BA.5 subvariants, July 2022. URL <https://www.medrxiv.org/content/10.1101/2022.07.11.22277448v1>. Pages: 2022.07.11.22277448.
- P. Arora, A. Kempf, I. Nehlmeier, S. R. Schulz, A. Cossman, M. V. Stankov, H.-M. Jäck, G. M. N. Behrens, S. Pöhlmann, and M. Hoffmann. Augmented neutralisation resistance of emerging omicron subvariants

- BA.2.12.1, BA.4, and BA.5. *The Lancet Infectious diseases*, 22(8):1117–1118, Aug. 2022. ISSN 1474-4457. doi: 10.1016/s1473-3099(22)00422-4. URL <https://europepmc.org/articles/PMC9239574>.
- H. H. Ayoub, M. Tomy, H. Chemaitlely, H. N. Altarawneh, P. Coyle, P. Tang, M. R. Hasan, Z. A. Kanaani, E. A. Kuwari, A. A. Butt, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, G. K. Nasrallah, F. M. Benslimane, H. A. A. Khatib, H. M. Yassine, M. G. A. Kuwari, H. E. A. Romaihi, H. F. Abdul-Rahim, M. H. Al-Thani, A. A. Khal, R. Bertollini, and L. J. Abu-Raddad. Estimating protection afforded by prior infection in preventing reinfection: Applying the test-negative study design, Jan. 2022. URL <https://www.medrxiv.org/content/10.1101/2022.01.02.22268622v1>. Pages: 2022.01.02.22268622.
- Y. Cao, A. Yisimayi, F. Jian, W. Song, T. Xiao, L. Wang, S. Du, J. Wang, Q. Li, X. Chen, Y. Yu, P. Wang, Z. Zhang, P. Liu, R. An, X. Hao, Y. Wang, J. Wang, R. Feng, H. Sun, L. Zhao, W. Zhang, D. Zhao, J. Zheng, L. Yu, C. Li, N. Zhang, R. Wang, X. Niu, S. Yang, X. Song, Y. Chai, Y. Hu, Y. Shi, L. Zheng, Z. Li, Q. Gu, F. Shao, W. Huang, R. Jin, Z. Shen, Y. Wang, X. Wang, J. Xiao, and X. S. Xie. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature*, 608(7923):593–602, Aug. 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04980-y. URL <https://www.nature.com/articles/s41586-022-04980-y>. Number: 7923 Publisher: Nature Publishing Group.
- H. Chemaitlely, H. H. Ayoub, P. Coyle, P. Tang, H. M. Yassine, H. A. Al-Khatib, M. K. Smatti, M. R. Hasan, Z. Al-Kanaani, E. Al-Kuwari, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, H. F. Abdul-Rahim, G. K. Nasrallah, M. G. Al-Kuwari, A. A. Butt, H. E. Al-Romaihi, M. H. Al-Thani, A. Al-Khal, R. Bertollini, and L. J. Abu-Raddad. Protection of Omicron sub-lineage infection against reinfection with another Omicron sub-lineage, Feb. 2022a. URL <https://www.medrxiv.org/content/10.1101/2022.02.24.22271440v1>. Pages: 2022.02.24.22271440.
- H. Chemaitlely, N. Nagelkerke, H. H. Ayoub, P. Coyle, P. Tang, H. M. Yassine, H. A. Al-Khatib, M. K. Smatti, M. R. Hasan, Z. Al-Kanaani, E. Al-Kuwari, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, H. F. Abdul-Rahim, G. K. Nasrallah, M. G. Al-Kuwari, A. A. Butt, H. E. Al-Romaihi, M. H. Al-Thani, A. Al-Khal, R. Bertollini, and L. J. Abu-Raddad. Duration of immune protection of SARS-CoV-2 natural infection against reinfection in Qatar, July 2022b. URL <https://www.medrxiv.org/content/10.1101/2022.07.06.22277306v1>. Pages: 2022.07.06.22277306.
- C.-. Forecasting Team and S. S. Lim. Past SARS-CoV-2 Infection Protection Against Reinfection: A Systematic Review and Meta-Analysis, July 2022. URL <https://papers.ssrn.com/abstract=4155225>.
- Instituto Nacional de Estatística. Statistics Portugal, 2022. URL <https://www.ine.pt/>.
- Instituto Nacional de Saúde Doutor Ricardo Jorge. National COVID-19 Serologic Survey in Portugal. Technical report, INSA, Lisbon, 2021. URL http://repositorio.insa.pt/bitstream/10400.18/7828/1/ISN_COVID19_relatório_3F_final.pdf.
- Instituto Nacional de Saúde Doutor Ricardo Jorge. Genetic diversity of the novel coronavirus SARS-CoV-2

- (COVID-19) in Portugal, 2022. URL <https://insaflu.insa.pt/covid19>.
- E. G. Levin, Y. Lustig, C. Cohen, R. Fluss, V. Indenbaum, S. Amit, R. Doolman, K. Asraf, E. Mendelson, A. Ziv, C. Rubin, L. Freedman, Y. Kreiss, and G. Regev-Yochay. Waning Immune Humoral Response to BNT162b2 Covid-19 Vaccine over 6 Months. *The New England Journal of Medicine*, 385(24):e84, Dec. 2021. ISSN 1533-4406. doi: 10.1056/NEJMoa2114583.
- P. Qu, J. Faraone, J. P. Evans, X. Zou, Y.-M. Zheng, C. Carlin, J. S. Bednash, G. Lozanski, R. K. Mallampalli, L. J. Saif, E. M. Oltz, P. J. Mohler, R. J. Gumina, and S.-L. Liu. Neutralization of the SARS-CoV-2 Omicron BA.4/5 and BA.2.12.1 Subvariants. *New England Journal of Medicine*, 386(26):2526–2528, June 2022. ISSN 0028-4793. doi: 10.1056/NEJMc2206725. URL <https://doi.org/10.1056/NEJMc2206725>. Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMc2206725>.
- World Health Organization. Public health surveillance for COVID-19: interim guidance. Interim guidance WHO/2019-nCoV/SurveillanceGuidance/2022.2, July 2022. URL <https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-SurveillanceGuidance-2022.2>.
- J. Yu, A.-r. Y. Collier, M. Rowe, F. Mardas, J. D. Ventura, H. Wan, J. Miller, O. Powers, B. Chung, M. Siamatu, N. P. Hachmann, N. Surve, F. Nampanya, A. Chandrashekhar, and D. H. Barouch. Neutralization of the SARS-CoV-2 Omicron BA.1 and BA.2 Variants. *New England Journal of Medicine*, 386(16):1579–1580, Apr. 2022. ISSN 0028-4793. doi: 10.1056/NEJMc2201849. URL <https://doi.org/10.1056/NEJMc2201849>. Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMc2201849>.

Chapter 10

Stability of hybrid versus vaccine immunity against BA.5 infection over 8 months

J. Malato, R.M. Ribeiro, E. Fernandes, P.P. Leite, P. Casaca, C. Antunes, V.R. Fonseca, M.C. Gomes, and L. Graca. Stability of hybrid versus vaccine immunity against BA.5 infection over 8 months. *The Lancet Infectious Diseases*. 2023; 23(2):148–150. doi: [https://doi.org/10.1016/S1473-3099\(22\)00833-7](https://doi.org/10.1016/S1473-3099(22)00833-7).

10.1 Introduction

The coverage of SARS-CoV-2 vaccination in large parts of the world, together with the high number of breakthrough infections, especially following the emergence of Omicron subvariants, makes hybrid immunity (resulting from vaccine and infection) common. Hybrid immunity, particularly after BA.1 or BA.2 infection, confers substantial protection against the BA.5 infection (Malato et al. 2022; Altarawneh et al. 2022; Hansen et al. 2023). However, although the waning of protection afforded by natural infection in non-vaccinated individuals or by vaccination has been well documented (Chemaitelly et al. 2021; Goldberg et al. 2022) the stability of hybrid immunity, specifically against the BA.5 subvariant, now dominant in many countries, has not been thoroughly addressed.

We used the Portuguese Covid-19 registry (SINAVE), which includes all notified cases of infection in the country on the basis of an official positive test and irrespective of clinical presentation, to investigate the risk of reinfection with BA.5 in a highly vaccinated population previously infected with BA.1 or BA.2 subvariants. We included the population aged 12 years or older, for whom the vaccination coverage was greater than 98% at the end of 2021 (additional

description in Section 10.2). The registry is very comprehensive due to legal requirements for compensation payment during mandatory isolation. We include infection data from the start of the pandemic until Sept 14, 2022.

We identified the periods of dominance (over 90% of the isolates) of BA.1 and BA.2 (Jan 1–Apr 17, 2022) and BA.5 infections (June 1–Sept 14, 2022) using the national SARS-CoV-2 genetic surveillance data and divided those periods into 15 day intervals (Figure 10.1A). We then calculated the relative risk (RR) of BA.5 infection in each interval for individuals that had the first infection during each BA.1 and BA.2 dominance subinterval, compared with individuals also vaccinated but without any previous documented infection. Reinfection was defined as two positive tests in the same individual, at least 90 days apart. We found that the RR increased from around 0.06 to around 0.35 between 3 months and 8 months post BA.1 or BA.2 infection (Figure 10.1B). Indeed, the RR initially increases rapidly, then more slowly, stabilising at around 0.37.

The present authors previously assessed the effect of unreported infections in the calculation of RR (Malato et al. 2022). Here, we mitigate this effect by calculating the RR for the same interval of BA.5 infection for individuals infected by BA.1 or BA.2 in distinct periods, thus with a constant frequency of unreported infections. In any case, our findings are consistent throughout the entire dataset. Our registry-based dataset includes data on essentially the whole population, but only includes data on positive tests. This feature precludes using a test-negative study design, which has been successfully used in other studies of RR (Altarawneh et al. 2022; Ayoub et al. 2022). However, previous reports indicate that the estimates of protection efficacy using the national registry are well aligned with studies that used a test-negative design, albeit in a different population (Malato et al. 2022; Altarawneh et al. 2022). Studies since 2021 have made clear the potential for immune imprinting, with one study (Chemaitlely et al. 2022) suggesting that protection against infection waned after the booster (relative to primary series). In our study, essentially the whole population is vaccinated with the booster dose, and therefore we cannot distinguish effects of booster versus primary series. However, our results of increased protection with hybrid immunity versus vaccine immunity, agrees with the overall conclusion of that study that “imprinting effects are unlikely to negate the overall public health value of

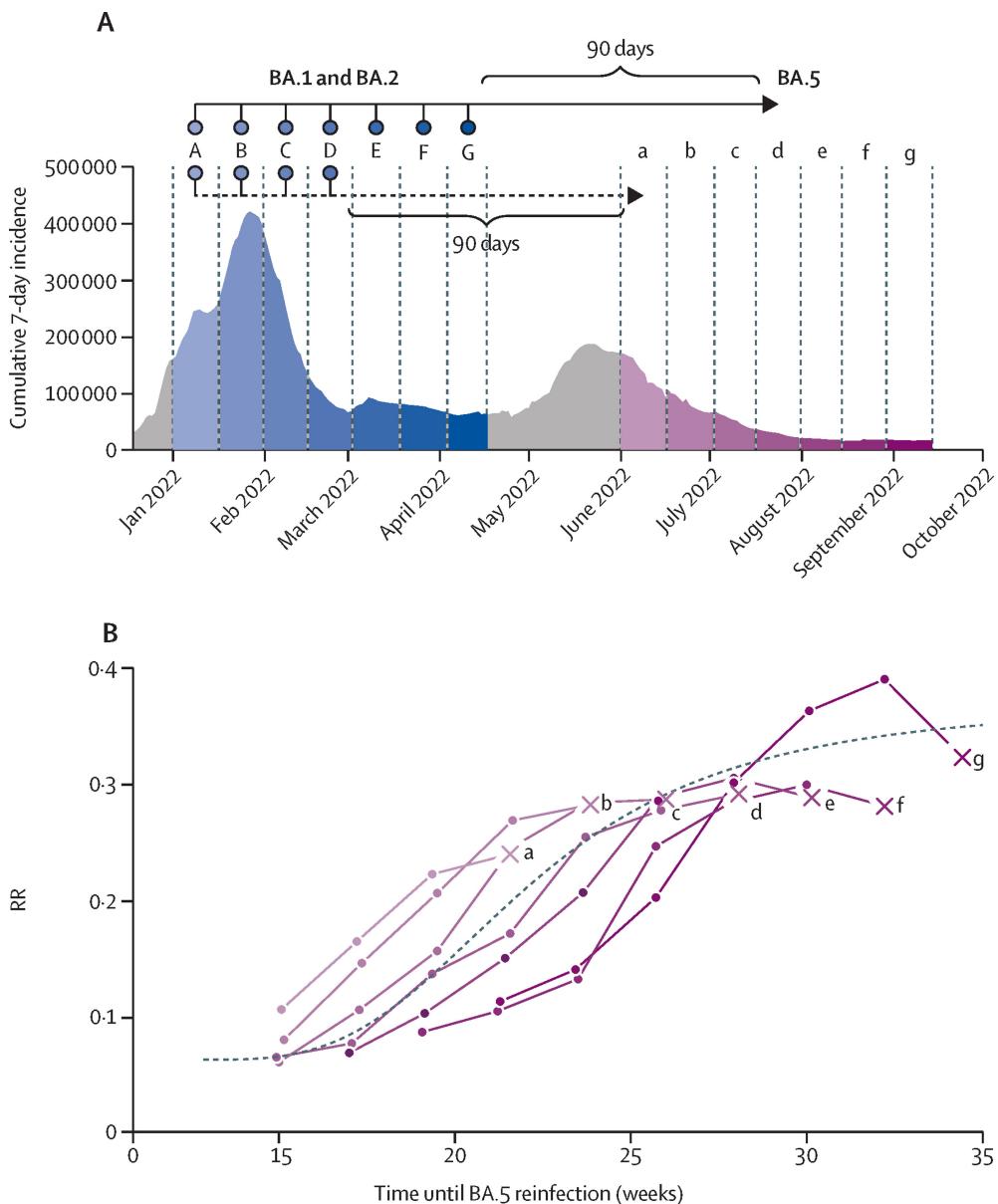


Figure 10.1: Stability of hybrid immunity protection against BA.5 infection following infection with BA.1 or BA.2 subvariants. (A) Incidence of documented SARS-CoV-2 infection overlaid with the period of dominance of the BA.1 and BA.2 variants, Jan 1–Apr 14, 2022, divided into 15-day sub-intervals (shades of blue), and the period of dominance of the BA.5 variant, Jun 1–Sep 14, 2022, also divided into 15-day sub-intervals (shades of purple). Two illustrative comparisons are represented. In period d (BA.5 dominance), the risk of infection was compared between individuals with a first documented infection in one of the seven subintervals of BA.1 and BA.2 dominance (A–G), represented with the solid arrow. In the second example with the dashed arrow, in period a of BA.5 dominance, the risk of infection was compared between individuals with a first documented infection in the first four periods of BA.1 and BA.2 dominance (A–D), as reinfections were only considered 90 days following the first infection. (B) RR of reinfection versus first infection in each subinterval of the period of BA.5 dominance (curves a–g, corresponding with the periods of the same letter as in (A) over time since the first infection. The increase in risk is well described by a saturating function (Equation (10.1)) as represented by the fitted line (dashed, black). RR, relative risk.

booster vaccinations” (Chemaitlely et al. 2022).

This study shows that hybrid immunity following infection with Omicron BA.1 or BA.2 when compared with vaccine-only immunity leads to substantially increased protection against BA.5 reinfection for up to 8 months.

10.2 Methods

10.2.1 Participant selection

We followed an approach similar to what we reported in a previous registry-based study (Malato et al. 2022). The population included in the study comprises: (1) All individuals resident in Portugal aged 12 years and older without a documented infection until the start of the follow-up period, which is June 1st to September 14th, 2022; and (2) All individuals resident in Portugal aged 12 years and older with a single documented infection between January 1st, 2022 (the initial period of dominance of BA.1) up to 90 days before the follow-up period and no other previous infection (see flowchart, Figure 10.2).

We used the national Covid-19 registry (SINAVE) to obtain information on all notified cases of infection, irrespective of clinical presentation. The “uninfected” population was defined as the population over 12 years of age without a documented infection in the registry at any time. The number of uninfected people on June 1st 2022 (the start of the BA.5 dominance period) was 5,325,097, representing 57% of the Portuguese population over 12 (data from the National Census 20212).

The data available in SINAVE include cases of positive tests (PCR tests and rapid antigen tests) performed by healthcare workers in accredited diagnostic facilities. Testing by an accredited facility was a requisite for access to social security compensation for days of isolation—this is a reason for the comprehensiveness of the registry and the inclusion of only validated tests. Only tests performing above the EU-defined minimum for test sensitivity and specificity are used in Portugal. The testing policy officially changed on October 1st, 2022 but with some relaxation of the implementation in the period following their announcement just before the

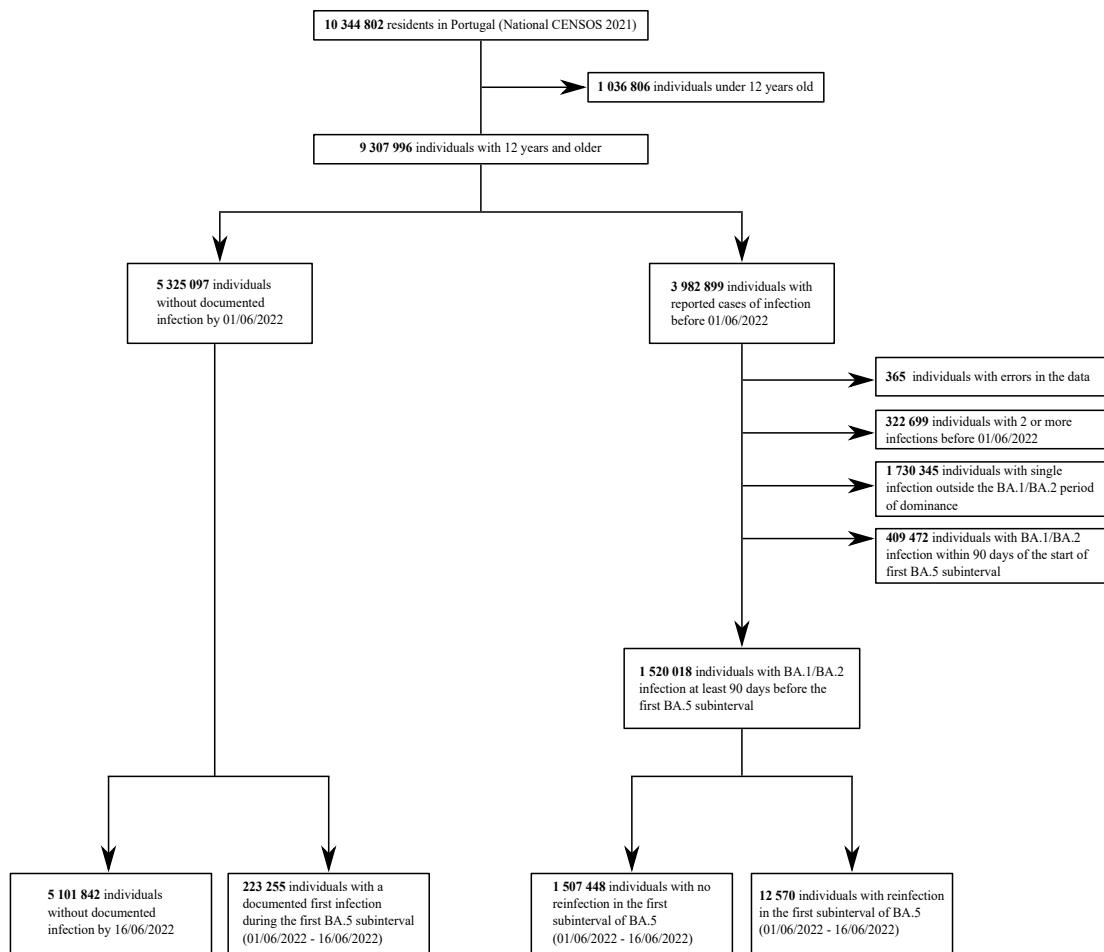


Figure 10.2: Flowchart describing the population selection. Representative flowchart representing the selection for the first subinterval of BA.5 reinfection. For later BA.5 intervals, the 90-day period prior to the start of the interval allowed inclusion of a further subperiod of BA.1/BA.2 (Figure 10.1A). Note: the date format is day/month/year.

official date. Therefore, we considered infections until September 14th, 2022 as the period with consistent implementation of comprehensive testing policies.

We used the national SARS-CoV-2 genetic surveillance database (Instituto Nacional de Saúde Doutor Ricardo Jorge 2022) to identify periods when one variant represented >90% of the sample isolates, as also defined and used in other studies (Malato et al. 2022; Altarawneh et al. 2022). We assigned infected individuals to the variants' dominance periods and excluded all individuals who had more than one infection before the study period (Figure 10.2). We pooled BA.1 and BA.2 infections, given the slow transition between the period of domi-

nance of these two subvariants. With this approach, we identified the periods of dominance of BA.1/BA.2 (January 1st to April 17th, 2022) and of BA.5 infections (June 1st to September 14th, 2022). We then divided those periods of dominance into approximately 15 days intervals (as seen in Figure 10.1A). Coincidentally, the period of BA.1/BA.2 dominance was divided into 7 sub-intervals, and the period of BA.5 was also divided into 7 sub-intervals (Table 10.1 and Table 10.2).

Table 10.1: Subintervals of BA.1/BA.2 and BA.5 dominance used in the study. Both periods of BA.1/BA.2 and BA.5 dominance were split in seven periods with approximately 2 weeks each. The fact that the two subvariants have the same number of intervals is a coincidence.

Variant	Interval	Start date	End date	Days
BA1/BA.2	A	01/01/22	16/01/22	15
	B	17/01/22	31/01/22	14
	C	01/02/22	15/02/22	14
	D	16/02/22	02/03/22	14
	E	03/03/22	18/03/22	15
	F	19/03/22	03/04/22	15
	G	04/04/22	17/04/22	13
BA.5	a	01/06/22	16/06/22	15
	b	17/06/22	02/07/22	15
	c	03/07/22	16/07/22	13
	d	17/07/22	31/07/22	14
	e	01/08/22	14/08/22	13
	f	15/08/22	29/08/22	14
	g	30/08/22	14/09/22	15

Reinfection was defined as two positive tests in the same individual, at least 90 days apart (World Health Organizaton 2022). Consequently, all cases of infection in the 90 days before the start of each sub-interval were not included, as these would not classify as “at risk of reinfection” for the entire duration of the subinterval under the definition above.

Given the high vaccine coverage, we compared one population with “hybrid immunity” (vaccination + infection with BA.1/BA.2) with a group of vaccinated individuals without

Table 10.2: Risk of omicron BA.5 infection at different intervals for individuals infected with BA.1/BA.2 in specific periods. We included in the study the population 12 years and older. Under “1st infection” is the number of individuals at risk for a second infection by BA.5 in the respective interval (i.e., without a second infection until that time). Note that the risk may depend on the epidemic situation and may differ in the BA.5 periods. RR, relative risk; CI, confidence interval.

BA.5 (a)	Date interval	Uninfected on Jun 1st 2022	1st infection	BA.5 infection	Absolute risk	RR (95% CI)	Protection efficacy, % (95% CI)
Uninfected		5325097	—	223255	0.042	—	—
BA.1/BA.2 A	01/01/22–16/01/22	—	421200	4130	0.010	0.240 (0.233, 0.248)	75.97 (75.22, 76.68)
B	17/01/22–31/01/22	—	620102	5488	0.009	0.223 (0.217, 0.229)	77.69 (77.09, 78.26)
C	01/02/22–15/02/22	—	341779	2328	0.007	0.165 (0.159, 0.172)	83.46 (82.77, 84.11)
D	16/02/22–02/03/22	—	136937	624	0.005	0.107 (0.099, 0.116)	89.29 (88.41, 90.10)
BA.5 (b)	Date interval	Uninfected on Jun 17th 2022	1st infection	BA.5 infection	Absolute risk	RR (95% CI)	Protection efficacy, % (95% CI)
Uninfected	—	5101842	—	135093	0.026	—	—
BA.1/BA.2 A	01/01/22–16/01/22	—	417070	3004	0.007	0.283 (0.273, 0.293)	71.73 (70.71, 72.72)
B	17/01/22–31/01/22	—	614614	4103	0.007	0.269 (0.261, 0.278)	73.07 (72.25, 73.87)
C	01/02/22–15/02/22	—	339451	1803	0.005	0.207 (0.198, 0.217)	79.31 (78.34, 80.24)
D	16/02/22–02/03/22	—	136313	530	0.004	0.147 (0.135, 0.160)	85.31 (84.01, 86.51)
E	03/03/22–18/03/22	—	166811	356	0.002	0.081 (0.073, 0.090)	91.89 (91.01, 92.69)
BA.5 (c)	Date interval	Uninfected on Jul 3rd 2022	1st infection	BA.5 infection	Absolute risk	RR (95% CI)	Protection efficacy, % (95% CI)
Uninfected	—	4966749	—	70757	0.014	—	—
BA.1/BA.2 A	01/01/22–16/01/22	—	414066	1616	0.004	0.287 (0.274, 0.302)	71.26 (69.84, 72.62)
B	17/01/22–31/01/22	—	610511	2295	0.004	0.284 (0.273, 0.296)	71.57 (70.40, 72.69)
C	01/02/22–15/02/22	—	337648	1140	0.003	0.247 (0.233, 0.261)	75.34 (73.88, 76.73)
D	16/02/22–02/03/22	—	135783	301	0.002	0.157 (0.141, 0.176)	84.27 (82.39, 85.95)
E	03/03/22–18/03/22	—	166455	249	0.001	0.107 (0.094, 0.121)	89.32 (87.91, 90.57)
F	19/03/22–03/04/22	—	133119	116	0.001	0.062 (0.052, 0.074)	93.81 (92.58, 94.84)
BA.5 (d)	Date interval	Uninfected on Jul 17th 2022	1st infection	BA.5 infection	Absolute risk	RR (95% CI)	Protection efficacy, % (95% CI)
Uninfected	—	4895992	—	41767	0.009	—	—
BA.1/BA.2 A	01/01/22–16/01/22	—	412450	975	0.002	0.292 (0.274, 0.311)	70.81 (68.94, 72.57)
B	17/01/22–31/01/22	—	608216	1331	0.002	0.278 (0.264, 0.293)	72.21 (70.70, 73.64)
C	01/02/22–15/02/22	—	336508	701	0.002	0.255 (0.237, 0.275)	74.49 (72.54, 76.30)
D	16/02/22–02/03/22	—	135482	196	0.001	0.172 (0.150, 0.198)	82.77 (80.19, 85.02)
E	03/03/22–18/03/22	—	166206	191	0.001	0.138 (0.119, 0.159)	86.23 (84.14, 88.05)
F	19/03/22–03/04/22	—	133003	87	0.001	0.078 (0.063, 0.096)	92.20 (90.38, 93.68)
G	04/04/22–17/04/22	—	109542	61	0.001	0.066 (0.051, 0.085)	93.39 (91.50, 94.85)
BA.5 (e)	Date interval	Uninfected on Aug 1st 2022	1st infection	BA.5 infection	Absolute risk	RR (95% CI)	Protection efficacy, % (95% CI)
Uninfected	—	4854225	—	28337	0.006	—	—
BA.1/BA.2 A	01/01/22–16/01/22	—	411475	657	0.002	0.289 (0.268, 0.312)	71.12 (68.85, 73.22)
B	17/01/22–31/01/22	—	606885	1000	0.002	0.306 (0.288, 0.325)	69.44 (67.52, 71.25)
C	01/02/22–15/02/22	—	335807	542	0.002	0.289 (0.266, 0.314)	71.11 (68.59, 73.42)
D	16/02/22–02/03/22	—	135286	161	0.001	0.207 (0.178, 0.242)	79.26 (75.80, 82.22)
E	03/03/22–18/03/22	—	166015	143	0.001	0.151 (0.128, 0.178)	84.89 (82.20, 87.17)
F	19/03/22–03/04/22	—	132916	79	0.001	0.104 (0.083, 0.129)	89.62 (87.06, 91.67)
G	04/04/22–17/04/22	—	109481	44	0.0004	0.070 (0.052, 0.094)	93.01 (90.61, 94.80)
BA.5 (f)	Date interval	Uninfected on Aug 15th 2022	1st infection	BA.5 infection	Absolute risk	RR (95% CI)	Protection efficacy, % (95% CI)
Uninfected	—	4825888	—	29100	0.006	—	—
BA.1/BA.2 A	01/01/22–16/01/22	—	410818	660	0.002	0.282 (0.261, 0.304)	71.85 (69.64, 73.89)
B	17/01/22–31/01/22	—	605885	1011	0.002	0.300 (0.282, 0.319)	70.02 (68.14, 71.78)
C	01/02/22–15/02/22	—	335265	551	0.002	0.285 (0.262, 0.309)	71.51 (69.05, 73.77)
D	16/02/22–02/03/22	—	135125	198	0.001	0.247 (0.215, 0.284)	75.30 (71.62, 78.50)
E	03/03/22–18/03/22	—	165872	130	0.001	0.133 (0.112, 0.158)	86.68 (84.19, 88.78)
F	19/03/22–03/04/22	—	132837	83	0.001	0.106 (0.085, 0.131)	89.44 (86.91, 91.48)
G	04/04/22–17/04/22	—	109437	57	0.001	0.088 (0.068, 0.114)	91.23 (88.63, 93.23)
BA.5 (g)	Date interval	Uninfected on Aug 30th 2022	1st infection	BA.5 infection	Absolute risk	RR (95% CI)	Protection efficacy, % (95% CI)
Uninfected	—	4796788	—	28125	0.006	—	—
BA.1/BA.2 A	01/01/22–16/01/22	—	410158	738	0.002	0.323 (0.301, 0.347)	67.66 (65.27, 69.89)
B	17/01/22–31/01/22	—	604874	1290	0.002	0.390 (0.370, 0.412)	60.97 (58.82, 63.00)
C	01/02/22–15/02/22	—	334714	685	0.002	0.363 (0.337, 0.391)	63.68 (60.89, 66.27)
D	16/02/22–02/03/22	—	134927	235	0.002	0.302 (0.266, 0.343)	69.83 (65.73, 73.44)
E	03/03/22–18/03/22	—	165742	193	0.001	0.203 (0.176, 0.234)	79.69 (76.62, 82.35)
F	19/03/22–03/04/22	—	132754	108	0.001	0.141 (0.117, 0.171)	85.87 (82.94, 88.29)
G	04/04/22–17/04/22	—	109380	72	0.001	0.114 (0.090, 0.144)	88.60 (85.65, 90.95)

infection. In other words, we assessed the stability of hybrid immunity (induced with Omicron BA.1/BA.2 infection + vaccines) versus vaccine immunity. The change in the relative risk that we report, translates the waning of such “additional” protection afforded by natural infection of the vaccinated individuals.

It is possible that the population we classified as “uninfected” contains individuals with a prior unnoticed infection (i.e., asymptomatic infection). In a previous publication, we have shown that considering a proportion (20%–40%) of unreported infections within the “uninfected” group (in line with data from the national serologic survey (Instituto Nacional de Saúde Doutor Ricardo Jorge 2021)) only has the effect of decreasing slightly the relative risk of BA.5 re-infection in each sub-interval (Malato et al. 2022), without changing our overall results. This is intuitive because if more people were infected (and moved out of the “uninfected group”), that inflates the absolute risk of first infection, and thus the relative risk of a second infection with BA.5 decreases.

In conclusion, the study design is similar to a prospective study but taking place in the past: the groups of interest were selected (i.e., individuals with no recorded infection or individuals with one infection in a defined period of time and without any additional infection reported until the start of the study period); and afterwards the individuals from the different groups were followed, under the same epidemiological conditions, for a pre-defined (and equal) number of days and their infections were recorded. We considered other study designs, such as test-negative study (Altarawneh et al. 2022; Ayoub et al. 2022; Hansen et al. 2023), but our registry-based dataset only includes information on positive tests, thus precluding the use of a test-negative design.

10.2.2 Vaccination coverage

The vaccine coverage with the primary vaccination series in the Portuguese residents over 12 years was >98% by the end of 2021. The primary series of the vaccination campaign used EU/EMA-authorized vaccines: Comirnaty (Pfizer/BioNTech), 69%; Spikevax (Moderna), 12%; Vaxzevria (AstraZeneca), 13%; and Janssen 6%.

While at the start of the BA.1/BA.2 period of dominance (January 1st, 2022), the coverage with the first booster was residual (mostly long-term care facility residents), at the start of the BA.5 period of dominance (June 1st), the coverage with the first booster was 82%. The vaccine boosters relied exclusively on mRNA vaccines (77% Comirnaty and 23% Spikevax). At the start of the BA.5 period, a second booster was not yet in use except for a highly specific (and very small) population of patients with severe immunosuppression.

10.2.3 Statistical analysis

We estimated the relative risk of BA.5 reinfection in each sub-interval using the modified Poisson regression method with a robust sandwich estimator for the variance as described previously (Zou 2004). We compared the risk of BA.5 infection for people with a previous single infection at different intervals, with the same risk for people without any previously recorded infection in the same interval periods (Table 10.2). Protection efficacy was estimated, in percentage, as $(1 - RR) \times 100\%$. Confidence intervals for the RR were calculated using the Wald normal approximation method, with the epitools R package (Aragon et al. 2020).

To ascertain the change in relative risk over time, we considered the sub-intervals of BA.5 infection as a blocking factor and used a mixed-effect approach to estimate the change in risk over time, where “sub-interval” was the random effect (Pinheiro and Bates 2000). We fitted the increase in risk with the following saturating function:

$$RR(t \geq 90) = rr_0 + (rr_A - rr_0) \frac{(t - 90)^n}{T_M^n + (t - 90)^n}. \quad (10.1)$$

In this equation, t represents time in days, which is larger than 90, since re-infections can only occur after that time, rr_0 is the initial relative risk when $t = 90$, and rr_A is the asymptotic risk when time is very large, T_M is the time, after 90 days, at which the relative risk is approximately $\frac{1}{2}$ of the asymptotic risk. Finally, n allows for different steepness in the increase of relative risk. This is a general saturation function, and it allowed us to test simpler versions, such as fixing $rr_A = 1$ or $n = 1$, which did not describe the data as well. We also

tested other saturation functions, such as logistic or generalised logistics, but the results were similar (i.e., the initial and the asymptotic relative risk). We used the software Monolix 2021R1 (Lixoft SAS, Antony, France) to fit this model using each sub-interval of BA.5 as the random effect (Figure 10.1B and Figure 10.3). The best fit included a random effect for T_M and n , but only fixed effects for rr_0 and rr_A . The population parameters for the best mixed-effect fit are $rr_0 = 0.064$ (95% CI = [0.056, 0.074]), $rr_A = 0.368$ (95% CI = [0.321, 0.424]), $T_M = 65.7$ (95% CI = [52.8, 81.9]), $n = 3.2$ (95% CI = [2.3, 4.4]). The stability of the estimated parameters for initial and asymptotic relative risk (with no random effect supported by the data) over the different time intervals strengthens our conclusions, as biases due to, for example, undocumented infections are unlikely to be important for the periods studied.

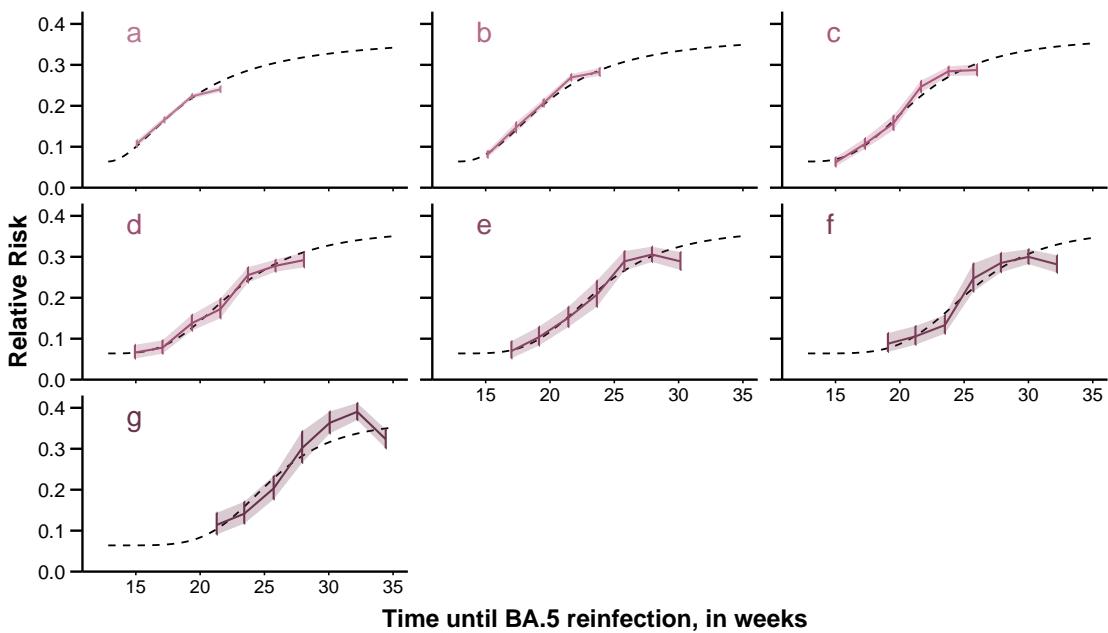


Figure 10.3: Variation of RR of protection against BA.5 infection over time since BA.1/BA.2 infection. Individual sub-interval fits of Equation (10.1) (dashed lines) to the different periods of BA.5 dominance (intervals a to g), corresponding to the population fit presented in Figure 10.1B. The RR calculated from the data is indicated in the solid lines with corresponding 95% confidence intervals (shaded area).

Bibliography

- H. N. Altarawneh, H. Chemaitley, H. H. Ayoub, M. R. Hasan, P. Coyle, H. M. Yassine, H. A. Al-Khatib, M. K. Smatti, Z. Al-Kanaani, E. Al-Kuwari, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, H. F. Abdul-Rahim, G. K. Nasrallah, M. G. Al-Kuwari, A. A. Butt, H. E. Al-Romaihi, M. H. Al-Thani, A. Al-Khal, R. Bertollini, P. Tang, and L. J. Abu-Raddad. Protective Effect of Previous SARS-CoV-2 Infection against Omicron BA.4 and BA.5 Subvariants. *New England Journal of Medicine*, 387(17):1620–1622, Oct. 2022. ISSN 0028-4793. doi: 10.1056/NEJMc2209306. URL <https://doi.org/10.1056/NEJMc2209306>. Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMc2209306>.
- T. J. Aragon, M. P. Fay, D. Wollschlaeger, and A. Omidpanah. epitools: Epidemiology Tools, Mar. 2020. URL <https://cran.r-project.org/web/packages/epitools/index.html>.
- H. H. Ayoub, M. Tomy, H. Chemaitley, H. N. Altarawneh, P. Coyle, P. Tang, M. R. Hasan, Z. A. Kanaani, E. A. Kuwari, A. A. Butt, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, G. K. Nasrallah, F. M. Benslimane, H. A. A. Khatib, H. M. Yassine, M. G. A. Kuwari, H. E. A. Romaihi, H. F. Abdul-Rahim, M. H. Al-Thani, A. A. Khal, R. Bertollini, and L. J. Abu-Raddad. Estimating protection afforded by prior infection in preventing reinfection: Applying the test-negative study design, Jan. 2022. URL <https://www.medrxiv.org/content/10.1101/2022.01.02.22268622v1>. Pages: 2022.01.02.22268622.
- H. Chemaitley, P. Tang, M. R. Hasan, S. AlMukdad, H. M. Yassine, F. M. Benslimane, H. A. Al Khatib, P. Coyle, H. H. Ayoub, Z. Al Kanaani, E. Al Kuwari, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, H. F. Abdul Rahim, G. K. Nasrallah, M. G. Al Kuwari, H. E. Al Romaihi, A. A. Butt, M. H. Al-Thani, A. Al Khal, R. Bertollini, and L. J. Abu-Raddad. Waning of BNT162b2 Vaccine Protection against SARS-CoV-2 Infection in Qatar. *New England Journal of Medicine*, 385(24):e83, Dec. 2021. ISSN 0028-4793. doi: 10.1056/NEJMoa2114114. URL <https://doi.org/10.1056/NEJMoa2114114>. Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMoa2114114>.
- H. Chemaitley, P. Tang, P. Coyle, H. M. Yassine, H. A. Al-Khatib, M. K. Smatti, M. R. Hasan, H. H. Ayoub, H. N. Altarawneh, Z. Al-Kanaani, E. Al-Kuwari, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, H. F. Abdul-Rahim, G. K. Nasrallah, M. G. Al-Kuwari, A. A. Butt, H. E. Al-Romaihi, M. H. Al-Thani, A. Al-Khal, R. Bertollini, and L. J. Abu-Raddad. Protection against reinfection with SARS-CoV-2 omicron BA.2.75* sublineage, Oct. 2022. URL <https://www.medrxiv.org/content/10.1101/2022.10.29.22281606v1>. ISSN: 2228-1606 Pages: 2022.10.29.22281606.
- Y. Goldberg, M. Mandel, Y. M. Bar-On, O. Bodenheimer, L. S. Freedman, N. Ash, S. Alroy-Preis, A. Huppert, and R. Milo. Protection and Waning of Natural and Hybrid Immunity to SARS-CoV-2. *New England Journal of Medicine*, 386(23):2201–2212, June 2022. ISSN 0028-4793. doi: 10.1056/NEJMoa2118946. URL <https://doi.org/10.1056/NEJMoa2118946>. Publisher: Massachusetts Medical Society _eprint: <https://doi.org/10.1056/NEJMoa2118946>.

- C. H. Hansen, N. U. Friis, P. Bager, M. Stegger, J. Fonager, A. Fomsgaard, M. A. Gram, L. E. Christiansen, S. Ethelberg, R. Legarth, T. G. Krause, H. Ullum, and P. Valentiner-Branth. Risk of reinfection, vaccine protection, and severity of infection with the BA.5 omicron subvariant: a nation-wide population-based study in Denmark. *The Lancet Infectious Diseases*, 23(2):167–176, Feb. 2023. ISSN 1473-3099, 1474-4457. doi: 10.1016/S1473-3099(22)00595-3. URL [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(22\)00595-3/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(22)00595-3/fulltext). Publisher: Elsevier.
- Instituto Nacional de Saúde Doutor Ricardo Jorge. National COVID-19 Serologic Survey in Portugal. Technical report, INSA, Lisbon, 2021. URL http://repositorio.insa.pt/bitstream/10400.18/7828/1/ISN_COVID19_relatório_3F_final.pdf.
- Instituto Nacional de Saúde Doutor Ricardo Jorge. Genetic diversity of the novel coronavirus SARS-CoV-2 (COVID-19) in Portugal, 2022. URL <https://insaflu.insa.pt/covid19>.
- J. a. Malato, R. M. Ribeiro, P. P. Leite, P. Casaca, E. Fernandes, C. Antunes, V. R. Fonseca, M. C. Gomes, and L. Graca. Risk of BA.5 Infection among Persons Exposed to Previous SARS-CoV-2 Variants. *New England Journal of Medicine*, 387(10):953–954, Sept. 2022. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMc2209479. URL <http://www.nejm.org/doi/10.1056/NEJMc2209479>.
- J. C. Pinheiro and D. M. Bates. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer-Verlag, New York, 2000. ISBN 978-0-387-98957-0. doi: 10.1007/b98882. URL <http://link.springer.com/10.1007/b98882>.
- World Health Organizaton. Public health surveillance for COVID-19: interim guidance. Interim guidance WHO/2019-nCoV/SurveillanceGuidance/2022.2, July 2022. URL <https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-SurveillanceGuidance-2022.2>.
- G. Zou. A modified poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology*, 159(7):702–706, Apr. 2004. ISSN 0002-9262. doi: 10.1093/aje/kwh090.

Part IV

GENERAL DISCUSSION

Chapter 11

General discussion

In this thesis, I have presented studies exploring different aspects of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) as a heterogeneous disease. I compared patient diagnosis agreement between four different case definitions and simulated case-control study scenarios under the assumption of misdiagnosis of patients to evaluate the implications in reproducibility of research findings (Chapter 2, Chapter 3, and Chapter 4). To study possible links between the immune origin of ME/CFS, patients were stratified into more homogeneous subgroups related to specific symptom profiles and infection triggers. They were then compared against healthy controls and multiple sclerosis (MS) patients using data from previously published IgG antibody responses towards antigens of six different herpesviruses and reported symptom assessments (Chapter 5, Chapter 6, and Chapter 7). This thesis was produced during the Covid-19 pandemic. As such, I extended the research on ME/CFS to study the increased risk of SARS-CoV-2 infections in ME/CFS patients (Chapter 8) and conducted population studies on the risk of SARS-CoV-2 Omicron BA.5 infection in the context of the Portuguese population, a cases study of a highly vaccinated community (Chapter 9 and Chapter 10).

11.1 Misdiagnosis and stratification of ME/CFS

The analysis of agreement among four of the most commonly used ME/CFS case definitions revealed that 37.1% had a disagreement in at least one diagnostic criterion (Chapter 2). This shows how different criteria have distinct focuses for the diagnosis of patients, with varying degrees of inclusiveness (Lim and Son 2020). This uncertainty surrounding diagnosis leads to (relative) misdiagnosis of patients, with an impact on research reproducibility (Nacul et al.

2019). Considering that currently there is no criterion considered to be a gold standard in either research or clinical practice, a proposed alternative to mitigate possible inconsistencies could be using more than one criterion together. In fact, when using at least two case definitions, the number of diagnosed individuals greatly improved.

In Chapter 2, Chapter 3, and Chapter 4, I studied how the inclusion of false positive individuals in ME/CFS cohorts and misclassification negatively impact the consistency of research, particularly in case-control studies. The findings of these studies highlight ME/CFS as an example of a challenging illness in terms of research success and reproducibility of results.

The first challenge is that ME/CFS patient cohorts can be composed of other subgroups, defined as misdiagnosis in a strict sense. This challenge is directly related to the difficulty in objectively characterising the diagnosed population under the idea that ME/CFS, being an umbrella term, might encompass distinct subtypes of the disease. In this sense, I identified the subgroup of false positive cases as having similar characteristics on the candidate causal factor of interest as the matched healthy controls, thus more clearly exemplifying the dilution effect on the statistical power to detect a putative disease association when increasing the misdiagnosis rate (Figure 11.1, path A). However, this model does not necessarily impose the idea that apparent cases should be viewed as healthy controls (see Section 4.4). Instead, this uncertain group can be viewed as other diseases or specific subgroups of ME/CFS, both characterised by reduced association or even predisposition to a different causal factor (Figure 11.1, path B) (Nacul et al. 2019). Under this consideration, the presence probability of the causal factor in suspected ME/CFS cases would be composed by other specific parametric values, and the 2×2 tables used to test the null hypothesis would be further augmented.

The second challenge stems from the notion that there are likely neither single nor strong associations between a candidate causal factor and ME/CFS (Dibble et al. 2020; Malato et al. 2023). Consequently—and aligning with the previous challenge—candidate markers for the disease should be tested on specific subsets of patients. Throughout the thesis, ME/CFS cases were stratified based on disease-triggering herpesviruses-related infections and symptoms grouped by mechanistic domains. One approach analysed data on antibody responses against specific peptides from Epstein–Barr virus (EBV) in ME/CFS patients and healthy controls

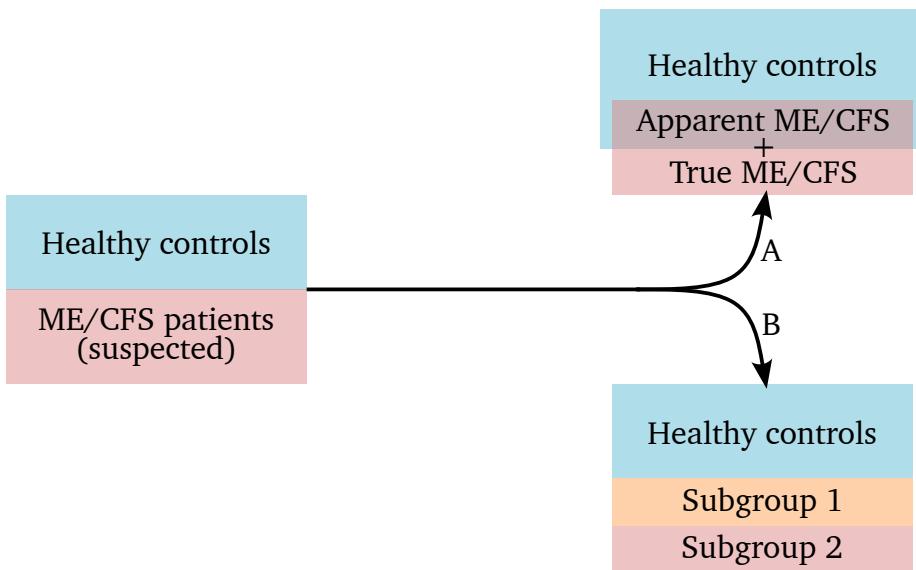


Figure 11.1: Alternative misdiagnosis assumptions to include patient stratification. From the initial sample of healthy controls and suspected ME/CFS cases (left box), assumptions can be made as to (A) assess the inclusion of apparent cases (false positives) that are similar to controls regarding the measured causal risk factor; or (B) stratify patients into distinct subgroups with specific distinct parametric values for the risk factor.

(Chapter 5). This analysis suggested the possible effect of molecular mimicry between viral and human antigens in patients with an infectious trigger. Moreover, relating these antibody responses to age and gender allowed for the development of classification models with good sensitivity and specificity that could be used in the diagnosis of this particular subgroup. Alternatively, I studied the differences between ME/CFS infection-triggered subgroups and MS (Chapter 6). The analysis supported the notion of distinct origins of both diseases with overlapping, albeit distinguishable, symptoms, which could potentially help improve exclusionary criteria. Finally, stratifying patients based on the severity of specific groups of symptoms related to particular domains revealed a preliminary association between IgG antibody responses against the neurotropic herpes simplex virus 1 (HSV1) and a more severe neurocognitive profile (Chapter 7). Although further research is needed to better characterise antibody-symptom associations in ME/CFS populations, patient stratification could provide insights into specific pathways for the disease.

Following the results from patient stratification, the third challenge in ME/CFS related to

the limitation of sample sizes in research, which usually do not surpass 300 individuals per study group. Given this constraint, the strategy for subsetting ME/CFS cohorts may seem unfeasible, as further reducing the size of compared groups can further diminish statistical power. Nevertheless, focusing on well-characterised subgroups is expected to strengthen possible associations, leading to more robust findings that will likely be replicated across independent studies (Jason et al. 2005).

Ultimately, addressing these challenges requires collaborative efforts and consensus for the implementation of standardised methods to objectively diagnose and stratify ME/CFS cases. These efforts are crucial to advance our understanding of ME/CFS aetiology and improve patient care (Pheby et al. 2020; Jason et al. 2023b). Furthermore, the concepts and methodologies proposed in the context of ME/CFS research can be readily transferred to other complex diseases, such as Long Covid, where there is evidence of misdiagnosis or suspicion of distinct disease subtypes (Zhang et al. 2023; Woodrow et al. 2023).

11.2 Results on Covid-19

A retrospective inspection of Covid-19 data analysis literature and information quality revealed three “waves” of research topics during the pandemic (Serio et al. 2022); Westermeier and Sepúlveda (2023) describe them as follows:

1. Basic epidemiology and clinical characterisation of the disease (e.g., prevalence, incidence, and transmission rate studies, and linking viral load with disease severity);
2. Covid-19 herd immunity, serologic testing, and asymptomatic characterisation;
3. Vaccination and therapeutic efficacy and comparability, with the identification of new variants of concern (VOC) and individuals at risk within the population.

The rapid progression between these research focuses over time evidences how collaborative efforts can quickly generate knowledge, to the benefit of all.

My research falls in the third wave of research. The findings helped to assess whether prior infection with SARS-CoV-2 VOC, particularly Omicron subvariants BA.1/BA.2, could

provide protection against the emerging BA.5, at a time when vaccination campaigns were well-established and adapted vaccines under clinical trials were based on BA.1 (Chapter 9). Furthermore, the study on immune waning showed that, despite the decline in efficacy against BA.5 infection over a period of up to eight months, hybrid immunity (vaccine + previous single infection) conferred substantial protection throughout, when compared with vaccinated only individuals. These findings on the stability of protection against severe disease can also be useful in potential vaccination booster campaigns in the near future (Chapter 10). In this sense, the Portuguese population was especially representative, as a large proportion of the population—particularly the younger age groups—had experienced a single past infection by the previous Omicron subvariants.

The rollout of Covid-19 vaccines has proven to have a large impact in reducing the morbidity and mortality induced by the SARS-CoV-2 virus. However, under the ever-evolving epidemiological implications of this virus, population studies should continue to work on the third “wave” of research, focusing on the risk of breakthrough infections by new VOC and on the identification and protection of more susceptible individuals.

11.3 ME/CFS and Long Covid

Intersecting the themes of ME/CFS and Covid-19 explored in this thesis is Long Covid.

After an acute SARS-CoV-2 infection, some patients suffer from what has been termed post-acute sequelae of SARS-CoV-2, post-acute Covid-19 syndrome, or simply Long Covid (Choutka et al. 2022). Overlapping with ME/CFS, Long Covid is characterised by symptoms such as post-exertional malaise (PEM) and limited exercise tolerance, accompanied by cognitive, autonomic, and sleep-related impairments. Common manifestations include dyspnea (59.7% of individuals diagnosed with Long Covid), myalgia (51.5%), extreme fatigue (39.6%), memory impairment (37.3%), and sleep disturbances (35.1%) (Logue et al. 2021; Sykes et al. 2021). The parallels between the two illnesses also extend to a similar female-to-male ratio (Jacobsen and Klein 2021) and proposed phenotypes and mechanistic pathways. These include impaired arginine metabolism, endothelial dysfunction and vascular function (McLaughlin et al. 2023),

association with oxidative damage, impaired autonomic ability (Hayes et al. 2023), and possible miscoordination between cellular and humoral adaptive immunity, leading to immune dysregulation and systemic inflammation (Komaroff and Lipkin 2021; Komaroff and Bateman 2021; Al-Hakeim et al. 2023; Yin et al. 2024). Additionally, EBV has been proposed as a risk factor for Long Covid, with reports of reactivation of this virus after infection by SARS-CoV-2. Studies indicate a significantly higher chance of active herpesvirus infections in Long Covid patients when compared with non-infected controls (Su et al. 2022; Bernal and Whitehurst 2023; Banko et al. 2023). Different subgroups and predispositions for Long Covid have also been proposed, suggesting heterogeneity in the pathophysiology of the disease (Zhang et al. 2023). This is also consistent with the fact that not all cases of Covid-19 lead to Long Covid, as is the case for the proposed viral incidences of ME/CFS. In this case, stratification of Long Covid based on specific criteria and hypotheses could also enhance the understanding of this complex disease.

Despite close similarities, ME/CFS and Long Covid exhibit key distinctions. Unlike ME/CFS, Long Covid has a clear pathophysiological association with previous infection and disease onset. Additionally, Long Covid patients experience unique symptoms such as changes in taste (ageusia) and smell, which are not generally observed in ME/CFS. These differences may suggest distinct mechanisms between the two illnesses. Furthermore, a study on endothelial dysfunction reported elevated concentrations of Edothelin-1 in ME/CFS and Long Covid patients when compared to healthy controls (Haffke et al. 2022). However, the same study found lower concentrations of Angiopoietin-2 in Long Covid alone. This growth factor is part of the Angiopoietin-Tie signalling pathway, involved in the regulation of endothelial homeostasis and angiogenesis (Akwii et al. 2019). Such findings suggest differences in (chronic) inflammation that could be a starting point for the differentiation between ME/CFS and Long Covid in terms of biomarker research.

Following the Covid-19 pandemic, ME/CFS has gained more attention due to the similarities with Long Covid. This has resulted in increased research funding and even policy changes, displaying the renewed interest in the disease (O'Neill 2022). However, as the number of Long Covid cases increases and new disease definitions are proposed, researchers could benefit from

the lessons learnt in ME/CFS over the years to promote research consistency (or avoid lack thereof) under the implementation of sound statistical methodologies right from the start, as a way to accelerate research in post-infectious syndromes (Westermeier et al. 2022; Jason et al. 2023a).

11.4 Concluding remarks

Misdiagnosis represents a transversal and significant bias in scientific research overall, either by increasing the number of sporadic false positive associations with a high rate of non-replication (Ioannidis 2005) or by diluting the true effect of putative disease associations (Westermeier and Sepúlveda 2023). In the context of ME/CFS, this issue can be particularly challenging, as evidenced by the limited statistical power to detect candidate causal factors in the analyses of misdiagnosis in case-control association studies. To address this, future studies on disease-specific biomarkers should prioritise adequately powered designs to improve consistency across results. This can be done by increasing sample sizes to at least 500 individuals per study group, integrating analysis of diagnostic accuracy in serological study designs, and considering potential misdiagnosis of patients (and even controls). Moreover, research should strive to be open, collaborative, and designed in accordance with strategies from international consortia or networks whenever possible (Scheibenberg et al. 2017; Nacul et al. 2021).

The multifactorial nature of ME/CFS required the identification of patient subgroups to enable the testing and interpretation of specific triggers and mechanisms underlying the disease's onset and expression (Jason et al. 2005). There is a growing body of evidence for an autoimmune component as origin and chronicity of ME/CFS, especially in cases following an acute viral infection (Blomberg et al. 2018; Sotzny et al. 2018). The identification of specific antibody responses against herpesviruses in these patients hints at the potential role of antigen mimicry in triggering autoimmune responses (Phelan et al. 2020; Sepúlveda et al. 2022). Research linking the immune component of the disease with the identification of characteristic symptom profiles may provide further insights into the pathogenesis of the disease. Additionally, the identification of key differences between ME/CFS and known autoimmune

diseases with overlapping symptoms such as MS and Long Covid, could help to improve and bring consensus to ME/CFS exclusionary criteria (Jason et al. 2023b). Future longitudinal studies with multiple timepoints should investigate the relationship between the observed antibody responses, controlling for symptom fluctuations and disease progression within stratified subgroups.

Despite the diagnostic and research challenges in ME/CFS, advancements in the underlying characteristics of this illness offer hope for the development of personalised care and treatment strategies to improve the clinical outcomes of patients.

Bibliography

- R. G. Akwii, M. S. Sajib, F. T. Zahra, and C. M. Mikelis. Role of Angiopoietin-2 in Vascular Physiology and Pathophysiology. *Cells*, 8(5):471, May 2019. ISSN 2073-4409. doi: 10.3390/cells8050471. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6562915/>.
- H. K. Al-Hakeim, H. T. Al-Rubaye, D. S. Al-Hadrawi, A. F. Almulla, and M. Maes. Long-COVID post-viral chronic fatigue and affective symptoms are associated with oxidative damage, lowered antioxidant defenses and inflammation: a proof of concept and mechanism study. *Molecular Psychiatry*, 28(2):564–578, Feb. 2023. ISSN 1476-5578. doi: 10.1038/s41380-022-01836-9. URL <https://www.nature.com/articles/s41380-022-01836-9>. Number: 2 Publisher: Nature Publishing Group.
- A. Banko, D. Miljanovic, and A. Cirkovic. Systematic review with meta-analysis of active herpesvirus infections in patients with COVID-19: Old players on the new field. *International Journal of Infectious Diseases*, 130:108–125, May 2023. ISSN 1201-9712. doi: 10.1016/j.ijid.2023.01.036. URL <https://www.sciencedirect.com/science/article/pii/S1201971223000371>.
- K. D. E. Bernal and C. B. Whitehurst. Incidence of Epstein-Barr virus reactivation is elevated in COVID-19 patients. *Virus Research*, 334, Sept. 2023. doi: 10.1016/j.virusres.2023.199157. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10292739/>. Publisher: Elsevier.
- J. Blomberg, C.-G. Gottfries, A. Elfaitouri, M. Rizwan, and A. Rosén. Infection Elicited Autoimmunity and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: An Explanatory Model. *Frontiers in Immunology*, 9, Feb. 2018. doi: 10.3389/fimmu.2018.00229.
- J. Choutka, V. Jansari, M. Hornig, and A. Iwasaki. Unexplained post-acute infection syndromes. *Nature Medicine*, 28(5):911–923, May 2022. ISSN 1546-170X. doi: 10.1038/s41591-022-01810-6. URL <https://www.nature.com/articles/s41591-022-01810-6>. Number: 5 Publisher: Nature Publishing Group.

- J. J. Dibble, S. J. McGrath, and C. P. Ponting. Genetic risk factors of ME/CFS: a critical review. *Human Molecular Genetics*, 29(R1):R117–R124, Sept. 2020. ISSN 0964-6906. doi: 10.1093/hmg/ddaa169. URL <https://doi.org/10.1093/hmg/ddaa169>.
- M. Haffke, H. Freitag, G. Rudolf, M. Seifert, W. Doechner, N. Scherbakov, L. Hanitsch, K. Wittke, S. Bauer, F. Konietzschke, F. Paul, J. Bellmann-Strobl, C. Kedor, C. Scheibenbogen, and F. Sotzny. Endothelial dysfunction and altered endothelial biomarkers in patients with post-COVID-19 syndrome and chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 20(1):138, Mar. 2022. ISSN 1479-5876. doi: 10.1186/s12967-022-03346-2. URL <https://doi.org/10.1186/s12967-022-03346-2>.
- L. D. Hayes, N. E. M. Sanal-Hayes, M. McLaughlin, E. C. J. Berry, and N. F. Sculthorpe. People with Long Covid and ME/CFS Exhibit Similarly Impaired Balance and Physical Capacity: A Case-Case-Control Study. *The American Journal of Medicine*, 0(0), July 2023. ISSN 0002-9343, 1555-7162. doi: 10.1016/j.amjmed.2023.06.028. URL [https://www.amjmed.com/article/S0002-9343\(23\)00465-5/fulltext](https://www.amjmed.com/article/S0002-9343(23)00465-5/fulltext). Publisher: Elsevier.
- J. P. A. Ioannidis. Why Most Published Research Findings Are False. *PLOS Medicine*, 2(8):e124, Aug. 2005. ISSN 1549-1676. doi: 10.1371/journal.pmed.0020124. URL <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124>. Publisher: Public Library of Science.
- H. Jacobsen and S. L. Klein. Sex Differences in Immunity to Viral Infections. *Frontiers in Immunology*, 12:720952, 2021. ISSN 1664-3224. doi: 10.3389/fimmu.2021.720952.
- L. A. Jason, K. Corradi, S. Torres-Harding, R. R. Taylor, and C. King. Chronic Fatigue Syndrome: The Need for Subtypes. *Neuropsychology Review*, 15(1):29–58, Mar. 2005. doi: 10.1007/s11065-005-3588-2.
- L. A. Jason, B. H. Natelson, H. Bonilla, Z. A. Sherif, S. D. Vernon, M. Verduzco Gutierrez, L. O'Brien, and E. Taylor. What Long COVID investigators can learn from four decades of ME/CFS research. *Brain Behavior and Immunity Integrative*, 4:100022, Dec. 2023a. ISSN 2949-8341. doi: 10.1016/j.bbii.2023.100022. URL <https://www.sciencedirect.com/science/article/pii/S2949834123000211>.
- L. A. Jason, S. Ravichandran, B. Z. Katz, B. H. Natelson, and H. F. Bonilla. Establishing a consensus on ME/CFS exclusionary illnesses. *Fatigue: Biomedicine, Health & Behavior*, 11(1):1–13, Jan. 2023b. ISSN 2164-1846, 2164-1862. doi: 10.1080/21641846.2022.2150487. URL <https://www.tandfonline.com/doi/full/10.1080/21641846.2022.2150487>.
- A. L. Komaroff and L. Bateman. Will COVID-19 Lead to Myalgic Encephalomyelitis/Chronic Fatigue Syndrome? *Frontiers in Medicine*, 7, Jan. 2021. doi: 10.3389/fmed.2020.606824.
- A. L. Komaroff and W. I. Lipkin. Insights from myalgic encephalomyelitis/chronic fatigue syndrome may help unravel the pathogenesis of postacute COVID-19 syndrome. *Trends in Molecular Medicine*, 27(9):895–906, Sept. 2021. ISSN 1471-499X. doi: 10.1016/j.molmed.2021.06.002.
- E. J. Lim and C.-G. Son. Review of case definitions for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 18(1):289, Dec. 2020. ISSN 1479-5876. doi: 10.1186/

- s12967-020-02455-0. URL <https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-020-02455-0>.
- J. K. Logue, N. M. Franko, D. J. McCulloch, D. McDonald, A. Magedson, C. R. Wolf, and H. Y. Chu. Sequelae in Adults at 6 Months After COVID-19 Infection. *JAMA Network Open*, 4(2):e210830, Feb. 2021. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2021.0830. URL <https://doi.org/10.1001/jamanetworkopen.2021.0830>.
- J. a. Malato, L. Graça, and N. Sepúlveda. Impact of Misdiagnosis in Case-Control Studies of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Diagnostics*, 13(3):531, Jan. 2023. ISSN 2075-4418. doi: 10.3390/diagnostics13030531. URL <https://www.mdpi.com/2075-4418/13/3/531>. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
- M. McLaughlin, N. E. M. Sanal-Hayes, L. D. Hayes, E. C. Berry, and N. F. Sculthorpe. People with Long COVID and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Exhibit Similarly Impaired Vascular Function. *The American Journal of Medicine*, Oct. 2023. ISSN 0002-9343. doi: 10.1016/j.amjmed.2023.09.013. URL <https://www.sciencedirect.com/science/article/pii/S0002934323006095>.
- L. Nacul, E. M. Lacerda, C. C. Kingdon, H. Curran, and E. W. Bowman. How have selection bias and disease misclassification undermined the validity of myalgic encephalomyelitis/chronic fatigue syndrome studies? *Journal of Health Psychology*, 24(12):1765–1769, Oct. 2019. ISSN 1461-7277. doi: 10.1177/1359105317695803.
- L. Nacul, F. J. Authier, C. Scheibenbogen, L. Lorusso, I. B. Helland, J. A. Martin, C. A. Sirbu, A. M. Mengshoel, O. Polo, U. Behrends, H. Nielsen, P. Grabowski, S. Sekulic, N. Sepulveda, F. Estévez-López, P. Zalewski, D. F. H. Pheby, J. Castro-Marrero, G. K. Sakkas, E. Capelli, I. Brundsdlund, J. Cullinan, A. Krumina, J. Bergquist, M. Murovska, R. C. W. Vermuelen, and E. M. Lacerda. European Network on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (EUROMENE): Expert Consensus on the Diagnosis, Service Provision, and Care of People with ME/CFS in Europe. *Medicina*, 57(5):510, May 2021. ISSN 1648-9144. doi: 10.3390/medicina57050510. URL <https://www.mdpi.com/1648-9144/57/5/510>.
- S. O'Neill. Sajid Javid promises radical action for patients debilitated by ME. May 2022. ISSN 0140-0460. URL <https://www.thetimes.co.uk/article/sajid-javid-promises-radical-action-for-patients-debilitated-by-me-9bkq6qf5g>.
- D. F. H. Pheby, D. Araja, U. Berkis, E. Brenna, J. Cullinan, J.-D. d. Korwin, L. Gitto, D. A. Hughes, R. M. Hunter, D. Trepel, and X. Wang-Steverding. The Development of a Consistent Europe-Wide Approach to Investigating the Economic Impact of Myalgic Encephalomyelitis (ME/CFS): A Report from the European Network on ME/CFS (EUROMENE). *Healthcare*, 8(2):88, Apr. 2020. doi: 10.3390/healthcare8020088.
- J. Phelan, A. D. Grabowska, and N. Sepúlveda. A potential antigenic mimicry between viral and human proteins linking Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) with autoimmunity: The case of HPV immunization. *Autoimmunity Reviews*, 19(4):102487, Apr. 2020. doi: 10.1016/j.autrev.2020.102487.

- C. Scheibenberg, H. Freitag, J. Blanco, E. Capelli, E. Lacerda, J. Authier, M. Meeus, J. C. Marrero, Z. Nora-Krukla, E. Oltra, and others. The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE. *Journal of translational medicine*, 15(1):162, 2017. doi: <https://doi.org/10.1186/s12967-017-1263-z>.
- N. Sepúlveda, J. a. Malato, F. Sotzny, A. D. Grabowska, A. Fonseca, C. Cordeiro, L. Graça, P. Biecek, U. Behrends, J. Mautner, F. Westermeier, E. M. Lacerda, and C. Scheibenberg. Revisiting IgG Antibody Reactivity to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Its Potential Application to Disease Diagnosis. *Frontiers in Medicine*, 9, 2022. ISSN 2296-858X. doi: 10.3389/fmed.2022.921101. URL <https://www.frontiersin.org/articles/10.3389/fmed.2022.921101>.
- C. D. Serio, A. Malgaroli, P. Ferrari, and R. S. Kenett. The reproducibility of COVID-19 data analysis: paradoxes, pitfalls, and future challenges. *PNAS Nexus*, 1(3):pgac125, July 2022. ISSN 2752-6542. doi: 10.1093/pnasnexus/pgac125. URL <https://doi.org/10.1093/pnasnexus/pgac125>.
- F. Sotzny, J. Blanco, E. Capelli, J. Castro-Marrero, S. Steiner, M. Murovska, and C. Scheibenberg. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome – Evidence for an autoimmune disease. *Autoimmunity Reviews*, 17(6):601–609, June 2018. doi: 10.1016/j.autrev.2018.01.009.
- Y. Su, D. Yuan, D. G. Chen, R. H. Ng, K. Wang, J. Choi, S. Li, S. Hong, R. Zhang, J. Xie, S. A. Kornilov, K. Scherler, A. J. Pavlovitch-Bedzyk, S. Dong, C. Lausted, I. Lee, S. Fallen, C. L. Dai, P. Baloni, B. Smith, V. R. Duvvuri, K. G. Anderson, J. Li, F. Yang, C. J. Duncombe, D. J. McCulloch, C. Rostomily, P. Troisch, J. Zhou, S. Mackay, Q. DeGottardi, D. H. May, R. Taniguchi, R. M. Gittelman, M. Klinger, T. M. Snyder, R. Roper, G. Wojciechowska, K. Murray, R. Edmark, S. Evans, L. Jones, Y. Zhou, L. Rowen, R. Liu, W. Chour, H. A. Algren, W. R. Berrington, J. A. Wallick, R. A. Cochran, M. E. Micikas, ISB-Swedish COVID-19 Biobanking Unit, T. Wrin, C. J. Petropoulos, H. R. Cole, T. D. Fischer, W. Wei, D. S. B. Hoon, N. D. Price, N. Subramanian, J. A. Hill, J. Hadlock, A. T. Magis, A. Ribas, L. L. Lanier, S. D. Boyd, J. A. Bluestone, H. Chu, L. Hood, R. Gottardo, P. D. Greenberg, M. M. Davis, J. D. Goldman, and J. R. Heath. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell*, 185(5): 881–895.e20, Mar. 2022. ISSN 1097-4172. doi: 10.1016/j.cell.2022.01.014.
- D. L. Sykes, L. Holdsworth, N. Jawad, P. Gunasekera, A. H. Morice, and M. G. Crooks. Post-COVID-19 Symptom Burden: What is Long-COVID and How Should We Manage It? *Lung*, 199(2):113–119, Apr. 2021. ISSN 1432-1750. doi: 10.1007/s00408-021-00423-z. URL <https://doi.org/10.1007/s00408-021-00423-z>.
- F. Westermeier and N. Sepúlveda. Editorial: Reproducibility and rigour in infectious diseases - surveillance, prevention and treatment. *Frontiers in Medicine*, 10, 2023. ISSN 2296-858X. URL <https://www.frontiersin.org/articles/10.3389/fmed.2023.1294969>.
- F. Westermeier, E. M. Lacerda, C. Scheibenberg, and N. Sepúlveda. Editorial: Current Insights Into Complex Post-infection Fatigue Syndromes With Unknown Aetiology: The Case of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Beyond. *Frontiers in Medicine*, 9, 2022. ISSN 2296-858X. URL <https://www.frontiersin.org>.

- [org/articles/10.3389/fmed.2022.862953](https://doi.org/10.3389/fmed.2022.862953).
- M. Woodrow, C. Carey, N. Ziauddeen, R. Thomas, A. Akrami, V. Lutje, D. C. Greenwood, and N. A. Alwan. Systematic Review of the Prevalence of Long COVID. *Open Forum Infectious Diseases*, 10(7):ofad233, July 2023. ISSN 2328-8957. doi: 10.1093/ofid/ofad233. URL <https://doi.org/10.1093/ofid/ofad233>.
- K. Yin, M. J. Peluso, X. Luo, R. Thomas, M.-G. Shin, J. Neidleman, A. Andrew, K. C. Young, T. Ma, R. Hoh, K. Anglin, B. Huang, U. Argueta, M. Lopez, D. Valdivieso, K. Asare, T.-M. Deveau, S. E. Munter, R. Ibrahim, L. Ständker, S. Lu, S. A. Goldberg, S. A. Lee, K. L. Lynch, J. D. Kelly, J. N. Martin, J. Münch, S. G. Deeks, T. J. Henrich, and N. R. Roan. Long COVID manifests with T cell dysregulation, inflammation and an uncoordinated adaptive immune response to SARS-CoV-2. *Nature Immunology*, pages 1–8, Jan. 2024. ISSN 1529-2916. doi: 10.1038/s41590-023-01724-6. URL <https://www.nature.com/articles/s41590-023-01724-6>. Publisher: Nature Publishing Group.
- H. Zhang, C. Zang, Z. Xu, Y. Zhang, J. Xu, J. Bian, D. Morozyuk, D. Khullar, Y. Zhang, A. S. Nordvig, E. J. Schenck, E. A. Shenkman, R. L. Rothman, J. P. Block, K. Lyman, M. G. Weiner, T. W. Carton, F. Wang, and R. Kaushal. Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. *Nature Medicine*, 29(1):226–235, 2023. ISSN 1078-8956. doi: 10.1038/s41591-022-02116-3. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9873564/>.

Part V

APPENDICES

Appendix A

Additional information to Chapter 4

A.1 Supplementary tables

Table A.1: Augmented version of the observed 2×2 contingency table in the presence of the misdiagnosis of ME/CFS cases for a classical case-control association study. Parameter θ_0 is the probability of the presence of the candidate causal factor shared across healthy controls and apparent (false positive) ME/CFS cases, θ_1^* is the true probability of the causal factor in the true ME/CFS patients. Misdiagnosis probability is given by the parameter γ .

Causal factor	Controls	ME/CFS-diagnosed cases	
		(Apparent)	(True)
Present	θ_0	$\gamma\theta_0$	$(1-\gamma)\theta_1^*$
Absent	$1-\theta_0$	$\gamma(1-\theta_0)$	$(1-\gamma)(1-\theta_1^*)$

Table A.2: Augmented version of the observable 2×2 contingency table in the case-control association study with possible misdiagnosis of ME/CFS cases and misclassification of the true serological status (seropositive, S^+ , and seronegative, S^-). Parameter θ_0 is the probability of the presence of the candidate causal factor in healthy controls and apparent (false positive) ME/CFS cases, θ_1^* is the true probability of the causal factor in the true ME/CFS patients. Misdiagnosis probability is modulated by the parameter γ . The true serological status is dependent on the sensitivity (π_{se}) and specificity (π_{sp}) of the serological test.

Estimated Serological status	True serological status	Controls	ME/CFS-diagnosed cases	
			(Apparent)	(True)
S^+	S^+	$\pi_{se}\theta_0$	$\pi_{se}\gamma\theta_0$	$\pi_{se}(1-\gamma)\theta_1^*$
	S^-	$(1-\pi_{sp})(1-\theta_0)$	$(1-\pi_{sp})\gamma(1-\theta_0)$	$(1-\pi_{sp})(1-\gamma)(1-\theta_1^*)$
S^-	S^+	$(1-\pi_{se})\theta_0$	$(1-\pi_{se})\gamma\theta_0$	$(1-\pi_{se})(1-\gamma)\theta_1^*$
	S^-	$\pi_{sp}(1-\theta_0)$	$\pi_{sp}\gamma(1-\theta_0)$	$\pi_{sp}(1-\gamma)(1-\theta_1^*)$

A.2 Supplementary equations

We constructed our analysis considering a classical epidemiological scenario where for a single putative risk factor, individuals can be divided into exposed versus non-exposed. This result can be summarised by a 2×2 contingency table, whose sampling distribution is the product of two independent Binomial distributions, one Binomial distribution per group,

$$f(x_i|n_i; \theta_i) = \prod_{i=0,1} \binom{n_i}{x_i} \theta_i^{x_i} (1 - \theta_i)^{n_i - x_i}, \quad (\text{A.1})$$

where x_0 and x_1 are the observed frequencies of healthy controls and suspected cases with presence of the candidate causal factor, respectively, n_0 and n_1 are the corresponding sample sizes of each group, and θ_0 and θ_1 are the probabilities for the presence of the candidate causal factor in healthy controls and suspected cases, respectively.

$$\theta_1^* = \frac{\theta_0 \Delta_T}{1 + \theta_0 (\Delta_T - 1)} \quad (\text{A.2})$$

Appendix B

Additional information to Chapter 5

B.1 Supplementary tables

Table B.1: The overall and per-EBV-strain number of 15-mer peptides (antigens) whose antibody responses analysed.

EBV Protein	Associated stage	Number of 15-mer peptides per EBV strain						
		Overall	AG876	B95.8	GD1	Cao	Raji	P3HR.1
BALF-2	Early lytic	290	278	278	278	0	0	0
BALF-5	Early lytic	256	250	250	250	0	0	0
BFRF-3	Late lytic	42	0	42	0	0	0	0
BLLF-1	Late lytic	273	204	202	199	0	0	204
BLLF-3	Early lytic	74	66	67	66	0	0	0
BLRF-2	Late lytic	41	38	38	38	0	0	0
BMRF-1	Early lytic	102	99	99	99	0	0	0
BZLF-1	Immediate early lytic	89	57	57	58	0	0	0
EBNA-1	Latency I, II, and III	182	98	107	111	0	0	0
EBNA-3	Latency III	446	223	226	224	0	0	0
EBNA-4	Latency III	469	229	221	224	0	0	0
EBNA-6	Latency III	461	254	234	230	0	0	0
LMP-1	Latency II and III	197	79	85	80	77	84	0
LMP-2	Latency II and	III	132	120	120	120	0	0

Table B.2: Comparison among different null models (including the covariates age and gender and their interaction) using the Akaike's information criterion (AIC). The best model for each analysis/comparison is shown in bold. ME/CFS_{all}, ME/CFS_{inf} and ME/CFS_{noinf} represent all the ME/CFS patients, ME/CFS patients with an infectious trigger, and ME/CFS patients with a non-infectious trigger, respectively.

Analysis/Comparison	Model (link function)	AIC	ROC (95% CI)
ME/CFS _{all} vs. Healthy controls	Logit	189.973	0.577 (0.478, 0.676)
	Probit	189.964	0.576 (0.478, 0.675)
	Clog-log	189.936	0.574 (0.475, 0.672)
ME/CFS _{inf} vs. Healthy controls	Logit	147.055	0.610 (0.500, 0.719)
	Probit	147.029	0.606 (0.496, 0.715)
	Clog-log	147.220	0.609 (0.499, 0.718)
ME/CFS _{noinf} vs. Healthy controls	Logit	127.619	0.556 (0.429, 0.683)
	Probit	127.629	0.559 (0.432, 0.687)
	Clog-log	127.547	0.556 (0.429, 0.683)
ME/CFS _{inf} vs. ME/CFS _{noinf}	Logit	129.205	0.596 (0.471, 0.720)
	Probit	129.236	0.597 (0.472, 0.721)
	Clog-log	129.529	0.596 (0.472, 0.721)

Table B.3: The top 5 most significant antibodies for each association analysis where ME/CFS_all, ME/CFS_inf and ME/CFS_noinf represent all ME/CFS patients, ME/CFS patients with an infectious trigger, and ME/CFS patients with a non-infectious trigger, respectively. For simplicity, the antibodies were identified by their peptide. Statistically significant findings were obtained for $-\log_{10}(\text{adjusted p-value}) > 1.30$ ($= -\log_{10}(0.05)$) controlling for false discovery rate of 5% using the Benjamini-Yekutieli procedure.

Analysis/Comparison	Peptide	$-\log_{10}(\text{adjusted p-value})$
ME/CFS _{all} vs. Healthy controls	EBNA6_0066	0.743
	BLRF2_0005	0.486
	EBNA4_0392	0.486
	EBNA4_0497	0.486
	EBNA4_0529	0.486
ME/CFS _{inf} vs. Healthy controls	EBNA6_0066	2.693
	EBNA6_0070	2.693
	EBNA4_0529	1.794
	EBNA3_0380	1.270
	EBNA6_0569	1.270
ME/CFS _{noinf} vs. Healthy controls	EBNA6_0782	1.193
	BALF2_0358	1.153
	BALF2_0765	1.153
	BALF5_0041	1.153
	BALF5_0206	1.153

B.2 Supplementary figures

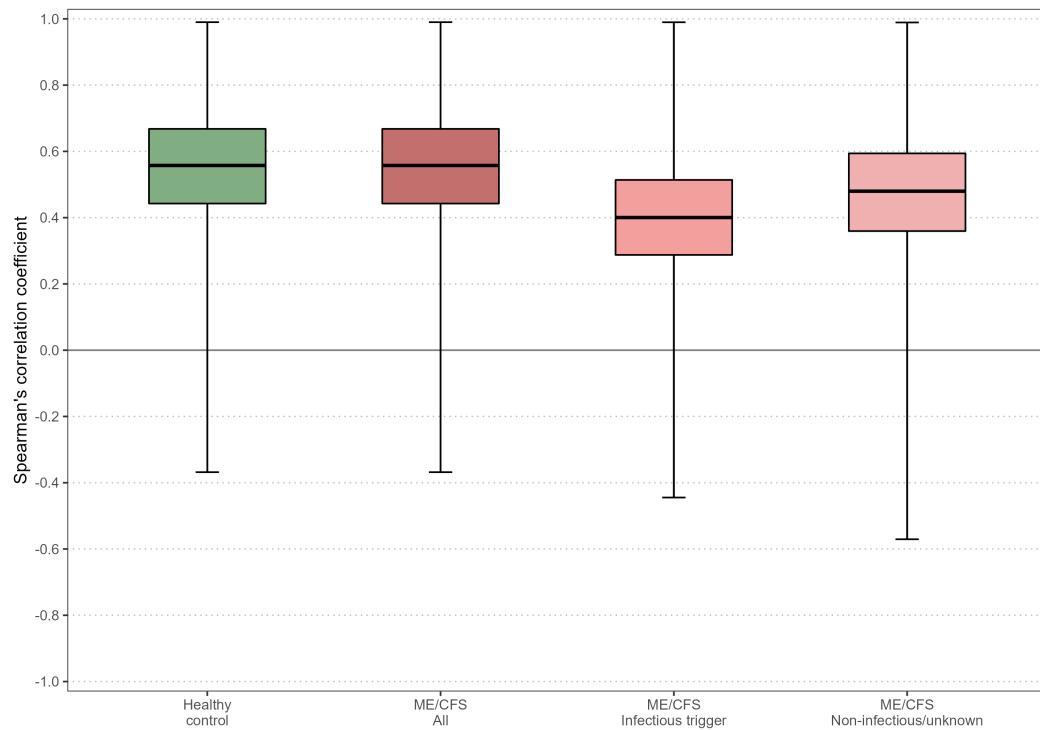


Figure B.1: Distributions of the Spearman's correlation coefficient between all the possible pairs of EBV-derived antibodies in healthy controls, all the ME/CFS patients, ME/CFS patients with an infectious trigger, and ME/CFS patients with a non-infectious or unknown trigger.

Appendix C

Additional information to Chapter 6

C.1 Supplementary tables

Table C.1: Description of the 59 (yes-no) symptoms available in the data and domain. Symptoms marked with an asterisk were removed due to a percentage of missing values higher than 25%.

Symptom	Description	Domain
airhunger	Air hunger, difficulty breathing, or shortness of breath on exertion/activity	Autonomic
bladderprob	Bladder problems	Autonomic
dizzystandup	Dizziness while standing up	Autonomic
extremepale	Being extremely pale	Autonomic
ibssymptoms	IBS symptoms	Autonomic
intolstandup	Intolerance to standing up	Autonomic
lightheaded	Feeling lightheaded	Autonomic
palpitations*	Palpitations	Autonomic
palpitstandup	Palpitations while standing up	Autonomic
sick_nausea*	Feeling sick/nausea	Autonomic
alcoholintoler*	Intolerance to alcohol	Immunological
feverchills	Fever/Chills	Immunological
fewviralinfect*	Fewer viral infections than before	Immunological
flusymptoms	Flu symptoms	Immunological
freqviralinfect	Frequent viral infections with long recovery periods	Immunological
newsensit	Worsen sensitivity to light	Immunological
sorethroat	Sore throat	Immunological
stiffnessmorn	Morning stiffness	Immunological
tenderglands	Tender glands	Immunological
coldhandsfeet*	Unusual cold hands or feet	Neuroendocrine
intolheat_cold	Intolerance to extremes of heat/cold	Neuroendocrine
sexualfunction	Decreased sexual function or interest	Neuroendocrine
unusualsweaty	Unusually sweaty	Neuroendocrine
weightchange*	Abnormal appetite or change in weight	Neuroendocrine
worsepoststress	Worsening of symptoms post stress	Neuroendocrine
backweak	Back weakness	Neurocognitive
brainfog	Brain fog or confusion	Neurocognitive
concentprob	Trouble concentrating	Neurocognitive
diffdecisions*	Difficulty making decisions	Neurocognitive
diffretaininfo	Difficulty retaining or recalling info	Neurocognitive
diffunderstand	Difficulty understanding things/thinking clearly	Neurocognitive
diffwords*	Difficulty finding or saying word	Neurocognitive
disorientation	Disorientation	Neurocognitive
eyesightdistub	Eyesight disturbance (temporary)	Neurocognitive
lossbalance	Loss of balance or unsteadiness while standing, unable to focus the vision	Neurocognitive
muscdisc	Muscle discomfort	Neurocognitive
muscweak	Muscle weakness	Neurocognitive
neckweak	Neck weakness	Neurocognitive
poorcoord	Poor coordination or unsteady movements (while walking)	Neurocognitive
senstlightnoise	Sensitivity to light/noise	Neurocognitive
shortmem	Short term memory problems	Neurocognitive
slowthink	Slow thinking	Neurocognitive
tingling	Tingling/numbness in arms and/or legs	Neurocognitive
twitching*	Muscle twitching	Neurocognitive
chestabdpain	Pain in chest or abdomen	Pain
diffmigraine	Migraine which is different/worse than before	Pain
jointpain*	Pain in two or more joints w/o inflamm.	Pain
jointpainnoinnfl	Joint pain w/o inflamm.	Pain
muscpain	Muscle pain	Pain
newheadache	Headaches which are new/worse than before	Pain
exerciseintol	Intolerance to exercise	Post-exertional malaise
fatiguediffaterill	Fatigue/exhaustion after activity that would not cause fatigue before	Post-exertional malaise
fatiguelast24h*	Fatigue/exhaustion after effort, lasting >24h	Post-exertional malaise
malaise24h	Malaise after exertion, lasting >24h	Post-exertional malaise
markedfatigexertion	Marked physical/mental fatigue/exhaustion after minimal effort, lasting >24h	Post-exertional malaise
painexertion	Pain after exertion/effort, lasting >24h	Post-exertional malaise
worsesympmore24h	Worsening of symptoms after exertion/effort, lasting >24h	Post-exertional malaise
sleepproblems	Problems in sleep, quality of duration; insomnia	Sleep
unrefsleep	Unrefreshing sleep	Sleep

Table C.2: Percentage of individuals with presence of each symptom, across multiple sclerosis (MS), ME/CFS as a single cohort, and ME/CFS separated into four distinct subgroups. Symptoms marked with an asterisk were removed due to a high percentage of missing values and percentages were not estimated. Differences in sample sizes from Table 6.1 reflect the removal of some individuals due to a large proportion of missing data across the symptoms.

Symptom code	MS (n = 39)	ME/CFS (n = 222)	ME/CFS_S0 (n = 41)	ME/CFS_S1 (n = 43)	ME/CFS_S2 (n = 92)	ME/CFS_S3 (n = 46)
airhunger	25.64	58.48	60.98	41.86	63.04	62.50
bladderprob	71.79	56.70	56.10	44.19	59.78	62.50
dizzystandup	17.95	68.30	65.85	58.14	73.91	68.75
extremepale	10.26	49.55	43.90	41.86	47.83	64.58
ibssymptoms	38.46	78.12	82.93	72.09	76.09	83.33
intolstandup	41.03	51.79	43.90	37.21	61.96	52.08
lightheaded	30.77	73.21	68.29	65.12	78.26	75.00
palpitations*	—	—	—	—	—	—
palpitstandup	5.13	33.93	26.83	32.56	39.13	31.25
sick_nausea*	—	—	—	—	—	—
alcoholintoler*	—	—	—	—	—	—
feverchills	7.69	57.59	60.98	39.53	65.22	56.25
fewviralinfect*	—	—	—	—	—	—
flusymptoms	10.26	71.88	73.17	53.49	78.26	75.00
freqviralinfect	5.13	52.68	56.10	32.56	57.61	58.33
newsensit	15.38	66.07	56.10	62.79	68.48	72.92
sorethroat	7.69	71.88	65.85	60.47	77.17	77.08
stiffnessmorn	66.67	70.98	65.85	72.09	70.65	75.00
tenderlands	12.82	75.45	75.61	58.14	81.52	79.17
coldhandsfeet*	—	—	—	—	—	—
intolheat_cold	58.97	74.55	80.49	69.77	73.91	75.00
sexualfunction	58.97	57.14	56.10	58.14	60.87	50.00
unusualsweaty	7.69	55.36	51.22	51.16	55.43	62.50
weightchange*	—	—	—	—	—	—
worsepoststress	71.79	89.29	85.37	86.05	90.22	93.75
backweak	30.77	62.95	65.85	62.79	59.78	66.67
brainfog	53.85	77.23	70.73	72.09	83.70	75.00
concentprob	61.54	95.98	90.24	97.67	96.74	97.92
diffdecisions*	—	—	—	—	—	—
diffretaininfo	53.85	81.70	70.73	81.40	89.13	77.08
diffunderstand	46.15	82.59	73.17	81.40	89.13	79.17
diffwords*	—	—	—	—	—	—
disorientation	25.64	50.45	48.78	34.88	58.70	50.00
eyesightdistub	66.67	60.71	60.98	48.84	64.13	64.58
lossbalance	89.74	73.66	73.17	67.44	79.35	68.75
muscdisc	48.72	86.16	87.80	83.72	85.87	87.50
muscweak	66.67	85.27	90.24	72.09	86.96	89.58
neckweak	25.64	54.91	53.66	55.81	53.26	58.33
poorcoord	71.79	62.05	58.54	55.81	66.30	62.50
senstlightnoise	30.77	77.68	65.85	65.12	89.13	77.08
shortmem	61.54	83.48	80.49	74.42	90.22	81.25
slowthink	41.03	75.89	65.85	60.47	84.78	81.25
tingling	82.05	69.64	68.29	62.79	67.39	81.25
twitching*	—	—	—	—	—	—
chestabdpain	35.90	77.23	78.05	81.40	75.00	77.08
diffmigraine	10.26	38.39	31.71	39.53	39.13	41.67
jointpain*	—	—	—	—	—	—
jointpainnoinnfl	20.51	55.36	51.22	53.49	56.52	58.33
muscpain	51.28	87.95	90.24	86.05	86.96	89.58
newheadache	28.21	77.23	75.61	79.07	77.17	77.08
exerciseintol	41.03	81.70	75.61	72.09	84.78	89.58
fatiguedifaterrill	71.79	96.43	90.24	100.00	96.74	97.92
fatiguelast24h*	—	—	—	—	—	—
malaise24h	30.77	95.98	92.68	90.70	97.83	100.00
markedfatigexertion	33.33	77.68	73.17	69.77	82.61	79.17
painexertion	25.64	75.89	70.73	72.09	75.00	85.42
worsesympmore24h	30.77	91.07	85.37	88.37	94.57	91.67
sleepproblems	51.28	85.71	78.05	81.40	90.22	87.50
unrefsleep	61.54	98.66	100.00	100.00	97.83	97.92

Table C.3: AUC and its 95% confidence interval (CI), optimal cutoff and associated sensitivity (Se) and specificity (Sp) to discriminate patients with multiple sclerosis (cases) from healthy controls. In the Direction column, the symbols “>” and “<” represent higher value in MS cases than in healthy controls and vice-versa, respectively. In the Cutoff column, the p-value within brackets came from the Pearson’s χ^2 test with Yates’ correction for 2×2 tables after being adjusted for a false discovery rate of 5% using the Benjamini-Hochberg procedure. In the AUC column, the symbol “*” denote the cases where there is evidence of AUC different from 0.50 (random guess).

Herpesvirus	Direction	AUC (95% CI)	Cutoff (p-value)	Se	Sp
CMV	controls > cases	0.513 (0.398, 0.629)	196.84 (0.0155)	0.9000	0.000
EBV-EBNA1	controls < cases	0.584 (0.488, 0.680)	20.515 (0.0073)	0.950	0.306
EBV-VCA	controls < cases	0.641 (0.538, 0.744)	139.235 (0.0073)	0.650	0.643
HHV6	controls < cases	0.539 (0.427, 0.650)	148.91 (0.0799)	0.125	0.969
HSV1	controls > cases	0.626 (0.531, 0.721)*	3.72 (0.0073)	0.950	0.327
HSV2	controls < cases	0.544 (0.437, 0.650)	6.875 (0.0799)	0.550	0.633
VZV	controls > cases	0.573 (0.461, 0.685)	172.575 (0.0595)	0.425	0.765

C.2 Supplementary figures

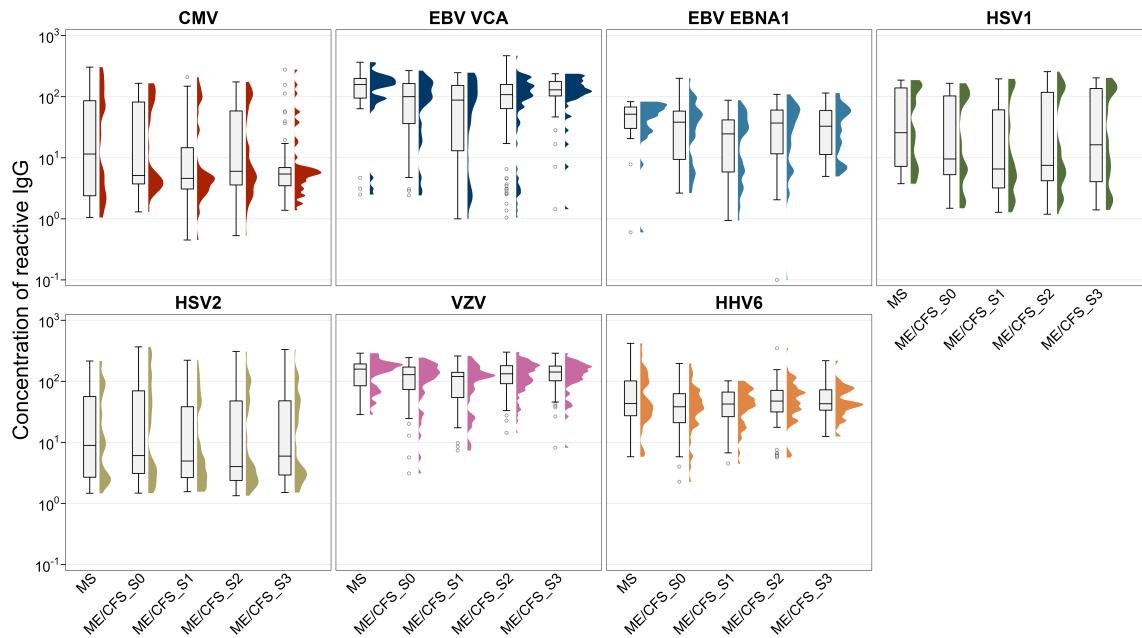


Figure C.1: Quantitative serology data per study group and herpesvirus/antigen.

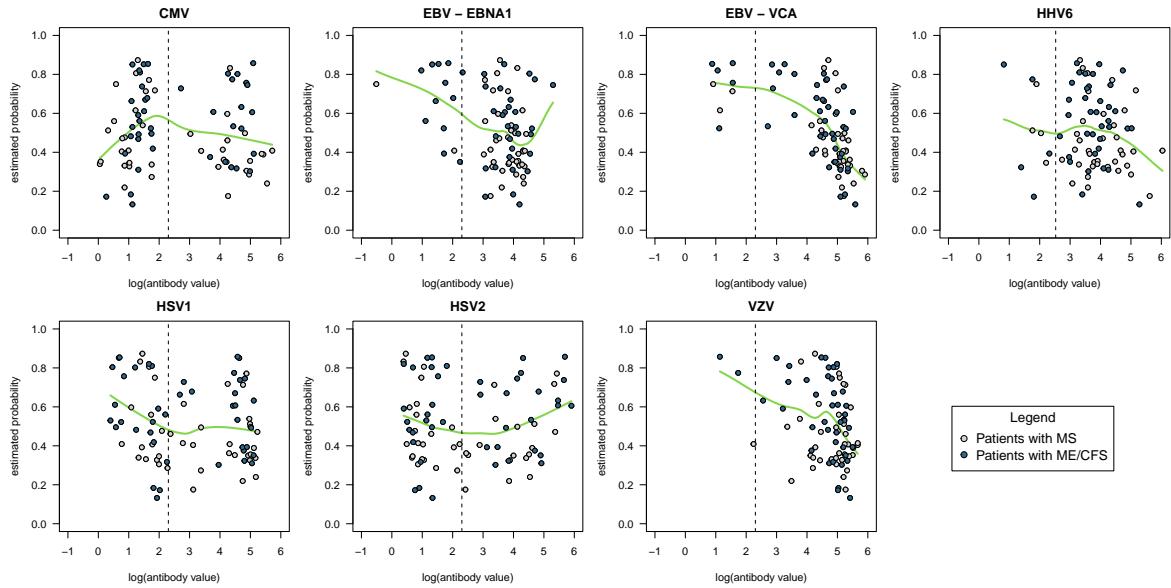


Figure C.2: Smooth-line approximations (green lines) of the relationship between $\log(\text{antibody value})$ and SL-estimated probability of ME/CFS_S0 patient when compared to patients with MS. Each dot represents a patient and the vertical dashed lines represent the cut-off for seropositivity according to the respective lab protocol.

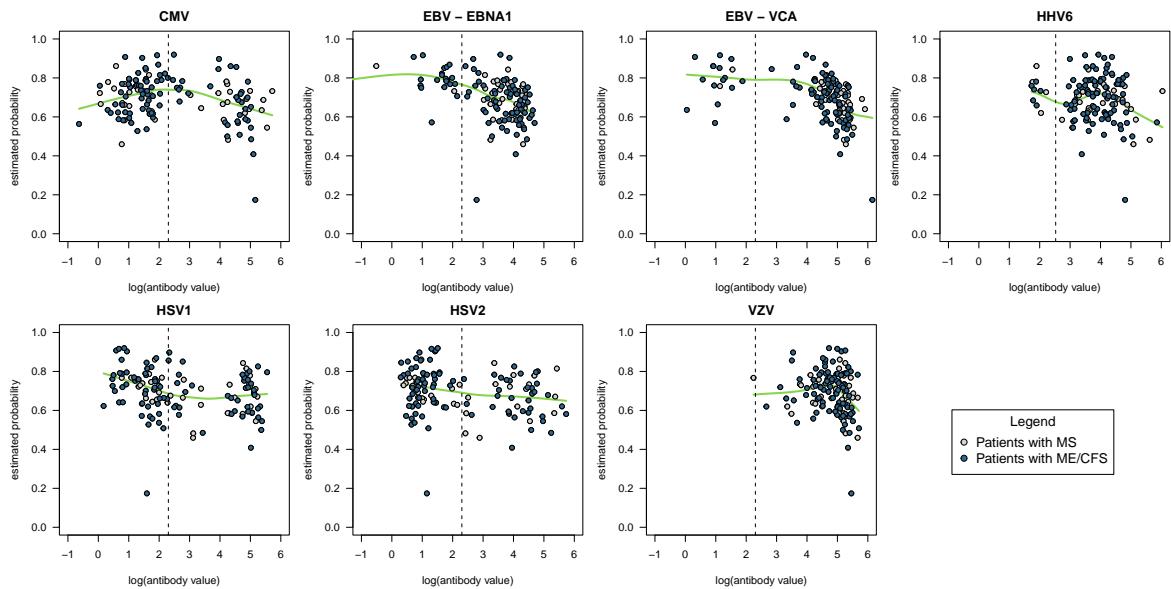


Figure C.3: Smooth-line approximations (green lines) of the relationship between $\log(\text{antibody value})$ and SL-estimated probability of ME/CFS_S2 patient when compared to patients with MS. Each dot represents a patient and the vertical dashed lines represent the cut-off for seropositivity according to the respective lab protocol.

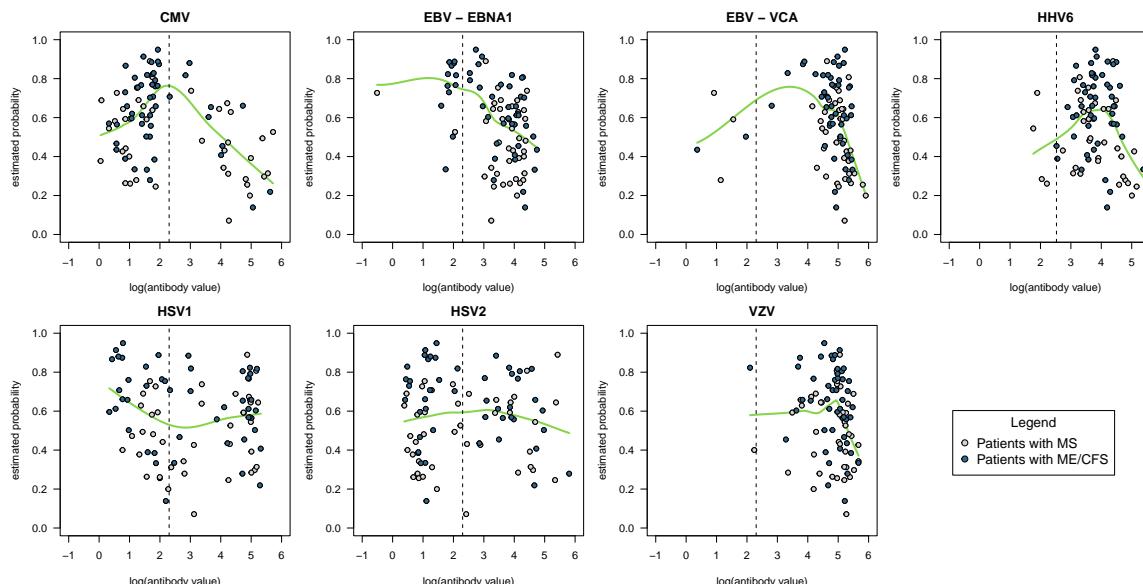


Figure C.4: Smooth-line approximations (green lines) of the relationship between $\log(\text{antibody values})$ and SL-estimated probability of ME/CFS_S3 patient when compared to patients with MS. Each dot represents a patient and the vertical dashed lines represent the cut-off for seropositivity according to the respective lab protocol.

Appendix D

Additional information to Chapter 7

D.1 Supplementary tables

Table D.1: Basic descriptive characteristics of healthy controls and ME/CFS patients. Patients were considered to be a single population, two subgroups based on the severity of symptoms (mild to moderate or severely affected), or four subgroups according to the responses about infection triggers given in the symptoms' assessment questionnaire (S0, S1, S2, or S3). P-values refer to Pearson's Chi-squared test for qualitative measures and the two-sided Kruskal-Wallis rank sum test on the quantitative medians. HC, Healthy control; S0, Do not know; S1, Non-infection trigger; S2, Infection trigger but not confirmed with a lab test; S3, Infection trigger confirmed with a lab test; SD, Standard deviation; Min, Minimum; Max, Maximum.

	HC (n = 106)	ME/CFS (n = 241)	P-value	Severity			Infection trigger				P-value, ME/CFS only	P-value	
				Mild/Moderate (n = 188)	Severe (n = 53)	P-value, ME/CFS only	P-value	S0 (n = 61)	S1 (n = 43)	S2 (n = 91)	S3 (n = 47)		
Sex			0.819			≈1.000	0.922					0.642	0.775
Female (%)	79 (74.5)	184 (76.3)		144 (76.6)	40 (75.5)			46 (75.4)	33 (76.7)	72 (80.0)	33 (70.2)		
Male (%)	27 (25.5)	57 (23.7)		44 (23.4)	13 (24.5)			15 (24.6)	10 (23.3)	18 (20.0)	14 (29.8)		
Age			0.827			0.799	0.947					0.338	0.497
Mean (SD)	41.7 (11.7)	42.2 (11.0)		42.0 (11.0)	42.5 (11.1)			43.5 (10.2)	40.2 (13.0)	43.1 (10.8)	40.5 (10.4)		
Median [Min, Max]	43.0 [18.0, 60.0]	43.0 [18.0, 60.0]		43.0 [18.0, 60.0]	46.0 [18.0, 59.0]			44.0 [22.0, 59.0]	41.0 [18.0, 60.0]	43.5 [18.0, 59.0]	41.0 [18.0, 57.0]		
Disease duration						0.001	—					0.495	—
Mean (SD)	—	12.4 (8.3)	—	11.4 (7.98)	15.8 (8.58)			12.3 (7.79)	11.6 (8.59)	12.0 (8.57)	13.7 (8.28)		
Median [Max, Min]	—	11.2 [0.2, 39.9]	—	9.8 [0.2, 38.0]	15.7 [1.5, 39.9]			11.1 [1.4, 37.0]	9.5 [1.6, 38.0]	11.0 [0.2, 39.9]	12.9 [1.1, 29.3]		
Missing (%)	106 (100)	5 (2.1)		5 (2.7)	0 (0.0)			2 (3.3)	2 (4.7)	1 (1.1)	0 (0.0)		
ME/CFS diagnosis						0.004	—					0.095	—
1994 CDC only (%)	—	33 (13.7)	—	33 (17.6)	0 (0.0)			12 (19.7)	10 (23.3)	6 (6.7)	5 (10.6)		
2003 CCC only (%)	—	3 (1.2)	—	2 (1.1)	1 (1.9)			0 (0.0)	1 (2.3)	1 (1.1)	1 (2.1)		
1994 CDC/2003 CCC (%)	—	205 (85.1)	—	153 (81.4)	52 (98.1)			49 (80.3)	32 (74.4)	83 (92.2)	41 (87.2)		

Table D.2: List of 57 ordinal symptoms available in the data with respective description, and domain. Symptoms marked with an asterisk were removed from the analysis due to the percentage of missing values across all patients surpassing 20% (see Figure 7.1).

Symptom	Description (simplified)	Domain
airhunger	Air hunger or dyspnea	Autonomic, Neurophysiological
bladderprob	Bladder problems	Autonomic
dizzystandup	Dizziness while standing up	Autonomic
extremepale	Being extremely pale	Autonomic
ibssymptoms	IBS symptoms	Autonomic
intolstandup	Intolerance to standing up on your feet	Autonomic, Neurophysiological
lightheaded	Feeling lightheaded	Autonomic, Neurophysiological
palpitations*	Palpitations	Autonomic, Neurophysiological
palpitstandup	Palpitations while standing up	Autonomic, Neurophysiological
sick_nausea*	Feeling sick/nauseated	Autonomic
alcoholintoler*	Intolerance to alcohol	Immunological
feverchills	Fever or chills	Immunological
flusymptoms	Flu-like symptoms	Immunological
freqviralinfect	Frequent viral infections	Immunological
newsensit	New sensitivities to food, medications, chemicals, odours, others	Immunological
sorethroat	Sore throat	Immunological
stiffnessmorn	Stiffness in the mornings	Immunological
tenderglands	Tender glands in neck or armpit	Immunological
backweak	Back weakness	Neurocognitive
brainfog	“Brain fog”/Confusion	Neurocognitive
concentprob	Trouble concentrating	Neurocognitive
diffdecisions*	Difficulty making decisions	Neurocognitive
diffretaininfo	Difficulty retaining information	Neurocognitive
diffunderstand	Difficulty in understanding things/thinking clearly	Neurocognitive
diffwords*	Difficulty finding/saying words	Neurocognitive
disorientation	Disorientation	Neurocognitive
eyesightdistub	Temporary disturbance in eyesight	Neurocognitive
lossbalance	Loss of balance/Unsteadiness of feet while standing up	Neurocognitive, Neurophysiological
muscdisc	Muscle discomfort	Neurocognitive
muscweak	Muscle weakness	Neurocognitive
neckweak	Neck weakness	Neurocognitive
poorcoord	Poor coordination/Unsteady movements	Neurocognitive
senstlightnoise	Unusual sensitivity to light/noise	Neurocognitive
shortmem	Short-term memory problems	Neurocognitive
slowthink	Slow thinking	Neurocognitive
tingling	Tingling/numbness in arms/legs	Neurocognitive
twitching*	Muscle twitching	Neurocognitive
coldhandsfeet*	Unusually cold hands/feet	Neuroendocrine
intolheat_cold	Intolerance to extremes of heat/cold	Neuroendocrine
sexualfunction	Decreased sexual interest/function	Neuroendocrine
unusualsweaty	Being unusually sweaty	Neuroendocrine
weightchange*	Abnormal appetite/Significant changes in weight	Neuroendocrine
worsepoststress	Worsening of symptoms with stress	Neuroendocrine
chestabdpain*	Pain in chest/abdomen	Pain
diffmigraine	Migraine different/worse than before	Pain
jointpain	Pain in ≥ 2 joints without swelling/redness	Pain
jointpainnoinnfl	Joint pains moving to different joints without swelling/redness	Pain
muscpain	Muscle pain	Pain
newheadache	Headaches new/different/worse than before	Pain
exerciseintol	Intolerance to exercise	PEM
fatiguelast24h*	Fatigue/Exhaustion after normal levels of activity, lasting 24 hours	PEM
malaise24h	Malaise after exertion/effort, lasting 24 hours	PEM
markedfatigexertion	Marked physical/mental fatigue/exhaustion after minimal exertion/effort, lasting 24 hours	PEM
painexertion	Pain after exertion/effort, lasting 24 hours	PEM
worsesympmore24h	Worsening of symptoms after exertion/effort, lasting 24 hours	PEM
sleepproblems	Problems in sleep quality/duration	Neurophysiological, Sleep
unrefsleep	Unrefreshing sleep	Neurophysiological, Sleep

Table D.3: Samples sizes (and rounded percentages, row-wise) of the 241 ME/CFS patients in each latent class estimated on each domain. For each domain, the latent classes are arranged by increasing level according to the severity profile from the response probabilities (Supplementary Figure D.2). Each domain is independent. The optimal number of classes on each domain was chosen based on the Bayesian information criterion (BIC). P-values refer to Pearson's χ^2 test.

Domain	Latent classes, n (%)			
	g1	g2	g3	P-value
Total	79 (32.78)	104 (43.15)	58 (24.07)	0.001
Immunological	138 (57.26)	103 (42.74)	—	0.024
Neuroendocrine	143 (59.34)	98 (40.66)	—	0.004
PEM	45 (18.67)	72 (29.88)	124 (51.45)	<0.001
Autonomic	108 (44.81)	89 (36.93)	44 (18.26)	<0.001
Neurocognitive	91 (37.76)	98 (40.66)	52 (21.58)	<0.001
Neurophysiological	113 (46.89)	92 (38.17)	36 (14.94)	<0.001
Pain	87 (36.10)	89 (36.93)	65 (26.97)	0.110

D.2 Supplementary figures

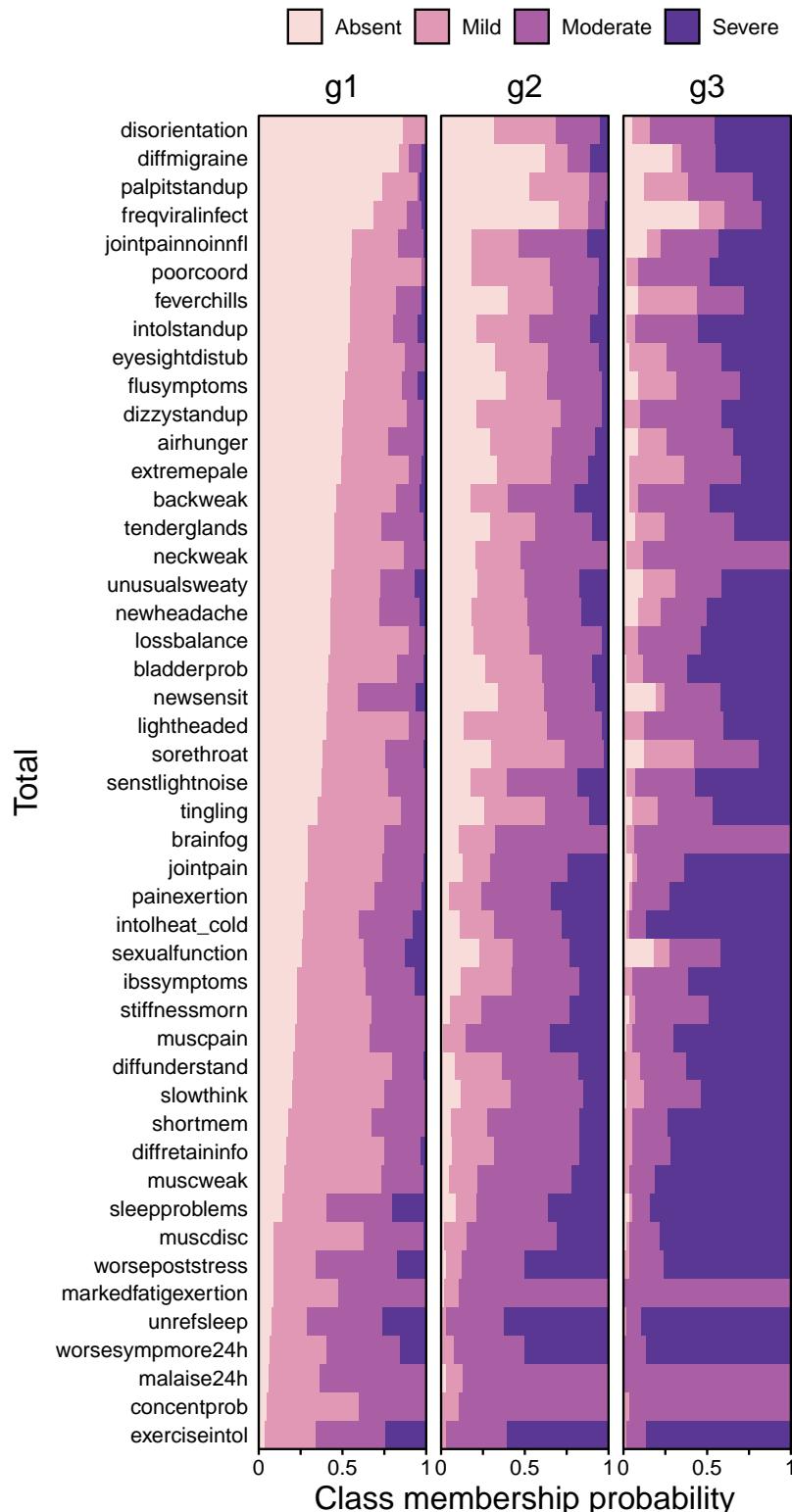


Figure D.1: Latent class analysis estimated class membership probabilities of symptom severities on different BIC-based subgroups (latent classes), when profiling ME/CFS patients with the totality of available symptoms. Subgroups were ordered by increasing severity in the response probabilities of symptoms. A more detailed description of each symptom can be found elsewhere (Supplementary Table D.2).

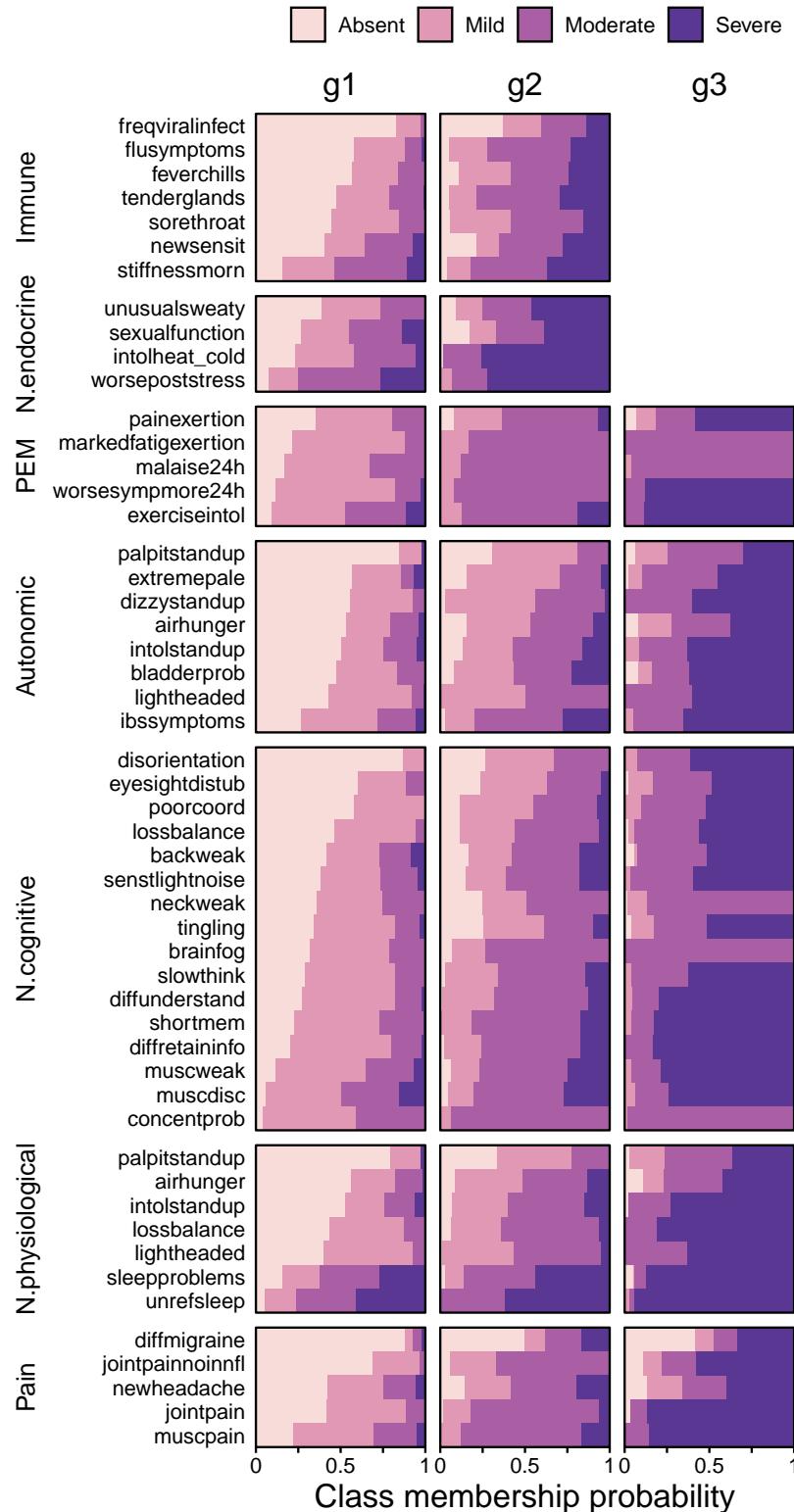


Figure D.2: Latent class analysis estimated class membership probabilities of symptom severities on different BIC-based subgroups (latent classes), across the different domains, autonomic, immunological, neurocognitive, neuroendocrine, neurophysiological, pain, and post-exertional malaise (PEM). For each domain, subgroups were ordered by increasing severity in the response probabilities of symptoms that make up each domain. A more detailed description of each symptom can be found elsewhere (Supplementary Table D.2).

Appendix E

Additional information to Chapter 8

E.1 Supplementary tables

Table E.1: Nineteen and eight CpG probes located in ACE and ACE2 and shared between Infinium HumanMethylation450K and Infinium HumanMethylationEPIC arrays by Illumina. The annotation was obtained from the packages *IlluminaHumanMethylation450kanno.ilmn12.hg19* and *IlluminaHumanMethylationEPICannoilm10b2.hg19* available from Bioconductor.

Gene (chr)	Probes	Position	Annotation
ACE (17)	cg09920557	61553938	TSS1500
	cg25054907	61553954	TSS1500
	cg02131967	61554106	TSS1500
	cg02440279	61554400	TSS200
	cg02040921	61554411	TSS200
	cg19354750	61554413	TSS200
	cg05952120	61554416	TSS200
	cg24877195	61554604	1stExon
	cg06751221	61554929	Body
	cg02261408	61559061	Body
	cg19802564	61561470	Body;TSS1500
	cg19826045	61561602	Body;TSS1500
	cg21796427	61562170	TSS200;Body
	cg04199256	61562197	Body;5'UTR;1stExon
	cg21094739	61572645	Body;Body
	cg01489398	61574335	Body;Body
	cg21881537	61574411	Body;Body
	cg21657705	61574500	Body;Body
	cg10468385	61574744	3'UTR;3'UTR
ACE2 (X)	cg23232263	15579482	3'UTR
	cg05039749	15583512	Body
	cg05748796	15619337	5'UTR
	cg16734967	15620103	5'UTR
	cg08559914	15620240	TSS200
	cg18877734	15621084	TSS1500
	cg21598868	15621167	TSS1500
	cg18458833	15621477	TSS1500

Table E.2: Summary data of the CpG probes including SNP or coincided with a polymorphic SNP.

Annotation	Probe ID	Chromosome	Position	SNP ID	Ref allele/ Alt allele	Minor allele frequency	
						North America	Europe
SNP within probes	cg04199256	17	61562197	rs191697444	C/T	0.00	0.00
	cg04199256	17	61562197	rs12720723	G/A	0.12	<0.01
	cg21881537	17	61574411	rs117135474	C/T	0.00	0.00
	cg21881537	17	61574411	rs200695691	TTGCC/T	0.03	0.00
	cg10468385	17	61574744	rs4365	G/A	0.00	0.04
	cg02481451	17	61594924	rs149678437	C/T	<0.01	<0.01
	cg16734967	X	15620103	rs182809041	A/C	<0.01	0.00
	Polymorphic SNP	X	15620240	rs186143966	C/T	0.00	0.00

Table E.3: Estimates of the best linear regression models for 5 significant CpG probes shown in Figure 8.1C. Healthy controls and the study of de Vega et al. (2014) were considered the reference effects of the disease status and of the study indicator variable, respectively. The respective data are shown in Figure 8.2.

Analysis	Coefficient	ACE						ACE2		
		cg09920557		cg19802564		cg21094739		cg10468385		
		Estimate (SE)	P-value	Estimate (SE)	P-value	Estimate (SE)	P-value	Estimate (SE)	P-value	
Overall	Intercept	-3.769 (0.065)	<0.001	1.551 (0.081)	<0.001	1.938 (0.096)	<0.001	3.082 (0.100)	<0.001	
	Disease status:ME/CFS	-0.060 (0.092)	0.519	-0.016 (0.115)	0.889	-0.064 (0.136)	0.638	-0.118 (0.142)	0.409	
	Study:de Vega et al.(2017)	-2.367 (0.079)	<0.001	0.7071 (0.098)	<0.001	1.419 (0.116)	<0.001	2.437 (0.121)	1.703 (0.085)	
	Study:Trivedi et al.(2018) ²	0.145 (0.092)	0.117	—	—	—	(0.142)	-0.625 (0.10334)	-0.88672 <0.001	
	Study:Herrera et al.(2018)	-2.245 (0.073)	<0.001	0.425 (0.091)	<0.001	0.740 (0.107)	<0.001	1.772 (0.112)	<0.001	
	Disease status:ME/CFS× Study: de Vega et al.(2017)	0.131 (0.108)	0.224	-0.107 (0.133)	0.425	-0.198 (0.158)	0.210	-0.159 (0.165)	0.336 (0.082)	
	Disease status:ME/CFS× Study:Trivedi et al.(2018)	-0.315 (0.130)	0.016	—	—	—	(0.199)	0.264 (0.157)	— <0.001	
	Disease status:ME/CFS× Study:Herrera et al.(2018)	0.048 (0.102)	0.638	0.072 (0.127)	0.571	0.157 (0.150)	0.295	0.181 (0.157)	0.250 <0.001	
	Females	Intercept	-3.769 (0.068)	<0.001	1.406 (0.107)	<0.001	1.938 (0.093)	<0.001	3.082 (0.092)	<0.001
	Disease status:ME/CFS	-0.060 (0.096)	0.535	-0.224 (0.152)	0.143	-0.064 (0.132)	0.629	-0.117 (0.130)	0.366	-0.141 (0.045)
	Study:de Vega et al.(2017)	-2.367 (0.082)	<0.001	0.499 (0.130)	<0.001	1.419 (0.113)	<0.001	2.437 (0.111)	<0.001	1.703 (0.075)
	Study:Trivedi et al.(2018) ²	0.145 (0.096)	0.131	—	—	—	(0.130)	-0.625 (0.090)	-0.887 <0.001	
	Study:Herrera et al.(2018)	-2.238 (0.079)	<0.001	-0.233 (0.124)	0.062	0.770 (0.108)	<0.001	1.786 (0.106)	<0.001	1.379 (0.073)
	Disease status:ME/CFS× Study: de Vega et al.(2017)	0.131 (0.112)	0.241	-0.112 (0.177)	0.526	-0.198 (0.153)	0.198	-0.159 (0.151)	0.293	— <0.001
	Disease status:ME/CFS× Study:Trivedi et al.(2018)	-0.315 (0.134)	0.020	—	—	—	(0.182)	0.264 (0.148)	0.148	— <0.001
	Disease status:ME/CFS× Study:Herrera et al.(2018)	0.029 (0.109)	0.791	0.258 (0.173)	0.138	0.166 (0.150)	0.270	0.241 (0.148)	0.104	— <0.001

Appendix F

Publications discussed in this thesis

F.1 Chapter 2

Statistical challenges of investigating a disease with a complex diagnosis

João Malato

Instituto de Medicina Molecular, Faculdade de Medicina & Centro de Estatística e Aplicações da Universidade de Lisboa, Lisboa, Portugal, *jmalato@medicina.ulisboa.pt*

Luís Graça

Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal, *lgraca@medicina.ulisboa.pt*

Nuno Sepúlveda

Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, United Kingdom & Centro de Estatística e Aplicações da Universidade de Lisboa, Lisboa, Portugal, *nuno.sepulveda@lshtm.ac.uk*

Keywords: Multidimensional scaling; Cluster analysis; Misclassification; Cohen's κ coefficient; Jacard's similarity index

Abstract: Given the absence of a disease-specific biomarker, there are more than 20 symptoms-based case definitions of myalgic encephalomyelitis/chronic fatigue syndrome. As a consequence, the diagnosis for a given patient could vary from one case definition to another. In this context, we analyse data from a biobank dedicated to this disease in order to study the agreement between different case definitions, the similarity between symptom's profile among all participants including healthy controls and patients with multiple sclerosis. We also investigate the impact of patients' misclassification on a hypothetical association analysis using data simulation.

1 Introduction

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a complex disease whose patients manifest unexplained fatigue lasting for more than six months [1] or suffer from post-exertional malaise that is not alleviated by rest [2]. Disease prevalence has been estimated between 0.4% and 1.0% affecting six women to one man [3]. The underlying pathological mechanisms remain poorly understood, but they are often associated with environmental stressors, including severe viral infections [4].

Until now there is no accurate biomarker for disease diagnosis. To overcome this problem, researchers and clinicians altogether have proposed more than 20 different case definitions based on patients' symptomatology while excluding known diseases that could explain the fatigue reported by suspected cases [5]. As a consequence, the diagnosis for a given patient can vary from one case definition to another. Therefore, research from ME/CFS could be affected by the inclusion of false positive cases in the respective data.

In the present paper, we discuss the problem of diagnosing ME/CFS using data from the United Kingdom ME/CFS Biobank (UKMEB). With this purpose, we first introduce the biobank and its data. We then assess the agreement between 4 common case definitions of ME/CFS in 275 suspected cases belonging to the UKMEB. We then estimate the similarity between symptom's severity profiles from suspected cases, patients with multiple sclerosis, and healthy controls. We also study the impact of patients' misclassification on the statistical power of a hypothetical association analysis. Finally, we conclude this paper with some final remarks.

2 The UKMEB

The UKMEB refers to a large data set of suspected cases of ME/CFS, healthy controls, and patients with multiple sclerosis included as an additional control group [6]. In terms of recruitment, suspected ca-

ses were identified in different institutions across the National Health Service from the United Kingdom and then referred to the CureMe group, a dedicated clinical research team based in the London School of Hygiene & Tropical Medicine and responsible for recruiting, managing, and curating the biobank. For this paper, the data set under analysis consists of a total of 523 participants divided into 275 suspected cases of ME/CFS, 136 healthy controls, and 112 patients with multiple sclerosis.

3 Diagnostic agreement analysis

After patients' referral for a possible integration in the biobank, suspected cases were comprehensively evaluated according to four case definitions of ME/CFS: Centre for Disease Control criteria (CDC-1994) [1], Canadian Consensus Criteria (CCC-2003) [2], Institute of Medicine Criteria (IOM-2005) [7], and International Consensus Criteria (ICC-2011) [8]. The CDC-1994 requires the patients to have unexplained fatigue for at least 6 months and at least four out of eight fatigue-related symptoms. The IOM-2005 is typically used by general practitioners and it requires the patients to show at least three main symptoms such as profound fatigue, post-exertional malaise, and unrefreshing sleep. The CCC-2003 requires the patients to manifest four or more fatigue specific symptoms, at least two neurological or cognitive ones, and at least one autoimmune, neuroendocrine, or immune symptom. Finally, the ICC-2011 is more focused on neuro-immune and cognitive symptoms, and on the inability to produce sufficient energy on demand (post-exertional neuroimmune exhaustion).

There were 269 (97.8%), 233 (84.7%), 229 (83.3%) and 213 (77.5%) out of 275 suspected cases whose symptoms agreed with CDC-1994, IOM-2005, CCC-2003, and ICC-2011, respectively (Table 1). This finding suggests that the general practitioners who referred the suspected cases to a possible integration in the biobank made their diagnosis based on the CDC-1994. Unsurprisingly, only 62.9% of

the suspected cases ($n = 173$) had a positive diagnosis across all the four case definitions. Therefore, the remaining suspected cases had at least one negative diagnosis.

Table 1: Frequency of suspected cases of ME/CFS according to their diagnostic outcomes using different case definitions. Percentages in the last row indicate the proportion of diagnosed cases by each case definition.

Case definition				N	% of total suspected cases
CDC-1994	IOM-2005	CCC-2003	ICC-2011		
+	+	+	+	173	62.9
+	+	+	-	32	11.6
+	-	+	+	16	5.8
+	+	-	+	16	5.8
+	-	-	-	14	5.1
+	+	-	-	10	3.6
+	-	+	-	5	1.8
+	-	-	+	3	1.1
-	-	+	+	3	1.1
-	-	-	+	1	0.4
-	+	-	-	1	0.4
-	+	-	+	1	0.4
97.8%	84.7%	83.3%	77.5%	275	100%

It is worth noting that there were no suspected cases who had a negative diagnosis across all case definitions. There were also three individuals whose symptoms agreed with ICC-2011 only, IOM-2005 only, or both criteria. These individuals were considered to be fatigued but non-ME/CFS patients given that they did not agree with either the CDC-1994 or the CCC-2003 as recommended for ME/CFS research [9].

To better understand the agreement between diagnostic outcomes obtained from different case definitions, we used the Jaccard's similarity index, J [10]. Note that this index is usually a measure used to compare objects with shared attributes. Here we instead applied this index to compare attributes themselves. For a pair of

case definitions (C_i, C_j), this index was estimated as

$$J(C_i, C_j) = \frac{S}{S_i + S_j - S}, \quad i, j = 1, \dots, 4, \quad (1)$$

where S_i and S_j are the number of suspected cases with a positive diagnosis by C_i and C_j , respectively, and S is the number of suspected cases with a positive diagnosis by both criteria. In theory, the index is defined between 0 and 1 (i.e., no and full agreement between C_i and C_j across all individuals, respectively).

The estimates of this index ranged from 0.752 (IOM-2005 versus ICC-2011) to 0.876 (CDC-1994 versus IOM-2005; CDC-1994 versus CCC-2003) (Table 2). The estimates showed the stringency and differences in scope of each case definition. In addition, these estimates showed that, even if the general practitioners applied two different case definitions of ME/CFS in their diagnosis, there could still be a fraction of suspected cases where the respective diagnostic outcomes might not agree with each other.

Table 2: Estimates of the Jaccard's similarity index for the four case definitions of ME/CFS using data from the UKMEB.

	CDC-1994	IOM-2005	CCC-2003	ICC-2011
CDC-1994	1.000	0.876	0.876	0.760
IOM-2005	0.876	1.000	0.840	0.752
CCC-2003	0.876	0.840	1.000	0.753
ICC-2011	0.760	0.752	0.753	1.000

4 Symptoms' similarity analysis

A major advantage of using data from the UKMEB is the comprehensive symptom's characterisation of all study participants. In particular, each participant had to report the severity of 57 symptoms occurred a month before data collection. Severity of each symptom was categorised into absence, mild, moderate, and severe. These invaluable data were then analysed to assess the similarity of

all participants in terms of their symptom's severity profile. With this purpose, we first computed all possible 4×4 contingency tables resulting from cross-tabulating the symptom's severity data for any given pair of participants (i, j) , $i, j = 1, \dots, 523$. We then calculated a similarity matrix between any given pair of individuals by estimating the Cohen's κ coefficient [11] in the corresponding 4×4 contingency tables, that is,

$$\kappa_{ij} = \frac{\sum_{k=1}^4 p_{ij,kk} - \sum_{k=1}^4 p_{ij,k} \cdot p_{ij,\cdot k}}{1 - \sum_{k=1}^4 p_{ij,k} \cdot p_{ij,\cdot k}}, \quad (2)$$

where $k = 1, \dots, 4$, $p_{ij,kk}$ is the proportion of symptoms with severity k reported by both individuals i and j , $p_{ij,k\cdot}$ is the proportion of symptoms with severity k reported by individual i , and $p_{ij,\cdot k}$ is the proportion of symptoms with severity k reported by individual j . The resulting similarity matrix was then analysed by classical multidimensional scaling (MDS; Figure 1A) and hierarchical cluster analysis using complete linkage (Figure 1B).

With respect to the classical MDS, the first two components could explain 33.1% of the total inertia (Figure 1A). More importantly, the first component clearly discriminated healthy controls from suspected cases of ME/CFS. In the same component, patients with multiple sclerosis and the three fatigued non-ME/CFS cases were located between these two groups with some overlap. As expected, healthy participants were the most homogeneous cohort due to an absence or, at most, mild severity of the different symptoms. In contrast, the suspected cases of ME/CFS consisted of a diverse group as evidenced by their wide spread in the plot. Interestingly, a few suspected cases of ME/CFS had symptom's severity profiles similar to the ones from healthy controls. In agreement with these observations, the hierarchical cluster analysis revealed that some suspected cases of ME/CFS could be placed in clusters together with healthy controls and patients with multiple sclerosis (Figure 1B); a detailed analysis on the optimal number of clusters will be done elsewhere. Therefore, it was reasonable to assume that some of the suspected

cases of ME/CFS, although agreeing with CDC-1994 or CCC-2003, could be in fact true cases of another disease, as discussed by Nacul et al. [12].

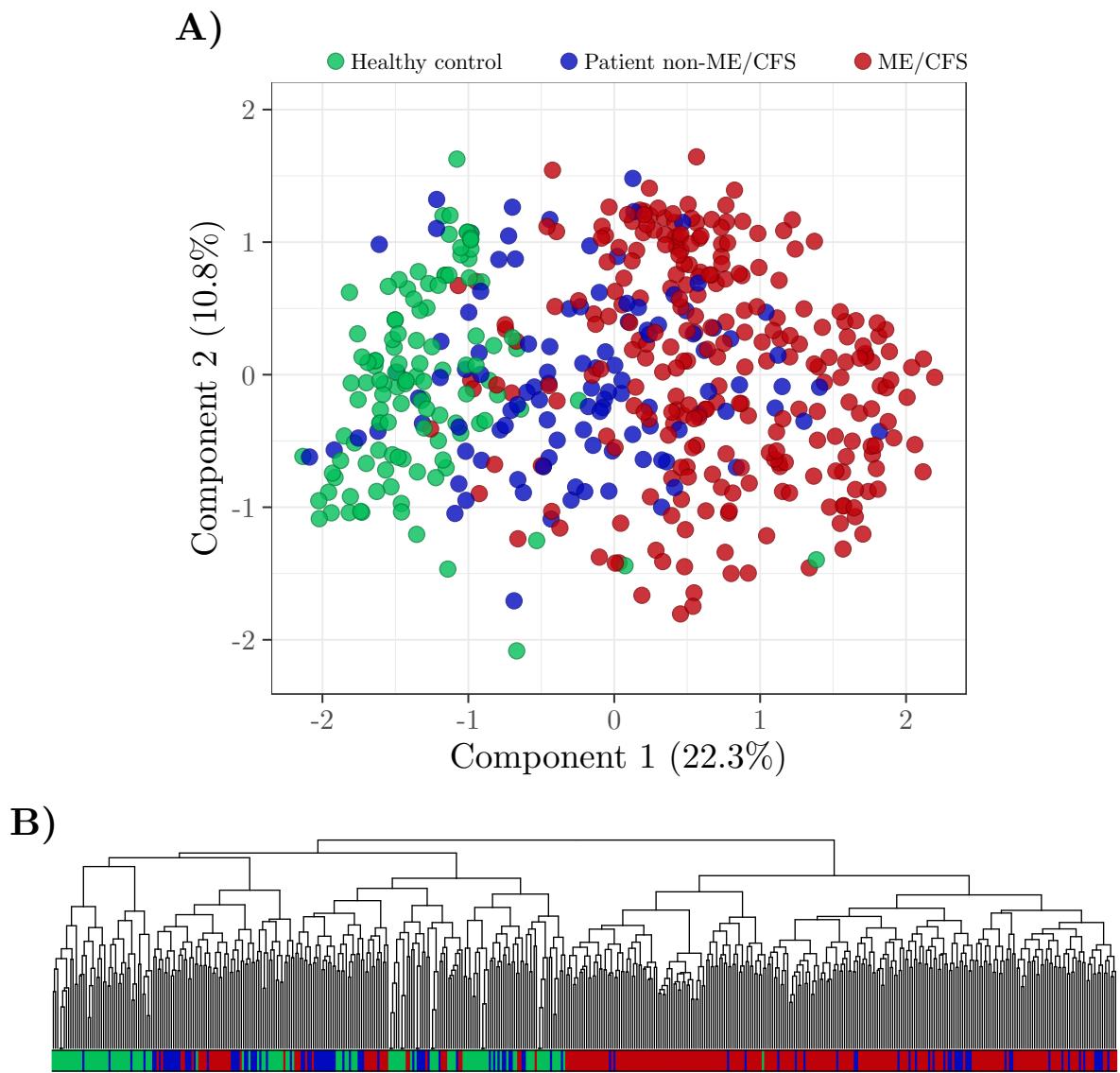


Figure 1: Symptom's similarity analysis based on the Cohen's κ coefficient: classical multidimensional scaling (A); dendrogram of hierarchical clustering analysis based on complete linkage (B) where the colour coding at the bottom is the same shown in A.

5 Impact of misclassification on an association analysis

Given the possibility of patients' misclassification, we performed a small simulation study to assess the reduction of statistical power attributed to this issue in the context of an association analysis. With this purpose, we simulated data from a case-control study with the aim to investigate a hypothetical association of a binary exposure variable (exposed versus not exposed) with ME/CFS. In this scenario, the observable data could be summarised by a 2×2 frequency table whose sampling distribution was given by the following product of two Binomial distributions,

$$f(x_0, x_1 | n_0, n_1; \theta_0, \theta_1) = \prod_{i=0,1} \binom{n_i}{x_i} \theta_i^{x_i} (1 - \theta_i)^{n_i - x_i}, \quad (3)$$

where x_0 and x_1 are the frequencies of exposed healthy controls and suspected cases, respectively, n_0 and n_1 are the associated sample sizes, and θ_0 and θ_1 are the corresponding probabilities of exposure in healthy controls and suspected cases.

To study the impact of a potential misclassification of suspected cases on the detection of a possible association, four main assumptions were considered for the simulated data: (i) suspected cases could be divided into apparent (or false positive) cases and true positive cases of ME/CFS; (ii) the apparent cases were deemed equivalent to healthy controls in terms of degree of exposure, i.e., the probability of exposure in these individuals was given by θ_0 ; (iii) there was an overall misclassification rate, γ , for the suspected cases; and (iv) misclassification was only dependent on the true clinical status of each suspected case. Under the assumption (ii) and the law of total probability, the probability of exposure associated with suspected cases could be written as

$$\theta_1 = \gamma\theta_0 + (1 - \gamma)\theta_1^*, \quad (4)$$

where θ_1^* is the probability of exposed true cases.

We then studied the power of rejecting the null hypothesis of lack of association (i.e., H_0 : odds ratio = 1) by the Pearson's χ^2 test for independence, when considering this simple misclassification scenario. Similar investigation could have been done using Fisher's exact test instead. With this purpose, we used simulation to estimate the number of times that H_0 could be rejected at a significance level of 5%.

We augmented the observable 2×2 frequency table where the suspected cases were subdivided into apparent and true positive cases (Table 3). In this case, we simulated data from healthy controls according to the Binomial distribution with a sample size of n_0 individuals and probability of success θ_0 . With respect to the suspected cases, we simulated data from a Multinomial distribution with a sample size of n_1 individuals and probability vector given by the probabilities shown in Table 3. Note that, given assumption (iv), the associated Multinomial distribution could be decomposed into the following Binomial distribution

$$n_{1,m}|n_1; \gamma \sim \text{Bin}(n_1, \gamma) , \quad (5)$$

referring to how many individuals were hypothetically misclassified as true positive cases, and two Binomial distributions conditional to $n_{1,m}$

$$X_{1,F}|n_{1,m}; \theta_0 \sim \text{Bin}(n_{1,m}, \theta_0) , \quad (6)$$

and

$$X_{1,T}|n_1 - n_{1,m}; \theta_1^* \sim \text{Bin}(n_1 - n_{1,m}, \theta_1^*) , \quad (7)$$

where $X_{1,F}$ and $X_{1,T}$ were the random variables referring to the number of exposed false positive and true positive cases, respectively. For illustrative purposes, we performed our simulation study with $n_0 = n_1 = 100$, $\theta_0 = 0.25$, and $\theta_1^* = 0.35$. According to this parameter specification, the odds ratio of true positive cases versus healthy controls was 1.62, a low but reasonable value for a putative association with ME/CFS, given that there is no disease-specific biomarker. To estimate the power of rejecting H_0 , we generated 10,000 data sets

for each value of γ , ranging from 0 (no misclassification) to 1 (full misclassification) with a lag of 0.01. In each data set, H_0 was rejected if the p-value of the Pearson's χ^2 test was less than 0.05. For a given parameter set, power was finally estimated as the proportion of simulated data sets in which H_0 was rejected.

Table 3: Augmented version of the observable 2×2 frequency table and the respective probabilities under a Binomial and a Multinomial distribution for healthy controls and suspected cases, respectively.

Exposure	Healthy Controls	Suspected Cases	
		False positive cases	True positive cases
1	θ_0	$\theta_0\gamma$	$\theta_1^*(1 - \gamma)$
0	$1 - \theta_0$	$(1 - \theta_0)\gamma$	$(1 - \theta_1^*)(1 - \gamma)$

As expected, the estimated power decreased with the misclassification rate γ (Figure 2). As a control scenario, when all suspected cases were considered to be false positives ($\gamma = 1$) and therefore the data sets were simulated from H_0 , the corresponding power was estimated at 5%, the significance level specified for the Pearson's χ^2 test. In opposition, when the suspected cases were all considered true positive cases ($\gamma = 0$), the power to detect a hypothetical association was estimated at 34%. This low power simply reflected the limited sample size to detect a weak association between exposure and the disease. In a less extreme case of misclassification, $\gamma = 10\%$ implied an estimated power of 29%, which reflected a decrease in 14.7% of the power estimated for the scenario with no misclassification.

6 Concluding remarks

In summary, our analysis showed that suspected cases of ME/CFS from the UKMEB did not fully agree with four main case definitions of the disease. In addition, some of these suspected cases showed

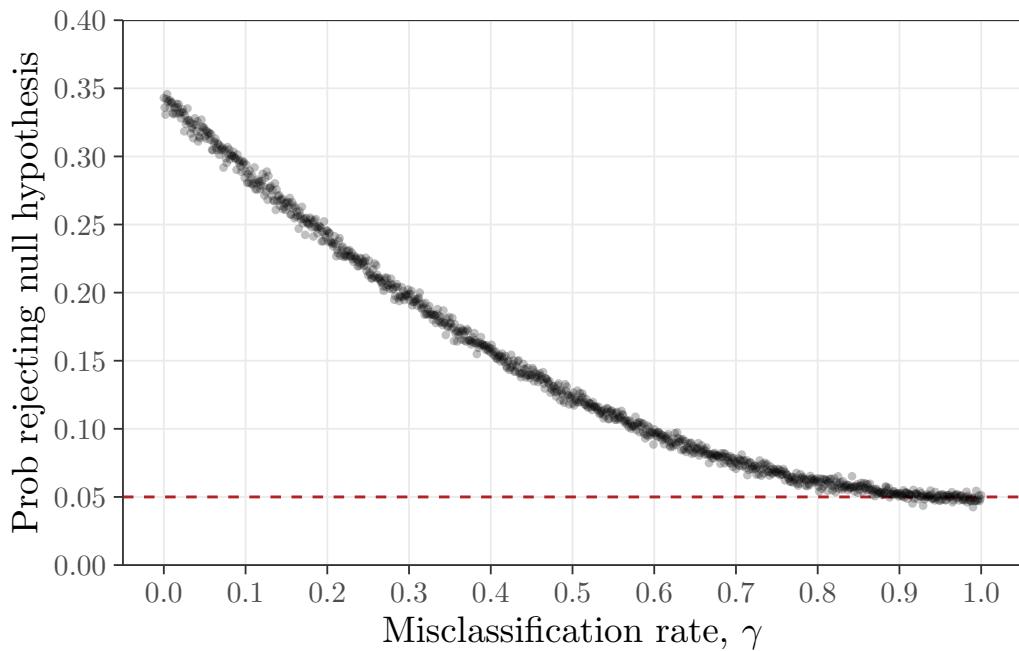


Figure 2: Estimated probability of rejecting H_0 (i.e., lack of association) as function of the misclassification rate γ .

symptom's severity profiles similar to healthy controls and patients with multiple sclerosis. These findings demonstrated the difficulty of diagnosing ME/CFS based on symptoms' assessment alone. To overcome this and other difficulties, there are currently efforts for a stronger collaboration among European researchers for accelerating the discovery of an objective disease-specific biomarker [13]. However, joint efforts for biomarker discovery are very likely to suffer from limited statistical power due to a possible misclassification of the suspected cases. A possible solution to this problem is to take into account for misclassification in the respective statistical analysis. Such a solution is also problematic because modelling misclassification leads to an eventual problem of overparameterisation. From a frequentist standpoint, overparameterization could be avoided by fixing the misclassification rate in a reasonable estimate for the sensitivity of the diagnostic test. A more elegant way of doing so is to use Bayesian analysis where the prior information about

the misclassification rate takes the form of a probability distribution. However, both frequentist and Bayesian solutions show a main hurdle for their implementation in the research of ME/CFS. Given the lack of a disease biomarker, it is unclear which reasonable value or probability distribution to choose for the sensitivity of current diagnostic tools of ME/CFS.

As a final remark, our formulation of the misclassification problem assumed that misclassification is only dependent on the true clinical status of the suspected cases. In practice, it is very likely that misclassification is dependent on the symptoms' severity profile of a given individual, or at least dependent on a given set of covariates. If so, Paulino et al. [14] provided a Bayesian solution for modelling misclassification in this scenario. Given its technical complexity, we envision some difficulties in a wide application of this statistical solution by researchers of ME/CFS who are typically not trained in such advanced statistical methodology. To overcome this potential problem, we recommend a strong collaboration between these researchers and biostatisticians who have in principle the technical skills needed.

Acknowledgements

We thank Eliana Lacerda and Luís Nacul from the CureME group for providing the data from the UKMEB. JM acknowledges a PhD fellowship by the Fundação para a Ciência e Tecnologia, Portugal (ref. SFRH/BD/149758/2019) and a travel grant awarded by the SPE to attend the XXIV SPE Congress in November 2019.

References

- [1] Fukuda, K., Straus, S.E., Hickie, I., Sharpe, M.C., Dobbins, J.G., Komaroff, A., Schluederberg, A., Jones, J.F., Lloyd, A.R., Wessely, S., Gantz, N.M., Holmes, G.P., Buchwald, D., Abbey,

- S., Rest, J., Levy, J.A., Jolson, H., Peterson, D.L., Vercoulen, J., Tirelli, U., Evengård, B., Natelson, B.H., Steele, L., Reyes, M., Reeves, W.C. (1994). The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Annals of Internal Medicine*, 121(12), 953–959.
- [2] Carruthers, B.M., Jain, A.K., Meirleir, K.L., Peterson, D.L., Klimas, N.G., Lerner, A.M., Bested, A.C., Flor-Henry, P., Joshi, P., Powles, A.C., Sherkey, J.A., van de Sande, M.I. (2003). Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Chronic Fatigue Syndrome*, 11, 7–115.
- [3] Morris, G., Maes, M. (2013). Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics. *BMC Medicine*, 11, 205.
- [4] Rasa, S., Nora-Kruk, Z., Henning, N., Eliassen, E., Shikova, E., Harrer, T., Scheibenbogen, C., Murovska, M., Prusty, B.K. (2018). Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) *Journal of Translational Medicine*, 16, 268.
- [5] Brurberg, K.G., Fønhus, M.S., Larun, L., Flottorp, S., Malterud, K. (2014). Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *BMJ Open* 4(2), e003973.
- [6] Lacerda, E.M., Mudie, K., Kingdon, C.C., Butterworth, J.D., O'Boyle, S., Nacul, L. (2018). The UK ME/CFS Biobank: A Disease-Specific Biobank for Advancing Clinical Research Into Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Neurology*, 9, 1026.
- [7] Committee on the Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome, Board on the Health of Select Populations, Institute of Medicine (2015). *Beyond*

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. National Academies Press, Washington (DC).

- [8] Carruthers, B.M., van de Sande, M.I., Meirleir, K.L., Klimas, N.G., Broderick, G., Mitchell, T., Staines, D., Powles, A.C., Speight, N., Vallings, R., Bateman, L., Baumgarten-Austrheim, B., Bell, D.S., Carlo-Stella, N., Chia, J., Darragh, A., Jo, D., Lewis, D., Light, A.R., Marshall-Gradisbik, S., Mena, I., Mikovits, J.A., Miwa, K., Murovska, M., Pall, M.L., Stevens, S. (2011). Myalgic encephalomyelitis: International Consensus Criteria. *Journal of Internal Medicine*, 270, 327–338.
- [9] Pheby, D.F., Araja, D., Berkis, U., Brenna, E., Cullinan, J., de Korwin, J.D., Gitto, L., Hughes, D.A., Hunter, R.M., Trepel, D., Wang-Steverding, X. (2020). The Development of a Consistent Europe-Wide Approach to Investigating the Economic Impact of Myalgic Encephalomyelitis (ME/CFS): A Report from the European Network on ME/CFS (EUROMENE). *Healthcare*, 8, 88.
- [10] Gower, J.C., Warrens, M.J. (2014). Similarity, dissimilarity, and distance, measures of. *Wiley StatsRef: Statistics Reference Online*, 1–11.
- [11] Agresti, A. (2002). *Categorical data analysis*. John Wiley & Sons, New Jersey.
- [12] Nacul, L., Lacerda, E.M., Kingdon, C.C., Curran, H., Bowman, E.W. (2017). How have selection bias and disease misclassification undermined the validity of myalgic encephalomyelitis/chronic fatigue syndrome studies? *Journal of Health Psychology*, 24, 1765–1769.
- [13] Scheibenbogen, C., Freitag, H., Blanco, J., Capelli, E., Lacerda E, Authier, J., Meeus, M., Castro-Marrero, J., Nora-Krukla, Z., Oltra, E., Bolle Strand, Elin, Shikova, E., Sekulic, S., Murovska,

- M. (2017). The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE. *Journal of Translational Medicine*, 15, 162.
- [14] Paulino, C.D., Soares, P., Neuhaus, J. (2003). Binomial Regression with Misclassification. *Biometrics*, 59, 670–675.

F.2 Chapter 3

Impact of misclassification and imperfect serological tests in association analyses of ME/CFS applied to COVID-19 data

João Malato (0000-0003-4389-1483)^{1,2}, Luís Graça (0000-0001-6935-8500)¹,
and Nuno Sepúlveda (0000-0002-8542-1706)^{2,3}

¹ Instituto de Medicina Molecular João Lobo Antunes, Faculty of Medicine,
University of Lisbon, Lisbon, Portugal
jmalato@medicina.ulisboa.pt,

² CEAUL—Centro de Estatística e Aplicações, University of Lisbon, Lisbon,
Portugal

³ Faculty of Mathematics and Information Science, Warsaw University of
Technology, Warsaw, Poland

Abstract. The diagnosis of ME/CFS is problematic due to the absence of a disease specific biomarker. As such, it is conducted under uncertainty using symptom-based criteria and the exclusion of known diseases. The possibility of misdiagnosing patients reduces the power to detect new and previously identified factors that can be associated with the disease. To investigate this problem, we previously conducted a simulation study to estimate the power of case-control association studies as a function of the misdiagnosed rate. Here we extended this simulation study to the more general situation where there is also the possibility of having misclassification in a binary factor related to a previous exposure to a given infection. Given the suggested link between ME/CFS and past viral infections including SARS-CoV-2 (that causes COVID-19), we performed the simulation study in the specific context of serological testing of this new coronavirus using published data from Portuguese, Spanish and Iranian seroepidemiological studies.

Keywords: misclassification, simulation, power studies, serology, myalgic encephalomyelitis/chronic fatigue (ME/CFS)

1 Background

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is one example of a complex disease with uncertainty in its diagnosis [1]. Patients diagnosed with this debilitating disorder manifest heterogeneous symptoms such as unexplained long-lasting fatigue [2], post-exertional malaise that arises after slight physical, or mental effort and is not alleviated by rest [3], accompanied by other symptoms. Its prevalence is estimated between 0.4% and 1% depending on the population, affecting more women than men, at a 6:1 ratio [4, 5].

The aetiology of ME/CFS has been proved difficult to determine. Different reported factors such as acute infections, genetic predisposition, or environmental stressors can serve as triggers for the disease onset [6, 7]. Moreover there is no biomarker, or combination of biomarkers, that characterise this heterogeneous disease, which ultimately leave its diagnosis to be mostly done on the basis of specific symptoms and exclusion of other diseases [8]. This further increases the uncertainty surrounding an objective diagnosis, which has resulted in more than 20 symptom-based criteria currently used to clinically diagnose ME/CFS [9]. Despite proposed protocols for criteria standardisation in ME/CFS research [10], distinct studies will inevitably define the cohort of patients differently, potentially with conflicting results [1]. This inherent level of misclassification—non-ME/CFS patients being incorrectly diagnosed as such—amongst ME/CFS cohorts has already been described in a study characterising the genome of suspected patients [11] and should be taken into account in order to minimise the negative effects on association studies [12].

Despite the pathomechanisms of ME/CFS remaining unknown, the disease has been described as having an autoimmune onset [13, 14]. This immune dysregulation often occurs after exposure to an acute viral infection [7, 15, 16], with multiple association studies relating the exposure to viruses as trigger for ME/CFS development [14, 17]. Serological surveys have thus been conducted to better understand the role of distinct viruses in this disease. However, so far there have not been replicable confirmed associations. Possible arguments for this can be the disparate cohorts of (inherently misclassified) suspected ME/CFS patients used and other factors related to study design such as the low sample sizes used [18] or the further stratification of patients into different subtypes [19]. Additionally, the serological tests used to assess exposure/non-exposure to the viruses are based on predetermined and arbitrary cutoff values to determine seropositive individuals [18]. This important but often overlooked aspect can potentially add an additional layer to the misclassification on ME/CFS, with impacts on the studies' reproducibility [20].

Previously, we studied the dissimilarity between different symptom diagnosis criteria and simulated the impacts of misclassification in a single scenario of potential misdiagnosis of suspected patients [12]. In the present paper we extended the proposed ideas on misclassification and studied its impact on the statistical power of serology hypothetical association studies. More recently, studies have related ME/CFS and the chronic post-viral syndrome developed after infection by the SARS-CoV-2 virus, responsible for the COVID-19 pandemic [21]. Despite the need for more extensive research on this topic, studies have reported that subset of patients following COVID-19 infection can develop a chronic syndrome that fulfils ME/CFS diagnostic criteria [22]. For illustrative purposes we extrapolated on the idea that there is in fact an association between COVID-19 and ME/CFS onset, however mild ($1.25 \leq \text{odds ratio} \leq 2.0$), and simulated multiple case-control association studies with different sample sizes, using results for seroprevalence surveys from three countries: Portugal [23], Spain [24], and Iran [25].

For each serology study, we hypothesised on the impact of misclassification, also accounting for the estimated levels of sensitivity and specificity.

2 Simulation study

2.1 Mathematical formulation of the problem

Following-up on the reported ideas on misclassification [12], the goal of the proposed hypothetical study was to assess the association of a binary exposure outcome (as exposed versus non-exposed) after a serological survey for COVID-19 with ME/CFS. This was accomplished by comparing a cohort of sampled patients suspected of ME/CFS to a cohort of sampled matched healthy controls. The sampling distribution of the designed case-control study was then, the product of two Binomial distributions given by the number of sampled individuals from the two cohorts, n_0 and n_1 , respectively for healthy controls and suspected ME/CFS patients, and the probability of exposure to the virus, θ_0 and θ_1 , respectively; with x_0 and x_1 being the observed frequencies of exposed healthy controls and suspected ME/CFS [12]. Altogether, the sampled populations can be summarised by a 2×2 frequency table that presents different outlines depending on the described parameters n_i and θ_i , $i = \{0, 1\}$. Testing the null hypothesis for lack of association to ME/CFS (i.e., $H_0 : \theta_0 = \theta_1$) was done through the Pearson's χ^2 test for independence. After testing, H_0 was rejected if the p-value for the Pearson's χ^2 test was less than the prespecified level of significance of 5%. Through simulation, and by repeating the inference multiple times under the same conditions, the power of the study was estimated as the overall proportion in which H_0 was rejected.

Previously [12], to account for the inherent misclassification as a diluting effect for the detection of a potential association, four assumptions were considered for the ME/CFS cohort: (i) sampled suspected ME/CFS cases can be divided into apparent (false positives) and true positive cases; (ii) the misclassified apparent cases are considered healthy controls, in the sense that they share the same probability of exposure to COVID-19, θ_0 ; (iii) there is an overall misclassification rate, γ , creating the two distinct possibilities of apparent and true cases within the cohort for suspected cases; and (iv) this misclassification rate is only dependent on the true clinical status of each of the suspected cases. Under the assumption (ii) and the law of total probability, the probability of exposure associated with the suspected cases was written as

$$\theta_1 = \gamma\theta_0 + (1 - \gamma)\theta_1^*, \quad (1)$$

where θ_1^* is the exposure probability of true ME/CFS cases.

However, this analysis does not account for the sensitivity and specificity of a serology test if the exposure to a given infection is determined this way. Therefore, four additional assumptions were considered for this study, with effects transversal to all data sets: (v) for each serology test performed, individuals can only be classified as seropositive or seronegative—in opposition to serology

tests where there are more than two possible outcomes; (vi) the levels of sensitivity, π_{se} , and specificity, π_{sp} , respectively determine the accuracy of a test to identify truly exposed and truly non-exposed individuals; (vii) these parameters related to the performance of the serology test create a category of undetected false positives and false negative for individuals poorly measured by the serology assessment; and (viii) the binary exposure outcomes given by π_{se} and π_{sp} are independent from the assessed cohort. Under these assumptions, the probability of exposure for suspected cases from Equation (1) can be extended to

$$\theta_1 = \pi_{se}\gamma\theta_0 + (1 - \pi_{sp})\gamma(1 - \theta_0) + \pi_{se}(1 - \gamma)\theta_1^* + (1 - \pi_{sp})(1 - \gamma)(1 - \theta_1^*) . \quad (2)$$

Under the eight assumptions, the observable 2×2 frequency table can be augmented, as the cohort for suspected ME/CFS is divided into apparent and true cases based on the misclassification rate, γ , and with sensitivity and specificity, respectively π_{se} and π_{sp} , defining the serology tests' overall accuracy to determine the seropositive (either true positive or false positive) and seronegative (both true and false negative) populations on both cohorts (Table 1).

Table 1. Augmented version of the observable 2×2 frequency table in the case-control association study scenario with possible misclassification of suspected ME/CFS cases (into apparent and true cases) and existence of false positive and false negative serological outcomes observed from serology tests done to assess exposure (confirmed by the true exposure indicator columns, with E for exposed individuals and \bar{E} for non-exposed).

Observed test outcome	True exposure indicator	Controls	Suspected cases	
			(Apparent)	(True)
Seropositive	E	$\pi_{se}\theta_0$	$\pi_{se}\gamma\theta_0$	$\pi_{se}(1 - \gamma)\theta_1^*$
	\bar{E}	$(1 - \pi_{sp})(1 - \theta_0)$	$(1 - \pi_{sp})\gamma(1 - \theta_0)$	$(1 - \pi_{sp})(1 - \gamma)(1 - \theta_1^*)$
Seronegative	E	$(1 - \pi_{se})\theta_0$	$(1 - \pi_{se})\gamma\theta_0$	$(1 - \pi_{se})(1 - \gamma)\theta_1^*$
	\bar{E}	$\pi_{sp}(1 - \theta_0)$	$\pi_{sp}\gamma(1 - \theta_0)$	$\pi_{sp}(1 - \gamma)(1 - \theta_1^*)$

Notes. Instead of the Pearson's χ^2 test, an analogous investigation could also been proposed using the Fisher's exact test to assess the null hypothesis for lack of association. Equation (2) includes parameters related to the accuracy of serology tests; based on this formulation, one can obtain Equation (1) by simply assuming $\pi_{se} = \pi_{sp} = 1$.

2.2 Parameterisation using real-word data

As example of real-life application, we looked at data from three distinct seroepidemiologic surveys: Portugal [23], Spain [24], and Iran [25]. The studies occurred between April and August 2020 and applied similar methods of estimation of

their populations' seroprevalence. Also, all surveys presented information regarding the sensitivity and specificity estimates for the serology tests performed. The estimated values for the mentioned parameters in each survey are presented in Table 2.

Table 2. Parameter values used in the study, where the probability of exposure to the virus and the sensitivity and specificity of the serology test are given by θ_0 , π_{se} , and π_{sp} , respectively.

Reference Country		θ_0	π_{se}	π_{sp}
[23]	Portugal	0.025	0.95	0.98
[24]	Spain	0.050	0.80	0.98
[25]	Iran	0.150	0.75	0.98

For the purpose of the study, we assumed the existence of an association between exposure to COVID-19 and ME/CFS onset. Despite few evidences thus far due to the novelty of the topic, some studies have mentioned this association based on the idea of immune dysregulation, linking the development of post-COVID-19 chronic symptoms with the autoimmune proposal for ME/CFS [14]. Since there are no biomarkers for ME/CFS diagnosis, we defined the association as a mild relation with three possible values of the overall true odds ratio, $\Delta_T = \{1.25, 1.5, 2\}$. Based on the values of θ_0 from the three surveys and the proposed Δ_T , the probability of exposure on true ME/CFS cases was determined by

$$\theta_1^* = \frac{\theta_0 \Delta_T}{1 + \theta_0 (\Delta_T - 1)} . \quad (3)$$

2.3 Simulation structure

The impact of inherent misclassification on the hypothetical case-control association studies was assessed through multiple simulations on different parametric values for θ_0 , π_{se} , π_{sp} , in accordance to each serological survey (Table 2), and Δ_T . For each combination of θ_0 and Δ_T , parameters θ_1 and θ_1^* were calculated from Equations (2) and (3), respectively. To illustrate how sample sizes also influence the overall power of a study, we performed our simulations considering cohort sample sizes of $n_0 = n_1 = \{100, 250, 500, 1000, 2500, 5000\}$.

To assess the power of rejecting H_0 , 10,000 data sets were generated for each value of γ , ranging from 0 (no misclassification) to 1 (no true ME/CFS patients in the cohort for suspected cases) with a lag of 0.01. As previously mentioned, H_0 was rejected at each data set if the p-value from the Pearson's χ^2 test was less than the usual level of significance. Finally, for each parameter set, power was estimated as the proportion of simulated data sets in which H_0 was rejected. All simulations and analyses were done using R statistical software, version 4.1.0 [26], using our own scripts, available upon request.

Notes. For the purpose of study consistency, the estimated seroprevalence values published on the serological surveys were considered as the probability of exposure in the cohort of matched healthy controls, θ_0 .

3 Simulation results

As expected, the estimated power to detect the hypothetical association decreased with misclassification rate (Figure 1). Looking at the extreme cases, the estimated power was highest when no misclassification was considered and all suspected ME/CFS cases were considered to be true positives ($\gamma = 0$). Irrespective of the scenario, as misclassification increases, the overall power is reduced towards 5% at the opposite most extreme value ($\gamma = 1$)—i.e., the significance level specified for the Pearson’s χ^2 test.

Along with gradually increasing the misclassification of suspected patients, the power to detect an association was estimated by varying the values of probability of exposure in healthy controls, θ_0 , and sensitivity of the serology test, π_{se} , for each country serological scenario. In all three illustrated scenarios, the specificity was the same and estimated at $\pi_{sp} = 0.98$ (Table 2).

Overall, only sample sizes of $n_i \geq 500$ individuals were able to reach a power of at least 80%—the specified power threshold to identify what can be considered as having acceptable reproducibility level (Table 3). Simulations with larger sample sizes granted a consistency to the reproducibility of the studies, with power remaining above the defined threshold for higher values of misclassification. Similarly, higher values of Δ_T also affected positively the overall power of each study (Table 3).

The Portuguese survey had the smallest estimate for the probability of virus exposure and the highest sensitivity [23]. For this scenario, only higher association values of $\Delta_T = 2$ and cohort sample sizes of 2500 and 5000 reached the power of at least 80%. Under these parametric conditions, the acceptable level of reproducibility was observable at $\gamma \leq 0.45$.

Compared to Portugal’s results, the Spain’s survey had θ_0 increased and the π_{se} decreased [24]. In this scenario there was an increase on the reproducibility, with more studies surpassing the 80% threshold. Nonetheless, this only occurred for sample sizes of $n_i \geq 1000$.

Lastly, Iran’s survey [25] had the highest estimate for θ_0 and lowest estimates for π_{se} . Despite the lower sensitivity, simulations under these parameters had higher power for the same sample sizes than the other two scenarios (Figure 1 and Table 3).

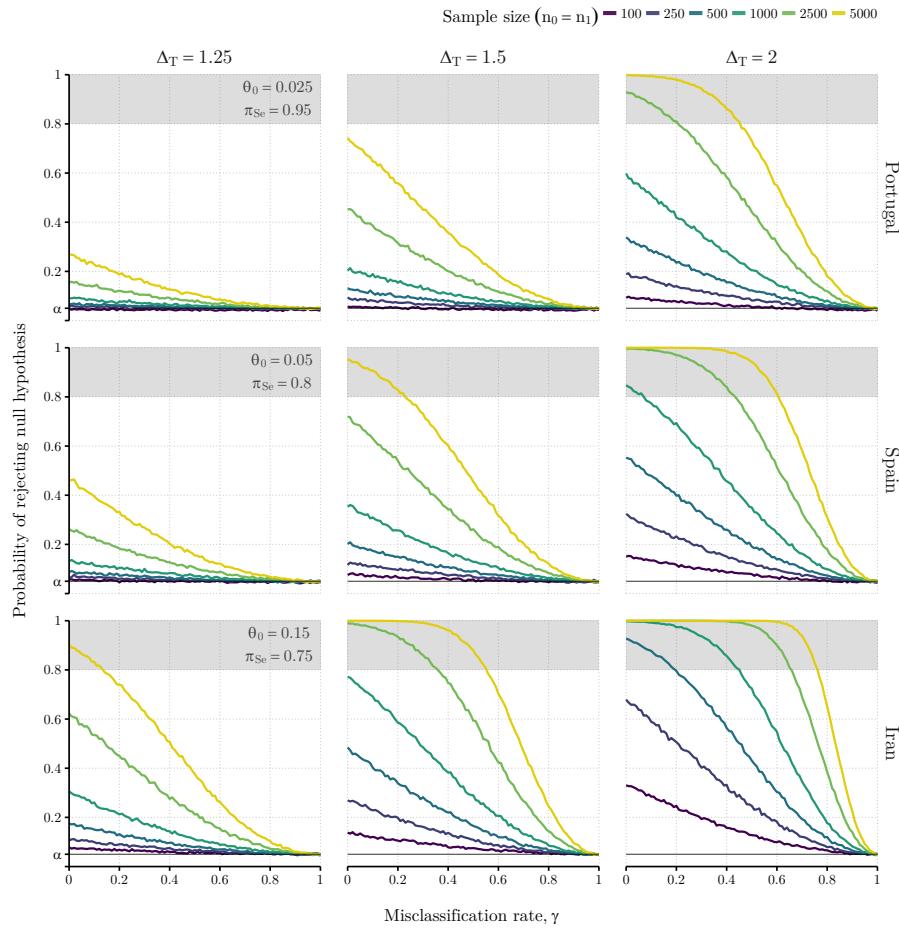


Fig. 1. Probabilities of rejecting the null hypothesis, i.e., absence of association between the two populations as function of the misclassification rate. Each column represents the values attributed to the true odds ratio for COVID-19 exposure and true ME/CFS, assessed between true positive cases and healthy controls. Each row indicates a country serologic survey, with distinct values of θ_0 and π_{Se} identified in the first column of each survey, and fixed $\pi_{sp} = 0.98$ across all simulations. Power analysis was estimated for different cohort sample sizes of 100, 250, 500, 1000, 2500, and 5000 individuals ($n_0 = n_1$), represented by the lines of different colours in each scenario. Gray filled area indicates scenarios where the probability of rejecting the null hypothesis is above 80%. Dark horizontal line indicates the level of significance used, $\alpha = 0.05$.

Table 3. Maximum values of misclassification rate, γ , that maintain power if at least 80% to reject the null hypothesis of lack of association, for different values of true odds ratio, Δ_T , country of serological survey, and sample sizes, n_i , $i = (0, 1)$. Cells with no value indicate the inability to reach the power threshold between cohort, even at $\gamma = 0$.

Country \ Δ_T	1.25	1.50	2.00	n_i
Portugal	–	–	–	
Spain	–	–	–	100
Iran	–	–	–	
Portugal	–	–	–	
Spain	–	–	–	250
Iran	–	–	–	
Portugal	–	–	–	
Spain	–	–	–	500
Iran	–	–	0.19	
Portugal	–	–	–	
Spain	–	–	0.07	1000
Iran	–	–	0.45	
Portugal	–	–	0.20	
Spain	–	–	0.43	2500
Iran	–	0.35	0.65	
Portugal	–	–	0.45	
Spain	–	0.22	0.60	5000
Iran	0.13	0.55	0.76	

4 Discussion

Focusing on ME/CFS, our simulation results showed how misclassification of patients poses an impact on the ability to consistently recognise true associations to a triggering viral exposure, prior to the disease onset. While still researching for biomarkers able discriminate the disease, the power is very likely to suffer from limited statistical power due to possible misclassification of the suspected ME/CFS cases. The proposed solution to this problem is to take into account for misclassification in the respective statistical analysis.

The results evidenced how increasing a study's sample size can increase its power. Until now, misclassification studies mostly focused on identifying the extent of misdiagnosed of patients when using distinct diagnosis criteria, not particularly looking at sample sizes [12]. With MEC/FS research being usually underfunded [28, 29], case-control studies are frequently performed on sample sizes below 250 patients. This allows for potential sporadic associations that ultimately cannot be replicated in follow-up studies. Throughout efforts to raise

awareness and laboratory collaborations, studies have been increasing their sampled populations. After all, our study showed that under the parameterised conditions, only cohorts with samples above 500 individuals were able to consistently reject the null hypothesis under some levels of misclassification (Table 3).

A more in depth study would be required to pose a more general conclusion on the influence in power caused by prevalence of exposure and the sensitivity and specificity of the serology test. One can argue that increasing the prevalence will make for better comparisons between cohorts through the Pearson's χ^2 test for independence, as it might improve the frequency distributions across the 2×2 contingency table cells. Whereas, sensitivity and specificity will produce a lessened effect, as serology tests keep improving—but still impactful, if not from the estimated π_{se} and π_{sp} , then because the majority of serological cutoff values for seropositivity used arise from inherently arbitrary choices if the researchers and manufacturers of the serology tests [20,30]. Nevertheless, diagnostic accuracy is still of extremely importance in the evaluation of medical diagnostic tests and should be taken into account when replication of a study—in this case, a scenario of a serology study—is necessary.

This hypothetical study was done in the context of the recent COVID-19 pandemic and the association of the long-term symptoms caused by the SARS-CoV-2 virus and ME/CFS diagnosis. With the lack of extensive information on this COVID-19 exposure–ME/CFS diagnosis relation premise, parameter Δ_T was defined within low-to-mild values as to not profoundly influence the simulation results. As more studies and serological surveys are published on the matter, focusing different populations or even focusing on serology tests for different specific antibodies against COVID-19, one could better parameterise the simulation study.

ME/CFS is a complex disease and there is still lack of understanding to the extension of the disease's aetiology and pathophysiology. Even under these uncertainties, accepting and accounting for a level of patient misclassification—however small—in association studies might help to improve the study designs and increase scientific reproducibility. Ultimately, the ability to replicate and reproduce the results proposed by a study is one of the most important aspects in research, and consistent results are what allows ideas to become postulates, continuously driving science forwards.

Acknowledgements

JM and NS acknowledge funding by the Fundação para a Ciência e Tecnologia, Portugal (ref. SFRH/BD/149758/2019 and UIDB/00006/2020, respectively). NS also acknowledges funding by the Polish National Agency for Academic Exchange (ref. PPN/ULM/2020/1/00069/U/00001).

References

- Nacul, L., Lacerda, E. M., Kingdon, C. C., Curran, H., Bowman, E. W.: How have selection bias and disease misclassification undermined the validity of myalgic

- encephalomyelitis/chronic fatigue syndrome studies? *Journal of Health Psychology*, SAGE Publications, 24, 1765-1769, (2017). 10.1177/1359105317695803
2. Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., Komaroff, A.: The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Annals of Internal Medicine*, American College of Physicians, 121, 953, (1994). 10.7326/0003-4819-121-12-199412150-00009
 3. Carruthers, B. M., Jain, A. K., Meirleir, K. L. D., Peterson, D. L., Klimas, N. G., Lerner, A. M., Bested, A. C., Flor-Henry, P., Joshi, P., Powles, A. C. P., Sherkey, J. A., van de Sande, M. I.: Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Chronic Fatigue Syndrome*, The Haworth Press, 11, 7-115, (2003). 10.1300/j092v11n01_02
 4. Lim, E.-J., Ahn, Y.-C., Jang, E.-S., Lee, S.-W., Lee, S.-H., Son, C.-G.: Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *Journal of Translational Medicine*, Springer Science and Business Media LLC, 18, (2020). 10.1186/s12967-020-02269-0
 5. Morris, G., Maes, M.: Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics. *BMC Medicine*, Springer Science and Business Media LLC, 11, (2013). 10.1186/1741-7015-11-205
 6. Lacerda, E. M., Geraghty, K., Kingdon, C. C., Palla, L., Nacul, L.: A logistic regression analysis of risk factors in ME/CFS pathogenesis. *BMC Neurology*, Springer Science and Business Media LLC, 19, (2019). 10.1186/s12883-019-1468-2
 7. Chu, L., Valencia, I. J., Garvert, D. W., Montoya, J. G.: Onset Patterns and Course of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Frontiers in Pediatrics*, Frontiers Media SA, 7, (2019). 10.3389/fped.2019.00012
 8. Smith, M. E. B., Nelson, H. D., Haney, E., Pappas, M., Daeges, M., Wasson, N., McDonagh, M.: Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Agency for Healthcare Research and Quality, (2014). 10.23970/ahrqpcerta219
 9. Brurberg, K. G., Fønhus, M. S., Larun, L., Flottorp, S., Malterud, K.: Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *BMJ Open*, BMJ, 4, e003973, (2014). 10.1136/bmjopen-2013-003973
 10. Pheby, D. F., Araja, D., Berkis, U., Brenna, E., Cullinan, J., de Korwin, J.-D., Gitto, L., Hughes, D. A., Hunter, R. M., Trepel, D., Wang-Steverding, X.: The Development of a Consistent Europe-Wide Approach to Investigating the Economic Impact of Myalgic Encephalomyelitis (ME/CFS): A Report from the European Network on ME/CFS (EUROMENE). *Healthcare*, MDPI AG, 8, 88, (2020). 10.3390/healthcare8020088
 11. Brown, D., Birch, C., Younger, J., Worthey, E.: ME/CFS: whole genome sequencing uncovers a misclassified case of glycogen storage disease type 13 previously diagnosed as ME/CFS. *Molecular Genetics and Metabolism*, Elsevier BV, 132, S194-S195, (2021). 10.1016/s1096-7192(21)00388-7
 12. Malato, J., Graça, L., Nacul, L., Lacerda, E., Sepúlveda, N.: Statistical challenges of investigating a disease with a complex diagnosis. In: Sociedade Portuguesa de Estatística (Ed.). *Estatística: Desafios Transversais às Ciências com Dados*, 153-167, (2021).
 13. Lorusso, L., Mikhaylova, S. V., Capelli, E., Ferrari, D., Ngonga, G. K., Ricevuti, G.: Immunological aspects of chronic fatigue syndrome. *Autoimmunity Reviews*, Elsevier BV, 8, 287-291, (2009). 10.1016/j.autrev.2008.08.003

14. Sotzny, F., Blanco, J., Capelli, E., Castro-Marrero, J., Steiner, S., Murovska, M., Scheibenbogen, C.: Myalgic Encephalomyelitis/Chronic Fatigue Syndrome – Evidence for an autoimmune disease. *Autoimmunity Reviews*, Elsevier BV, 17, 601-609, (2018). 10.1016/j.autrev.2018.01.009
15. Rasa, S., Nora-Krukla, Z., Henning, N., Eliassen, E., Shikova, E., Harrer, T., Scheibenbogen, C., Murovska, M., Prusty, B. K.: Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, Springer Science and Business Media LLC, 16, (2018). 10.1186/s12967-018-1644-y
16. Blomberg, J., Gottfries, C.-G., Elfaitouri, A., Rizwan, M., Rosén, A.: Infection Elicited Autoimmunity and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: An Explanatory Model. *Frontiers in Immunology*, Frontiers Media SA, 9, (2018). 10.3389/fimmu.2018.00229
17. Bansal, A., Bradley, A., Bishop, K., Kiani-Alikhan, S., Ford, B.: Chronic fatigue syndrome, the immune system and viral infection. Elsevier BV, 26, 24-31, (2012). 10.1016/j.bbci.2011.06.016
18. Scheibenbogen, C., Freitag, H., Blanco, J., Capelli, E., Lacerda, E., Authier, J., Meeus, M., Marrero, J. C., Nora-Krukla, Z., Oltra, E., et al.: The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE. *Journal of translational medicine*, Springer, 15, 162, (2017). <https://doi.org/10.1186/s12967-017-1263-z>
19. Jason, L. A., Corradi, K., Torres-Harding, S., Taylor, R. R., King, C.: Chronic Fatigue Syndrome: The Need for Subtypes. *Neuropsychology Review*, Springer Science and Business Media LLC, 15, 29-58, (2005). 10.1007/s11065-005-3588-2
20. Domingues, T. D., Grabowska, A. D., Lee, J.-S., Ameijeiras-Alonso, J., Westermeyer, F., Scheibenbogen, C., Cliff, J. M., Nacul, L., Lacerda, E. M., Mourão, H., Sepúlveda, N.: Herpesviruses Serology Distinguishes Different Subgroups of Patients From the United Kingdom Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Biobank. *Frontiers in Medicine*, Frontiers Media SA, 8, (2021). 10.3389/fmed.2021.686736
21. Komaroff, A. L., Bateman, L.: Will COVID-19 Lead to Myalgic Encephalomyelitis/Chronic Fatigue Syndrome? *Frontiers in Medicine*, Frontiers Media SA, 7, (2021). 10.3389/fmed.2020.606824
22. Kedor, C., Freitag, H., Meyer-Arndt, L., Wittke, K., Zoller, T., Steinbeis, F., Hafke, M., Rudolf, G., Heidecker, B., Volk, H., Skurk, C., Paul, F., Bellmann-Strobl, J., Scheibenbogen, C.: Chronic COVID-19 Syndrome and Chronic Fatigue Syndrome (ME/CFS) following the first pandemic wave in Germany – a first analysis of a prospective observational study. *Cold Spring Harbor Laboratory*, (2021). 10.1101/2021.02.06.21249256
23. Kislaya, I., Gonçalves, P., Barreto, M., Sousa, R. D., Garcia, A. C., Matos, R., Guiomar, R., Rodrigues, A. P., behalf of ISNCOVID-19 Group, O.: Seroprevalence of SARS-CoV-2 Infection in Portugal in May-July 2020: Results of the First National Serological Survey (ISNCOVID-19). *Ordem dos Medicos*, 34, 87, (2021). 10.20344/amp.15122
24. Pollán, M., Pérez-Gómez, B., Pastor-Barriuso, R., Oteo, J., Hernán, M. A., Pérez-Olmeda, M., et al.: Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. Elsevier BV, 396, 535-544, (2020). 10.1016/s0140-6736(20)31483-5
25. Khalagi, K., Gharibzadeh, S., Khalili, D., Mansournia, M. A., Samiee, S. M., Aghamohamadi, S., Roodaki, M. M.-M.-A., Hashemi, S. M., Tayeri, K., Tabar,

- H. N., Azadmanesh, K., Tabrizi, J. S., Mohammad, K., Hajipour, F., Namaki, S., Raeisi, A., Ostovar, A.: Prevalence of COVID-19 in Iran: results of the first survey of the Iranian COVID-19 Serological Surveillance programme. Elsevier BV, 27, 1666-1671, (2021). 10.1016/j.cmi.2021.06.002
- 26. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for statistical Computing, Vienna, Austria, (2020). <https://www.r-project.org/>
 - 27. Greenland, S.: Basic methods for sensitivity analysis of biases. International journal of epidemiology, Oxford University Press, 25, 1107-1116, (1996). <https://doi.org/10.1093/ije/25.6.1107-a>
 - 28. Dimmock, M. E., Mirin, A. A., Jason, L. A.: Estimating the disease burden of ME/CFS in the United States and its relation to research funding. Journal of Medicine and Therapeutics, Open Access Text Pvt, Ltd., 1, (2016). 10.15761/jmt.1000102
 - 29. Mirin, A. A., Dimmock, M. E., Jason, L. A.: Research update: The relation between ME/CFS disease burden and research funding in the USA. Work, IOS Press, 66, 277-282, (2020). 10.3233/WOR-203173
 - 30. Domingues, T. D., Mourão, H., Sepúlveda, N.: Analysis of antibody data using Finite Mixture Models based on Scale Mixtures of Skew-Normal distributions. medRxiv, Cold Spring Harbor Laboratory Press, (2021). 10.1101/2021.03.08.21252807

F.3 Chapter 4

Article

Impact of Misdiagnosis in Case-Control Studies of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

João Malato ^{1,2} , Luís Graça ¹  and Nuno Sepúlveda ^{2,3,*} 

¹ Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal

² CEAUL—Centro de Estatística e Aplicações da Universidade de Lisboa, 1749-016 Lisboa, Portugal

³ Faculty of Mathematics and Information Science, Warsaw University of Technology, 00-662 Warszawa, Poland

* Correspondence: n.sepulveda@mini.pw.edu.pl

Abstract: Misdiagnosis of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) can occur when different case definitions are used by clinicians (relative misdiagnosis) or when failing the genuine diagnosis of another disease (misdiagnosis in a strict sense). This problem translates to a recurrent difficulty in reproducing research findings. To tackle this problem, we simulated data from case-control studies under misdiagnosis in a strict sense. We then estimated the power to detect a genuine association between a potential causal factor and ME/CFS. A minimum power of 80% was obtained for studies with more than 500 individuals per study group. When the simulation study was extended to the situation where the potential causal factor could not be determined perfectly (e.g., seropositive/seronegative in serological association studies), the minimum power of 80% could only be achieved in studies with more than 1000 individuals per group. In conclusion, current ME/CFS studies have suboptimal power under the assumption of misdiagnosis. This power can be improved by increasing the overall sample size using multi-centric studies, reporting the excluded illnesses and their exclusion criteria, or focusing on a homogeneous cohort of ME/CFS patients with a specific pathological mechanism where the chance of misdiagnosis is reduced.



Citation: Malato, J.; Graça, L.; Sepúlveda, N. Impact of Misdiagnosis in Case-Control Studies of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Diagnostics* **2023**, *13*, 531. <https://doi.org/10.3390/diagnostics13030531>

Academic Editor: César Fernández De Las Peñas

Received: 10 December 2022

Revised: 20 January 2023

Accepted: 28 January 2023

Published: 1 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a heterogeneous disease whose hallmark symptom is unexplained persistent fatigue [1] or post-exertional malaise upon minimal physical or mental effort [2]. Disease heterogeneity derives from the coexistence of multiple pathological mechanisms in the same patient. Examples of these mechanisms are leaky gut [3], the presence of deleterious autoantibodies [4], oxidative stress [5,6], persisting viral infections [7,8], and severe longstanding stress [9]. Unsurprisingly, research efforts to find a biomarker for disease diagnosis have failed over the years.

Current diagnosis of ME/CFS is performed via multiple polythetic disease definitions where some but not all core symptoms should be present in a suspected case [10]. A differential diagnosis should also be made by excluding known diseases that could explain fatigue and other major symptoms (e.g., multiple sclerosis and diabetes). Given the multiplicity of existing disease definitions, it is possible to diagnose a suspected case of ME/CFS by a consensual case definition but not by an alternative one [11]. This situation is here referred to as relative misdiagnosis because it is only admissible when considering the outcome of a given case definition relative to the one from another case definition. This type of misdiagnosis is typically present when comparing or combining data from studies using different case definitions. Given that consensual definitions for ME/CFS are both difficult to find and suboptimal to patient/control discrimination [12], some efforts were made to investigate empirical approaches to ME/CFS diagnosis [13–15].

Misdiagnosis, in a strict sense, arises from the situation where ME/CFS-diagnosed individuals, irrespective of the case definition, are genuine patients of another disease. This has been illustrated in a patient initially diagnosed with ME/CFS but was found to have a rare autosomal adult-onset disorder [16]. This misdiagnosis can result from random fluctuations in the natural, pathological process of the exclusionary disease (e.g., low-graded remitting/relapsing multiple sclerosis). It can also emerge from limited resources to run the battery of tests necessary to exclude all known diseases that could explain fatigue; for example, not performing whole-genome sequencing to exclude rare genetic diseases. There is also ambiguity around the exclusionary criteria themselves, which leaves clinicians unsure of what illnesses should be actually excluded [17]. Therefore, this type of misdiagnosis seems inevitably present in ME/CFS studies [18].

In this paper, we performed a simulation study to determine the statistical power of detecting associations with ME/CFS under misdiagnosis in a strict sense; relative misdiagnosis is beyond the scope of this paper because it is more related to a discussion about the different case definitions, as made elsewhere [19,20]. We also investigated the impact of imperfect sensitivity/specificity for the presence of a given antibody that could be causing ME/CFS. Finally, we extended our analysis to discuss the statistical power of two published studies [21,22].

2. Statistical Methodology

2.1. Formulation of the Problem

Let us assume a typical case-control study in which diagnosed ME/CFS patients and healthy controls were matched for possible confounding factors, such as age, gender, and body mass index. The main objective of this study is to investigate the association between a candidate causal factor (e.g., a genetic factor or the occurrence of a given infection) and ME/CFS. For simplicity, let us assume that this factor has only two possible values, present and absent; the probabilities for that factor being present in healthy controls and suspected cases are represented by θ_0 and θ_1 , respectively. In general, the respective data are given by a two-way contingency table (Table 1).

Table 1. Two-way contingency table of a typical case-control study where θ_0 and θ_1 are the probabilities of the candidate causal factor being present in healthy controls and ME/CFS-diagnosed cases, respectively.

Causal Factor	Controls	ME/CFS-Diagnosed Cases
Present	θ_0	θ_1
Absent	$1 - \theta_0$	$1 - \theta_1$

Statistically speaking, we aim to investigate the evidence for an association between ME/CFS and the causal factor. This is translated to the following hypotheses

$$H_0 : \theta_0 = \theta_1 \text{ versus } H_1 : \theta_0 \neq \theta_1 .$$

One can then use the classical Pearson's χ^2 test, where p -values < 0.05 indicate a significant association at the 5% significance level.

In this scenario, our objective is to study the impact of misdiagnosis on the power of the Pearson's χ^2 test to detect an association with the disease. With this objective, we considered seven simplifying assumptions:

- I. ME/CFS-diagnosed cases are a mix of apparent and genuine patients of the disease;
- II. The causal factor is only associated with genuine ME/CFS patients;
- III. Apparent cases are similar to healthy controls as far as the association with the causal factor is concerned;
- VI. The chance of an ME/CFS misdiagnosis is only dependent on the true clinical status of the cases and not on the confounding factors;

- V. The true association is independent of disease duration and disease triggers, among other factors occurring during the disease course;
- VI. Healthy controls were not misdiagnosed as such;
- VII. The value of the candidate causal factor can be determined perfectly in each individual.

The first assumption is simply the invocation of misdiagnosis in a strict sense (i.e., they are actually patients of another disease). The second assumption determines that there is a true association between the causal factor and ME/CFS. In the third assumption, we determine that the apparent ME/CFS cases share with healthy controls the same probability of the causal factor being present, θ_0 . The fourth and fifth assumptions simplify the determination of what a misdiagnosed case can be, linking it exclusively to the true/apparent category, thus, rejecting other potential disease-related factors that may influence the disease association. The sixth assumption aims at excluding the situation in which healthy controls could include undiagnosed genuine ME/CFS patients.

Note that the above assumptions are for mathematical convenience and represent the minimal set of conditions that enable the derivation of simple formulas for the probability of the causal factor being present in putative cases. As a consequence, the data simulation procedure is simplified. Additional assumptions can be invoked, but they would lead to a more-complex data simulation procedure. This is the case of also assuming that genuine cases are divided into several sub-types with different degrees of association with the causal factor. This situation, although more realistic, is beyond the scope of this paper due to its higher modeling complexity. On the other hand, the apparently different assumption in which misdiagnosis does not depend on the clinic and the clinicians who performed the diagnoses falls under the umbrella of the fourth assumption, where putative confounding factors would be given by a confounding factor referring to the participating clinics if applicable and another one referring to the clinicians.

Based on the above assumptions, the probability of the causal factor being present in ME/CFS-diagnosed cases can be expressed as follows

$$\theta_1 = \gamma\theta_0 + (1 - \gamma)\theta_1^*, \quad (1)$$

where γ is the probability of misdiagnosing an apparent case as a genuine one, and θ_1^* is the probability of the candidate causal factor being present in genuine ME/CFS cases. If misdiagnosis could be an observable outcome, the above 2×2 contingency table could be augmented as shown in Table A1 (Appendix A).

A more complex situation emerges from the previous scenario where the candidate causal factor cannot be determined perfectly in each individual. As a consequence, there is the possibility of having misdiagnosis together with misclassification of the causal factor. This is particularly relevant to serological studies that aim at investigating whether the presence of specific antibodies is associated with ME/CFS [23] or whether these antibodies can be used for disease diagnosis [24]. Note that the serological evaluation of a suspected case is not mandatory by consensual definitions of ME/CFS [1].

To model this new situation, the above assumption VII is replaced with two additional assumptions:

- VII. There are only two possible serological outcomes for each individual: seronegative or seropositive;
- VIII. The sensitivity and specificity of the serological classification are identical for all of the individuals.

The revised assumption VII excludes the situation where the serological classification can contemplate an indeterminate status due to the laboratory protocol [22] or the presence of multiple serological populations [25]. Similarly to assumption V for misdiagnosis, the new assumption VIII intends to disregard the effect of confounders (i.e., age or gender) and disease-related factors (i.e., disease duration or disease severity) on the performance of the serological classification.

Under the validity of assumptions I–VIII, the probability of the candidate causal factor being present in a ME/CFS-diagnosed patient can be extended to

$$\theta_1 = \pi_{se}\gamma\theta_0 + (1 - \pi_{sp})\gamma(1 - \theta_0) + \pi_{se}(1 - \gamma)\theta_1^* + (1 - \pi_{sp})(1 - \gamma)(1 - \theta_1^*) , \quad (2)$$

where π_{se} and π_{sp} are the sensitivity and specificity for the serological classification, respectively; see Table A2 (Appendix A) for details. Note that when $\pi_{se} = \pi_{sp} = 1$ (perfect serological testing), the above formula converts to Equation (1).

2.2. Simulation Study

To investigate the impact of the above misdiagnosis scenarios, we performed a comprehensive simulation study using the R statistical software, version 4.1.0 [26]. Individuals from each group were selected in accordance with the study's sampling distribution, as shown in Equation (A1) (Appendix B). We assumed the same sample size for ME/CFS patients and healthy controls (i.e., $n_0 = n_1$, with n_0 and n_1 being the sample sizes for healthy controls and ME/CFS-diagnosed patients in each simulated scenario, respectively). We considered the following sample sizes per study group: 100, 250, 500, 1000, 2500, and 5000.

To parameterize the simulation study, we first specified the association between the candidate causal factor and genuine ME/CFS patients by the odds ratio (hereafter denoted as Δ_T) and the probability of the presence of the causal candidate factor in healthy controls and apparent ME/CFS cases (θ_0). We considered the true association (i.e., Δ_T) between genuine ME/CFS cases and the causal factor to vary from weak to strong values (i.e., $\Delta_T \in \{1.25, 1.5, 2, 3, 5, 10\}$). We also specified $\theta_0 \in \{0.05, 0.1, 0.25, 0.5\}$. If data comes from a genetic association study, θ_0 could represent the minor allele frequency of a given single nucleotide variant in the healthy population. Note that, having θ_0 and Δ_T fixed in the respective values, the value of θ_1^* can be estimated, as shown in Equation (A2) (Appendix B). The misdiagnosis probability (or rate) γ was varied from 0 to 1 (all diagnosed individuals are genuine and apparent ME/CFS cases, respectively) with a lag of 0.01.

To simulate data from the second misdiagnosis scenario, we considered fixed parameters $\Delta_T = 3$ and $\theta_0 = 0.25$. For parameters π_{se} and π_{sp} , we considered all possible combinations of 0.80, 0.90, 0.925, 0.975, and 1.0, where $\pi_{se} = \pi_{sp} = 1$ corresponded to the first scenario.

For each misdiagnosis scenario, parameter set, and sample size, we simulated 10,000 data sets to estimate the power of detecting an association under the presence of misdiagnosis. A detailed description of the simulation procedure can be found elsewhere [11,27]. In each data set, we rejected the presence of association if the *p*-value of Pearson's χ^2 test was greater than the usual 5% level of significance. For each parameter combination, the power ($1 - \beta$) was estimated by the proportion of the simulated data sets in which an association was detected. To facilitate the understanding of the simulation results, we specified a target power of at least 80%.

2.3. Application to Two ME/CFS Studies

We also studied the impact of misdiagnosis on published data from a candidate gene association study and an immunological evaluation study. The first study recruited 201 healthy controls and 305 ME/CFS patients whose symptoms complied with the Canadian Consensus Criteria [21]. Five single-nucleotide polymorphisms (SNPs) were evaluated in all participants. The study found significant associations of rs2476601 and rs3087243 with ME/CFS whose onset was triggered by an acute infection.

The second study refers to serological data on 251 ME/CFS patients and 107 healthy controls from the UK ME/CFS Biobank [22]. These serological data referred to antibody positivity to each of six different herpesviruses: human cytomegalovirus (CMV), Epstein-Barr virus (EBV), herpes simplex virus 1 and 2 (HSV1 and HSV2), varicella-zoster virus (VZV), and human herpesvirus (HHV6). Antibody positivity per herpesvirus was previously determined by different lab protocols that did not provide any information about the specificity and sensitivity of the resulting serological classification.

In both studies, we estimated the power of detecting an association as a function of misdiagnosis probability, γ , using simulated data generated from the reported associations, as explained later.

3. Results

3.1. Simulation Study: Impact of ME/CFS Misdiagnosis

The power to detect an association with ME/CFS decreased with the misdiagnosis probability (Figures 1 and 2). The maximum power was achieved when the diagnosed individuals were all genuine ME/CFS cases ($\gamma = 0$). When the diagnosed individuals were all apparent ME/CFS cases ($\gamma = 1$), the corresponding power matched the 5% significance level. This result was a direct consequence of assumption III, in which the misdiagnosed cases were considered identical to healthy controls as far as the association with the candidate causal factor was concerned.

As expected, the most optimistic scenarios were associated with $\Delta_T = 5$ or 10 (i.e., strong associations between the candidate causal factor and ME/CFS). In these scenarios, one could find a maximum misdiagnosis probability for which the power of 80% was achieved (Table 2). For $\Delta_T = 10$, a misdiagnosis probability of 0.53 was sufficient to ensure the desired power for sample sizes greater than or equal to 100 individuals per study group ($n_i \geq 100$), irrespective of θ_0 . This minimum probability was reduced to 0.24 for $\Delta_T = 5$.

Table 2. Maximum values of misdiagnosis probability γ that ensure the minimum power of 80% to detect a genuine association Δ_T as a function of θ_0 , and sample size n per group. Cells with no value indicate that the minimum power could not be reached in the respective parameter combination.

Δ_T	θ_0	0.05	0.1	0.25	0.5	n (per Group)
10	0.59	0.65	0.64	0.53	100	
	0.24	0.43	0.50	0.42		
	—	0.02	0.25	0.23		
	—	—	—	—		
	—	—	—	—		
	—	—	—	—		
5	0.77	0.79	0.77	0.70	250	
	0.56	0.66	0.69	0.63		
	0.20	0.41	0.53	0.50		
	—	—	0.23	0.26		
	—	—	—	—		
	—	—	—	—		
3	0.84	0.86	0.84	0.78	500	
	0.70	0.76	0.78	0.73		
	0.47	0.60	0.67	0.65		
	—	0.27	0.46	0.47		
	—	—	0.04	0.13		
	—	—	—	—		
2	0.89	0.90	0.89	0.84	1000	
	0.80	0.84	0.85	0.81		
	0.64	0.72	0.77	0.75		
	0.32	0.50	0.62	0.62		
	—	0.05	0.32	0.38		
	—	—	—	—		
1.5	—	—	—	—	1000	
	—	—	—	—		
	—	—	—	—		
	—	—	—	—		
	—	—	—	—		
	—	—	—	—		
1.25	—	—	—	—	1000	
	—	—	—	—		
	—	—	—	—		
	—	—	—	—		
	—	—	—	—		
	—	—	—	—		

Table 2. Cont.

$\Delta_T \backslash \theta_0$	0.05	0.1	0.25	0.5	n (per Group)
10	0.93	0.94	0.93	0.90	2500
5	0.88	0.90	0.90	0.88	
3	0.78	0.83	0.85	0.84	
2	0.58	0.69	0.76	0.76	
1.5	0.18	0.42	0.58	0.59	
1.25	—	—	0.20	0.28	
10	0.95	0.95	0.95	0.93	5000
5	0.91	0.93	0.93	0.91	
3	0.84	0.88	0.90	0.88	
2	0.71	0.78	0.83	0.83	
1.5	0.44	0.59	0.70	0.72	
1.25	—	0.20	0.44	0.49	

Similar optimistic scenarios were observed for sample sizes of 2500 and 5000 individuals per study group with the exception of the case of lowest $\Delta_T = 1.25$. Combining these large sample sizes with strong associations between the candidate causal factor and the true ME/CFS cases, failing to achieve the target power only occurred when almost all the cases were misdiagnosed (with misdiagnosis probability greater than or equal to 0.88).

Unsurprisingly, the most pessimistic situations were related to $\Delta_T = 1.25, 1.5, n_0 = n_1 = 100$, or a combination of the two. When $\Delta_T = 1.25$, the sample size had to increase to 2500 or 5000 individuals per group in order to achieve the target power. Therefore, for this weak association, the chance of finding reproducible results was very low, even under the assumption of a perfect diagnosis. As a consequence, testing the “common disease, common variant hypothesis” in ME/CFS is likely to fail in future genetic associations. Finally, the case of $n_0 = n_1 = 100$ was particularly problematic given that it was not possible to find any value misdiagnosis probability in which the desired power could be achieved for $\Delta_T \leq 2$ (Figure 1).

3.2. Simulation Study: Impact of ME/CFS Misdiagnosis and Misclassification on the Candidate Causal Factor

We then simulated the data of a hypothetical association study in which there were both imperfect diagnoses and misclassification of the candidate causal factor (Figure 2). This situation underpins any serological association study in ME/CFS, given the estimation of seropositivity of all individuals could be affected by the sensitivity and specificity associated with the classification rule used. At this point, it was clear that for values of $\Delta_T = 1.25, 1.5$, and 2, the desired power was not often achieved for sample sizes smaller than 500 individuals per group in the case of perfect classification of the causal factor. Therefore, the additional assumption of imperfect classification of the candidate causal factor would make the previously estimated power even worse. Because of that, we only performed our simulation study on the more optimistic scenario in which $\Delta_T = 3$ (Table 3).

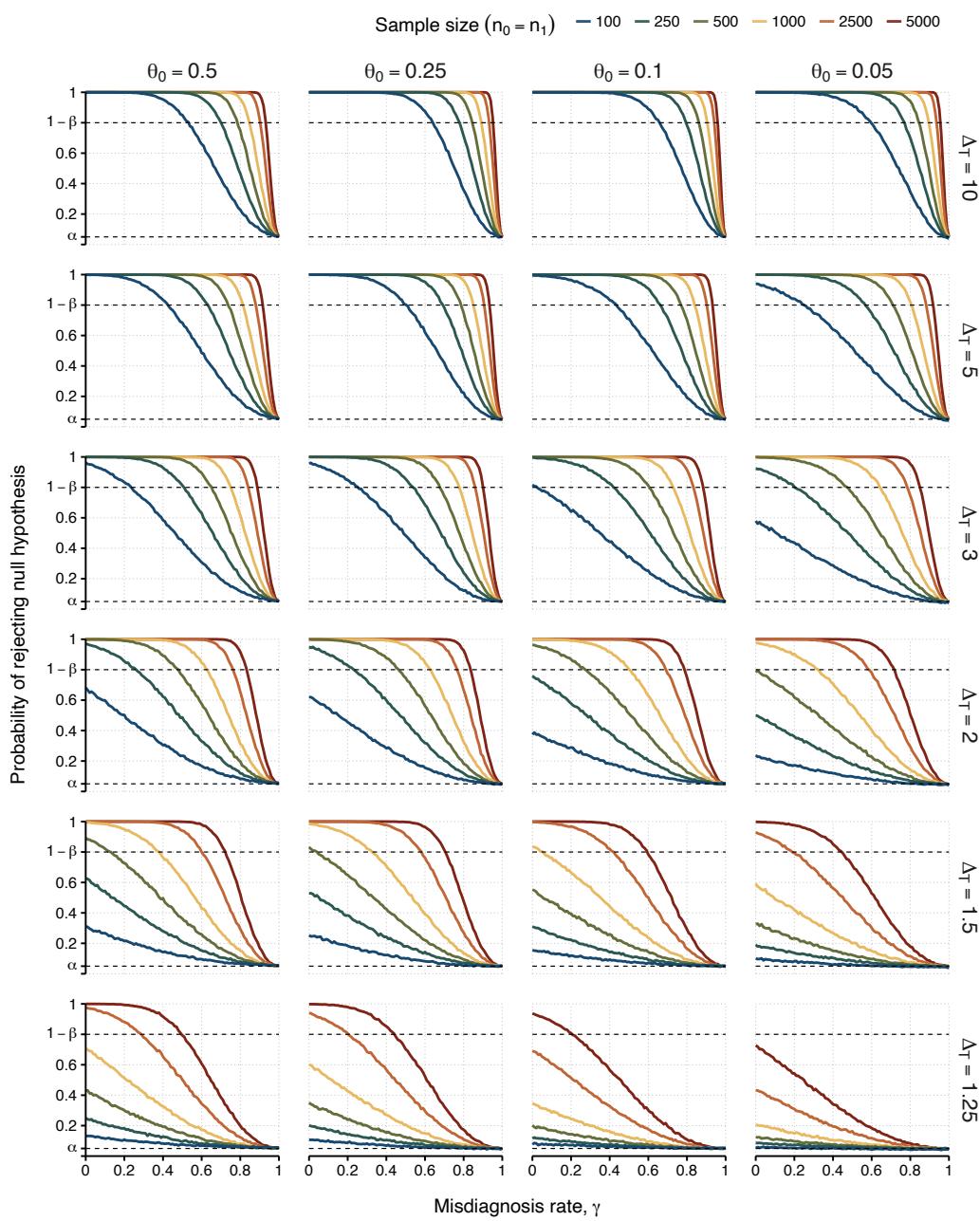


Figure 1. Probabilities of detecting an association (i.e., rejecting H_0) as a function of the misdiagnosis rate. Each column represents the values attributed to the risk allele frequency found in matched healthy controls and false positive ME/CFS cases ($\theta_0 \in \{0.05, 0.1, 0.25, 0.5\}$). Each row varies the true odds ratio for the association between risk allele frequency assessed between true positive cases and healthy controls ($\Delta_T \in \{1.25, 1.5, 2, 3, 5, 10\}$). Power was estimated for different sample sizes of 100, 250, 500, 1000, 2500, and 5000 ($n_0 = n_1$), represented by lines with different colours in each scenario. The upper dashed line indicates the target power of 80% (i.e., $1 - \beta = 0.80$). The lower dashed line indicates the 5% significance level.

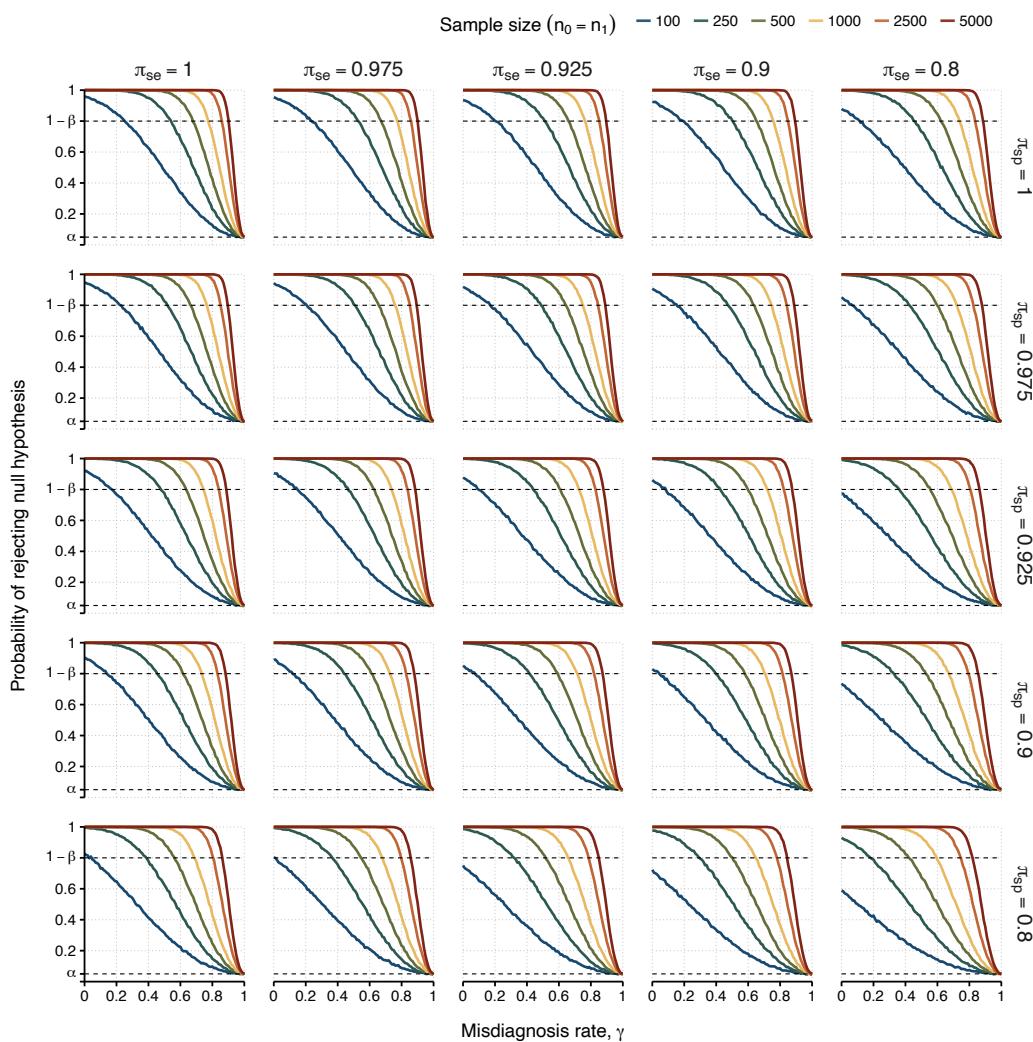


Figure 2. Probabilities of detecting an association (i.e., rejecting H_0) as a function of the misdiagnosis rate. Each scenario represents simulated results with a different combination of sensitivity (π_{se}) and specificity π_{sp} for the serological test for columns and rows, respectively. Power was estimated for different sample sizes of 100, 250, 500, 1000, 2500, and 5000 ($n_0 = n_1$), represented by lines with different colours in each scenario, with the probability of exposure in healthy controls fixed as $\theta_0 = 0.25$ and true odds ratio $\Delta_T = 3$. The upper dashed line indicates the target power of 80% (i.e., $1 - \beta = 0.80$). The lower dashed line indicates the 5% significance level.

Table 3. Maximum values of misdiagnosis probability γ that still ensure a power of rejecting the null hypotheses of at least 80% for $\Delta_T = 3$ and $\theta_0 = 0.25$, where π_{se} and π_{sp} represent sensitivity and specificity associated with the classification of the candidate, respectively. See Table 2 for more information.

π_{sp}	π_{se}	1	0.975	0.925	0.9	0.8	n (per Group)
1	1	0.25	0.23	0.20	0.19	0.11	
0.975	1	0.22	0.20	0.17	0.15	0.06	
0.925	1	0.17	0.14	0.09	0.08	—	
0.9	1	0.13	0.11	0.07	0.04	—	100
0.8	1	0.03	—	—	—	—	

Table 3. Cont.

π_{sp}	π_{se}	1	0.975	0.925	0.9	0.8	n (per Group)
1	0.53	0.52	0.51	0.50	0.45		
	0.975	0.51	0.50	0.48	0.47	0.42	
	0.925	0.47	0.46	0.43	0.42	0.36	250
	0.9	0.45	0.43	0.41	0.39	0.32	
	0.8	0.38	0.36	0.31	0.29	0.18	
0.975	0.67	0.67	0.66	0.65	0.62		
	0.66	0.65	0.64	0.63	0.59		
	0.63	0.62	0.60	0.59	0.55		500
	0.61	0.61	0.59	0.57	0.52		
	0.56	0.54	0.51	0.50	0.42		
0.925	0.77	0.77	0.76	0.75	0.73		
	0.76	0.75	0.74	0.74	0.72		
	0.74	0.73	0.72	0.71	0.68		1000
	0.73	0.72	0.71	0.70	0.67		
	0.68	0.67	0.65	0.64	0.59		
0.9	0.85	0.85	0.85	0.84	0.83		
	0.85	0.85	0.84	0.84	0.82		
	0.84	0.83	0.82	0.82	0.80		2500
	0.83	0.82	0.81	0.81	0.79		
	0.80	0.79	0.78	0.78	0.74		
0.8	0.90	0.90	0.89	0.89	0.88		
	0.89	0.89	0.89	0.88	0.87		
	0.88	0.88	0.87	0.87	0.86		5000
	0.88	0.87	0.87	0.87	0.85		
	0.86	0.85	0.84	0.84	0.81		

3.3. Application to Data from Two ME/CFS Studies

We illustrated the problem of misdiagnosis in data from two ME/CFS studies [21,22]. We started with data from a candidate gene association study [21]. In this study, some genetic associations were only found to be significant when comparing healthy controls to ME/CFS patients with an infectious disease trigger onset (Table 4). The estimated allele-related odds ratios varied from 0.84 [95%CI = (0.56, 1.27)] (rs1799724, TNF) to 1.63 [95%CI = (1.04, 2.55)] (rs2476601, PTPN22). In our re-analysis, we investigated the impact of misdiagnosis if a replication study were conducted in a similar population. In line with the original study, no genotyping errors were assumed for the genetic data. The reported odds ratios were assumed to be the true ones for the population, and data were simulated with the same allele frequencies as reported in the original study.

Table 4. Reported associations of a candidate gene association study [21] where $\hat{\theta}_0$ represents the frequencies of the non-reference allele for healthy controls and $\hat{\Delta}_T$ is the odds ratio of these allele frequencies when comparing ME/CFS patients with an infectious disease trigger to healthy controls. *p*-values are associated with the Pearson's χ^2 test for two-way contingency tables.

SNP	Gene	$\hat{\theta}_0$	$\hat{\Delta}_T$	95% CI ($\hat{\Delta}_T$)	<i>p</i> -Value
rs3087243	CTLA4	0.56	1.54	(1.17, 2.03)	0.002
rs2476601	PTPN22	0.08	1.63	(1.04, 2.55)	0.033
rs1799724	TNF	0.13	0.84	(0.56, 1.27)	0.409
rs1800629	TNF	0.16	0.89	(0.61, 1.30)	0.551
rs3807306	IRF5	0.51	0.94	(0.72, 1.22)	0.637

Again, the estimated probability of detecting an association decreased with the misdiagnosis probability (Figure 3A). More importantly, when the misdiagnosis probability

was low ($\gamma < 0.09$), it was possible to achieve the minimum power of 80% for the allele association reported for rs3087243 in *CTLA4*. Therefore, the target power cannot be ensured for $\gamma > 0.09$. For the remaining SNPs, the target power was never achieved, irrespective of the misdiagnosis probability. This is particularly problematic for rs2476601 in *PTPN22* whose association was reported to be significant at the 5% significance level. For this SNP, the misdiagnosis probability of approximately 0.10 had an estimated power of about 50%. This result implies that the chance of replicating the reported association was no better than flipping a coin.

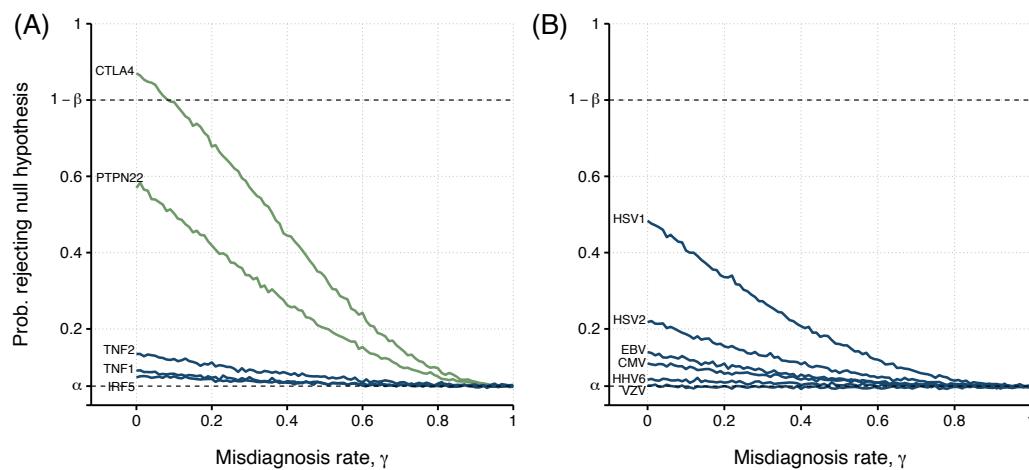


Figure 3. The relationship between the misdiagnosis probability (or rate) and the probability of detecting an association (i.e., rejecting the H_0) estimated from simulated data from two previously published studies: (A). Data from five different SNPs (genes *PTPN22*, *CTLA4*, *TNF* (TNF1-rs1799724 and TNF2-rs1800629), and *IRF5*); (B). Data of antibody positivity related to six human herpesviruses (CMV, EBV, HSV1 and HSV2, VZV, and HHV6). For each study, risk allele frequencies or the probability of exposure and true odds ratio were determined by Steiner et al. [21] ($n_0 = 201$; $n_1 = 305$) and Cliff et al. [22] ($n_0 = 107$; $n_1 = 251$; $\pi_{se} = \pi_{sp} = 0.975$), with determined values shown in Tables 4 and 5, respectively. Green lines indicate candidate risk factors where a significant association with the disease was found in the original study. Blue lines show non-significant ME/CFS risk factors. The upper dashed line indicates the target power, where the probability of rejecting the null hypothesis is $1 - \beta = 0.80$. The lower dashed line indicates the significance level used, $\alpha = 0.05$.

The second study referred to putative associations of six herpes virus infections with ME/CFS using antibody positivity data [22]. In these data, all individuals were classified as seronegative or seropositive for each antibody used. Under the assumption of perfect serological classification and diagnosis, the associations of these serological data with severely affected ME/CFS patients ranged from 0.65 [95%CI = (0.21, 1.97)] to 1.60 [95%CI = (0.83, 3.09)] for EBV and HSV1, respectively (Table 5). In this study, no association was deemed significant at the usual significance level of 5%, according to the original study (p -values ≥ 0.16).

The original serological classification was based on a cut-off in the antibody levels determined by the 2σ rule; the cut-off is the mean plus twice the standard deviation of a known or hypothetical seronegative population. Under the assumption of a normal distribution for the seronegative population, the expected specificity of the serological specificity is approximately 0.975 [28]. We assumed this value for π_{sp} . For simplicity, we assumed $\pi_{se} = \pi_{sp}$. Again, we simulated data from this scenario as the original study and estimated the probability of detecting an association as a function of the misdiagnosis probability.

In this study, the minimum power of 80% could not be reached for any of the antibodies (Figure 3B). The best case was the antibody data related to HSV1. In this case, the maximal power was around 0.50 in the absence of misdiagnosis. This power dropped to 0.30 when $\gamma = 0.25$. For the remaining cases, the power was almost less than 0.20. This could partially

be explained by the fact that θ_0 is higher than 0.93 for antibody data related to EBV, HHV6, and VZV.

Table 5. Summary of serological findings from [22], where $\hat{\theta}_0$ represents the seroprevalence of healthy controls, and $\hat{\Delta}_T$ refers to the odds ratio for being seropositive when comparing severely affected ME/CFS patients to healthy controls. The 95% CI ($\hat{\Delta}_T$) and *p*-values are associated with the Pearson's χ^2 test for two-way contingency tables.

Herpes Virus	$\hat{\theta}_0$	$\hat{\Delta}_T$	95% CI ($\hat{\Delta}_T$)	<i>p</i> -Value
HSV1	0.42	1.60	(0.83, 3.09)	0.163
HSV2	0.34	1.36	(0.69, 2.66)	0.377
EBV	0.93	0.65	(0.21, 1.97)	0.442
CMV	0.37	0.84	(0.42, 1.67)	0.613
VZV	0.97	0.75	(0.12, 4.63)	0.757
HHV6	0.95	1.27	(0.24, 6.79)	0.776

4. Discussion

This study investigated the impact of misdiagnosis on the reproducibility of ME/CFS association studies. Our simulation study showed that strong associations with ME/CFS can be detected with reasonable power even under a non-negligible misdiagnosis rate. However, strong associations might not be the case of ME/CFS given the difficulty in finding a disease biomarker [29] and a clear genetic signature of the disease [30–33].

Studies with sample sizes larger than 500 individuals per study group are able to compensate for the reduction in power due to misdiagnosis alone. This minimum sample size increases when, besides misdiagnosis, there is also the possibility of not determining the presence of the causal factor perfectly. In general, large studies are becoming common in well-known and highly-funded diseases, such as cancer, cardiovascular diseases [34], and autoimmune disorders [35,36]. However, large ME/CFS studies are currently unfeasible due to limited funding and poor societal recognition of the disease [37]. This problem can be somehow minimized by using data from the United Kingdom ME/CFS Biobank that includes biological samples of more than 500 individuals [38]. Another solution is to conduct multi-centric studies [29]. Increasing sample size via data from self-reported ME/CFS cases (as performed in studies based on the UK Biobank) does not seem a viable solution because the chance of misdiagnosis is too high for obtaining reliable results. This problem is clearly illustrated in a Polish study where 1400 individuals were believed to be suffering from ME/CFS, but only 69 individuals actually complied with a consensual ME/CFS case definition [39].

Current serological association studies of ME/CFS neglect the possibility of misclassifying seropositive individuals. In addition, it is common to leave the sensitivity and specificity of the respective serological classification unreported. This research practice adds to the list of other factors that can contribute to the lack of reproducibility of ME/CFS serological studies [40]. Genetic association studies of ME/CFS also neglect the possibility of misclassifying the genotypes of the individuals. This neglect is reasonable in most studies given that genotype error rates are often below 1%, and rare genetic markers with higher genotype errors are typically excluded from the analysis [33,41,42].

Our results are based on the assumption that disease association is independent of possible confounding factors. This assumption seems appropriate for randomized clinical studies or studies based on the analysis of specific subgroups, such as only focusing on adult women with an infection at the disease onset. However, it is also known that age, gender, and exposure to a given infectious agent can affect the results [43,44]. Therefore, the assumption might not be true in general.

Our results are also based on the assumption that the controls are indeed healthy. Interestingly, ME/CFS patients and some healthy controls might have the same symptoms profile and similar levels of fatigue [11,45]. More importantly, the use of self-reported healthy controls [44,46] or control samples from existing blood banks [47–49] are also

common practices in ME/CFS research. According to these research practices, a more realistic assumption is to divide healthy controls into genuine and apparent controls. However, we anticipate that the statistical power to detect a putative disease association is further reduced in this more general scenario. To avoid this scenario, a thorough clinical assessment should also be performed in putative healthy controls.

This study was framed in terms of ME/CFS misdiagnosis in a strict sense. However, from a modelling standpoint, this framing is mathematically equivalent to the situation where ME/CFS-diagnosed cases can be partitioned into two subgroups of genuine patients but with distinct pathological mechanisms and where the association is only present in one of these subgroups. Therefore, our results are directly applicable to this alternative situation but with caution. As alluded to in the introduction, ME/CFS might not be one but several diseases under the same umbrella term, as suggested by genomic data [50,51]. Having said that, a more realistic situation is to have multiple subgroups with different degrees of association with the potential causal factor. Therefore, there is a need to extend our simulation study to this situation.

In conclusion, current case-control association studies of ME/CFS seem to have limited power to mitigate the effect of misdiagnosis in the detection of putative disease associations. A sample size of 500 or 1000 individuals per study group is a minimal requirement to detect mild-to-moderate associations with a high power under the assumption of misdiagnosis. These sample sizes are attainable from multi-centric studies; these studies require extensive collaboration among ME/CFS researchers. Under the impossibility of increasing sample size, research efforts should be made towards reducing the rate of strict misdiagnosis. This can be achieved by following existing recommendations for research reports of ME/CFS, such as reporting the screening laboratory tests and the cut-off values for exclusion [52]. It can also be achieved by the continued search for alternative diagnoses and co-morbidities [53]. In the end, a better understanding of multiple disease pathways leading to ME/CFS leads to better diagnoses, and, therefore, one should ultimately aim to study homogeneous cohorts of patients where the chance of strict misdiagnosis is reduced.

Author Contributions: Conceptualisation, J.M. and N.S.; methodology, J.M. and N.S.; software, J.M.; validation, J.M. and N.S.; formal analysis, J.M.; investigation, J.M., L.G. and N.S.; resources, J.M. and N.S.; data curation, J.M.; writing—original draft preparation, J.M.; writing—review and editing, J.M., L.G. and N.S.; visualisation, J.M.; supervision, L.G. and N.S.; project administration, N.S. All authors have read and agreed to the published version of the manuscript.

Funding: J.M. acknowledges funding from Fundação para a Ciência e Tecnologia, Portugal (grant ref. SFRH/BD/149758/2019 and UIDB/00006/2020). Research in L.G. lab is funded by Fundação para a Ciência e Tecnologia, Portugal, through 081_596653860, and Fundacion la Caixa, Spain (grant number HR22-00741). N.S. acknowledges funding from Fundação para a Ciência e Tecnologia, Portugal (ref. UIDB/00006/2020), and the Polish National Agency for Academic Exchange, Poland (ref. PPN/ULM/2020/1/00069/U/00001).

Data Availability Statement: The source codes and all the simulated data can be downloaded freely from <https://github.com/jtmalato/misclassification-simulations> (accessed on 10 October 2022).

Acknowledgments: J.M. would like to thank Przemysław Biecek and the MI² DataLab group at the Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland, for their support during part of the analysis and discussion of this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ME/CFS	Myalgic Encephalomyelitis/Chronic Fatigue Syndrome
SNP	Single-nucleotide polymorphism
PTPN22	Tyrosine phosphatase non-receptor type 22
CTLA4	Cytotoxic T-lymphocyte-associated protein 4

TNF	Tumor necrosis factor
IRF5	Interferon regulatory factor 5
CMV	Human cytomegalovirus
EBV	Epstein–Barr virus
HSV1	Herpes simplex virus 1
HSV2	Herpes simplex virus 2
VZV	Varicella-zoster virus
HHV6	Human herpesvirus
CI	Confidence interval

Appendix A. Appendix Tables

Table A1. Augmented version of the observed 2×2 contingency table in the presence of the misdiagnosis of ME/CFS cases for a classical case-control association study. Parameter θ_0 is the probability of the presence of the candidate causal factor shared across healthy controls and apparent (false positive) ME/CFS cases, θ_1^* is the true probability of the causal factor in the true ME/CFS patients. Misdiagnosis probability is given by the parameter γ .

Causal Factor	Controls	ME/CFS-Diagnosed Cases	
		(Apparent)	(True)
Present	θ_0	$\gamma\theta_0$	$(1 - \gamma)\theta_1^*$
Absent	$1 - \theta_0$	$\gamma(1 - \theta_0)$	$(1 - \gamma)(1 - \theta_1^*)$

Table A2. Augmented version of the observable 2×2 contingency table in the case-control association study with possible misdiagnosis of ME/CFS cases and misclassification of the true serological status (seropositive, S^+ , and seronegative, S^-). Parameter θ_0 is the probability of the presence of the candidate causal factor in healthy controls and apparent (false positive) ME/CFS cases, θ_1^* is the true probability of the causal factor in the true ME/CFS patients. Misdiagnosis probability is modulated by the parameter γ . The true serological status is dependent on the sensitivity (π_{se}) and specificity (π_{sp}) of the serological test.

Estimated Serological Status	True Serological Status	Controls	ME/CFS-Diagnosed Cases	
			(Apparent)	(True)
S^+	S^+	$\pi_{se}\theta_0$	$\pi_{se}\gamma\theta_0$	$\pi_{se}(1 - \gamma)\theta_1^*$
	S^-	$(1 - \pi_{sp})(1 - \theta_0)$	$(1 - \pi_{sp})\gamma(1 - \theta_0)$	$(1 - \pi_{sp})(1 - \gamma)(1 - \theta_1^*)$
S^-	S^+	$(1 - \pi_{se})\theta_0$	$(1 - \pi_{se})\gamma\theta_0$	$(1 - \pi_{se})(1 - \gamma)\theta_1^*$
	S^-	$\pi_{sp}(1 - \theta_0)$	$\pi_{sp}\gamma(1 - \theta_0)$	$\pi_{sp}(1 - \gamma)(1 - \theta_1^*)$

Appendix B. Mathematical Formulation

Appendix B.1. Sampling Distribution

We constructed our analysis considering a classical epidemiological scenario where for a single putative risk factor, individuals can be divided into exposed versus non-exposed. This result can be summarised by a 2×2 contingency table, whose sampling distribution is the product of two independent Binomial distributions, one Binomial distribution per group,

$$f(x_i|n_i; \theta_i) = \prod_{i=0,1} \binom{n_i}{x_i} \theta_i^{x_i} (1 - \theta_i)^{n_i - x_i}, \quad (\text{A1})$$

where x_0 and x_1 are the observed frequencies of healthy controls and suspected cases with the presence of the candidate causal factor, respectively, n_0 and n_1 are the corresponding sample sizes of each group, and θ_0 and θ_1 are the probabilities for the presence of the candidate causal factor in healthy controls and suspected cases, respectively.

Appendix B.2. Simulation Study Estimation of Parameter θ_1^*

$$\theta_1^* = \frac{\theta_0 \Delta_T}{1 + \theta_0 (\Delta_T - 1)} \quad (\text{A2})$$

References

- Fukuda, K.; Straus, S.E.; Hickie, I.; Sharpe, M.C.; Dobbins, J.G.; Komaroff, A. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. *Ann. Intern. Med.* **1994**, *121*, 953. [CrossRef] [PubMed]
- Carruthers, B.M.; Jain, A.K.; De Meirlier, K.L.; Peterson, D.L.; Klimas, N.G.; Lerner, A.M.; Bested, A.C.; Flor-Henry, P.; Joshi, P.; Powles, A.C.P.; et al. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Clinical Working Case Definition, Diagnostic and Treatment Protocols. *J. Chronic Fatigue Syndr.* **2003**, *11*, 7–115. [CrossRef]
- König, R.S.; Albrich, W.C.; Kahlert, C.R.; Bahr, L.S.; Löber, U.; Vernazza, P.; Scheibenbogen, C.; Forslund, S.K. The Gut Microbiome in Myalgic Encephalomyelitis (ME)/Chronic Fatigue Syndrome (CFS). *Front. Immunol.* **2022**, *12*, 628741. [CrossRef] [PubMed]
- Wirth, K.; Scheibenbogen, C. A Unifying Hypothesis of the Pathophysiology of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS): Recognitions from the finding of autoantibodies against β 2-adrenergic receptors. *Autoimmun. Rev.* **2020**, *19*, 102527. [CrossRef] [PubMed]
- Wood, E.; Hall, K.H.; Tate, W. Role of mitochondria, oxidative stress and the response to antioxidants in myalgic encephalomyelitis/chronic fatigue syndrome: A possible approach to SARS-CoV-2 ‘long-haulers’? *Chronic Dis. Transl. Med.* **2021**, *7*, 14–26. [CrossRef] [PubMed]
- Castro-Marrero, J.; Cordero, M.D.; Sáez-Francas, N.; Jimenez-Gutierrez, C.; Aguilar-Montilla, F.J.; Aliste, L.; Alegre-Martin, J. Could Mitochondrial Dysfunction Be a Differentiating Marker Between Chronic Fatigue Syndrome and Fibromyalgia? *Antioxidants Redox Signal.* **2013**, *19*, 1855–1860. [CrossRef]
- Rasa, S.; Nora-Krukla, Z.; Henning, N.; Eliassen, E.; Shikova, E.; Harrer, T.; Scheibenbogen, C.; Murovska, M.; Prusty, B.K. Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *J. Transl. Med.* **2018**, *16*, 268. [CrossRef]
- Sepúlveda, N.; Carneiro, J.; Lacerda, E.; Nacul, L. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome as a Hyper-Regulated Immune System Driven by an Interplay Between Regulatory T Cells and Chronic Human Herpesvirus Infections. *Front. Immunol.* **2019**, *10*, 2684. [CrossRef]
- Cortes Rivera, M.; Mastronardi, C.; Silva-Aldana, C.; Arcos-Burgos, M.; Lidbury, B. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: A Comprehensive Review. *Diagnostics* **2019**, *9*, 91. [CrossRef]
- Smith, M.E.B.; Nelson, H.D.; Haney, E.; Pappas, M.; Daeges, M.; Wasson, N.; McDonagh, M. Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Evid. Rep. Technol. Assess.* **2014**, *219*, 1–433. [CrossRef]
- Malato, J.; Graça, L.; Nacul, L.; Lacerda, E.; Sepúlveda, N. Statistical challenges of investigating a disease with a complex diagnosis. In *Estatística: Desafios Transversais às Ciências com Dados*; Milheiro, P., Pacheco, A., de Sousa, B., Alves, I.F., Pereira, I., Polidoro, M.J., Ramos, S., Eds.; Sociedade Portuguesa de Estatística: Lisboa, Portugal, 2021; pp. 153–167.
- Jason, L.A.; Sunnquist, M.; Brown, A.; Evans, M.; Vernon, S.D.; Furst, J.D.; Simonis, V. Examining case definition criteria for chronic fatigue syndrome and myalgic encephalomyelitis. *Fatigue Biomed. Health Behav.* **2014**, *2*, 40–56.
- Reeves, W.C.; Wagner, D.; Nisenbaum, R.; Jones, J.F.; Gurbaxani, B.; Solomon, L.; Papanicolaou, D.A.; Unger, E.R.; Vernon, S.D.; Heim, C. Chronic Fatigue Syndrome – A clinically empirical approach to its definition and study. *BMC Med.* **2005**, *3*, 19. [CrossRef] [PubMed]
- Jason, L.A.; Kot, B.; Sunnquist, M.; Brown, A.; Reed, J.; Furst, J.; Newton, J.L.; Strand, E.B.; Vernon, S.D. Comparing and contrasting consensus versus empirical domains. *Fatigue Biomed. Health Behav.* **2015**, *3*, 63–74. [CrossRef] [PubMed]
- Conroy, K.E.; Islam, M.F.; Jason, L.A. Evaluating case diagnostic criteria for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS): Toward an empirical case definition. *Disabil. Rehabil.* **2022**, *1*–8. [CrossRef]
- Brown, D.; Birch, C.; Younger, J.; Worthey, E. ME/CFS: Whole genome sequencing uncovers a misclassified case of glycogen storage disease type 13 previously diagnosed as ME/CFS. *Mol. Genet. Metab.* **2021**, *132*, S194–S195. [CrossRef]
- Jason, L.A.; Ravichandran, S.; Katz, B.Z.; Natelson, B.H.; Bonilla, H.F. Establishing a consensus on ME/CFS exclusionary illnesses. *Fatigue Biomed. Health Behav.* **2023**, *11*, 1–13.
- Nacul, L.; Lacerda, E.M.; Kingdon, C.C.; Curran, H.; Bowman, E.W. How have selection bias and disease misclassification undermined the validity of myalgic encephalomyelitis/chronic fatigue syndrome studies? *J. Health Psychol.* **2017**, *24*, 1765–1769. [CrossRef] [PubMed]
- Brurberg, K.G.; Fønhus, M.S.; Larun, L.; Flottorp, S.; Malterud, K. Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): A systematic review. *BMJ Open* **2014**, *4*, e003973. [CrossRef] [PubMed]
- Lim, E.J.; Son, C.G. Review of case definitions for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *J. Transl. Med.* **2020**, *18*, 289. [CrossRef]
- Steiner, S.; Becker, S.C.; Hartwig, J.; Sotzny, F.; Lorenz, S.; Bauer, S.; Löbel, M.; Stitrich, A.B.; Grabowski, P.; Scheibenbogen, C. Autoimmunity-Related Risk Variants in PTPN22 and CTLA4 Are Associated With ME/CFS With Infectious Onset. *Front. Immunol.* **2020**, *11*, 578. [CrossRef]

22. Cliff, J.M.; King, E.C.; Lee, J.S.; Sepúlveda, N.; Wolf, A.S.; Kingdon, C.; Bowman, E.; Dockrell, H.M.; Nacul, L.; Lacerda, E.; et al. Cellular Immune Function in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Front. Immunol.* **2019**, *10*, 796. [[CrossRef](#)] [[PubMed](#)]
23. Ruiz-Pablos, M.; Paiva, B.; Montero-Mateo, R.; Garcia, N.; Zabaleta, A. Epstein-Barr Virus and the Origin of Myalgic Encephalomyelitis or Chronic Fatigue Syndrome. *Front. Immunol.* **2021**, *12*, 656797. [[CrossRef](#)] [[PubMed](#)]
24. Sepúlveda, N.; Malato, J.T.; Sotzny, F.; Grabowska, A.D.; Fonseca, A.; Cordeiro, C.; Graça, L.; Biecek, P.; Behrends, U.; Mautner, J.; et al. Revisiting IgG antibody reactivity to Epstein-Barr virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and its potential application to disease diagnosis. *Front. Med.* **2022**, *9*, 921101. [[CrossRef](#)]
25. Sepúlveda, N.; Stresman, G.; White, M.T.; Drakeley, C.J. Current Mathematical Models for Analyzing Anti-Malarial Antibody Data with an Eye to Malaria Elimination and Eradication. *J. Immunol. Res.* **2015**, *2015*, 1–21. [[CrossRef](#)] [[PubMed](#)]
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
27. Malato, J.; Graça, L.; Sepúlveda, N. Impact of Misclassification and Imperfect Serological Tests in Association Analyses of ME/CFS Applied to COVID-19 Data. In *Recent Developments in Statistics and Data Science*; Bispo, R., Henriques-Rodrigues, L., Alpizar-Jara, R., de Carvalho, M., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 215–225.
28. Domingues, T.D.; Mourão, H.; Sepúlveda, N. Analysis of antibody data using Finite Mixture Models based on Scale Mixtures of Skew-Normal distributions. *medRxiv* **2021**, medRxiv:2021.03.08.21252807. [[CrossRef](#)]
29. Scheibenbogen, C.; Freitag, H.; Blanco, J.; Capelli, E.; Lacerda, E.; Authier, J.; Meeus, M.; Marrero, J.C.; Nora-Krukle, Z.; Oltra, E.; et al. The European ME/CFS Biomarker Landscape project: An initiative of the European network EUROMENE. *J. Transl. Med.* **2017**, *15*, 162. [[CrossRef](#)]
30. Herrera, S.; de Vega, W.C.; Ashbrook, D.; Vernon, S.D.; McGowan, P.O. Genome-epigenome interactions associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Epigenetics* **2018**, *13*, 1174–1190. [[CrossRef](#)]
31. Tanigawa, Y.; Li, J.; Justesen, J.M.; Horn, H.; Aguirre, M.; DeBoever, C.; Chang, C.; Narasimhan, B.; Lage, K.; Hastie, T.; et al. Components of genetic associations across 2,138 phenotypes in the UK Biobank highlight adipocyte biology. *Nat. Commun.* **2019**, *10*, 4064. [[CrossRef](#)]
32. Dibble, J.J.; McGrath, S.J.; Ponting, C.P. Genetic risk factors of ME/CFS: A critical review. *Hum. Mol. Genet.* **2020**, *29*, R117–R124. [[CrossRef](#)]
33. Hajdarevic, R.; Lande, A.; Mehlsen, J.; Rydland, A.; Sosa, D.D.; Strand, E.B.; Mella, O.; Pociot, F.; Fluge, Ø.; Lie, B.A.; et al. Genetic association study in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) identifies several potential risk loci. *Brain, Behav. Immun.* **2022**, *102*, 362–369. [[CrossRef](#)]
34. Giri, A.; Hellwege, J.N.; Keaton, J.M.; Park, J.; Qiu, C.; Warren, H.R.; Torstenson, E.S.; Kovesdy, C.P.; Sun, Y.V.; Wilson, O.D.; et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **2018**, *51*, 51–62. [[CrossRef](#)]
35. International Multiple Sclerosis Genetics Consortium (IMSGC). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **2013**, *45*, 1353–1360. [[CrossRef](#)] [[PubMed](#)]
36. Bjornevik, K.; Cortese, M.; Healy, B.C.; Kuhle, J.; Mina, M.J.; Leng, Y.; Elledge, S.J.; Niebuhr, D.W.; Scher, A.I.; Munger, K.L.; et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* **2022**, *375*, 296–301. [[CrossRef](#)] [[PubMed](#)]
37. Pheby, D.F.H.; Araja, D.; Berkis, U.; Brenna, E.; Cullinan, J.; de Korwin, J.D.; Gitto, L.; Hughes, D.A.; Hunter, R.M.; Trepel, D.; et al. A Literature Review of GP Knowledge and Understanding of ME/CFS: A Report from the Socioeconomic Working Group of the European Network on ME/CFS (EUROMENE). *Medicina* **2020**, *57*, 7. [[CrossRef](#)]
38. Lacerda, E.M.; Mudie, K.; Kingdon, C.C.; Butterworth, J.D.; O'Boyle, S.; Nacul, L. The UK ME/CFS Biobank: A Disease-Specific Biobank for Advancing Clinical Research Into Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Front. Neurol.* **2018**, *9*, 1026. [[CrossRef](#)] [[PubMed](#)]
39. Słomko, J.; Newton, J.L.; Kujawski, S.; Tafil-Klawe, M.; Klawe, J.; Staines, D.; Marshall-Gradisnik, S.; Zalewski, P. Prevalence and characteristics of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) in Poland: A cross-sectional study. *BMJ Open* **2019**, *9*, e023955. [[CrossRef](#)] [[PubMed](#)]
40. Ariza, M.E. Commentary: Antibodies to Human Herpesviruses in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Patients. *Front. Immunol.* **2020**, *11*, 1400. [[CrossRef](#)]
41. Grabowska, A.D.; Lacerda, E.M.; Nacul, L.; Sepúlveda, N. Review of the Quality Control Checks Performed by Current Genome-Wide and Targeted-Genome Association Studies on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Front. Pediatr.* **2020**, *8*, 293. [[CrossRef](#)]
42. Hajdarevic, R.; Lande, A.; Rekeland, I.; Rydland, A.; Strand, E.B.; Sosa, D.D.; Creary, L.E.; Mella, O.; Egeland, T.; Saugstad, O.D.; et al. Fine mapping of the major histocompatibility complex (MHC) in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) suggests involvement of both HLA class I and class II loci. *Brain, Behav. Immun.* **2021**, *98*, 101–109. [[CrossRef](#)]
43. Domingues, T.D.; Grabowska, A.D.; Lee, J.S.; Ameijeiras-Alonso, J.; Westermeier, F.; Scheibenbogen, C.; Cliff, J.M.; Nacul, L.; Lacerda, E.M.; Mourão, H.; et al. Herpesviruses Serology Distinguishes Different Subgroups of Patients From the United Kingdom Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Biobank. *Front. Med.* **2021**, *8*, 686736. [[CrossRef](#)]

44. Szklarski, M.; Freitag, H.; Lorenz, S.; Becker, S.C.; Sotzny, F.; Bauer, S.; Hartwig, J.; Heidecke, H.; Wittke, K.; Kedor, C.; et al. Delin-eating the Association Between Soluble CD26 and Autoantibodies Against G-Protein Coupled Receptors, Immunological and Cardiovascular Parameters Identifies Distinct Patterns in Post-Infectious vs. Non-Infection-Triggered Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Front. Immunol.* **2021**, *12*, 644548. [[CrossRef](#)] [[PubMed](#)]
45. Celli, M.; Chalder, T. Measuring fatigue in clinical and community settings. *J. Psychosom. Res.* **2010**, *69*, 17–22. [[CrossRef](#)] [[PubMed](#)]
46. Loebel, M.; Eckey, M.; Sotzny, F.; Hahn, E.; Bauer, S.; Grabowski, P.; Zerweck, J.; Holenya, P.; Hanitsch, L.G.; Wittke, K.; et al. Serological profiling of the EBV immune response in Chronic Fatigue Syndrome using a peptide microarray. *PLoS ONE* **2017**, *12*, e0179124. [[CrossRef](#)]
47. Kaushik, N. Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. *J. Clin. Pathol.* **2005**, *58*, 826–832. [[CrossRef](#)] [[PubMed](#)]
48. Johnston, S.; Staines, D.; Klein, A.; Marshall-Gradisnik, S. A targeted genome association study examining transient receptor potential ion channels, acetylcholine receptors, and adrenergic receptors in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *BMC Med. Genet.* **2016**, *17*, 79. [[CrossRef](#)] [[PubMed](#)]
49. Lande, A.; Fluge, Ø.; Strand, E.B.; Flåm, S.T.; Sosa, D.D.; Mella, O.; Egeland, T.; Saugstad, O.D.; Lie, B.A.; Viken, M.K. Human Leukocyte Antigen alleles associated with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *Sci. Rep.* **2020**, *10*, 5267. [[CrossRef](#)]
50. Kerr, J.R.; Petty, R.; Burke, B.; Gough, J.; Fear, D.; Sinclair, L.I.; Matthey, D.L.; Richards, S.C.M.; Montgomery, J.; Baldwin, D.A.; et al. Gene Expression Subtypes in Patients with Chronic Fatigue Syndrome/Myalgic Encephalomyelitis. *J. Infect. Dis.* **2008**, *197*, 1171–1184. [[CrossRef](#)]
51. Zhang, L.; Gough, J.; Christmas, D.; Matthey, D.L.; Richards, S.C.M.; Main, J.; Enlander, D.; Honeybourne, D.; Ayres, J.G.; Nutt, D.J.; et al. Microbial infections in eight genomic subtypes of chronic fatigue syndrome/myalgic encephalomyelitis. *J. Clin. Pathol.* **2009**, *63*, 156–164. [[CrossRef](#)]
52. Jason, L.A.; Unger, E.R.; Dimitrakoff, J.D.; Fagin, A.P.; Houghton, M.; Cook, D.B.; Marshall, G.D.; Klimas, N.; Snell, C. Minimum data elements for research reports on CFS. *Brain Behav. Immun.* **2012**, *26*, 401–406. [[CrossRef](#)]
53. Nacul, L.; Authier, F.J.; Scheibenbogen, C.; Lorusso, L.; Helland, I.B.; Martin, J.A.; Sirbu, C.A.; Mengshoel, A.M.; Polo, O.; Behrends, U.; et al. European Network on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (EUROMENE): Expert Consensus on the Diagnosis, Service Provision, and Care of People with ME/CFS in Europe. *Medicina* **2021**, *57*, 510. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

F.4 Chapter 5



Revisiting IgG Antibody Reactivity to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Its Potential Application to Disease Diagnosis

OPEN ACCESS

Edited by:

Giusto Trevisan,
University of Trieste, Italy

Reviewed by:

Katiúscia Nan,
Department of Dermatology, Italy
Michele Gardarelli,
CFU Italia odv, Italy

***Correspondence:**

Nuno Sepúlveda
n.sepulveda@mini.pw.edu.pl
Carmen Scheibenbogen
carmen.scheibenbogen@charite.de

Specialty section:

This article was submitted to
Infectious Diseases - Surveillance,
Prevention and Treatment,
a section of the journal
Frontiers in Medicine

Received: 15 April 2022

Accepted: 26 May 2022

Published: 24 June 2022

Citation:

Sepúlveda N, Malato J, Sotzny F, Grabowska AD, Fonseca A, Cordeiro C, Graça L, Biecek P, Behrends U, Mautner J, Westermeier F, Lacerda EM and Scheibenbogen C (2022) Revisiting IgG Antibody Reactivity to Epstein-Barr Virus in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome and Its Potential Application to Disease Diagnosis. *Front. Med.* 9:921101.
doi: 10.3389/fmed.2022.921101

Nuno Sepúlveda^{1,2*}, João Malato^{2,3}, Franziska Sotzny⁴, Anna D. Grabowska⁵, André Fonseca^{2,6}, Clara Cordeiro^{2,6}, Luís Graça³, Przemysław Biecek¹, Uta Behrends^{7,8}, Josef Mautner^{7,8}, Francisco Westermeier^{9,10}, Eliana M. Lacerda¹¹ and Carmen Scheibenbogen^{4*}

¹ Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland, ² CEAUL – Centro de Estatística e Aplicações da Universidade de Lisboa, Lisbon, Portugal, ³ Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal, ⁴ Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt Universität zu Berlin, Institute of Medical Immunology, Berlin, Germany, ⁵ Department of Biophysics, Physiology, and Pathophysiology, Medical University of Warsaw, Warsaw, Poland, ⁶ Faculdade de Ciências e Tecnologia, Universidade do Algarve, Faro, Portugal, ⁷ Technical University of Munich, School of Medicine, Childrens' Hospital, Munich, Germany, ⁸ German Center for Infection Research (DZIF), Braunschweig, Germany, ⁹ Department of Health Studies, Institute of Biomedical Science, FH Joanneum University of Applied Sciences, Graz, Austria, ¹⁰ Centro Integrativo de Biología y Química Aplicada (CIBQA), Universidad Bernardo O'Higgins, Santiago, Chile, ¹¹ Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

Infections by the Epstein-Barr virus (EBV) are often at the disease onset of patients suffering from Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). However, serological analyses of these infections remain inconclusive when comparing patients with healthy controls (HCs). In particular, it is unclear if certain EBV-derived antigens eliciting antibody responses have a biomarker potential for disease diagnosis. With this purpose, we re-analyzed a previously published microarray data on the IgG antibody responses against 3,054 EBV-related antigens in 92 patients with ME/CFS and 50 HCs. This re-analysis consisted of constructing different regression models for binary outcomes with the ability to classify patients and HCs. In these models, we tested for a possible interaction of different antibodies with age and gender. When analyzing the whole data set, there were no antibody responses that could distinguish patients from healthy controls. A similar finding was obtained when comparing patients with non-infectious or unknown disease trigger with healthy controls. However, when data analysis was restricted to the comparison between HCs and patients with a putative infection at their disease onset, we could identify stronger antibody responses against two candidate antigens (EBNA4_0529 and EBNA6_0070). Using antibody responses to these two antigens together with age and gender, the final classification model had an estimated sensitivity and specificity of 0.833 and 0.720, respectively. This reliable case-control discrimination suggested the use of the antibody levels related to these

candidate viral epitopes as biomarkers for disease diagnosis in this subgroup of patients. To confirm this finding, a follow-up study will be conducted in a separate cohort of patients.

Keywords: Epstein-Barr virus, Myalgic Encephalomyelitis/Chronic Fatigue Syndrome, antigen mimicry, biomarker discovery, patient stratification

INTRODUCTION

Infections by the ubiquitous Epstein-Barr virus (EBV) are linked to multiple sclerosis, rheumatoid arthritis, systemic erythematosus lupus, lymphomas, among other known diseases (1–3). A less-known disease where EBV infections are also important is Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) (4–6). The hallmark symptom of this condition is an unexplained but persistent fatigue that cannot be alleviated by rest and that can increase upon minimal physical and emotional effort (7, 8). In ME/CFS, acute EBV infections are reported by a subset of patients at the onset of their symptoms (9, 10). Reactivation of latent EBV infections has also been described during the disease course (11). However, current evidence remains inconclusive on whether the prevalence of these reactivations is either higher or lower in patients than in healthy controls (12). This conflicting evidence notwithstanding, ME/CFS patients show deficient B- and T-cell responses against EBV and altered antibody profiles when compared with healthy controls (10, 13–15). Finally, CD4+ T cells recognizing self-peptides on HLA-DR15, the strongest genetic risk factor for multiple sclerosis, have been shown to cross-react with peptides derived from EBV (16). Multiple sclerosis patients share many symptoms with the ones suffering from ME/CFS (17–19). EBV antigens were also reported to share sequence homology with human peptides derived from the myelin basic protein (20–22), lactoperoxidase (23), and anoctamin-2 (24, 25). These observations suggest that molecular mimicry between human and EBV-derived antigens could play a role in the pathogenesis of ME/CFS. This suggestion is in line with our recent hypothesis that links the pathogenesis of ME/CFS to chronically activated immune responses (26). Our assumption raises the possibility that the immune system of some ME/CFS patients is oscillating between an activation state that attempts controlling latent herpesviruses infections and the suppression of deleterious autoimmune responses *via* the activation of regulatory T cells (26). Thus, considering the growing body of evidence that links EBV infection to the pathogenesis of ME/CFS, studies that aim at elucidating underlying mechanisms are needed.

A major problem in investigating ME/CFS is the inexistence of a robust biomarker that could ascertain the disease diagnosis. In the past, different discovery studies suggested certain cytokines, antibodies against self and non-self-antigens, microRNAs, and methylation markers as potential disease biomarkers (27). Antibodies against EBV antigens are of particular interest as disease biomarkers given the above evidence connecting this virus with the disease and routine application of serological assays in the clinical practice. However, EBV

antigens included in commercial kits are mostly markers of exposure to the infection and are unable to distinguish between patients with ME/CFS and healthy controls (28). This distinction can only be made when comparing a subset of clinically diagnosed ME/CFS patients with an EBV infection trigger to healthy controls (10). A serological evaluation of antibodies against less-studied EBV antigens did not identify any that could be used as a specific disease biomarker (29). However, this antibody evaluation was done using a limited number of EBV-derived antigens and no subgroup analysis was performed. The lack of patient stratification in ME/CFS studies reduces the chance of reproducing the same findings in follow-up studies (27, 30). Therefore, it is still possible to identify alternative antigens whose antibody responses could be used as disease biomarkers for a subgroup of patients.

Recently, we analyzed antibody responses against more than 3,000 overlapping antigens derived from 14 EBV proteins (23). The aim of this study was to extract an antibody signature against EBV in ME/CFS patients when compared to healthy controls. In the present study, we extended the analysis of the obtained data with the specific objective of optimizing biomarker discovery. In particular, we compared patients with or without an infectious trigger at disease onset to healthy controls in order to discover EBV-derived antigens whose antibody responses could be used for ME/CFS diagnosis.

MATERIALS AND METHODS

Study Participants

Ninety-two ME/CFS patients were recruited between 2011 and 2015 at the Charité outpatient clinic for immunodeficiencies at the Institute of Medical Immunology in the Charité Universitätsmedizin Berlin, Germany. Additional fifty individuals were recruited from the employees of the same clinic, who self-reported to be healthy and to not suffer from fatigue. However, neither clinical nor laboratory assessment was performed to confirm the healthy status of those individuals. ME/CFS patients and healthy controls were matched for gender and age (Table 1) with 50% of women and an overall average of ~43 years of age. Fifty-four out of 92 patients (58.7%) reported an acute infection at their disease onset, whilst the remaining 38 patients (41.3%) reported either a disease trigger other than an infection, did not know their disease onset or the information about the disease trigger was missing. These two subgroups were also matched for age and gender (Table 1).

TABLE 1 | Basic characteristics of ME/CFS patients and healthy controls, where *p*-values refer to the comparison between ME/CFS groups and healthy controls.

Group	Female		Age, years		<i>P</i> -value
	<i>N</i>	%	<i>P</i> -value	Mean (age range)	
Healthy controls	50	50.0	N/A	42.4 (25–61)	N/A
ME/CFS (all)	92	51.1	0.901	43.7 (25–66)	0.453
With infectious trigger	54	50.0	~1.000	43.2 (17–66)	0.585
Unknown trigger or without infectious trigger	38	52.6	0.807	44.4 (24–66)	0.679

Peptide Array

Data under analyses refer to the signal intensities derived from IgG antibody responses to 3,054 EBV-associated peptides measured by a seroarray described in detail in the original study (23). These peptides consisted of partially overlapping 15 amino acids (15-mer) and covered the full length of the following proteins (**Supplementary Table 1**): BALF-2, BALF-5, BFRF-3, BLLF-1, BLLF-3, BLRF-2, BMRF-1, BZLF-1, EBNA-1, EBNA-3, EBNA-4, EBNA-6, LMP-1, and LMP-2. The 15-mer peptides overlapped in 11 amino acids. The amino-acid sequences of these peptides were representative of the following EBV strains: AG876 (West Africa, EBV type 2), B95-8 (USA, EBV type 1), GD1 (China, EBV type 1), Cao (China, EBV type 1), Raji (Nigeria, EBV type 1), and P3HR-1 (Nigeria, EBV type 2). These data are freely available in Supplementary File S1 of the original study (23).

Statistical Analysis

We used the Chi-square test to compare ME/CFS patients to healthy controls in terms of gender distribution. The non-parametric Mann-Whitney test was used to compare the medians of the respective age distributions. There was evidence for age- and gender-matched distributions if the *p*-values of these tests were greater than the significance level of 0.05.

We first performed a multivariate analysis using (i) the classical principal component analysis (PCA) and (ii) computing different correlation matrices using Spearman's correlation coefficient (which is invariant to monotonic changes in the scale of the data, is robust against the presence of outliers, and does not depend on the normality assumption). We then performed linear discriminant analyses (LDA) to determine the best linear combination of all the antibody responses that could distinguish ME/CFS patients and their subgroups from healthy individuals. A similar analysis was done to compare the two subgroups of ME/CFS patients.

The outcome of each LDA was the estimated classification probability for each individual. These estimated probabilities were then analyzed by the respective receiver operating characteristic (ROC) curve where 1-specificity and sensitivity are plotted against each other as a function of the cutoff of the underlying classification probability. After computing each ROC curve, we calculated the respective area under the

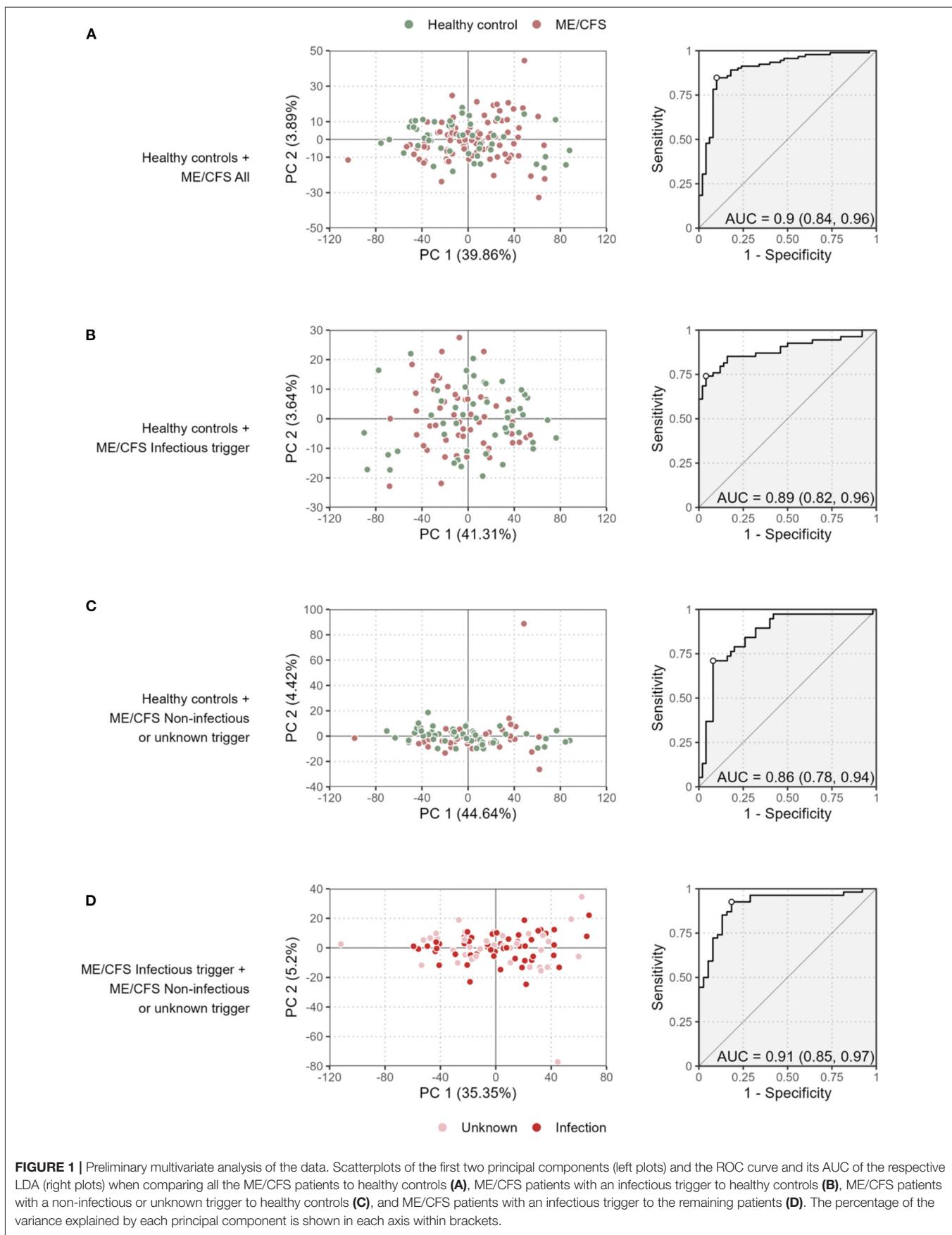
curve (AUC) and its 95% confidence interval to determine the accuracy of the classification irrespective of the cut-off used. In general, an $AUC = 0.50$ is indicative of a complete random classification of the individuals, while $AUC = 1.00$ implies that the constructed classifier perfectly predicts the true class membership of each individual.

We performed further antibody-wide association analyses related to the following comparisons (or classification exercises): (i) healthy controls versus all the ME/CFS patients; (ii) healthy controls versus ME/CFS patients with an infectious trigger; (iii) healthy controls versus ME/CFS patients with a non-infectious or unknown trigger; and (iv) ME/CFS patients with an infectious trigger versus the remaining ME/CFS patients. In each association analysis, we first estimated three regression models: logistic model, probit model, and complementary log-log model. In these models, the disease status was the outcome variable, age and gender were the respective covariates. To determine the best link function for the outcome variable, we selected the model with the lowest Akaike's information criterion (AIC). For the best link function ("the null model"), we estimated the respective ROC and its AUC as described above.

We fitted five different logistic models, including the main effects and all the interaction terms related to age, gender, and the antibody response under analysis: (i) a model with main effects only and no interaction terms; (ii) a model with an interaction term between age and the antibody response; (iii) a model with an interaction term between gender and the antibody response; (iv) a model with two interaction terms between age and the antibody response and between gender and the antibody response; (v) a model with all two-way and three-way interaction terms related to age, gender, and the antibody response. We compared each of these models with the null one using Wilks's likelihood ratio test, where low *p*-values provide evidence for these models, including effects of an antibody response. We reported the minimum *p*-value obtained from these model comparisons. Finally, we adjusted the minimum *p*-values of each analysis. This adjustment was made using the Benjamini-Yekutieli procedure ensuring a global false discovery rate (FDR) of 5% under the assumption of dependent tests (31). In this analysis, adjusted *p*-values < 0.05 indicated statistically significant results.

To filter out redundant antibody responses, we pooled all the significant antibody responses in a single model. The effect and interaction terms of these antibody responses were defined according to the most significant model obtained in the previous stage of analysis. We performed a backward stepwise model selection. The resulting model was finally evaluated in terms of predictive performance using ROC analysis as described above.

The above analysis was primarily done for the whole data set irrespective of the ME/CFS subgroups. We repeated the same analysis to compare each subgroup of ME/CFS patients (with infectious and non-infectious or unknown disease trigger) with the healthy controls. Finally, we repeated the analysis to compare the two subgroups of ME/CFS patients.



Statistical Software

The statistical analysis was performed in the R software version 4.0.3 with core functions and the following packages: MASS v7.3-56 to perform stepwise model selection (32), pROC v1.18.0 to estimate the ROC curve and the respective AUC (33), OptimalCutpoints v1.1-5 to estimate the optimal cutoff and the associated sensitivity/specificity (34). The full reproducible code is freely available from NS or JMal upon request.

RESULTS

Principal Component and Linear Discriminant Analyses

We first performed a PCA to discriminate patients with ME/CFS and their subgroups from healthy controls (**Figures 1A–C**). A similar analysis was done for discriminating patients with an infectious trigger from the remaining patients (**Figure 1D**).

The proportion of variance explained by the first principal component varied from 35.4% (**Figure 1D**) to 44.6% (**Figure 1C**) referring to the comparisons between the two subgroups of ME/CFS patients, and between healthy controls and patients with non-infectious or unknown disease trigger, respectively. These high estimates suggested that different antibody levels were correlated with each other. This interpretation was confirmed by determining the distributions of Spearman's correlation coefficient between all possible pairs of antibodies using data from each study group (**Supplementary Figure 1**). In particular, the antibody levels were positively correlated with each other with median correlation estimates of 0.56, 0.56, 0.40, and 0.48 for healthy controls, all the ME/CFS patients, ME/CFS patients with an infectious disease trigger, and the remaining ME/CFS patients, respectively. Interestingly, the median correlation estimate was decreased in ME/CFS patients with an infectious trigger when compared to other study groups. This finding suggested that the production of the antibodies against the EBV-derived antigens could be reduced in these patients when compared to healthy controls or patients with non-infectious or unknown disease trigger.

The visualization of the first two components did not reveal a clear discrimination between healthy controls and ME/CFS patients (or their subgroups). To improve this analysis, we then performed different LDAs in search of a linear combination of the antibody measurements that could be used for disease diagnosis. The performance of the constructed classifiers ranged from 0.86 (**Figure 1C**) to 0.91 (**Figure 1D**) referring to the classification of healthy controls and ME/CFS patients with non-infectious or unknown disease trigger and the classification of the two subgroups of ME/CFS patients, respectively. Therefore, the results of this analysis indicate that the antibody data could discriminate different study groups.

Antibody-Wide Association Analysis

The next step of the analysis was to identify specific antibody responses that could be used to discriminate the different study groups. With this purpose, we first determined the best "null" model among the logistic, probit, and complementary log-log models. All of them included age and gender and their interaction as covariates for each comparison between any two

TABLE 2 | Estimates of the final complementary log-log model to discriminate ME/CFS patients with an infectious disease trigger from healthy controls.

Model term	Coefficient estimate (SE)	P-value
Intercept	10.67 (10.33)	0.302
Age (in years)	-0.49 (0.26)	0.060
Gender (Woman)	-17.33 (6.85)	0.011
EBNA4_0529	2.25 (1.09)	0.039
EBNA6_0070	-5.62 (3.09)	0.069
Age × Gender	0.07 (0.04)	0.070
Gender × EBNA6_0070	4.05 (1.75)	0.021
Age × EBNA6_0070	0.15 (0.08)	0.062

study groups (**Supplementary Table 2**). The best "null" models were the following: (i) complementary log-log - comparison between healthy controls and all the ME/CFS patients [AUC = 0.574; 95% CI = (0.475; 0.672)]; (ii) probit-comparison between healthy controls and ME/CFS patients with an infectious trigger [AUC = 0.606; 95% CI = (0.496; 0.715)]; (iii) complementary log-log - comparison between healthy controls and ME/CFS patients with a non-infectious or unknown trigger [AUC = 0.556; 95% CI = (0.429; 0.683)]; and (iv) logit - comparison between the two subgroups of ME/CFS groups [AUC = 0.596; 95% CI = (0.471; 0.720)]. The 95% confidence interval for the AUC of these null models included 0.50 and therefore, the respective predicted classification was consistent with a random guess. Such a result was in agreement with the age and gender matching between different study groups and healthy controls (**Table 1**).

We performed further antibody-wide association analyses controlling for a global FDR of 5%. The comparison between healthy controls and all the ME/CFS patients did not identify any significant antibody associations with the disease (**Figure 2A**). The top 5 antibodies, although not statistically significant, were EBNA6_0066, BLRF2_0005, EBNA4_0392, EBNA4_0497, and EBNA4_0529 (adjusted p-values = 0.181, 0.326, 0.326, 0.326, and 0.326, respectively).

When the comparison was limited to healthy controls and ME/CFS patients with an infectious trigger, we identified three significant antibodies related to the following antigens (**Figure 2B**): EBNA6_0066, EBNA6_0070, and EBNA4_0529 (adjusted p-values = 0.005, 0.005, 0.038, respectively). The first two antigens were shared between AG876, B95-8, and GD1 strains, while the third one was derived from the B95-8 strain. We compared ME/CFS patients with non-infectious or unknown disease trigger to healthy controls, and found no significant differences in the antibody responses (**Figure 2C**). The same finding was obtained when we compared the two subgroups of ME/CFS patients (**Figure 2D**). The top 5 antibodies related to these analyses can be found in **Supplementary Table 3**.

Analysis of Candidate Antigens for Classifying ME/CFS Patients With an Infectious Trigger

We then analyzed in detail the impact of the antibody levels against the three candidate antigens on the classification of ME/CFS patients with an infectious trigger. Antibody levels

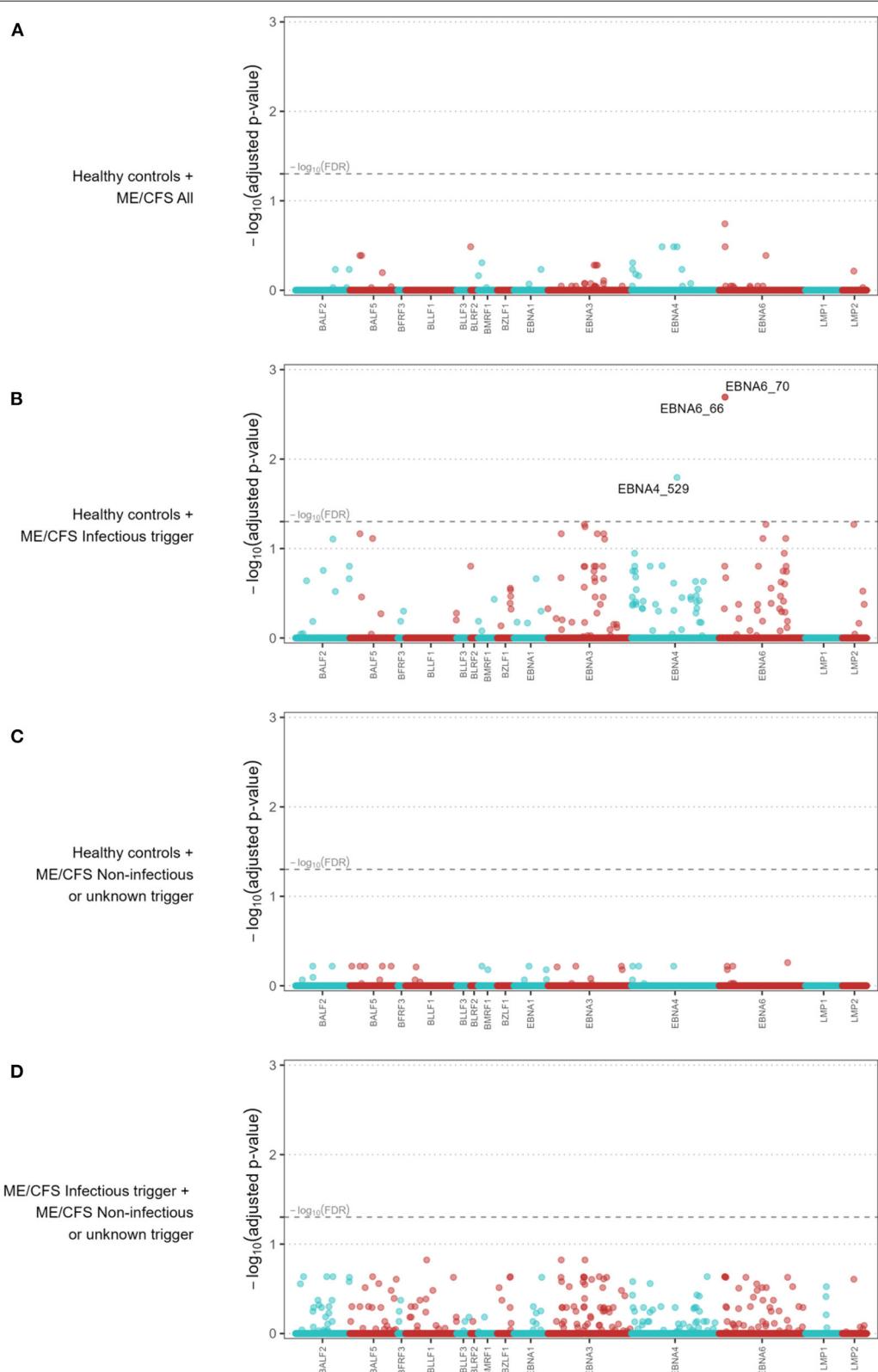


FIGURE 2 | Antibody-wide association analyses when comparing all the ME/CFS patients to healthy controls **(A)**, ME/CFS patients with an infectious trigger to healthy controls **(B)**, ME/CFS patients with a noninfectious or unknown trigger to healthy controls **(C)**, and ME/CFS patients with an infectious trigger to the remaining patients **(D)**. The x-axes comprise each antibody while the y-axes represent the $-\log_{10}(\text{adjusted } p\text{-value})$ of the respective association. In the x-axes, the antibodies

(Continued)

FIGURE 2 | were ordered alphabetically first by the protein name and then by the starting point of the antigen within the protein. Adjusted p-values were calculated according to the Benjamini-Yekutieli procedure for a global FDR of 5% under the assumption of dependent data. Dashed line represents the threshold for statistical significance (i.e., $-\log_{10}(\text{FDR}) = 0.05$) and $-\log_{10}(\text{adjusted } p\text{-values}) > 1.30$ were considered statistically significant.

were increased in this subgroup of ME/CFS patients when compared to healthy controls (**Figure 3A**). The same evidence could not be found when comparing all the ME/CFS patients to healthy controls (**Figure 3A**). Data related to EBNA4_0529, EBNA6_0066 and EBNA6_0070 were significantly correlated with each other (Spearman's correlation coefficients higher than 0.58; **Figure 3B**). The correlation between the levels of antibodies against EBNA6_0066 and EBNA6_0070 could be explained by the fact that these two peptides are 15-mers overlapping 11 amino acids with each other (23). In contrast, it was unclear why the levels of antibodies against EBNA4_0529 and EBNA6_0066 were highly correlated (Spearman's correlation coefficient = 0.79), considering that these antigens did not share a high sequence homology (**Figure 3C**).

Given the high correlation between antibody levels related to these antigens, a statistical redundancy was expected when using their data for patients' classification purpose. This redundancy was confirmed when the three candidate antibodies were included as covariates in the same model. A stepwise variable selection procedure led to the exclusion of the antibody levels related to EBNA6_0066 from the final classification model.

The final model included the main effects of antibodies to EBNA4_0529 and EBNA6_0070 and the two-way interaction of the latter with age and gender (**Table 2**). On the one hand, the \log_{10} -levels of antibodies related to EBNA4 increased the probability of being a patient (coefficient estimate = 2.25, Standard error = 1.09). In particular, the odds of being a patient were estimated to increase ~ 9.5 ($e^{2.25}$) times per fold-change in the levels of these antibodies. On the other hand, the effects of antibody levels related to EBNA6_0070 on the probability of an individual being an ME/CFS patients were not so trivial to ascertain (**Figure 4A**). In particular, women with high EBNA6_0070 antibody levels showed an increasing estimated probability of being a patient with increasing age. In contrast, the probability profile of being patient was different in men. In that case, younger men with low EBNA6_0070 antibody levels or older men with high EBNA6 antibody levels had a higher probability of being a patient.

The AUC of the classification predicted by the final model was estimated at 0.835 with a 95% CI = (0.759; 0.911; **Figure 4B**). This estimate suggested that the combination of these two antibodies together with age and gender could be used for the diagnosis of patients with an infectious trigger. The optimal sensitivity and specificity were estimated at 0.833 and 0.720, respectively. Therefore, ME/CFS patients were better discriminated than healthy controls by this model.

When the same classification model was applied to the whole cohort of ME/CFS patients, the AUC decreased to 0.731 with a 95% CI = (0.648, 0.814). This could be explained by the cohort of patients with a non-infectious or unknown trigger in which the performance of the classification

model was close to a random guess [AUC = 0.583; 95% CI = (0.461; 0.705)].

DISCUSSION

This study, based on previously published data, aimed to discover EBV-derived antigens that could elicit distinct antibody responses in ME/CFS patients when compared to healthy controls. The key finding was the identification of two candidate antigens inducing increased antibody responses in ME/CFS patients with an infectious trigger. The high sensitivity and specificity of our classification model including these antibodies suggest their potential for diagnosis of this subgroup of affected individuals. For ME/CFS patients without an infectious trigger, we could not find any antigens causing antibody responses that could be used for diagnostic purposes. This finding is in agreement with an extensive serological investigation of different herpesviruses in ME/CFS patients (29). This negative finding supports the hypothesis that EBV plays a role in the group of ME/CFS patients with an infectious trigger. In a subset of patients, infectious mononucleosis caused by primary EBV infection can be documented as a trigger (10). In many others, no infection with a specific pathogen could be associated with the disease onset (5). A tempting hypothesis from our finding is that EBV reactivation which can occur during other infections may play until now an underestimated role in triggering ME/CFS. In line with this concept, a recent study showed that EBV reactivation during COVID-19 is a risk factor for Post COVID Syndrome which also includes ME/CFS (35). Alternatively, the responses to the EBNA6 peptides are due to a cross-reactivity to other pathogens, as outlined below.

Other findings of this study pointed to three key challenges associated with the discovery of a biomarker. Firstly, it is difficult to identify a disease-specific biomarker for all the ME/CFS patients. Thus, given the heterogeneous nature of ME/CFS, it is pivotal to stratify patients adequately (30), based on age, gender, and disease trigger for biomarker discovery (27). In this regard, the identification of antibody patterns specific to ME/CFS patients with an infectious trigger was in agreement with other studies where significant results could be found for the same subgroup of patients (10, 36, 37). However, given the vast number of infectious agents associated with ME/CFS (5, 38), it is worth noting that this subgroup of patients could be further subdivided according to the nature of the causative infection. In this regard, the data about the infectious agents that could have initiated ME/CFS are either inconclusive or simply based on self-reported history in most patients, as demonstrated by the data from the United Kingdom ME/CFS Biobank, where only a minority of patients had their infection confirmed with the lab test (10). Secondly, the final classification model included non-trivial statistical interactions of antibodies against EBNA6_0070

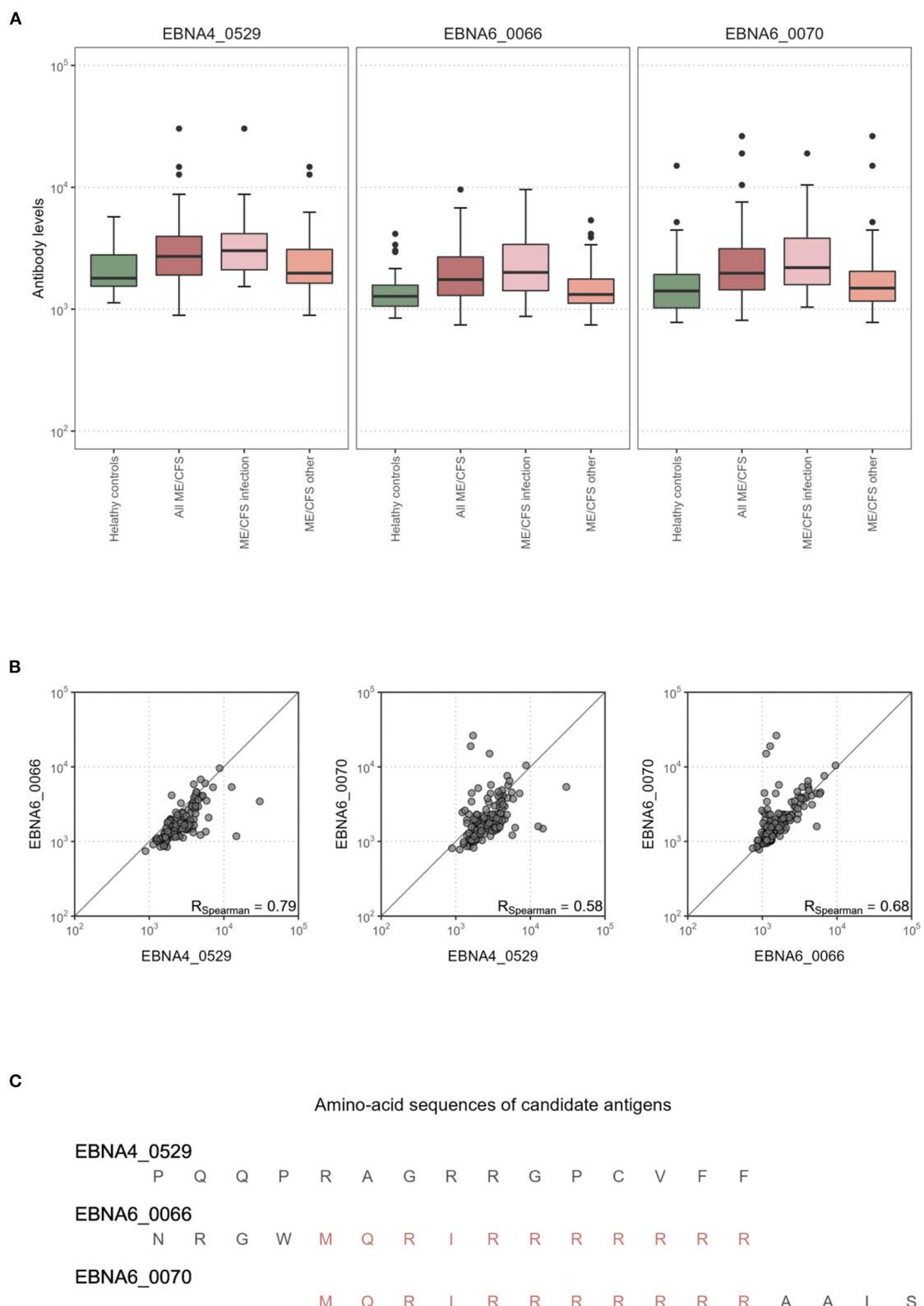


FIGURE 3 | Statistical analysis of the antibody levels related to EBNA4_0529, EBNA6_0066, and EBNA6_0070. **(A)** Boxplots of the data per study group. **(B)** Scatterplots and the respectively Spearman's correlation coefficients (R) in the whole dataset. **(C)** Amino acid sequences of EBNA4_0529, EBNA6_0066, and EBNA6_0070.

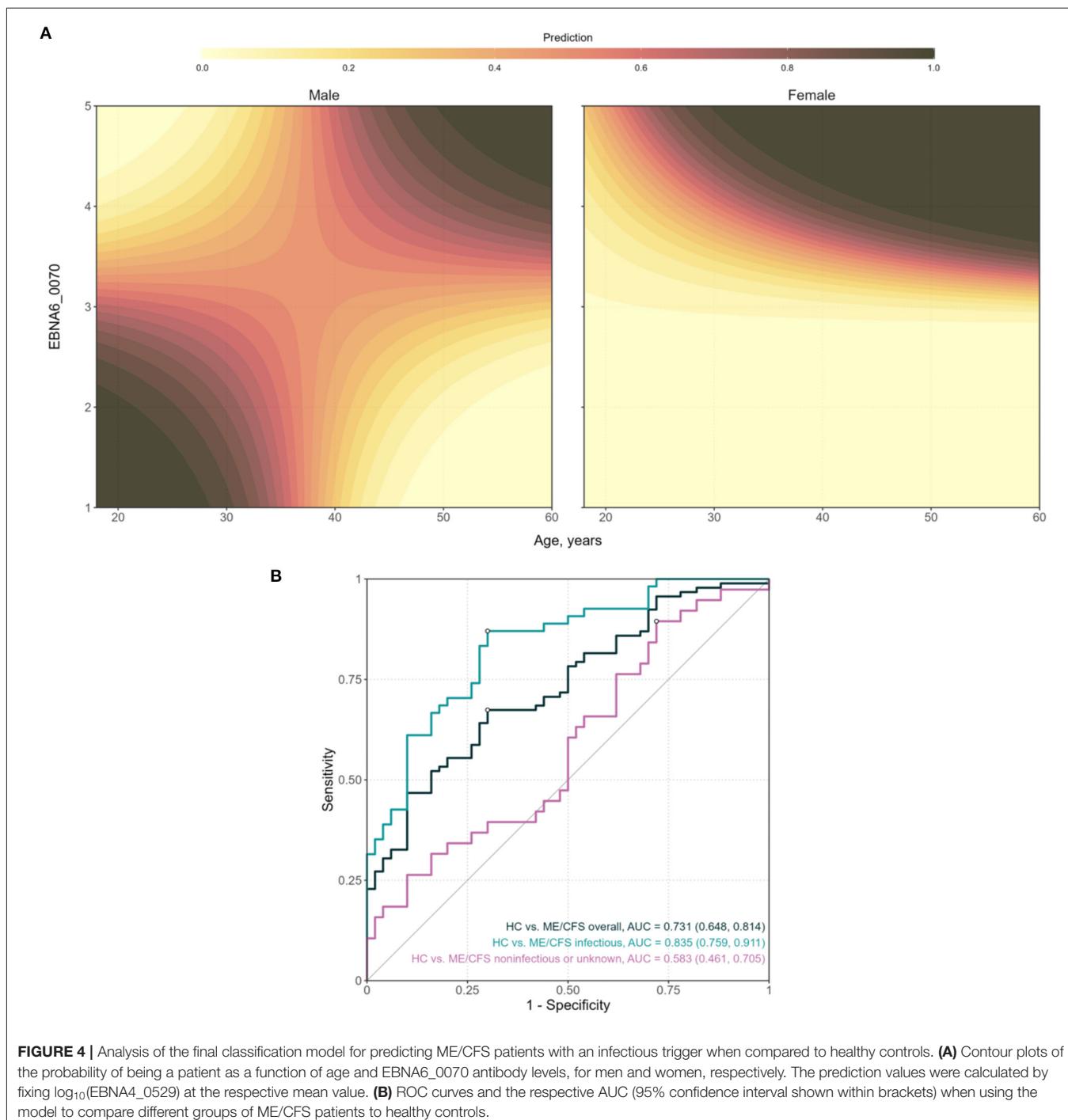


FIGURE 4 | Analysis of the final classification model for predicting ME/CFS patients with an infectious trigger when compared to healthy controls. **(A)** Contour plots of the probability of being a patient as a function of age and EBNA6_0070 antibody levels, for men and women, respectively. The prediction values were calculated by fixing $\log_{10}(\text{EBNA4}_0529)$ at the respective mean value. **(B)** ROC curves and the respective AUC (95% confidence interval shown within brackets) when using the model to compare different groups of ME/CFS patients to healthy controls.

with both age and gender. This finding implies that significant interactions between candidate biomarkers and confounding factors might be overlooked by analysts or, even when tested, they are likely to be discarded due to the small sample sizes to detect them. The presence of these interactions might be yet another factor that contributes to the lack of reproducibility between biomarker studies on ME/CFS. A proposed strategy to overcome this limitation is to conduct more advanced statistical analyses

including the application of machine learning techniques which intrinsically consider the complexity of a large set of clinical and biological data, as demonstrated in drug discovery (39). Thirdly, the interaction between the candidate antibodies against EBNA6_0070 and gender implied a remarkable distinct antibody signature between male and female patients. Again, this finding is in line with gender differences in immunity to viral infection (40). In particular, men have typically lower antibody responses

when vaccinated and are more susceptible to infections than women (41). In this regard, our study suggested that the higher probability of younger man being an ME/CFS patient is associated with lower levels of antibodies against the antigen EBNA6_0070. In contrast, female and male patients seemed to be at higher risk with higher antibodies at increasing age suggesting that at least a subset develop these antibody responses later in life. An implication of having a different antibody profiling between men and women is that analysis of each gender should be performed separately. At the same time, it is important to note that epidemiological data on ME/CFS suggested approximately a disease ratio of three women to one man (42–44). Therefore, if gender is an important stratification factor for biomarker discovery, studies should be designed toward a more balanced gender ratio. Similar sample sizes between male and female cohorts ensure comparable statistical power when analyzing data from each sex separately.

Both EBNA4_0529 and EBNA6_0070 antigens are derived from proteins whose genetic expression typically occurs during the EBV type III latency. Therefore, the acquisition of the respective antibodies might have occurred during initial B-cell transformation and immortalization. It could also be acquired slowly over time, given that the type III latency pattern can be detected sporadically in lymphoid follicles where EBV-infected B cells can proliferate and mimic a germinal center reaction program (45). We can hypothesize from our data that both male and female patients developing higher antibody responses against this antigen later in life are at an increased risk of developing ME/CFS suggesting that reactivation of EBV plays a role. In male patients a subgroup with lower EBNA6 antibodies early in life is at risk of developing ME/CFS, too. Using the recent analytical framework of ME/CFS natural progression (46), antibodies against these antigens are more likely to be biomarkers of patients suffering from ME/CFS more than 2 years of disease rather than the ones either in prodromal period or at early stages in line with our findings. Based on that assumption, these antibodies seemed more appropriate for diagnosing putative patients with delayed disease diagnosis rather than early suspected cases. However, it is known that the delay of ME/CFS diagnosis is a recurrent problem in the clinic (8, 47). As such, we anticipate a higher utility of these antibodies when redeployed to real-world screening. Another practical implication of using these antibodies as biomarkers is the possibility of developing routine ELISA kits that can be standardized across different laboratories and easily scalable for large population screenings. Notwithstanding these promising practical expectations, it is important to emphasize that past studies also suggested potential disease biomarkers (27) and, therefore, it is imperative to replicate the findings of this study with different cohorts of patients.

An interesting observation is that both EBNA6_0066 and EBNA6_0070 contain an arginine-repeat sequence. Such a sequence has homologies with putative epitopes from several human proteins (48). Such homologies suggest a potential molecular mimicry between the viral and human antigens. Molecular mimicry can trigger deleterious autoimmune responses as hypothesized for ME/CFS pathogenesis (38, 49). Molecular mimicry between human and microbial antigens

has been also hypothesized for several autoimmune diseases (50), such as multiple sclerosis and rheumatoid arthritis, and Post COVID syndrome, whose patients share similar symptoms with ME/CFS ones (19, 51–53). Interestingly, T cell clones recognizing such arginine-repeat sequences were isolated from a patient with multiple sclerosis supporting our concept of epitope mimicry (48). Finally, arginine-repeat sequences are found in various other pathogens including enteroviruses and human papillomavirus which are also triggers of ME/CFS (5).

Further we can hypothesize that peptides highly enriched in arginine residues might be particularly susceptible to citrullination, in which arginine residues are post-translationally converted to citrulline. These post-translational modifications occur during cell death under normal physiological conditions. However, under chronic inflammation, the accumulation of citrullinated (auto)antigens in inflamed sites might lead to deleterious autoimmune responses, thus, promoting the onset of different autoimmune diseases (54). A potential cross-reactivity between microbial and citrullinated human antigens could also be a mechanism by which an autoimmune disease can be triggered. In rheumatoid arthritis, antibodies against EBNA-1 peptides were shown to cross-react with denatured collagen and keratin (55). However, in the present study, we could not find any antibodies against EBNA-1-derived peptides to be associated with ME/CFS. Interestingly, the serum levels of citrulline were reported to be elevated in ME/CFS patients when compared to healthy controls (56). However, another study could not confirm this finding, but instead provided evidence for increased plasma levels of arginine residues (57). Another source of antigen modification is the process of generating new and more immunogenic epitopes from ubiquitous molecules upon oxidative and nitrosative stress. In ME/CFS, IgM antibodies against several of these neoepitopes, including NO-Arginine, were increased in patients (58). In all of these possible scenarios, it is imperative to investigate the stability of this candidate biomarker antigen to post-translational modifications that could be occurred and eventually increased during the disease course.

In conclusion, this study identified two candidate antigens whose antibodies could be used to identify ME/CFS patients with an infectious trigger. To strengthen our findings, two other cohorts of patients are currently studied, including the well-characterized ME/CFS patients with different disease triggers and healthy controls from the United Kingdom ME/CFS biobank (10).

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: Loebel et al. (23).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Charité Universitätsmedizin Berlin in accordance with the 1964 World Medical Association Declaration of Helsinki and its later amendments (23). The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NS and CS conceived this research. NS, JMal, AG, and AF performed the data analysis. FS, UB, EML, and CS collected and provided the data. All authors interpreted and discussed the results. NS wrote the paper. All authors have read, revised, and approved the final version of the manuscript.

FUNDING

NS, FW, EML, and CS received funding from ME Research UK (SCIO Charity Number SC036942) for this project. NS acknowledges funding from the Fundação para a Ciência e Tecnologia, Portugal (ref. UIDB/00006/2020) and the Polish National Agency for Academic Exchange, Poland

(ref. PPN/ULM/2020/1/00069/U/00001). JMal and AF were funded by the Fundação para a Ciência e Tecnologia, Portugal (ref. SFRH/BD/149758/2019 and SFRH/BD/147629/2019, respectively). CC work was partially financed by national funds through FCT – Fundação para a Ciência e a Tecnologia under the project UIDB/00006/2020. EML acknowledges funding from the National Institutes of Health (ref. NIH 2R01AI103629) and the ME Association for the maintenance of the UK ME/CFS Biobank (UKMEB – ME Association (Grant PF8947_ME Association). CS acknowledges funding from the Weidenhammer Zoebel and Lost Voices Foundation, Germany.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.921101/full#supplementary-material>

REFERENCES

- Houen G, Trier NH. Epstein-barr virus and systemic autoimmune diseases. *Front Immunol.* (2020) 11:587380. doi: 10.3389/fimmu.2020.587380
- Shannon-Lowe C, Rickinson AB, Bell AI. Epstein-Barr virus-associated lymphomas. *Philos Trans R Soc B Biol Sci.* (2017) 372:20160271. doi: 10.1098/rstb.2016.0271
- Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science.* (2022) 375:296–301. doi: 10.1126/science.abj8222
- Koo D. Chronic fatigue syndrome. A critical appraisal of the role of Epstein-Barr virus. *West J Med.* (1989) 150:590–6.
- Rasa S, Nora-Krukle Z, Henning N, Eliassen E, Shikova E, Harrer T, et al. Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *J Transl Med.* (2018) 16:268. doi: 10.1186/s12967-018-1644-y
- Ruiz-Pablos M, Paiva B, Montero-Mateo R, Garcia N, Zabaleta A. Epstein-barr virus and the origin of myalgic encephalomyelitis or chronic fatigue syndrome. *Front Immunol.* (2021) 12:656797. doi: 10.3389/fimmu.2021.656797
- Rivera MC, Mastronardi C, Silva-Aldana CT, Arcos-Burgos M, Lidbury BA. Myalgic encephalomyelitis/chronic fatigue syndrome: a comprehensive review. *Diagnostics.* (2019) 9:91. doi: 10.3390/diagnostics9030091
- Bateman L, Bested AC, Bonilla HF, Chheda BV, Chu L, Curtin JM, et al. Myalgic encephalomyelitis/chronic fatigue syndrome: essentials of diagnosis and management. *Mayo Clin Proc.* (2021) 96:2861–78. doi: 10.1016/j.mayocp.2021.07.004
- Hickie I, Davenport T, Wakefield D, Vollmer-Conna U, Cameron B, Vernon SD, et al. Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: prospective cohort study. *BMJ.* (2006) 333:575. doi: 10.1136/bmj.38933.585764.AE
- Domingues TD, Grabowska AD, Lee JS, Ameijeiras-Alonso J, Westermeier F, Scheibenbogen C, et al. Herpesviruses serology distinguishes different subgroups of patients from the United Kingdom myalgic encephalomyelitis/chronic fatigue syndrome biobank. *Front Med.* (2021) 8:686736. doi: 10.3389/fmed.2021.686736
- Shikova E, Reshkova V, Kumanova A, Raleva S, Alexandrova D, Capo N, Murovska MB. Cytomegalovirus, Epstein-Barr virus, and human herpesvirus-6 infections in patients with myalgic encephalomyelitis/chronic fatigue syndrome. *J Med Virol.* (2020) 92:3682. doi: 10.1002/jmv.25744
- Lee JS, Lacerda EM, Naclul L, Kingdon CC, Norris J, O'Boyle S, et al. Salivary DNA loads for human herpesviruses 6 and 7 are correlated with disease phenotype in myalgic encephalomyelitis/chronic fatigue syndrome. *Front Med.* (2021) 8:1129. doi: 10.3389/fmed.2021.656692
- Kerr JR. Epstein-Barr virus induced gene-2 upregulation identifies a particular subtype of chronic fatigue syndrome/myalgic encephalomyelitis. *Front Pediatr.* (2019) 7:59. doi: 10.3389/fped.2019.00059
- Lerner AM, Ariza ME, Williams M, Jason L, Beqaj S, Fitzgerald JT, et al. Antibody to Epstein-Barr virus deoxyuridine triphosphate nucleotidohydrolase and deoxyribonucleotide polymerase in a chronic fatigue syndrome subset. *PLoS ONE.* (2012) 7:e47891. doi: 10.1371/journal.pone.0047891
- Loebel M, Strohschein K, Giannini C, Koelsch U, Bauer S, Doebris C, et al. Deficient EBV-specific B- and T-cell response in patients with Chronic Fatigue Syndrome. *PLoS ONE.* (2014) 9:e85387. doi: 10.1371/journal.pone.0085387
- Wang J, Jelcic I, Mühlénbruch L, Haunerdinger V, Toussaint NC, Zhao Y, et al. HLA-DR15 molecules jointly shape an autoreactive T cell repertoire in multiple sclerosis. *Cell.* (2020) 183:1264–81.e20. doi: 10.1016/j.cell.2020.09.054
- Malato J, Graça L, Naclul L, Lacerda E, Sepúlveda N. Statistical challenges of investigating a disease with a complex diagnosis. *medRxiv.* (2021) 2021:2021.03.19.21253905. doi: 10.1101/2021.03.19.21253905
- Morris G, Maes M. Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics. *BMC Med.* (2013) 11:205. doi: 10.1186/1741-7015-11-205
- Gaber TAZK, Oo WW, Ringrose H. Multiple sclerosis/chronic fatigue syndrome overlap: when two common disorders collide. *NeuroRehabilitation.* (2014) 35:529–34. doi: 10.3233/NRE-141146
- Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell.* (1995) 80:695–705. doi: 10.1016/0092-8674(95)90348-8
- Holmøy T, Kvale EØ, Vartdal F. Cerebrospinal fluid CD4+ T cells from a multiple sclerosis patient cross-recognize Epstein-Barr virus and myelin basic protein. *J NeuroVirology.* (2004) 105:278–83. doi: 10.1080/13550280490499524
- Lüemann JD, Jelčić I, Roberts S, Lutterotti A, Tackenberg B, Martin R, et al. EBNA1-specific T cells from patients with multiple

- sclerosis cross react with myelin antigens and co-produce IFN- γ and IL-2. *J Exp Med.* (2008) 205:1763–73. doi: 10.1084/jem.20072397
23. Loebel M, Eckey M, Sotzny F, Hahn E, Bauer S, Grabowski P, et al. Serological profiling of the EBV immune response in Chronic Fatigue Syndrome using a peptide microarray. *PLoS ONE.* (2017) 12:e0179124. doi: 10.1371/journal.pone.0179124
24. Tengvall K, Huang J, Hellström C, Kammer P, Biström M, Ayoglu B, et al. Molecular mimicry between Anoctamin 2 and Epstein-Barr virus nuclear antigen 1 associates with multiple sclerosis risk. *Proc Natl Acad Sci USA.* (2019) 116:16955–60. doi: 10.1073/pnas.1902616116
25. Sepúlveda N. Impact of genetic variation on the molecular mimicry between Anoctamin-2 and Epstein-Barr virus nuclear antigen 1 in Multiple Sclerosis. *Immunol Lett.* (2021) 238:29–31. doi: 10.1016/j.imlet.2021.07.007
26. Sepúlveda N, Carneiro J, Lacerda E, Nacul L. Myalgic encephalomyelitis/chronic fatigue syndrome as a hyper-regulated immune system driven by an interplay between regulatory T cells and chronic human herpesvirus infections. *Front Immunol.* (2019) 10:2684. doi: 10.3389/fimmu.2019.02684
27. Scheibenbogen C, Freitag H, Blanco J, Capelli E, Lacerda E, Authier J, et al. The European ME/CFS biomarker landscape project: an initiative of the European network EUROMENE. *J Transl Med.* (2017) 15:162. doi: 10.1186/s12967-017-1263-z
28. Cliff JM, King EC, Lee JS, Sepúlveda N, Wolf AS, Kingdon C, et al. Cellular immune function in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Front Immunol.* (2019) 10:796. doi: 10.3389/fimmu.2019.00796
29. Blomberg J, Rizwan M, Böhlén-Wiener A, Elfaitouri A, Julin P, Zachrisson O, et al. Antibodies to human herpesviruses in myalgic encephalomyelitis/chronic fatigue syndrome patients. *Front Immunol.* (2019) 10:1946. doi: 10.3389/fimmu.2019.01946
30. Jason LA, Corradi K, Torres-Harding S, Taylor RR, King C. Chronic fatigue syndrome: the need for subtypes. *Neuropsychol Rev.* (2005) 15:29–58. doi: 10.1007/s11065-005-3588-2
31. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat.* (2001) 29:1165–88. doi: 10.1214/aos/101369998
32. Venables WN, Ripley BD. *Modern Applied Statistics with S. Fourth.* New York, NY: Springer (2002). Available online at: <https://www.stats.ox.ac.uk/pub/MASS4/> (accessed May 28, 2021)
33. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* (2011) 12:1–8. doi: 10.1186/1471-2105-12-77
34. López-Ratón M, Rodríguez-Alvarez MX, Cadarso-Suárez C, Gude-Sampedro F. Optimalcutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *J Stat Softw.* (2014) 61:1–36. doi: 10.18637/jss.v061.i08
35. Su Y, Yuan D, Chen DG, Ng RH, Wang K, Choi J, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell.* (2022) 185:881–95.e20. doi: 10.1016/j.cell.2022.01.014
36. Steiner S, Becker SC, Hartwig J, Sotzny F, Lorenz S, Bauer S, et al. Autoimmunity-related risk variants in PTPN22 and CTLA4 are associated with ME/CFS with infectious onset. *Front Immunol.* (2020) 11:578. doi: 10.3389/fimmu.2020.00578
37. Szklarski M, Freitag H, Lorenz S, Becker SC, Sotzny F, Bauer S, et al. Delineating the association between soluble CD26 and autoantibodies against G-protein coupled receptors, immunological and cardiovascular parameters identifies distinct patterns in post-infectious vs. non-infection-triggered myalgic encephalomyelitis/chronic fatigue syndrome. *Front Immunol.* (2021) 12:1077. doi: 10.3389/fimmu.2021.44548
38. Blomberg J, Gottfries CG, Elfaitouri A, Rizwan M, Rosén A. Infection elicited autoimmunity and Myalgic encephalomyelitis/chronic fatigue syndrome: an explanatory model. *Front Immunol.* (2018) 9:229. doi: 10.3389/fimmu.2018.00229
39. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol Divers.* (2021) 25:1315–60. doi: 10.1007/s11030-021-10217-3
40. Jacobsen H, Klein SL. Sex differences in immunity to viral infections. *Front Immunol.* (2021) 12:3483. doi: 10.3389/fimmu.2021.720952
41. Aaby P, Benn CS, Flanagan KL, Klein SL, Kollmann TR, Lynn DJ, et al. The non-specific and sex-differential effects of vaccines. *Nat Rev Immunol.* (2020) 20:464–70. doi: 10.1038/s41577-020-0338-x
42. Chu L, Valencia IJ, Garvert DW, Montoya JG. Onset patterns and course of myalgic encephalomyelitis/chronic fatigue syndrome. *Front Pediatr.* (2019) 7:12. doi: 10.3389/fped.2019.00012
43. Johnston SC, Staines DR, Marshall-Gradisnik SM. Epidemiological characteristics of chronic fatigue syndrome/myalgic encephalomyelitis in Australian patients. *Clin Epidemiol.* (2016) 8:97–107. doi: 10.2147/CLEP.S96797
44. Nacul LC, Lacerda EM, Pheby D, Campion P, Molokhia M, Fayyaz S, et al. Prevalence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in three regions of England: a repeated cross-sectional study in primary care. *BMC Med.* (2011) 9:91. doi: 10.1186/1741-7015-9-91
45. Thorley-Lawson DA. EBV persistence—introducing the virus. *Curr Top Microbiol Immunol.* (2015) 390:151. doi: 10.1007/978-3-319-22822-8_8
46. Nacul L, O’Boyle S, Palla L, Nacul FE, Mudie K, Kingdon CC, et al. How myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) progresses: the natural history of ME/CFS. *Front Neurol.* (2020) 11:826. doi: 10.3389/fneur.2020.00826
47. Nacul L, Authier FJ, Scheibenbogen C, Lorusso L, Helland IB, Martin JA, et al. European network on myalgic encephalomyelitis/chronic fatigue syndrome (EUROMENE): expert consensus on the diagnosis, service provision, and care of people with ME/CFS in Europe. *Medicina.* (2021) 57:510. doi: 10.3390/medicina57050510
48. Sospedra M, Zhao Y, Hausen H, Muraro PA, Hamashin C, De Villiers EM, et al. Recognition of conserved amino acid motifs of common viruses and its role in autoimmunity. *PLoS Pathog.* (2005) 1:0335–48. doi: 10.1371/journal.ppat.0010041
49. Phelan J, Grabowska AD, Sepúlveda N. A potential antigenic mimicry between viral and human proteins linking Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) with autoimmunity: the case of HPV immunization. *Autoimmun Rev.* (2020) 19:102487. doi: 10.1016/j.autrev.2020.102487
50. Rojas M, Restrepo-Jiménez P, Monsalve DM, Pacheco Y, Acosta-Ampudia Y, Ramírez-Santana C, et al. Molecular mimicry and autoimmunity. *J Autoimmun.* (2018) 95:100–23. doi: 10.1016/j.jaut.2018.10.012
51. Ali S, Matcham F, Irving K, Chalder T. Fatigue and psychosocial variables in autoimmune rheumatic disease and chronic fatigue syndrome: a cross-sectional comparison. *J Psychosom Res.* (2017) 92:1–8. doi: 10.1016/j.jpsychores.2016.11.002
52. Moss-Morris R, Chalder T. Illness perceptions and levels of disability in patients with chronic fatigue syndrome and rheumatoid arthritis. *J Psychosom Res.* (2003) 55:305–8. doi: 10.1016/S0022-3999(03)00013-8
53. Komaroff AL, Lipkin WI. Insights from myalgic encephalomyelitis/chronic fatigue syndrome may help unravel the pathogenesis of postacute COVID-19 syndrome. *Trends Mol Med.* (2021) 27:895–906. doi: 10.1016/j.molmed.2021.06.002
54. Alghamdi M, Alasmari D, Assiri A, Mattar E, Aljaddawi AA, Alattas SG, et al. An overview of the intrinsic role of citrullination in autoimmune disorders. *J Immunol Res.* (2019) 2019:7592851. doi: 10.1155/2019/7592851
55. Birkenfeld P, Haratz N, Klein G, Sulitzeanu D. Cross-reactivity between the EBNA-1 p107 peptide, collagen, and keratin: implications for the pathogenesis of rheumatoid arthritis. *Clin Immunol Immunopathol.* (1990) 54:14–25. doi: 10.1016/0090-1229(90)90002-8
56. Pall ML. Levels of nitric oxide synthase product citrulline are elevated in sera of chronic fatigue syndrome patients. *J Chronic Fatigue Syndr.* (2002) 10:37–41. doi: 10.1300/J092v10n03_04
57. Naviaux RK, Naviaux JC Li K, Bright AT, Alaynick WA, Wang L, et al. Metabolic features of chronic fatigue syndrome. *Proc Natl Acad Sci USA.* (2016) 113:E5472–80. doi: 10.1073/pnas.1607571113

58. Maes M, Mihaylova I, Leunis J. Chronic fatigue syndrome is accompanied by an IgM-related immune response directed against neoepitopes formed by oxidative or nitrosative damage to lipids. *Neuro Endocrinol Lett.* (2006) 27:615–21. doi: 10.1097/YCO.0b013e32831a4728

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of

the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sepúlveda, Malato, Sotzny, Grabowska, Fonseca, Cordeiro, Graça, Biecek, Behrends, Mautner, Westermeier, Lacerda and Scheibenbogen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

F.5 Chapter 6



Association analysis between symptomology and herpesvirus IgG antibody concentrations in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and multiple sclerosis

Tiago Dias Domingues ^{a,b,1}, João Malato ^{b,c,1}, Anna D. Grabowska ^d, Ji-Sook Lee ^e, Jose Ameijeiras-Alonso ^f, Przemysław Biecek ^g, Luís Graça ^c, Helena Mourão ^{a,b}, Carmen Scheibenbogen ^h, Francisco Westermeier ^{i,j}, Luis Nacul ^{k,l}, Jacqueline M. Cliff ^m, Eliana Lacerda ^k, Nuno Sepúlveda ^{b,g,*}

^a Departamento de Estatística e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

^b Faculty of Mathematics & Information Science, Warsaw University of Technology, Warsaw, Poland

^c Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal

^d Department of Biophysics, Physiology, And Pathophysiology, Medical University of Warsaw, Warsaw, Poland

^e Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

^f Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

^g Faculty of Mathematics & Information Science, Warsaw University of Technology, Warsaw, Poland

^h Institute of Medical Immunology, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt Universität zu Berlin and Berlin Institute of Health, Berlin, Germany

ⁱ Department of Health Studies, Institute of Biomedical Science, FH Joanneum University of Applied Sciences, Graz, Austria

^j Centro Integrativo de Biología y Química Aplicada (CIBQA), Universidad Bernardo O'Higgins, Santiago, Chile

^k Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

^l BC Women's Hospital, Vancouver, BC V6H 3N1, Canada

^m Department of Life Sciences and Centre for Inflammation Research and Translational Medicine, Brunel University London, United Kingdom

ARTICLE INFO

Keywords:

Enzyme-linked immunosorbent assay

Epstein-barr virus

Cytomegalovirus

Human herpesvirus-6

Varicella-zoster virus

Herpes simplex virus-1 and -2

SuperLearner

United Kingdom ME/CFS biobank

ABSTRACT

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and multiple sclerosis (MS) are two complex and multifactorial diseases whose patients experience persistent fatigue, cognitive impairment, among other shared symptoms. The onset of these diseases has also been linked to acute herpesvirus infections or their reactivations. In this work, we re-analyzed a previously-described dataset related to IgG antibody responses to 6 herpesviruses (CMV – cytomegalovirus; EBV – Epstein-Barr virus; HHV6 – human herpesvirus-6; HSV1 and HSV2 – herpes simplex virus-1 and -2, respectively; VZV – varicella-zoster virus) from the United Kingdom ME/CFS biobank. The primary goal was to report the underlying symptomology and its association with herpesvirus IgG antibodies using data from 4 disease-trigger-based subgroups of ME/CFS patients ($n = 222$) and patients with MS ($n = 46$). The secondary objective was to assess whether serological data could distinguish ME/CFS and its subgroup from MS using a SuperLearner (SL) algorithm. There was evidence for a significant negative association between temporary eye insight disturbance and CMV antibody concentrations and for a significant positive association between

* Corresponding author. Faculty of Mathematics & Information Science, Warsaw University of Technology, Warsaw, Poland

E-mail address: N.Sepulveda@mini.pw.edu.pl (N. Sepúlveda).

¹ Authors contributed equally to this work.

<https://doi.org/10.1016/j.heliyon.2023.e18250>

Received 12 June 2023; Received in revised form 10 July 2023; Accepted 12 July 2023

Available online 13 July 2023

2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

bladder problems and EBV antibody concentrations in the MS group. In the ME/CFS or its subgroups, the most significant antibody-symptom association was obtained for increasing HSV1 antibody concentration and brain fog, a finding in line with a negative impact of HSV1 exposure on cognitive outcomes in both healthy and disease conditions. There was also evidence for a higher number of significant antibody-symptom associations in the MS group than in the ME/CFS group. When we combined all the serological data in an SL algorithm, we could distinguish three ME/CFS subgroups (unknown disease trigger, non-infection trigger, and an infection disease trigger confirmed in the lab at the time of the event) from the MS group. However, we could not find the same for the remaining ME/CFS group (related to an unconfirmed infection disease). In conclusion, IgG antibody data explains more the symptomatology of MS patients than the one of ME/CFS patients. Given the fluctuating nature of symptoms in ME/CFS patients, the clinical implication of these findings remains to be determined with a longitudinal study. This study is likely to ascertain the robustness of the associations during natural disease course.

1. Introduction

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a widely neglected disease characterized by persistent fatigue, post-exertional malaise (PEM), unrefreshing sleep, among other symptoms related to multiple body systems [1]. Despite the research efforts [2], there are no objective biomarkers for the diagnosis and prognosis of ME/CFS. The absence of this key clinical tool delays disease diagnosis and subsequent treatment [3]. It also slows down research progress due to lack of finding's reproducibility across studies [4,5]. Notwithstanding all these problems, there is a growing body of evidence for an autoimmune component for the origin and chronicity of the disease, especially in patients with an acute infection at their disease onset [6–9]. The main candidate proteins for this deleterious autoimmune phenomenon are the adrenergic receptors [10–12]. However, other human proteins, such as Anoctamin-2 and thyroid peroxidase, have also been suggested [13,14].

The autoimmune hypothesis for the etiology of ME/CFS directed research efforts towards the identification of key differences between ME/CFS and different autoimmune diseases, including multiple sclerosis (MS) [13,15]. In this regard, patients with MS were deemed an important disease control group given that they experience chronic fatigue as a major manifestation of their disease. These patients and the ME/CFS ones also share some neurological symptoms, such as brain fog, memory loss, cognitive impairment, and photosensitivity/photophobia [16]. In addition, the root cause of both diseases has been linked to the infections by herpesviruses [17–21], such as Epstein-Barr virus (EBV) [22–24] and human herpesvirus-6 (HHV6) [25–28]. Finally, it was recently hypothesized that ME/CFS and MS share reduced craniospinal compliance and dilated pressured bridging cortical veins [29].

The similarity between these two diseases also prompted the creation of the United Kingdom ME/CFS biobank (UKMEB), a large sample resource for the research community, where patients with MS were included as a disease control group [30,31]. Until now, most of the UKMEB-based studies compared ME/CFS patients to healthy controls [32–37] with a few exceptions where the MS group was also included in the analysis [38–41]. For example, one study reported significant differences in the frequency of several immune-cell populations between MS and ME/CFS patients [41]; however, some of these differences were not found when comparing these groups to healthy controls. The same study showed that MS patients had a significantly higher seroprevalence to EBV-derived nuclear antigen-1 (EBNA1) antigen when comparing to both healthy controls and ME/CFS patients divided according to their disease severity. Another study reported evidence for similar average concentration of the stress-related growth/differentiation factor 15 peptide between MS and ME/CFS group [39]. Therefore, it is important to increase current knowledge about the pathological differences between ME/CFS and MS using data from the UKMEB.

In a previous study, we divided the cohort of ME/CFS patients from the UKMEB into 4 subgroups according to their putative disease triggers and compared them to healthy controls in terms of IgG antibody responses to 6 herpesviruses [42]. We found reductions in antibody concentrations to EBV in the ME/CFS subgroup without a putative infection trigger and to CMV in the ME/CFS subgroup with a confirmed infection at disease onset. In the present study, we analyzed symptoms data from the UKMEB using the MS cohort as the control group and assessed the respective association with the same herpesvirus IgG antibody data mentioned above. We also integrated the IgG antibody data in a SuperLearner (SL) algorithm with the objective of assessing the classification power to distinguish each ME/CFS subgroup from the MS one. Overall, our analyses aimed at providing new data on what differentiates ME/CFS and its subgroups from MS. Given that there are more published studies related to the role of herpesviruses on MS than on ME/CFS, most of the findings reported here are not particularly novel (e.g., high EBV IgG antibody concentrations) for this disease.

2. Materials and methods

2.1. Study participants

The study participants were 222 adult patients with ME/CFS and 46 adult patients with MS whose serological data were available from the UKMEB. As in our previous study [42], we divided patients with ME/CFS into the following four subgroups according to the disease trigger: S0 - an unknown trigger ($n = 42$); S1 - a non-infectious trigger ($n = 42$); S2 - an infection that was not evaluated by a lab test at the time of the event ($n = 92$); and S3 - an infection that was confirmed by a lab test at the time of the event ($n = 46$). The diagnosis of these patients complied with either the 1994 Centers for Disease Control and Prevention (1994 CDC) criteria [43] or the

2003 Canadian Consensus Criteria (2003 CCC) [44]. All putative study participants were excluded if they had (i) taken antiviral medication or drugs known to alter immune function in the preceding 3 months; (ii) had any vaccinations in the preceding 3 months; (iii) had a history of acute and chronic infectious diseases, such as hepatitis B and C, tuberculosis, HIV (but not herpes virus or other retrovirus infection); (iv) another chronic disease such as cancer, coronary heart disease, or uncontrolled diabetes; (v) a severe mood disorder; (vi) been pregnant or breastfeeding in the preceding 12 months; or (vii) were morbidly obese ($BMI \geq 40$). Further information on the UKMEB can be found elsewhere [30,31]. The severity of ME/CFS patients was divided into mild/moderate and severely affected (home or bed bound). MS patients had a previous diagnosis made by neurologists from the UK National Health System. No assessment of disease severity was made for these patients.

2.2. Symptomology assessment

At the recruitment, all the study participants were asked to report their disease duration and disease course. They also answered a symptom's assessment questionnaire (supplied as Supplementary Material, question #13) on the presence or absence of 58 different symptoms in the previous 7 days (Supplementary Table S1). The symptoms were related to the following domains: autonomic ($n = 10$); immunological ($n = 9$); neuroendocrine ($n = 6$); neurocognitive ($n = 19$); pain ($n = 6$); post-exertional malaise ($n = 6$); and sleep function ($n = 2$). The frequency of each symptom per study group can be found in Supplementary Table S2.

2.3. Herpesvirus IgG antibodies

As mentioned in the Introduction, we re-analyzed the same herpesvirus IgG antibody concentration dataset, as described in the original and follow-up study [41,42]. Briefly, antibody quantification of each participant was performed at recruitment using different commercial ELISAs. Data referred to plasma concentrations (in arbitrary units per milliliter, U/ml) of IgG antibodies against human cytomegalovirus (CMV), EBV (EBNA1 and EBV-derived Viral Capsid Antigen, EBV-VCA), HHV6, herpes simplex virus-1 (HSV1), herpes simplex virus-2 (HSV2), and varicella-zoster virus (VZV). In the manufacturer's protocols, the suggested cutoff values for seropositivity were 12 U/ml for HSV1, HSV 2, VZV, CMV and EBV, and 12.5 U/ml for HHV6. The maximum of plasma concentrations per antibody was the following: CMV – 304 U/ml; EBV-EBNA1 - 200 U/ml; EBV-VCA - 468 U/ml; HHV6 - 419 U/ml; HSV1 - 257 U/ml; HSV2 - 367 U/ml; VZV – 301 U/ml. Given that these maximum values translated into a narrow fold change in log10 scale beyond the seropositivity cut-off values, we chose to present some of the results using log2 scale.

2.4. Statistical analysis

Basic quantitative characteristics (e.g., age and disease duration) were summarized by means, median, min, max, and standard deviation. Absolute frequencies and percentages were used for summarizing basic categorical variables (e.g., gender and disease course). To compare different study groups in terms of these data, we used the non-parametric Kruskal-Wallis test and the Pearson's χ^2 test for quantitative and categorical variables, respectively. We used a significance level of 5% in these tests.

Symptomology data from the UKMEB were here analyzed for the first time. In this analysis, we removed 9 symptoms, because the frequency of missing data was higher than 25% in each disease group or overall. For each symptom, we calculated an age-adjusted odds ratio (OR) and the respective confidence interval using a logistic regression model where age was included as a confounding factor and MS was used as the reference group. The level of each confidence interval was adjusted by the Bonferroni correction to ensure a global confidence level of at least 95%. This correction was used due to its tendency to be more conservative towards the null hypothesis (i.e., absence of association/homogeneity between groups).

With respect to serological data, we first estimated seroprevalences in each study group by the proportion of individuals above the cutoff points recommended by the respective ELISA protocol. We calculated the respective 95% confidence interval using the Pearson-Clopper exact method. We performed the Pearson's χ^2 test to assess the homogeneity of seroprevalences to a given herpesvirus. For the respective quantitative data, we reported the median and the interquartile range (IQR) in each study group and herpesvirus antibody. To compare the median antibody concentrations across all study groups, we performed the non-parametric Kruskal-Wallis tests using a significance level of 5%.

Similarly to our previous study [42], we computed the Receiver Operating Characteristic (ROC) curve for each serological data, using all possible cut-off values to discriminate cases (ME/CFS subgroups) from controls (MS patients). We then estimated the respective areas under these curves (AUC) and computed their 95% confidence interval. For this estimation, we checked whether antibody concentrations of cases were either higher or lower than controls and chose the direction that provided the maximum AUC. In this analysis, there was evidence for a random classification if the confidence intervals for AUC included the value 0.5. We finally computed the optimal cut-point for each of these ROC curves by maximizing the significance of association provided by the Pearson's χ^2 test. We finally adjusted the p-values associated with a given serological variable for multiple testing using the Benjamini-Hochberg procedure [45]. This adjustment ensured a global false discovery rate (FDR) of 5%.

In contrast with our previous study, we now used an SL algorithm [46] to combine the serological data of multiple herpesviruses and determined the power of these data in discriminating patients of each ME/CFS subgroup from MS patients. In this algorithm, we estimated 4 classifiers for binary outcomes: Elastic-Net logistic regression, linear discriminant analysis, quadratic discriminant analysis, and random forests. For each pairwise comparison, these classifiers were trained using 10-fold cross-validation. The estimated probabilities from each individual model were then pooled together by a weighted average estimated by the SL algorithm. We finally performed a similar ROC-based analysis of these results, as described above.

The final step of the study contemplated the association analysis between symptomology and herpesvirus IgG antibodies in each study group. With this purpose, we used the non-parametric Mann-Whitney test to compare the concentration of a given herpesvirus antibody in absence and presence of the symptom under analysis in each study group. When analyzing data from a given study group, we adjusted the raw p-values using the Benjamini-Hochberg procedure in order to ensure a global FDR of 5%.

All the analyses were conducted in the R software version 4.2.2 [47] using the following packages: **Table 1** to summarize the baseline characteristics [48], **pROC** to perform the ROC analyses [49], **OptimalCutpoints** to estimate the optimal cutpoints in the ROC curves [50], and **SuperLearner** to perform the analysis based on the SL algorithm [51].

2.5. Ethical approval

Ethical approval was granted by the London School of Hygiene & Tropical Medicine Ethics Committee (Ref. 6123) and the National Research Ethics Service (NRES) London-Bloomsbury Research Ethics Committee (REC ref. October 11, 1760, IRAS ID: 77765). All participants provided written informed consent for data collection (questionnaire, clinical measurement and laboratory tests), and for allowing their samples to be available to any research receiving ethical approval (33).

3. Results

3.1. Basic characteristics of the study participants

A summary of the basic characteristics can be found in **Table 1**. The four ME/CFS subgroups did not differ substantially in terms of the age distribution, with average values of 44.5, 41.0, 43.4, and 41.1 years old for S0, S1, S2, and S3, respectively. Patients with MS were significantly older (average of 48.8 years old) than the ME/CFS patients (Kruskal-Wallis test, $p = 0.007$). Hence, we used age-adjusted ORs for the symptomology data. The percentages of female patients ranged from 69.6% (S3) to 82.6% (MS) across the five study groups. This range was not statistically significant (Pearson's χ^2 test, $p = 0.578$).

Most of ME/CFS patients had a diagnosis complying with both 1994 CDC/2003 CCC. The differences in proportion of different diagnostic combinations were not statistically significant among the ME/CFS (Pearson's χ^2 test, $p = 0.144$). The percentage of severely-affected ME/CFS patients was 9.5% in both S0 and S1. This percentage was significantly higher in the remaining subgroups (30.4% in S2 and 26.1% in S3; Pearson's χ^2 test, $p = 0.007$). The disease duration was approximately the same across the ME/CFS

Table 1

Basic characteristics of the study participants where P-values refer to the Kruskal-Wallis test and the Pearson's χ^2 test for quantitative and categorical variables, respectively.

	ME/CFS_S0 (n = 42)	ME/CFS_S1 (n = 42)	ME/CFS_S2 (n = 92)	ME/CFS_S3 (n = 46)	MS (n = 46)	P-value
Age						0.007
Mean (SD)	44.5 (10.7)	41.0 (13.0)	43.4 (10.9)	41.1 (10.5)	48.8 (7.10)	
Median [Min, Max]	44.0 [23.0, 60.0]	42.5 [18.0, 60.0]	44.0 [18.0, 59.0]	42.0 [19.0, 57.0]	50.5 [31.0, 58.0]	
Gender						0.578
Female	33 (78.6%)	32 (76.2%)	74 (80.4%)	32 (69.6%)	38 (82.6%)	
Male	9 (21.4%)	10 (23.8%)	18 (19.6%)	14 (30.4%)	8 (17.4%)	
ME/CFS Diagnosis						0.143
1994 CDC Only	9 (21.4%)	9 (21.4%)	6 (6.5%)	6 (13.0%)	NA	
2003 CCC Only	0 (0%)	1 (2.4%)	1 (1.1%)	1 (2.2%)	NA	
1994 CDC/2003 CCC	33 (78.6%)	32 (76.2%)	85 (92.4%)	39 (84.8%)	NA	
Disease course						0.317
Constantly getting worse	4 (9.5%)	4 (9.5%)	16 (17.4%)	8 (17.4%)	10 (21.7%)	
Constantly improving	0 (0%)	1 (2.4%)	0 (0%)	1 (2.2%)	0 (0%)	
No change	1 (2.4%)	1 (2.4%)	7 (7.6%)	2 (4.3%)	2 (4.3%)	
Relapsing and remitting	3 (7.1%)	4 (9.5%)	6 (6.5%)	8 (17.4%)	10 (21.7%)	
Fluctuating	16 (38.1%)	15 (35.7%)	30 (32.6%)	11 (23.9%)	14 (30.4%)	
Missing Data	18 (42.9%)	17 (40.5%)	33 (35.9%)	16 (34.8%)	10 (21.7%)	
Disease Severity						0.007
Mild/Moderate	38 (90.5%)	38 (90.5%)	64 (69.6%)	34 (73.9%)	NA	
Severe	4 (9.5%)	4 (9.5%)	28 (30.4%)	12 (26.1%)	NA	
Disease Onset (Infection)						NA
Don't know	42 (100%)	0 (0%)	0 (0%)	0 (0%)	14 (30.4%)	
No, did not have an infection at onset	0 (0%)	42 (100%)	0 (0%)	0 (0%)	22 (47.8%)	
Yes, but not confirmed by a lab test	0 (0%)	0 (0%)	92 (100%)	0 (0%)	1 (2.2%)	
Yes, confirmed by a lab test	0 (0%)	0 (0%)	0 (0%)	46 (100%)	0 (0%)	
Missing Data	0 (0%)	0 (0%)	0 (0%)	0 (0%)	9 (19.6%)	
Disease Duration						0.754
Mean (SD)	12.7 (8.39)	11.8 (8.60)	12.1 (8.52)	13.2 (8.22)	NA (NA)	
Median [Min, Max]	11.4 [1.40, 37.0]	10.3 [1.60, 38.0]	11.2 [0.200, 39.9]	12.0 [1.10, 29.3]	NA [NA, NA]	
Missing	1 (2.4%)	2 (4.8%)	2 (2.2%)	0 (0%)	46 (100%)	

subgroups (mean values between 11.8 and 13.2 years; Kruskal-Wallis test, $p = 0.754$). No information was available on disease duration for the MS patients.

Finally, in terms of disease course, there was a large proportion of missing data across the 5 study groups. Notwithstanding this data limitation, the most prevalent disease course was “fluctuating symptoms” in both ME/CFS and MS cohorts. However, differences in proportion of distinct disease course profiles were not statistically significant (Pearson’s χ^2 test, $p = 0.317$).

3.2. Analysis of symptomology

An initial analysis of symptoms’ data revealed that the age-adjusted ORs for the presence of each symptom when comparing the overall ME/CFS group to the MS one tended to be significantly higher than 1 (Figs. 1 and 2). This observation suggested that ME/CFS

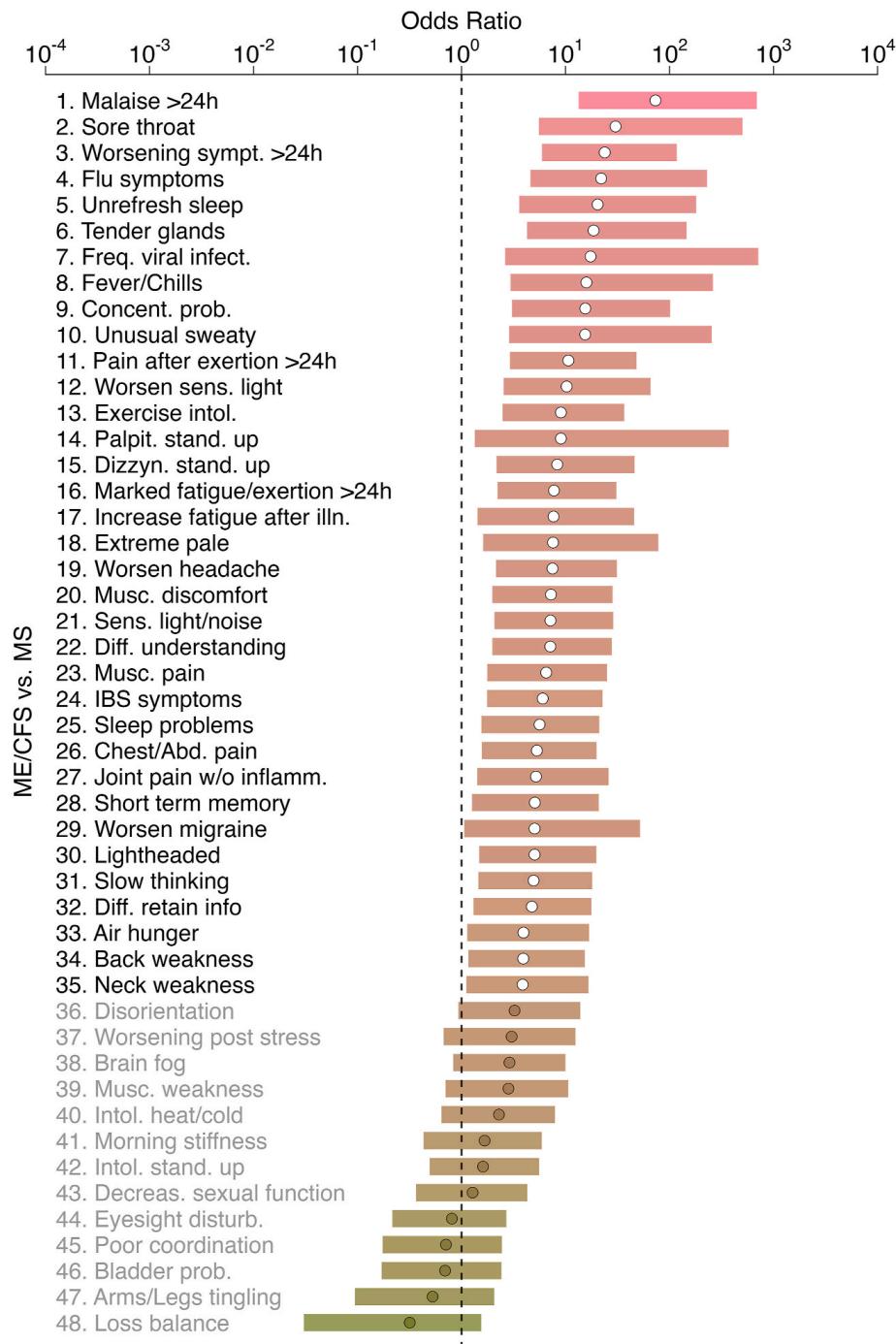


Fig. 1. Age-adjusted odds ratios (dots) ordered by magnitude and their Bonferroni-adjusted 95% confidence intervals (horizontal bars) for the presence of each of 48 symptoms when comparing the whole ME/CFS group to the MS group (reference). White-filled dots refer to symptoms where there was evidence of a higher frequency in the ME/CFS group than in the MS group. Grey-filled dots refer to symptoms where there was evidence for the same frequency in both cohorts.

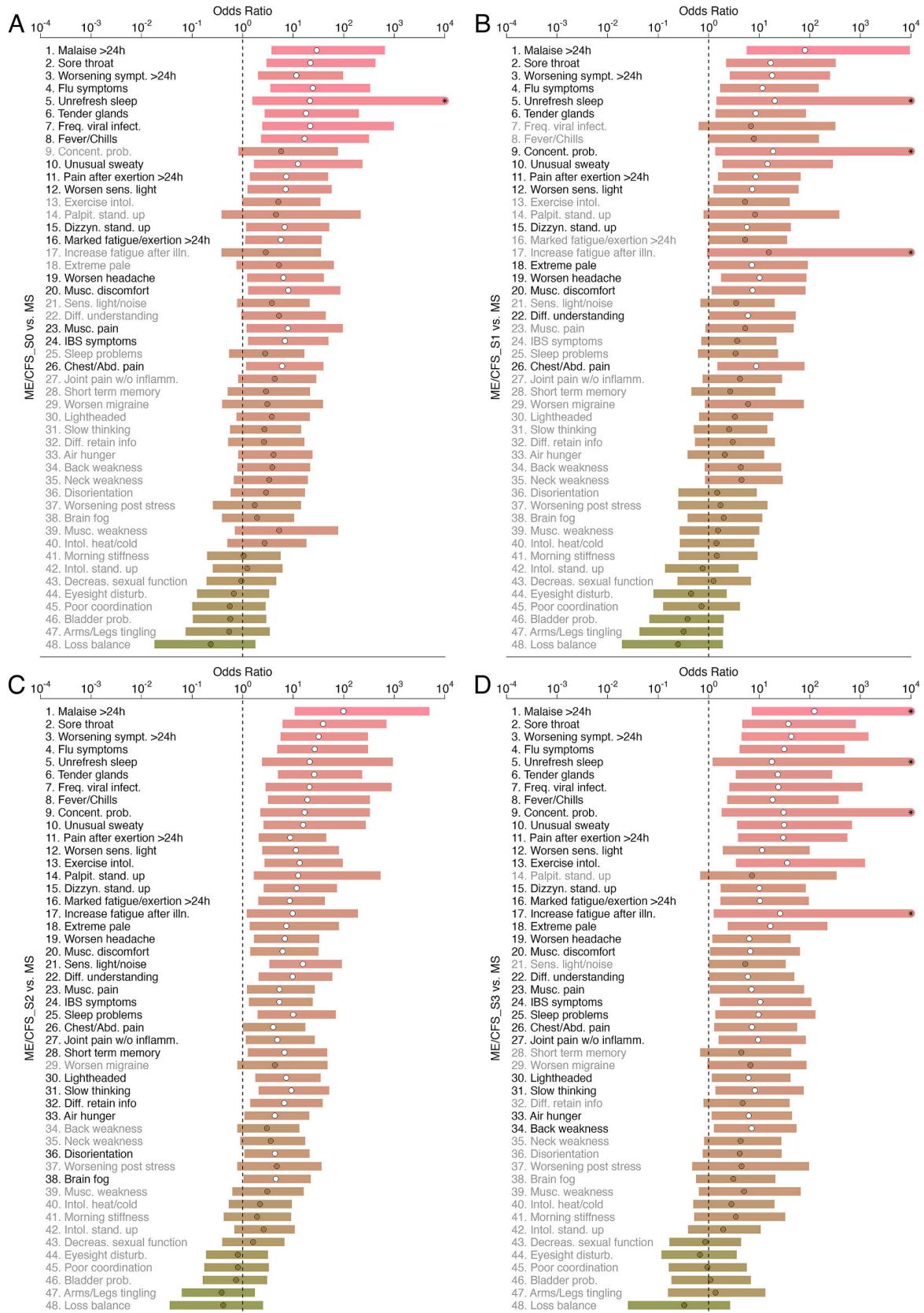


Fig. 2. Age-adjusted odds ratios (dots) and their Bonferroni-adjusted 95% confidence intervals (horizontal bars) for the presence of each of 48 symptoms when comparing ME/CFS_S0 (A), ME/CFS_S1 (B), ME/CFS_S2 (C), and ME/CFS_S3 (D) to the MS group. The “**” symbol in some of the bars denotes an upper limit beyond the maximum value used for the xx axis. See Fig. 1 legend for further information.

cohort as a whole or divided in its 4 subgroups had symptoms with a higher frequency than patients with MS. In all pairwise comparisons, the highest age-adjusted ORs were obtained for PEM lasting more than 24 h. This finding was expected given that PEM is a key symptom for the ME/CFS diagnosis according to 2003 CCC. Thirteen symptoms had similar frequencies in ME/CFS and MS cohorts: disorientation, worsening post-stress, brain fog, muscle weakness, intolerance to heat/cold, muscle stiffness in the morning, intolerance to standing up, decreased in sexual function, temporary eyesight disturbance, poor coordination, bladder problems, tingling or numbness in arms and/or legs, and loss of balance (Fig. 1).

The number of significant age-adjusted ORs reduced in the subgroup analysis, probably due to a reduction in sample size per ME/CFS subgroup (Fig. 2A–D). On the one hand, ME/CFS_S0 and ME/CFS_S1 had 18 and 16 symptoms whose frequency was significantly increased in relation to the MS group, respectively. Thirteen symptoms were shared between these two ME/CFS groups. In contrast, this number of significant symptoms was 34 and 29 for ME/CFS_S2 and ME/CFS_S3, respectively. Hence, both ME/CFS subgroups not related to a putative infectious disease trigger were the most similar ones to the MS group. Given that these subgroups had a lower frequency of severely-affected patients than the remaining groups, the MS group seems to be composed of patients with a mild/moderate symptomatology.

Finally, there was no evidence for any symptom whose presence was significantly more prevalent in the MS group than in the overall ME/CFS group or its subgroups (Figs. 1 and 2).

3.3. Univariate analysis of herpesvirus IgG antibody data

As stated in the Introduction, EBV and HHV6 infections are thought to be at the origin and a risk factor for both MS [22,25] and ME/CFS [23]. Given this evidence, one could hypothesize that IgG antibody concentrations related to these viruses could be elevated in both diseases. In this regard, we previously found evidence for similar seroprevalence to each herpesvirus, including EBV and HHV6, when comparing mild/moderate and severely affected ME/CFS patients, MS patients, and healthy controls [41]. Similar evidence was found for the respective antibody concentration levels except a higher EBNA1-VCA IgG antibody concentration in the MS cohort.

Here, we repeated a similar analysis using data of seroprevalence and IgG antibody concentration levels for 5 study groups to test whether a disease-trigger stratification of the ME/CFS group could provide additional information (Table 2 and Supplementary Fig. S1). Interestingly, the most significant differences remained to be related to both EBV-related IgG antibodies. These differences were in part driven by a higher seroprevalence or a higher median IgG antibody concentration in the MS group, even when compared to the ME/CFS_S3 group where confirmed EBV infections were reported to be the cause of the disease [42]. Therefore, our re-analysis using a different ME/CFS patient's stratification was once again against the prior expectation that IgG antibody concentrations related to different herpesviruses are equivalent in ME/CFS and MS.

In agreement with our previous analysis when we compared ME/CFS subgroups to MS [42], differences in CMV seroprevalence groups were in the vicinity of statistical significance. This result was due to lower seroprevalences to this virus in ME/CFS_S1 and ME/CFS_S3. For the remaining herpesvirus antibodies, there were no significant differences among the study groups in terms of seroprevalence or IgG antibody concentrations.

We extended this analysis with the objective of discriminating patients with MS from healthy controls combining these data with a ROC curve approach (Supplementary Table S3). The most significant AUC result was found for the antibodies against EBV-VCA. In this case, the respective antibody concentrations were significantly increased in patients with MS. The estimate of AUC was 0.641 (95% CI=(0.538; 0.744)) with an optimal sensitivity and specificity of approximately 0.650. Another significant result was found for antibodies against HSV1 (0.626, 95%CI=(0.531; 0.721)). Interestingly, these antibodies were in higher concentrations in healthy controls than in MS patients. The corresponding sensitivity and specificity were estimated at 0.950 and 0.327, respectively. This result indicated that MS patients are well characterized by a low HSV1 antibody concentration. The data from the remaining antibodies did not provide any evidence for a significant result.

Table 2

Seroprevalence (SeroP, in percentage), median concentration and the respective IQR per study group and herpesvirus IgG antibody, where P-value refers to the Pearson's χ^2 test for seroprevalence-based analysis and the Kruskal-Wallis test for the corresponding quantitative data.

Herpesvirus (Antigen)	Analysis	ME/CFS_S0	ME/CFS_S1	ME/CFS_S2	ME/CFS_S3	MS	P- value
CMV	SeroP (95% CI)	42.9 (27.7; 59.0)	25.6 (13.5; 41.2)	37.0 (27.1; 47.7)	18.8 (8.9; 32.6)	45.0 (29.3; 61.5)	0.035
	Median (IQR)	5.4 (78.5)	4.6 (13.3)	6 (59.8)	5.4 (1.5)	5.8 (64.4)	0.407
EBV (EBNA1)	SeroP (95% CI)	71.4 (55.4; 84.3)	58.1 (42.1; 73.0)	75.0 (64.9; 83.4)	75.0 (60.4; 86.4)	95.0 (83.1; 99.4)	0.004
	Median (IQR)	39.3 (17.8)	24.6 (17.1)	37.2 (23.5)	32.9 (26.4)	46.9 (17.8)	0.006
EBV (VCA)	SeroP (95% CI)	88.1 (74.4; 96.0)	74.4 (58.8; 86.5)	84.8 (75.8; 91.4)	95.8 (85.7; 99.5)	92.5 (79.6; 98.4)	0.030
	Median (IQR)	99.4 (64.4)	88.0 (64.8)	108.9 (49.9)	129.5 (48.6)	155.4 (33.8)	0.003
HHV6	SeroP (95% CI)	85.7 (71.5; 94.6)	90.7 (77.9; 97.4)	92.4 (84.9; 96.9)	100.0 (92.6; 100.0)	90.0 (76.3; 97.2)	0.146
	Median (IQR)	38.3 (21.1)	42.5 (24.3)	47.5 (25.2)	43.0 (29.7)	43.4 (50.8)	0.418
HSV1	SeroP (95% CI)	50.0 (34.2; 65.8)	34.9 (21.0; 50.9)	43.5 (33.2; 54.2)	54.2 (39.2; 68.6)	57.5 (40.9; 73.0)	0.207
	Median (IQR)	12.1 (89.5)	6.6 (56.5)	7.6 (108.8)	16.5 (119.7)	22.6 (109.6)	0.273
HSV2	SeroP (95% CI)	50.0 (34.2; 65.8)	30.2 (17.2; 46.6)	35.9 (26.1; 46.5)	45.8 (31.4; 60.8)	40.0 (24.9; 56.7)	0.309
	Median (IQR)	12.2 (61.4)	5.0 (33.5)	4.1 (45.0)	6.2 (41.8)	8.1 (41.8)	0.650
VZV	SeroP (95% CI)	95.2 (83.8; 99.4)	93.0 (80.9; 98.5)	100.0 (96.1; 100.0)	97.9 (88.9; 99.9)	97.5 (86.8; 99.9)	0.169
	Median (IQR)	129.5 (46.2)	119.7 (20.8)	135.7 (48.4)	142.2 (35.8)	156.2 (36.7)	0.069

We then performed a classification exercise of each subgroup of ME/CFS patients using patients with MS as the control group (Table 3). In general, the MS cohort tended to have higher antibody concentrations than the ME/CFS patients, irrespective of the herpesvirus under analysis apart from a single exception, ME/CFS_S0 Vs MS for CMV. However, the classification power of each antibody was not significant as illustrated by the 95% CI for the AUC that contained the value 0.50. The most optimistic scenarios were observed for the pairwise comparisons involving the EBV-related antigens. In particular, the best classification was obtained for ME/CFS_S1 vs MS for EBNA1 (AUC = 0.741, 95% CI=(0.634; 0.848)). In this case, the sensitivity and specificity were estimated at 0.465 and 0.950, respectively, using a cutoff of 20.0 U/ml. These estimates suggested that high antibody concentrations to this EBV antigen were able to discriminate MS patients almost perfectly. However, the same could not be said for detecting cases from this ME/CFS subgroup.

ME/CFS_S1 seemed the most different one from the MS group in terms of herpesvirus IgG antibodies (Table 3). This interpretation was supported by a significant AUC for antibodies against EBNA1 (as reported above), EBV-VCA (AUC = 0.695, 95%CI=(0.580;0.811), Se = 0.442, Sp = 0.900), HSV1 (AUC = 0.635, 95%CI=(0.513;0.756), Se = 0.256, Sp = 0.975), and VZV (AUC = 0.667, 95%CI=(0.548;0.786), Se = 0.744, Sp = 0.630). This result suggested that combining data from multiple antibodies could help discriminating these two groups of patients.

3.4. Combined analysis of IgG antibody data using an SL algorithm

In this step of the analysis, we integrated all IgG antibody data in several classifiers to distinguish patients of each ME/CFS subgroup from patients with MS. For each pairwise comparison, these classifiers were then assembled into a final classifier using an SL algorithm (Fig. 3).

This combined data analysis provided evidence for an AUC significantly different from 0.5 (random guess) for ME/CFS_S0 (AUC = 0.658, 95%CI=(0.536;0.779)), ME/CFS_S1 (AUC = 0.731, 95%CI=(0.622;0.841)), and ME/CFS_S3 (AUC = 0.707, 95%CI=(0.599;0.816)) subgroups using the MS group as a control. The highest sensitivity was obtained for the ME/CFS_S2 group (0.772), but at the cost of a poor specificity related to the MS group (0.075) (Fig. 3). The most balanced sensitivity and specificity estimates were observed for ME/CFS_S0 vs MS (0.619 and 0.725, respectively). In the remaining pairwise comparisons (ME/CFS_S1 and ME/CFS_S3), there was evidence of a high specificity (higher than 0.90) but a modest sensitivity (up to 0.512).

As expected from the analysis based on a single antibody, the best AUC was obtained for the comparison between ME/CFS_S1 and MS. When compared to the MS group, the probability of having a patient from this ME/CFS subgroup decreased with the antibody

Table 3

Area under the Receiver Operating Characteristic curve (AUC) and its 95% confidence interval (95% CI), optimal cutoff and associated sensitivity (Se) and specificity (Sp) to discriminate ME/CFS_S0, ME/CFS_S1, ME/CFS_S2, ME/CFS_S3 subgroups (cases) from patients with multiple sclerosis (MS) used as controls. In the Direction column, the symbols “>” and “<” represent higher value in MS cases than in healthy controls and vice-versa, respectively. In the Cutoff column, the p-value within parenthesis is associated with the Pearson's χ^2 test for 2×2 tables after adjusting for an FDR of 5%. In the AUC column, the symbol “**” denote the cases where there was evidence of an AUC different from 0.50 (random guess).

Herpesvirus (Antigen)	Comparison (Vs MS)	Direction	AUC (95% CI)	Cutoff in U/ml (P-value)	Se	Sp
CMV	ME/CFS_S0	controls < cases	0.551 (0.421; 0.680)	2.845 (0.0214)	0.952	0.325
	ME/CFS_S1	controls > cases	0.547 (0.418; 0.677)	29.21 (0.0701)	0.791	0.425
	ME/CFS_S2	controls < cases	0.507 (0.389; 0.624)	194.91 (0.0256)	0.000	0.900
	ME/CFS_S3	controls > cases	0.545 (0.413; 0.678)	58.755 (0.0028)	0.938	0.375
EBV (EBNA1)	ME/CFS_S0	controls > cases	0.590 (0.463; 0.716)	15.415 (0.0257)	0.286	0.950
	ME/CFS_S1	controls > cases	0.741 (0.634; 0.848) *	20.035 (0.0004)	0.465	0.950
	ME/CFS_S2	controls > cases	0.602 (0.505; 0.700) *	20.505 (0.0035)	0.359	0.950
	ME/CFS_S3	controls > cases	0.621 (0.503; 0.739) *	19.985 (0.0028)	0.396	0.950
EBV (VCA)	ME/CFS_S0	controls > cases	0.668 (0.550; 0.787) *	132.96 (0.0257)	0.643	0.675
	ME/CFS_S1	controls > cases	0.695 (0.580; 0.811) *	71.435 (0.0029)	0.442	0.900
	ME/CFS_S2	controls > cases	0.645 (0.543; 0.746) *	150.345 (0.0144)	0.707	0.575
	ME/CFS_S3	controls > cases	0.566 (0.443; 0.690)	139.315 (0.0687)	0.583	0.650
HHV6	ME/CFS_S0	controls > cases	0.590 (0.465; 0.714)	78.105 (0.0374)	0.881	0.350
	ME/CFS_S1	controls > cases	0.558 (0.429; 0.687)	103.94 (0.0057)	1.000	0.225
	ME/CFS_S2	controls > cases	0.518 (0.403; 0.632)	130.625 (0.0256)	0.978	0.150
	ME/CFS_S3	controls > cases	0.520 (0.391; 0.649)	91.82 (0.0271)	0.938	0.275
HSV1	ME/CFS_S0	controls > cases	0.579 (0.454; 0.704)	2.14 (0.0371)	0.167	1.000
	ME/CFS_S1	controls > cases	0.635 (0.513; 0.756) *	2.84 (0.0104)	0.256	0.975
	ME/CFS_S2	controls > cases	0.596 (0.497; 0.696)	2.755 (0.0303)	0.196	0.975
	ME/CFS_S3	controls > cases	0.550 (0.429; 0.671)	2.8 (0.0182)	0.250	0.975
HSV2	ME/CFS_S0	controls < cases	0.557 (0.432; 0.683)	231.395 (0.0737)	0.119	1.000
	ME/CFS_S1	controls > cases	0.536 (0.409; 0.663)	6.81 (0.1600)	0.628	0.55
	ME/CFS_S2	controls > cases	0.520 (0.410; 0.629)	7.03 (0.1343)	0.609	0.550
	ME/CFS_S3	controls > cases	0.527 (0.402; 0.651)	2.895 (0.1341)	0.771	0.400
VZV	ME/CFS_S0	controls > cases	0.596 (0.472; 0.721)	141.005 (0.0737)	0.571	0.650
	ME/CFS_S1	controls > cases	0.667 (0.548; 0.786) *	139.495 (0.0025)	0.744	0.650
	ME/CFS_S2	controls > cases	0.532 (0.418; 0.646)	148.24 (0.5054)	0.554	0.550
	ME/CFS_S3	controls > cases	0.546 (0.421; 0.672)	180.995 (0.0981)	0.792	0.400

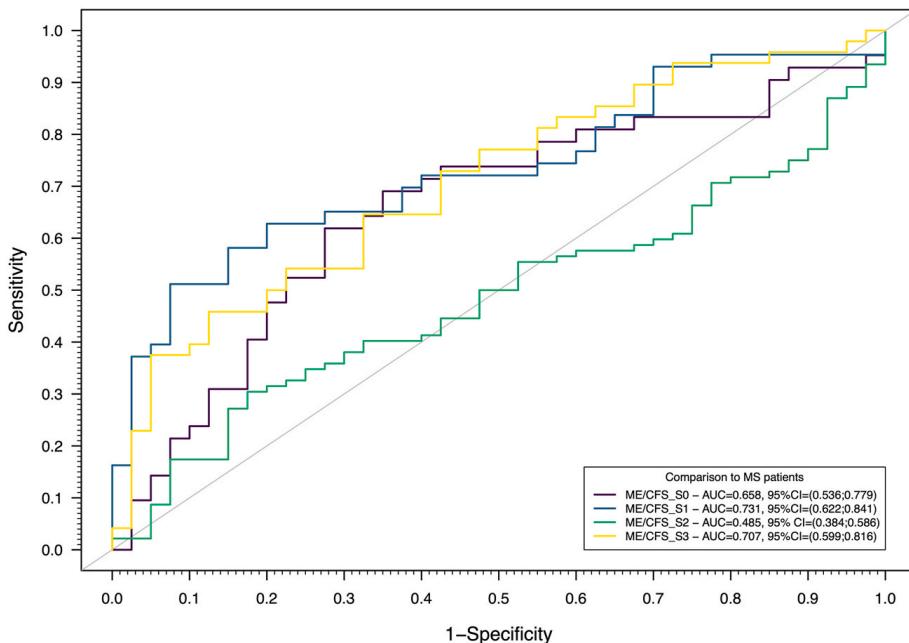


Fig. 3. ROC curves for the predictions based on an SL algorithm trained with 4 different classifiers (Elastic-Net Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Random Forest) and 10-fold cross-validation using antibody data and patients with MS as the controls. Optimal sensitivities and specificities were estimated at 0.619 and 0.725 for ME/CFS_S0, 0.512 and 0.925 for ME/CFS_S1, 0.772 and 0.075 for ME/CFS_S2, 0.375 and 0.950 for ME/CFS_S3, when compared to the MS group.

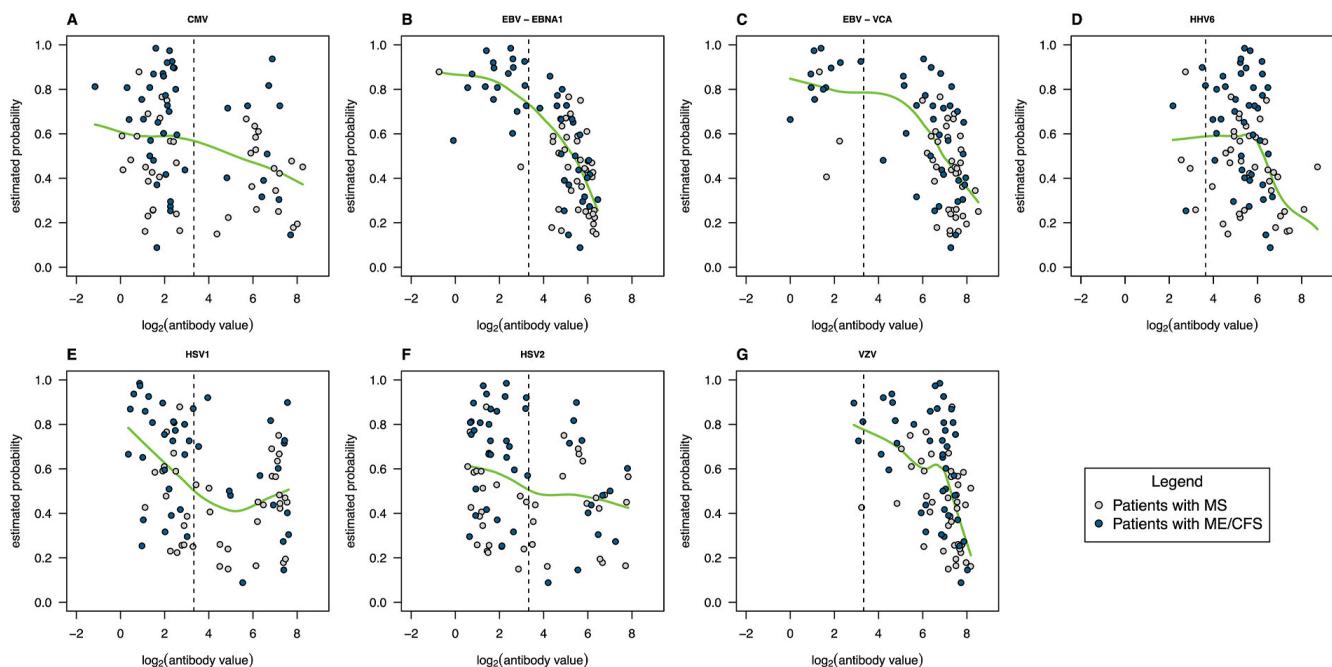


Fig. 4. Smooth-line approximations (green lines) of the relationship between $\log_2(\text{antibody concentrations})$ and SL-estimated probability of ME/CFS_S1 patient when compared to patients with MS (A – CMV, B – EBV-EBNA1, C – EBV-VCA, D – HHV-6, E – HSV1, F – HSV2, G – VZV). In the plots, each dot represents a patient and the vertical dashed line represents the cut-off value for seropositivity according to the respective lab protocol.

concentrations related to different herpesviruses including EBV (Fig. 4A–G); see the respective probability profiles for the remaining pairwise comparison in [Supplementary Figs. S2, S3, and S4](#) (ME/CFS_S0 vs MS, ME/CFS_S2 vs MS, ME/CFS_S3 vs MS, respectively). Note that, given the poor discrimination between ME/CFS_S2 and MS, the probability profile of having a patient from this ME/CFS subgroup was almost constant across all the herpesviruses antibodies ([Supplementary Fig. S3](#)).

3.5. Association analysis between symptomology and herpesvirus IgG antibodies

Lastly, we correlated the data of presence/absence of each symptom with data of each herpesvirus IgG antibody (Fig. 5A–F). In the case of the MS cohort, HHV6 antibody concentration was significantly and negatively associated with difficulties in understanding, and worsening of symptoms after stress. Similar significant negative association was found between CMV IgG antibody concentrations and eyesight disturbances. The only significant positive association was found for EBV-EBNA1 antibody concentrations and bladder problems.

For the overall ME/CFS group, a positive and a negative association reached statistical significance between brain fog and HSV1 IgG antibody concentrations and between chest/abdominal pain and HHV6 antibody concentrations, respectively. These significant associations could not be confirmed by the subsequent subgroup analysis. For the ME/CFS_S0 subgroup, two significant negative associations were found (neck weakness/HSV1 antibody concentrations and fever chills/EBV-VCA antibody concentrations). In the ME/CFS_S1 subgroup, there was a significant negative association between difficulties in retaining/recalling information and VZV IgG antibody concentrations. In the ME/CFS_S2 subgroup, negative and positive associations were found between EBV-VCA antibody concentration and worsening of symptoms after exertion lasting more than 24 h and between EBNA1 and short-term memory problems, respectively. Finally, feeling lightheaded was significantly and negatively associated with CMV IgG antibody concentrations in the ME/CFS_S3 subgroup. In opposition, bladder problems were significantly and positively associated with VZV antibody concentrations in the same ME/CFS subgroup.

4. Discussion

Our study showed that the overall ME/CFS cohort had more symptoms whose frequency was higher than the one observed for the MS patients. This finding was in agreement with a higher frequency of functional symptoms reported by an online self-report survey with the aim at finding the key differentiating symptoms of these two diseases [52]. Interestingly, sore throat, tender glands, and flu-like symptoms from the immunological domain were also at the top differentiators of both ME/CFS_S2 and ME/CFS_S3 ME/CFS (i.e.

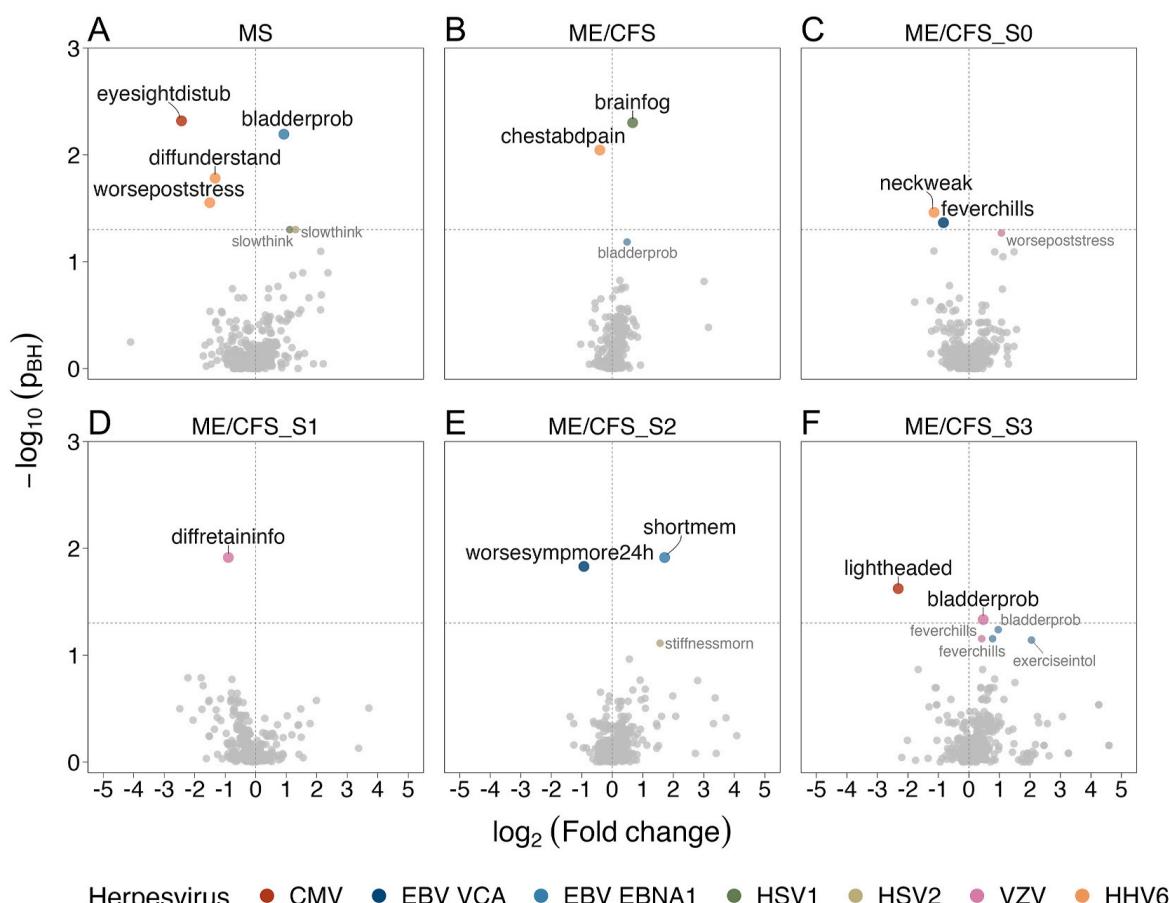


Fig. 5. Association analysis between symptomology and IgG antibody data for the MS group (A), the overall ME/CFS (B), the ME/CFS_S0 subgroup (C), ME/CFS_S1 subgroup (D), ME/CFS_S2 subgroup (E), and ME/CFS_S3 subgroup (F), where the xx axis refers to $\log_2(\text{mean fold-change})$ between individuals with and without a given symptom, respectively, and the yy axis refers to the logarithm in base 10 of the p-values derived from the Mann-Whitney test and adjusted for an FDR of 5% using the Benjamini-Hochberg procedure ($-\log_{10}(p_{\text{BH}})$). See Supplementary Table S1 for linking the symptom codes presented in each plot to the respective symptom descriptions.

e., subgroups related to a putative infectious trigger) and MS. Again, this finding is in line with another study where tender lymph nodes and flu-like symptoms could correctly distinguish MS from ME/CFS 81% of the time [53]. Therefore, our study provided additional evidence that, besides the presence of hallmark persistent fatigue and PEM, symptoms from the immunological domain are crucial to diagnose ME/CFS and differentiate from MS. This is reflected in the 1994 CDC criterion and 2003 CCC, two recommended case definitions for ME/CFS diagnosis, especially in the research setting [3]. However, we cannot rule out that difference in symptoms' frequency might be simply due to mild MS cases who did not undergo any treatment together with the presence of severely affected patients in the different ME/CFS subgroups. In this regard, it is important to emphasize that severe MS cases (most likely undergoing any immune-therapy) were excluded from participation in the UKMEB.

We found several positive and negative associations between herpesvirus antibody concentrations and different symptoms in each study group, even after adjusting for multiple testing. This finding is remarkable given the dozens of symptoms evaluated, the modest sample size of each study group, and the use of a non-parametric test that usually reduces the statistical power to detect group differences. However, we cannot rule out that these associations might be affected by patient's subjective perception of each symptom. At the same time, casting doubts on this perception is indirectly supporting theories that ME/CFS is a psychosomatic rather than an organic or a physical condition. A more plausible impact on the robustness of these associations is that they could have resulted from random fluctuations in symptomatology occurring over the natural disease course. To confirm or refute this last explanation, one could conduct a longitudinal study with multiple timepoints and check whether the associations remain valid during follow-up.

The MS group had the highest number of significant antibody-symptom associations (or close to statistical significance). Therefore, infections by these herpesviruses seem to have a higher impact on this group than on the ME/CFS groups. The strongest association was observed for CMV antibody concentrations and the presence of eyesight disturbances. This is an interesting association given that photophobia and photosensitivity were commonly reported by MS patients [54] and CMV is also known to cause retinitis in immunocompromised individuals [55]. In this scenario, a reduction in antibodies against CMV might result in a low-grade ocular infection in MS patients. The second strongest association was found for EBNA1 antibody concentrations and bladder problems. This association is according to the view that bladder dysfunction, also a common symptom observed in MS, is linked to autoantibodies (for example, against muscarinic receptor-3) [56]. Our finding suggests that the origin of these autoantibodies could be a cross-reactive antibody response to an EBNA1 peptide mimicking a human protein, as demonstrated for MS [57–59]. Interestingly, a positive association was also found for bladder problems in the ME/CFS_S3 group but related to VZV antibody concentrations. Given that infections by VZV can also cause bladder dysfunction [60], we speculate that patients from this ME/CFS subgroup experienced a recent reactivation of the virus. Notwithstanding the recent observation of VZV reactivation after SARS-CoV-2 infections [61] or under stressful conditions [62], there is weak evidence for VZV reactivation in ME/CFS patients [26,63]. The remaining significant antibody-symptoms associations in the MS group were both negative and related to HHV6 IgG antibody concentrations. These associations are in opposition to the evidence that antibody concentrations to HHV6 are positively associated with the risk of developing MS [27]. A possible explanation for this contradicting finding is that a lower HHV6 antibody concentration triggers the reactivation of this virus, as observed in individuals after cord blood transplantation [64]. Unfortunately, the quantification of viral DNA was not performed in this study and, therefore, this explanation could not be tested with data.

As mentioned above, we found a lesser number of significant antibody-symptom associations in ME/CFS. Interestingly, the strongest association was found for increasing HSV1 antibody concentrations and brain fog. It is worth noting that brain fog is a colloquial term used by patients that seems to gather a constellation of symptoms mostly concerning deficits in cognition, such as reduced speed and efficiency of information processing, attention, concentration, and working memory [65–67]. If brain fog is taken in this sense, then the positive association with HSV1 antibody levels seems interesting given that this virus is known to be neurotropic during latency [68]. The presence and concentration levels of HSV1 IgG antibodies, as measured here, have been positively associated with cognitive deficits in both healthy individuals [69–71] and patients suffering from Alzheimer's disease [72], schizophrenia [73], and bipolar disorder [74]. The same association could be significantly found for ME/CFS_S1 and ME/CFS_S2 before adjusting for multiple testing. Hence, a reduction in the statistical power by dividing patients in different strata is a plausible explanation for not finding this association in the ME/CFS subgroup analysis. At this point, it is reasonable to raise the question of why similar association with brain fog was not detected for antibody concentrations related to HHV6, another neurotropic virus whose viral miRNA was found in post-mortem brain biopsies of ME/CFS patients [28]. A possible answer for this question is that, given that exposure to HHV6 is more prevalent in human populations than to HSV1, the respective antibody quantification has shorter dynamic range due to a high percentage of seropositive individuals (Table 2 and Supplementary Fig. S1). In theory, this shorter dynamic range implies a higher sample size to detect smaller differences between patients reporting or not brain fog.

In the ME/CFS subgroup analysis, the detected antibody-symptom associations were not consistent across the 4 ME/CFS subgroups. This result suggested a large heterogeneity of different ME/CFS groups in terms of symptoms and their relationship with the underlying herpesviruses IgG antibody concentrations. An analysis based on disease severity could provide a clearer insight into the relationship between these two sets of variables. This possibility is motivated by a longitudinal study from the UKMEB where disease scores and different symptoms were associated with herpesvirus reactivation [26]. This alternative analysis by disease severity was beyond the scope of this study, but it will be done in a near future.

Our study also showed that the IgG antibody data were statistically significant to discriminate three ME/CFS subgroups from MS. Such a discrimination capability could be attributed to increased IgG antibody concentrations to multiple herpesviruses in the MS group, as previously found for EBV antigens [13]. This capability, although reaching statistical significance, is likely to have a diminished impact in the clinic, because it neither have a high sensitivity to classify these ME/CFS subgroups nor a high specificity to classify MS patients. This observation suggests that one should look for alternative disease-specific biomarkers or to combine these basic antibody measurements with more specific herpesvirus-related antibodies, such as those against EBNA4_0529 and EBNA6_0070

peptides derived from EBV [75]. Notwithstanding this observation, it is worth mentioning that the reporting of a less promising finding helps the research community to focus on more promising biomarkers, thus, avoiding the waste of valuable research efforts and resources. It is also in line with the expectations of the UKMEB participants who are less concerned about breakthroughs in medical science and more interested in incremental steps and collaborative efforts which might one day lead to effective diagnosis, treatment, or cure [76].

We found that serological data including the EBV antigens were unable to discriminate MS patients from the ME/CFS subgroup of patients who self-reported an infection (not evaluated by a lab test) at their disease onset. This subgroup is the largest in size and the one with the highest frequency of patients who self-reported a flu-like infection (22%) or an unspecified viral infection (35%) at their disease onset [42]. Therefore, this subgroup seems very heterogeneous and non-specific. Given that many factors can contribute to MS pathogenesis [68], the lack of evidence for a serological discrimination between these two groups might be attributed to the presence of multiple factors shared between this heterogeneous subgroup of ME/CFS patients and ME patients. In this line of thought, one could make the case for MS-related drugs such as cyclophosphamide [77] and rituximab [78] to be deployed to treat ME/CFS patients, as made elsewhere [79,80]. However, this deployment is likely to be more successful if targeting not all but only patients from this ME/CFS subgroup. However, future studies should be conducted to ascertain further similarities between this subgroup of ME/CFS patients and patients with MS.

There are four major limitations of this study. As already highlighted, the main limitation is that our re-analysis focused on data portraying a single snapshot of these patients. Given the fluctuating clinical course of ME/CFS (Table 1), it is likely that the reported antibody-symptom associations might render irreproducible. It is also possible that non-significant associations would become statistically (and clinically) significant in latter time points. A longitudinal study including the measurements at multiple time points would resolve this limitation. In practice, such a study is expensive and affected by other problems, such as patients' compliance or the presence of missing data due to drop-outs. Patient's exhaustion (often triggered by data collection) is also another limiting factor in the context of ME/CFS research. As such, the common and pragmatic approach is to conduct association analyses between different biomarkers and symptoms (or, alternatively, symptom scores) using cross-sectional data, as illustrated not only for ME/CFS but also for post-COVID [12,32,81,82].

The second limitation is that IgG antibodies are often used as biomarkers of a past active infection. As such, these antibodies have limited power to judge when the last active infection occurred and whether these herpesviruses are currently active or simply latent. In this regard, our study would have benefited from a quantification of IgM antibodies in the same biological samples. However, additional serological testing was beyond the scope of this study due to lack of funding and inexistence of biological material for many of the study participants. In this scenario, our study should be viewed as an opportunistic re-analysis of already existing data with its own limitations.

The third limitation is that the MS cohort was conveniently enrolled in the UKMEB due to the criterion of excluding any patient following any immune therapy. In this perspective, the recruited MS patients can be seen as "pure patients" or simply "mild cases". On the one hand, the recruitment of patients without any treatment suggests a selective bias towards sampling patients with less severe symptoms, thus, explaining the higher number of symptoms that are more prevalent in the ME/CFS group than in the MS one. This putative bias was also suggested by a higher number of symptoms whose frequency was similar in MS and ME/CFS_S0/ME/CFS_S1 in which the frequency of severely affected patients was less than 10%. However, low adherence to modifying disease therapy by MS patients is considered a major concern in the management of the disease [83,84] and, therefore, the recruited MS patients are in part representative of this clinical population who do not adhere to any treatment.

The fourth limitation is that disease course profiles do not comply with the standard classification used in MS research; this standard classification is reviewed elsewhere [85]. These profiles, if following such a standard, would have helped to understand whether the proportions of sampled MS patients at different disease courses would match those found across epidemiological studies of MS, thus, increasing the interpretation on the representativeness of the MS cohort. Notwithstanding this limitation, 22% of the MS patients reported constantly getting worse. This frequency anecdotally coincides with the prevalence of a primary progressive disease course (15%–22%) in MS patients from different parts of the UK [86]. In addition, all study groups were well matched for the alternative disease course categorization used here and, therefore, there is evidence that our findings are less likely to be driven by differences in putative disease courses of the patients.

In conclusion, some candidate antibody-symptom associations were identified for ME/CFS patients. However, the clinical impact of these association remains to be determined given the fluctuating nature of symptoms in ME/CFS patients and the nature of IgG antibodies in simply detecting exposure. To understand the true impact of our findings in the clinic, we propose a future longitudinal study with at least three time points. This study is likely to help discerning the robustness of the reported associations during disease course and whether they could be targeted by any existing drug treatment. This study, although theoretical appealing, might be difficult to design with an appropriate statistical power and to execute in the clinic given the lack of funding affecting ME/CFS research, poor societal recognition of the disease, and possible patient's exhaustion during follow-up.

Funding information

This research was funded by: FCT - Fundação para a Ciência e Tecnologia, Portugal, ref. grant: SFRH/BD/149758/2019 (J.M.), and UIDB/00006/2020 (T.D.D., J.M., H.M., NS); the Polish National Agency for Academic Exchange (NAWA), ref. grant: PPN/ULM/2020/1/00069/U/00001 (N.S.). The UKMEB was established with a joint grant from the charities ME Association (including continuing support), ME Research UK and Action for ME, as well as private donors. Research reported in this manuscript was supported by the National Institutes of Health (NIH) under award number 2R01AI103629.

Data statement

The data set used in this study is freely available from the UKMEB upon application.

Author contribution statement

Tiago Dias Domingues, João Malato, Anna Daria Grabowska, Jose Ameijeiras-Alonso, Luís Graça, Helena Mourão, Carmen Scheibenbogen, Francisco Westermeier: Analyzed and interpreted the data; Wrote the paper.

Ji-Sook Lee: Performed the experiments; Contributed reagents, materials tools or data; Wrote the paper.

Przemysław Biecek: Analyzed and interpreted the data; Contributed reagents, materials tools or data; Wrote the paper.

Jacqueline M Cliff and Luis Nacul: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials tools or data; Wrote the paper.

Eliana M Lacerda: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials tools or data; Wrote the paper.

Nuno Sepúlveda: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Data availability statement

The data set used in this study is freely available from the United Kingdom ME/CFS biobank upon application.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank all the participants who have generously contribute to the study by donating samples, time, and energy.

Abbreviations

1994 CDC	1994 Centers for Disease Control and Prevention criteria
2003 CCC	2003 Canadian Consensus criteria
CMV	Cytomegalovirus
EBNA-1	EBV-derived nuclear antigen-1
EBV	Epstein-Barr virus
EBV-VCA	EBV-derived Viral Capsid Antigen
FDR	False discovery rate
HSV1	Herpes Simplex virus-1
HSV2	Herpes Simplex virus-2
HHV6	Human Herpesvirus-6
ME/CFS	Myalgic Encephalomyelitis/Chronic Fatigue Syndrome
MS	Multiple Sclerosis
SL	SuperLearner
UKMEB	United Kingdom ME/CFS biobank
VZV	Varicella-Zoster virus

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2023.e18250>.

References

- [1] M.C. Rivera, C. Mastronardi, C.T. Silva-Aldana, M. Arcos-Burgos, B.A. Lidbury, Myalgic encephalomyelitis/chronic fatigue syndrome: a comprehensive review, *Diagnostics* 9 (2019) 91, <https://doi.org/10.3390/diagnostics9030091>.
- [2] C. Scheibenbogen, H. Freitag, J. Blanco, E. Capelli, E. Lacerda, J. Authier, M. Meeus, J. Castro Marrero, Z. Nora-Kruk, E. Oltra, E.B. Strand, E. Shikova, S. Sekulic, M. Murovska, The European ME/CFS Biomarker Landscape project: an initiative of the European network EUROMENE, *J. Transl. Med.* 15 (2017) 162, <https://doi.org/10.1186/s12967-017-1263-z>.
- [3] L. Nacul, F.J. Authier, C. Scheibenbogen, L. Lorusso, I.B. Helland, J.A. Martin, C.A. Sirbu, A.M. Mengshoel, O. Polo, U. Behrends, H. Nielsen, P. Grabowski, S. Sekulic, N. Sepulveda, F. Estévez-López, P. Zalewski, D.F.H. Pheby, J. Castro-Marrero, G.K. Sakkas, E. Capelli, I. Brundsdlund, J. Cullinan, A. Krumina,

- J. Bergquist, M. Murovska, R.C.W. Vermuelen, E.M. Lacerda, European network on myalgic encephalomyelitis/chronic fatigue syndrome (EUROMENE): expert Consensus on the diagnosis, Service provision, and care of people with ME/CFS in europe, Medicina (Kaunas) 57 (2021) 510, <https://doi.org/10.3390/medicina57050510>.
- [4] L. Nacul, E.M. Lacerda, C.C. Kingdon, H. Curran, E.W. Bowman, How have selection bias and disease misclassification undermined the validity of myalgic encephalomyelitis/chronic fatigue syndrome studies? *J. Health Psychol.* 24 (2019) 1765–1769, <https://doi.org/10.1177/1359105317695803>.
- [5] J. Malato, L. Graça, N. Sepúlveda, Impact of misdiagnosis in case-control studies of myalgic encephalomyelitis/chronic fatigue syndrome, *Diagnostics* 13 (2023) 531, <https://doi.org/10.3390/diagnostics13030531>.
- [6] F. Sotzny, J. Blanco, E. Capelli, J. Castro-Marrero, S. Steiner, M. Murovska, C. Scheibenbogen, Myalgic encephalomyelitis/chronic fatigue syndrome – evidence for an autoimmune disease, *Autoimmun. Rev.* 17 (2018) 601–609, <https://doi.org/10.1016/j.autrev.2018.01.009>.
- [7] S. Steiner, S.C. Becker, J. Hartwig, F. Sotzny, S. Lorenz, S. Bauer, M. Löbel, A.B. Stittrich, P. Grabowski, C. Scheibenbogen, Autoimmunity-related risk variants in PTPN22 and CTLA4 are associated with ME/CFS with infectious onset, *Front. Immunol.* 11 (2020) 578, <https://doi.org/10.3389/fimmu.2020.00578>.
- [8] N. Sepúlveda, J. Carneiro, E. Lacerda, L. Nacul, Myalgic encephalomyelitis/chronic fatigue syndrome as a hyper-regulated immune system driven by an interplay between regulatory T cells and chronic human herpesvirus infections, *Front. Immunol.* 10 (2019) 2684, <https://doi.org/10.3389/fimmu.2019.02684>.
- [9] J. Blomberg, C.G. Gottfries, A. Elfaitouri, M. Rizwan, A. Rosén, Infection elicited autoimmunity and Myalgic encephalomyelitis/chronic fatigue syndrome: an explanatory model, *Front. Immunol.* 9 (2018) 229, <https://doi.org/10.3389/fimmu.2018.00229>.
- [10] M. Loebel, P. Grabowski, H. Heidecke, S. Bauer, L.G. Hanitsch, K. Wittke, C. Meisel, P. Reinke, H.D. Volk, O. Fluge, O. Mella, C. Scheibenbogen, Antibodies to β adrenergic and muscarinic cholinergic receptors in patients with Chronic Fatigue Syndrome, *Brain Behav. Immun.* 52 (2016) 32–39, <https://doi.org/10.1016/J.BBI.2015.09.013>.
- [11] A. Bynke, P. Julin, C.-G. Gottfries, H. Heidecke, C. Scheibenbogen, J. Bergquist, Autoantibodies to beta-adrenergic and muscarinic cholinergic receptors in Myalgic Encephalomyelitis (ME) patients – a validation study in plasma and cerebrospinal fluid from two Swedish cohorts, *Brain, Behav. Immun. - Heal.* 7 (2020), 100107, <https://doi.org/10.1016/J.BBIH.2020.100107>.
- [12] H. Freitag, M. Szklarski, S. Lorenz, F. Sotzny, S. Bauer, A. Philippe, C. Kedor, P. Grabowski, T. Lange, G. Riemekasten, H. Heidecke, C. Scheibenbogen, Autoantibodies to vasoregulatory g-protein-coupled receptors correlate with symptom severity, autonomic dysfunction and disability in myalgic encephalomyelitis/chronic fatigue syndrome, *J. Clin. Med.* 10 (2021) 3675, <https://doi.org/10.3390/JCM10163675/51>.
- [13] M. Loebel, M. Eckey, F. Sotzny, E. Hahn, S. Bauer, P. Grabowski, J. Zerweck, P. Holenya, L.G. Hanitsch, K. Wittke, P. Borchmann, J.U. Rüffer, F. Hiepe, K. Ruprecht, U. Behrends, C. Meindl, H.D. Volk, U. Reimer, C. Scheibenbogen, Serological profiling of the EBV immune response in Chronic Fatigue Syndrome using a peptide microarray, *PLoS One* 12 (2017), e0179124, <https://doi.org/10.1371/journal.pone.0179124>.
- [14] N. Sepúlveda, Impact of genetic variation on the molecular mimicry between Anoctamin-2 and Epstein-Barr virus nuclear antigen 1 in Multiple Sclerosis, *Immunol. Lett.* 238 (2021) 29–31, <https://doi.org/10.1016/j.imlet.2021.07.007>.
- [15] S. Ramos, E. Bremu, S. Broadley, R. Kwiatek, J. Ng, T. Nguyen, S. Freeman, D. Staines, S. Marshall-Gradisnik, Regulatory T, natural killer T and γδ T cells in multiple sclerosis and chronic fatigue syndrome/myalgic encephalomyelitis: a comparison, *Asian Pac. J. Allergy Immunol.* 34 (2016) 300–305, <https://doi.org/10.12932/AP0733>.
- [16] G. Morris, M. Maes, Myalgic encephalomyelitis/chronic fatigue syndrome and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in phenomenology and neuroimmune characteristics, *BMC Med.* 11 (2013) 205, <https://doi.org/10.1186/1741-7015-11-205>.
- [17] S. Rasa, Z. Nora-Krukla, N. Henning, E. Eliassen, E. Shikova, T. Harrer, C. Scheibenbogen, M. Murovska, B.K. Prusty, Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), *J. Transl. Med.* 16 (2018) 268, <https://doi.org/10.1186/s12967-018-1644-y>.
- [18] S. Rasa-Dzelzkaleja, A. Krumina, S. Capenko, Z. Nora-Krukla, S. Gravelsina, A. Vilmane, L. Ievina, Y. Shoenfeld, M. Murovska, The persistent viral infections in the development and severity of myalgic encephalomyelitis/chronic fatigue syndrome, *J. Transl. Med.* 21 (2023), <https://doi.org/10.1186/S12967-023-03887-0>.
- [19] Z. Khalesi, V. Tamrchi, M.H. Razizadeh, A. Letaifati, P. Moradi, A. Habibi, N. Habibi, J. Heidari, M. Noori, M. Nahid Samiei, Z. Azarash, M. Hoseini, H. Saadati, A. Bahavar, M. Farajzadeh, S. Saeb, M. Hadadi, M. Sorouri Majd, S. Mothaghzadeh, P. Fazli, K. Asgari, S.J. Kiani, S. Ghorbani, Association between human herpesviruses and multiple sclerosis: a systematic review and meta-analysis, *Microb. Pathog.* 177 (2023), 106031, <https://doi.org/10.1016/J.MICPATH.2023.106031>.
- [20] S. Sedighi, O. Gholizadeh, S. Yasamineh, S. Akbarzadeh, P. Amini, P. Favakehi, H. Afkhami, A. Firouzi-Amandi, D. Pahlevan, M. Eslami, B. Yousefi, V. Poortahmasebi, M. Dadashpour, Comprehensive investigations relationship between viral infections and multiple sclerosis pathogenesis, *Curr. Microbiol.* 80 (2023) 1–11, <https://doi.org/10.1007/S00284-022-03112-Z/FIGURES/1>.
- [21] M.E. Ariza, Myalgic encephalomyelitis/chronic fatigue syndrome: the human herpesviruses are back, *Biomolecules* 11 (2021) 1–17, <https://doi.org/10.3390/biom11020185>.
- [22] S.S. Soldan, P.M. Lieberman, Epstein-Barr virus and multiple sclerosis, *Nat. Rev. Microbiol.* 21 (2023) 51–64, <https://doi.org/10.1038/s41579-022-00770-5>.
- [23] M. Ruiz-Pablos, B. Paiva, R. Montero-Mateo, N. Garcia, A. Zabaleta, Epstein-barr virus and the origin of myalgic encephalomyelitis or chronic fatigue syndrome, *Front. Immunol.* 12 (2021), 656797, <https://doi.org/10.3389/fimmu.2021.656797>.
- [24] K. Bjornevik, M. Cortese, B.C. Healy, J. Kuhle, M.J. Mina, Y. Leng, S.J. Elledge, D.W. Niebuhr, A.I. Scher, K.L. Munger, A. Ascherio, Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis, *Science* 375 (2022) 296–301, <https://doi.org/10.1126/science.abj8222>.
- [25] W. Lundström, R. Gustafsson, Human herpesvirus 6A is a risk factor for multiple sclerosis, *Front. Immunol.* 13 (2022) 840753, <https://doi.org/10.3389/FIMMU.2022.840753>.
- [26] J.S. Lee, E.M. Lacerda, L. Nacul, C.C. Kingdon, J. Norris, S. O'Boyle, C.h. Roberts, L. Palla, E.M. Riley, J.M. Cliff, Salivary DNA loads for human herpesviruses 6 and 7 are correlated with disease phenotype in myalgic encephalomyelitis/chronic fatigue syndrome, *Front. Med.* 8 (2021) 1129, <https://doi.org/10.3389/fmed.2021.656692>.
- [27] E. Engdahl, R. Gustafsson, J. Huang, M. Biström, I. Lima Bomfim, P. Stridh, M. Khademi, N. Brenner, J. Butt, A. Michel, D. Jons, M. Hortlund, L. Alonso-Magdalena, A.K. Hedström, L. Flamand, M. Ihira, T. Yoshikawa, O. Andersen, J. Hillert, L. Alfredsson, T. Waterboer, P. Sundström, T. Olsson, I. Kockum, A. Fogdell-Hahn, Increased serological response against human herpesvirus 6A is associated with risk for multiple sclerosis, *Front. Immunol.* 10 (2019) 2715, <https://doi.org/10.3389/fimmu.2019.02715>.
- [28] F. Kasimir, D. Toomey, Z. Liu, A.C. Kaiping, M.E. Ariza, B.K. Prusty, Tissue specific signature of HHV-6 infection in ME/CFS, *Front. Mol. Biosci.* 9 (2022), 1044964, <https://doi.org/10.3389/fmolb.2022.1044964>.
- [29] G. Bateman, A. Bateman, Chronic Fatigue Syndrome and Multiple Sclerosis Have Reduced Craniospinal Compliance and Dilated Pressurized Bridging Cortical Veins: A Hypothesis Illustrated with Two Case Studies. *Preprints.org*, 2023, p. 2023052264, <https://doi.org/10.20944/PREPRINTS202305.2264.V1>.
- [30] E.M. Lacerda, E.W. Bowman, J.M. Cliff, C.C. Kingdon, E.C. King, J.-S. Lee, T.G. Clark, H.M. Dockrell, E.M. Riley, H. Curran, L. Nacul, The UK ME/CFS biobank for biomedical research on myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and multiple sclerosis, *Open J. Biore sour.* 4 (2017) 4, <https://doi.org/10.5334/ojb.28>.
- [31] E.M. Lacerda, K. Mudie, C.C. Kingdon, J.D. Butterworth, S. O'Boyle, L. Nacul, The UK ME/CFS biobank: a disease-specific biobank for advancing clinical research into myalgic encephalomyelitis/chronic fatigue syndrome, *Front. Neurol.* 9 (2018) 1026, <https://doi.org/10.3389/fneur.2018.01026>.
- [32] J. Blauensteiner, R. Bertinat, L.E. León, M. Riederer, N. Sepúlveda, F. Westermeier, Altered endothelial dysfunction-related miRs in plasma from ME/CFS patients, *Sci. Rep.* 11 (2021), 10604, <https://doi.org/10.1038/S41598-021-89834-9>.
- [33] R. Bertinat, R. Villalobos-Labra, L. Hofmann, J. Blauensteiner, N. Sepúlveda, F. Westermeier, Decreased NO production in endothelial cells exposed to plasma from ME/CFS patients, *Vasc. Pharmacol.* 143 (2022), 106953, <https://doi.org/10.1016/J.VPH.2022.106953>.
- [34] T. Vogl, I.N. Kalka, S. Klompus, S. Levitan, A. Weinberger, E. Segal, Systemic antibody responses against human microbiota flagellins are overrepresented in chronic fatigue syndrome patients, *Sci. Adv.* 8 (2022) 2422, <https://doi.org/10.1126/sciadv.abq2422>.
- [35] L.S. Rodrigues, L.H. Da Silva Nali, C.O.D. Leal, E.C. Sabino, E.M. Lacerda, C.C. Kingdon, L. Nacul, C.M. Romano, HERV-K and HERV-W transcriptional activity in myalgic encephalomyelitis/chronic fatigue syndrome, *Autoimmun. Highlights.* 10 (2019) 1–5, <https://doi.org/10.1186/s13317-019-0122-8>.

- [36] A. González-Cebrián, E. Almenar-Pérez, J. Xu, T. Yu, W.E. Huang, K. Giménez-Orenga, S. Hutchinson, T. Lodge, L. Nathanson, K.J. Morten, A. Ferrer, E. Oltra, Diagnosis of myalgic encephalomyelitis/chronic fatigue syndrome with partial least squares discriminant analysis: relevance of blood extracellular vesicles, *Front. Med.* 9 (2022), 842991, <https://doi.org/10.3389/fmed.2022.842991>.
- [37] E. Almenar-Pérez, L. Sarría, L. Nathanson, E. Oltra, Assessing diagnostic value of microRNAs from peripheral blood mononuclear cells and extracellular vesicles in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome, *Sci. Rep.* 10 (2020) 2064, <https://doi.org/10.1038/s41598-020-58506-5>.
- [38] E.M. Lacerda, K. Geraghty, C.C. Kingdon, L. Palla, L. Nacul, A logistic regression analysis of risk factors in ME/CFS pathogenesis, *BMC Neurol.* 19 (2019) 275, <https://doi.org/10.1186/s12883-019-1468-2>.
- [39] A. Melvin, E. Lacerda, H.M. Dockrell, S. O'Rahilly, L. Nacul, Circulating levels of GDF15 in patients with myalgic encephalomyelitis/chronic fatigue syndrome, *J. Transl. Med.* 17 (2019) 409, <https://doi.org/10.1186/s12967-019-02153-6>.
- [40] V. Jain, A. Arunkumar, C. Kingdon, E. Lacerda, L. Nacul, Prevalence of and risk factors for severe cognitive and sleep symptoms in ME/CFS and MS, *BMC Neurol.* 17 (2017) 117, <https://doi.org/10.1186/S12883-017-0896-0>.
- [41] J.M. Cliff, E.C. King, J.S. Lee, N. Sepúlveda, A.S. Wolf, C. Kingdon, E. Bowman, H.M. Dockrell, L. Nacul, E. Lacerda, E.M. Riley, Cellular immune function in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), *Front. Immunol.* 10 (2019) 796, <https://doi.org/10.3389/fimmu.2019.00796>.
- [42] T.D. Domingues, A.D. Grabowska, J.S. Lee, J. Ameijeiras-Alonso, F. Westermeier, C. Scheibenbogen, J.M. Cliff, L. Nacul, E.M. Lacerda, H. Mourão, N. Sepúlveda, Herpesviruses serology distinguishes different subgroups of patients from the United Kingdom myalgic encephalomyelitis/chronic fatigue syndrome biobank, *Front. Med.* 8 (2021), 686736, <https://doi.org/10.3389/fmed.2021.686736>.
- [43] K. Fukuda, S.E. Straus, I. Hickie, M.C. Sharpe, J.G. Dobbins, A. Komaroff, The chronic fatigue syndrome: a comprehensive approach to its definition and study, *Ann. Intern. Med.* 121 (1994) 953–959, <https://doi.org/10.7326/0003-4819-121-12-199412150-00009>.
- [44] B.M. Carruthers, A.K. Jain, K.L. De Meirlier, D.L. Peterson, N.G. Klimas, A.M. Lemer, A.C. Bested, P. Flor-Henry, P. Joshi, A.C.P. Powles, J.A. Sherkey, M.I. Van de Sande, Myalgic encephalomyelitis/chronic fatigue syndrome: clinical working case definition, diagnostic and treatment protocols, *J. Chronic Fatigue Syndrome* 11 (2003) 7–115, https://doi.org/10.1300/J092v11n01_02.
- [45] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B* 57 (1995) 289–300, <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- [46] M.J. Van Der Laan, E.C. Polley, A.E. Hubbard, Super learner, *Stat. Appl. Genet. Mol. Biol.* 6 (2007) 25, <https://doi.org/10.2202/1544-6115.1309>.
- [47] R Core Team, R: A Language and Environment for Statistical Computing, 2022. <https://www.r-project.org/>.
- [48] B. Rich, table 1: Tables of Descriptive Statistics in HTML, 2023. <https://cran.r-project.org/package=table1>.
- [49] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, M. Müller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, *BMC Bioinf.* 12 (2011) 71, <https://doi.org/10.1186/1471-2105-12-77>.
- [50] M. López-Ratón, M.X. Rodríguez-Alvarez, C. Cadarso-Suárez, F. Gude-Sampedro, Optimalcutpoints: an R package for selecting optimal cutpoints in diagnostic tests, *J. Stat. Software* 61 (2014) 1–36, <https://doi.org/10.18637/jss.v061.i08>.
- [51] E. Polley, E. LeDell, C. Kennedy, M. van der Laan, SuperLearner: Super Learner Prediction, 2021. <https://cran.r-project.org/package=SuperLearner>.
- [52] L. Jason, D. Ohanian, A. Brown, M. Sunquist, S. McManimen, L. Klebek, P. Fox, M. Sorenson, Differentiating multiple sclerosis from myalgic encephalomyelitis and chronic fatigue syndrome, *Insights Biomed* 2 (2017) 11, <https://doi.org/10.21767/2572-5610.10027>.
- [53] D. Ohanian, A. Brown, M. Sunnquist, J. Furst, L. Nicholson, L. Klebek, L.A. Jason, Identifying key symptoms differentiating myalgic encephalomyelitis and chronic fatigue syndrome from multiple sclerosis, *Neurol.* 4 (2016) 41. [https://pmc.articles/PMC5214344/](https://pmc/articles/PMC5214344/). (Accessed 27 February 2023).
- [54] A. Cortese, A. Conte, G. Ferrazzano, E. Sgarlata, E. Millefiorini, M. Frontoni, A. Berardelli, Photophobia in multiple sclerosis, *Mult. Scler. Relat. Disord.* 26 (2018) 55–57, <https://doi.org/10.1016/j.msard.2018.09.005>.
- [55] A.D. Port, A. Orlin, S. Kiss, S. Patel, D.J. D'Amico, M.P. Gupta, Cytomegalovirus retinitis: a review, *J. Ocul. Pharmacol. Therapeut.* 33 (2017) 224–234, <https://doi.org/10.1089/JOP.2016.0140>.
- [56] P.A. McCombe, T.P. Gordon, M.W. Jackson, Bladder dysfunction in multiple sclerosis, *Expert Rev. Neurother.* 9 (2009) 331–340, <https://doi.org/10.1586/14737175.9.3.331>.
- [57] B. Ayoglu, N. Mitsios, I. Kockum, M. Khademi, A. Zandian, R. Sjöberg, B. Forsström, J. Bredenberg, I.L. Bomfim, E. Holmgren, H. Grönlund, A.O. Guerreiro-Cacais, N. Abdelmagid, M. Uhlén, T. Waterboer, L. Alfredsson, J. Mulder, J.M. Schwenk, T. Olsson, P. Nilsson, Anoactamin 2 identified as an autoimmune target in multiple sclerosis, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 2188–2193, <https://doi.org/10.1073/pnas.1518553113>.
- [58] T.V. Lanz, R.C. Brewer, P.P. Ho, J.S. Moon, K.M. Jude, D. Fernandez, R.A. Fernandes, A.M. Gomez, G.S. Nadj, C.M. Bartley, R.D. Schubert, I.A. Hawes, S. E. Vazquez, M. Iyer, J.B. Zuchero, B. Teegen, J.E. Dunn, C.B. Lock, L.B. Kipp, V.C. Coatham, B.M. Ueberheide, B.T. Aftab, M.S. Anderson, J.L. DeRisi, M.R. Wilson, R.J.M. Bashford-Rogers, M. Platten, K.C. Garcia, L. Steinman, W.H. Robinson, Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM, *Nature* 603 (2022) 321–327, <https://doi.org/10.1038/S41586-022-04432-7>.
- [59] K. Tengvall, J. Huang, C. Hellström, P. Kammer, M. Biström, B. Ayoglu, I.L. Bomfim, P. Stridh, J. Butt, N. Brenner, A. Michel, K. Lundberg, L. Padyukov, I. E. Lundberg, E. Svenungsson, I. Ernberg, S. Olafsson, A.T. Dilthey, J. Hillert, L. Alfredsson, P. Sundström, P. Nilsson, T. Waterboer, T. Olsson, I. Kockum, Molecular mimicry between Anoactamin 2 and Epstein-Barr virus nuclear antigen 1 associates with multiple sclerosis risk, *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019) 16955–16960, <https://doi.org/10.1073/pnas.1902623116>.
- [60] R. Sakakibara, S. Sawai, T. Ogata, Varicella-zoster virus infection and autonomic dysfunction, *Auton. Neurosci.* 242 (2022), 103018, <https://doi.org/10.1016/J.AUTNEU.2022.103018>.
- [61] R. Martinez-Reviriejo, S. Tejada, G.A.R. Adebanjo, C. Chello, M.C. Machado, F.R. Parisella, M. Campins, A. Tammaro, J. Rello, Varicella-Zoster virus reactivation following severe acute respiratory syndrome coronavirus 2 vaccination or infection: new insights, *Eur. J. Intern. Med.* 104 (2022) 73, <https://doi.org/10.1016/J.EJIM.2022.07.022>.
- [62] B.V. Rooney, B.E. Crucian, D.L. Pierson, M.L. Laudenslager, S.K. Mehta, Herpes virus reactivation in astronauts during spaceflight and its application on earth, *Front. Microbiol.* 10 (2019) 16, <https://doi.org/10.3389/fmicb.2019.00016>.
- [63] D.M. Koelle, S. Baryc, M.L. Huang, R.L. Ashley, L. Corey, J. Zeh, S. Ashton, D. Buchwald, Markers of viral infection in monozygotic twins discordant for chronic fatigue syndrome, *Clin. Infect. Dis.* 35 (2002) 518–525, <https://doi.org/10.1086/341774>.
- [64] H. Nakayama, R. Yamazaki, J. Kato, Y. Koda, M. Sakurai, T. Mori, Impact of specific antibody level on human herpesvirus 6 reactivation after allogeneic stem cell transplantation, *Transplant. Cell. Ther.* 27 (2021) 174.e1–174.e5, <https://doi.org/10.1016/J.JTCT.2020.10.011>.
- [65] A.J. Ross, M.S. Medow, P.C. Rowe, J.M. Stewart, What is brain fog? An evaluation of the symptom in postural tachycardia syndrome, *Clin. Auton. Res.* 23 (2013) 305–311, <https://doi.org/10.1007/S10286-013-0212-Z>.
- [66] A.J. Ocon, Caught in the thickness of brain fog: exploring the cognitive symptoms of Chronic Fatigue Syndrome, *Front. Physiol.* 4 (2013) 63, <https://doi.org/10.3389/FPHYS.2013.00063>.
- [67] K. Krishnan, Y. Lin, K.-R.M. Prewitt, D.A. Potter, Multidisciplinary approach to brain fog and related persisting symptoms post COVID-19, *J. Heal. Serv. Psychol.* 48 (2022) 31, <https://doi.org/10.1007/S42843-022-00056-7>.
- [68] M.E. Marcocci, G. Napoletani, V. Protto, O. Kolesova, R. Piacentini, D.D. Li Puma, P. Lomonte, C. Grassi, A.T. Palamara, G. De Chiara, Herpes simplex virus-1 in the brain: the dark side of a sneaky infection, *Trends Microbiol.* 28 (2020) 808–820, <https://doi.org/10.1016/J.TIM.2020.03.003>.
- [69] E. Fruchter, S. Goldberg, D. Fenchel, I. Grotto, K. Ginat, M. Weiser, The impact of Herpes simplex virus type 1 on cognitive impairments in young, healthy individuals - a historical prospective study, *Schizophr. Res.* 168 (2015) 292–296, <https://doi.org/10.1016/J.SCHRES.2015.08.036>.
- [70] I. Jonker, H.C. Klein, H.E. Duivis, R.H. Yolken, J.G.M. Rosmalen, R.A. Schoevers, Association between exposure to HSV1 and cognitive functioning in a general population of adolescents. The TRAILS study, *PLoS One* 9 (2014), e101549, <https://doi.org/10.1371/JOURNAL.PONE.0101549>.
- [71] K.D. Tarter, A.M. Simanek, J.B. Dowd, A.E. Aiello, Persistent viral pathogens and cognitive impairment across the life course in the third national health and nutrition examination survey, *J. Infect. Dis.* 209 (2014) 837–844, <https://doi.org/10.1093/INFDIS/JIT616>.
- [72] M.J. Murphy, L. Fani, M.K. Ikram, M. Ghanbari, M.A. Ikram, Herpes simplex virus 1 and the risk of dementia: a population-based study, *Sci. Rep.* 11 (2021) 8691, <https://doi.org/10.1038/S41598-021-87963-9>.

- [73] F. Dickerson, J.R. Schroeder, V. Nimgaonkar, J. Gold, R. Yolken, The association between exposure to herpes simplex virus type 1 (HSV-1) and cognitive functioning in schizophrenia: a meta-analysis, *Psychiatr. Res.* 291 (2020), 113157, <https://doi.org/10.1016/J.PSYCHRES.2020.113157>.
- [74] J.D. Tucker, A.S. Bertke, Assessment of cognitive impairment in HSV-1 positive schizophrenia and bipolar patients: systematic review and meta-analysis, *Schizophr. Res.* 209 (2019) 40–47, <https://doi.org/10.1016/J.SCHRES.2019.01.001>.
- [75] N. Sepúlveda, J. Malato, F. Sotzny, A.D. Grabowska, A. Fonseca, C. Cordeiro, L. Graça, P. Bicek, U. Behrends, J. Mautner, F. Westermeier, E.M. Lacerda, C. Scheibenbogen, Revisiting IgG antibody reactivity to epstein-barr virus in myalgic encephalomyelitis/chronic fatigue syndrome and its potential application to disease diagnosis, *Front. Med.* 9 (2022), 921101, <https://doi.org/10.3389/FMED.2022.921101>.
- [76] E.M. Lacerda, C. McDermott, C.C. Kingdon, J. Butterworth, J.M. Cliff, L. Nacul, Hope, disappointment and perseverance: reflections of people with Myalgic encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) and Multiple Sclerosis participating in biomedical research. A qualitative focus group study, *Health Expect.* 22 (2019) 373–384, <https://doi.org/10.1111/HEX.12857>.
- [77] Patti F., Lo Fermo S., Lights and shadows of cyclophosphamide in the treatment of multiple sclerosis, *Autoimmune Dis.* 2011 (2011) 14, doi:10.4061/2011/961702.
- [78] S. Brancati, L. Gozzo, L. Longo, D.C. Vitale, F. Drago, Rituximab in multiple sclerosis: are we ready for regulatory approval? *Front. Immunol.* 12 (2021) 2600, <https://doi.org/10.3389/fimmu.2021.661882>.
- [79] I.G. Rekeland, A. Fosså, A. Lande, I. Ktoridou-Valen, K. Sørland, M. Holsen, K.J. Tronstad, K. Risa, K. Alme, M.K. Viken, B.A. Lie, O. Dahl, O. Mella, Ø. Fluge, Intravenous cyclophosphamide in myalgic encephalomyelitis/chronic fatigue syndrome. An open-label phase II study, *Front. Med.* 7 (2020) 162, <https://doi.org/10.3389/FMED.2020.00162>.
- [80] Ø. Fluge, I.G. Rekeland, K. Lien, H. Thürmer, P.C. Borchgrevink, C. Schäfer, K. Sørland, J. Aßmus, I. Ktoridou-Valen, I. Herder, M.E. Gotaas, Ø. Kvammen, K. A. Baranowska, L.M.L.J. Bohnen, S.S. Martinsen, A.E. Lonar, A.E.H. Solvang, A.E.S. Gya, O. Bruland, K. Risa, K. Alme, O. Dahl, O. Mella, B-lymphocyte depletion in patients with myalgic encephalomyelitis/chronic fatigue syndrome: a randomized, double-blind, placebo-controlled trial, *Ann. Intern. Med.* 170 (2019) 585–593, <https://doi.org/10.7326/M18-1451>.
- [81] M. Szklarski, H. Freitag, S. Lorenz, S.C. Becker, F. Sotzny, S. Bauer, J. Hartwig, H. Heidecke, K. Wittke, C. Kedor, L.G. Hanitsch, P. Grabowski, N. Sepúlveda, C. Scheibenbogen, Delineating the association between soluble CD26 and autoantibodies against G-protein coupled receptors, immunological and cardiovascular parameters identifies distinct patterns in post-infectious vs. Non-Infection-Triggered myalgic encephalomyelitis/chro, *Front. Immunol.* 12 (2021) 1077, <https://doi.org/10.3389/fimmu.2021.644548>.
- [82] F. Sotzny, I.S. Filgueiras, C. Kedor, H. Freitag, K. Wittke, S. Bauer, N. Sepúlveda, D.L. Mathias da Fonseca, G.C. Baiocchi, A.H.C. Marques, M. Kim, T. Lange, D. R. Plaça, F. Luebber, F.M. Paulus, R. De Vito, I. Jurisica, K. Schulze-Forster, F. Paul, J. Bellmann-Strobl, R. Rust, U. Hoppmann, Y. Shoenfeld, G. Riemeckasten, H. Heidecke, O. Cabral-Marques, C. Scheibenbogen, Dysregulated autoantibodies targeting vaso- and immunoregulatory receptors in Post COVID Syndrome correlate with symptom severity, *Front. Immunol.* 13 (2022) 981532, <https://doi.org/10.3389/FIMMU.2022.981532>.
- [83] P. Rieckmann, A. Boyko, D. Centonze, I. Elovaara, G. Giovannoni, E. Havrdová, O. Hommes, J. Kesselring, G. Kobelt, D. Langdon, J. Lelorier, S.A. Morrow, C. Oreja-Guevara, S. Schippling, C. Thalheim, H. Thompson, P. Vermersch, Achieving patient engagement in multiple sclerosis: a perspective from the multiple sclerosis in the 21st Century Steering Group, *Mult. Scler. Relat. Disord.* 4 (2015) 202–218, <https://doi.org/10.1016/J.MSARD.2015.02.005>.
- [84] M. Cerghet, E. Dobie, J.E. Lafata, L. Schultz, S. Elias, M. Pladevall, J. Reuther, Adherence to disease-modifying agents and association with quality of life among patients with relapsing-remitting multiple sclerosis, *Int. J. MS Care.* 12 (2010) 51–58, <https://doi.org/10.7224/1537-2073-12.2.51>.
- [85] S. Klineova, F.D. Lublin, Clinical course of multiple sclerosis, *Cold Spring Harb. Perspect. Med.* 8 (2018) a028928, <https://doi.org/10.1101/CSHPERSPECT.A028928>.
- [86] M. Pugliatti, G. Rosati, H. Carton, T. Riise, J. Drulovic, L. Vécsei, I. Milanov, The epidemiology of multiple sclerosis in Europe, *Eur. J. Neurol.* 13 (2006) 700–722, <https://doi.org/10.1111/J.1468-1331.2006.01342.X>.

F.6 Chapter 8



Research article

The SARS-CoV-2 receptor angiotensin-converting enzyme 2 (ACE2) in myalgic encephalomyelitis/chronic fatigue syndrome: A meta-analysis of public DNA methylation and gene expression data



João Malato^{a,b}, Franziska Sotzny^{c,**}, Sandra Bauer^c, Helma Freitag^c, André Fonseca^d, Anna D. Grabowska^e, Luís Graça^a, Clara Cordeiro^{b,d}, Luís Nacul^{f,g}, Eliana M. Lacerda^f, Jesus Castro-Marrero^h, Carmen Scheibenbogen^c, Francisco Westermeier^{i,j}, Nuno Sepúlveda^{b,c,k,*}

^a Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal^b CEAUL – Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal^c Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt Universität zu Berlin and Berlin Institute of Health, Institute of Medical Immunology, Berlin, Germany^d Faculdade de Ciências e Tecnologia, Universidade do Algarve, Faro, Portugal^e Department of Biophysics, Physiology, and Pathophysiology, Medical University of Warsaw, Warsaw, Poland^f Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom^g Complex Chronic Diseases Program, British Columbia Women's Hospital and Health Centre, Vancouver, British Columbia, Canada^h Vall d'Hebron Hospital Research Institute, Division of Rheumatology, ME/CFS Unit, Universitat Autònoma de Barcelona, Barcelona, Spainⁱ Institute of Biomedical Science, Department of Health Studies, FH Joanneum University of Applied Sciences, Graz, Austria^j Centro Integrativo de Biología y Química Aplicada (CIBQA), Universidad Bernardo OHiggins, Santiago, Chile^k Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom

ARTICLE INFO

ABSTRACT

Keywords:

Myalgic encephalomyelitis/chronic fatigue syndrome
SARS-CoV-2
ACE2
Gene expression
DNA methylation

People with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) often report a high frequency of viral infections and flu-like symptoms during their disease course. Given that this reporting agrees with different immunological abnormalities and altered gene expression profiles observed in the disease, we aimed at answering whether the expression of the human angiotensin-converting enzyme 2 (ACE2), the major cell entry receptor for SARS-CoV-2, is also altered in these patients. In particular, a low expression of ACE2 could be indicative of a high risk of developing COVID-19. We then performed a meta-analysis of public data on CpG DNA methylation and gene expression of this enzyme and its homologous ACE protein in peripheral blood mononuclear cells and related subsets. We found that patients with ME/CFS have decreased methylation levels of four CpG probes in the ACE locus (cg09920557, cg19802564, cg21094739, and cg10468385) and of another probe in the promoter region of the ACE2 gene (cg08559914). We also found a decreased expression of ACE2 but not of ACE in patients when compared to healthy controls. Accordingly, in newly collected data, there was evidence for a significant higher proportion of samples with an ACE2 expression below the limit of detection in patients than healthy controls. Altogether, patients with ME/CFS can be at a higher COVID-19 risk and, if so, they should be considered a priority group for vaccination by public health authorities. To further support this conclusion, similar research is recommended for other human cell entry receptors and cell types, namely, those cells targeted by the virus.

* Corresponding author.

** Corresponding author.

E-mail addresses: franziska.sotzny@charite.de (F. Sotzny), nunosep@gmail.com (N. Sepúlveda).

1. Introduction

Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) is a multifactorial and complex disease characterized by two key symptoms: (1) persistent but unexplained fatigue that is not alleviated by rest; and (2) post-exertional malaise upon minimal physical or even mental effort [1, 2]. Although its cause remains unknown, a growing body of evidence strongly associates ME/CFS with several microbial and viral infections, as potential triggering factors [3, 4]. In addition, it is currently hypothesized that reactivations of dormant viral infections also play a role [5, 6] due to several immunological abnormalities [7, 8, 9]. On the molecular basis of the disease, peripheral blood mononuclear cells (PBMCs) have altered gene expression profiles [10], including a decreased abundance of the human angiotensin-converting enzyme 2 (ACE2) [11], the main receptor of the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) for cell invasion [12, 13, 14]. Altogether, this evidence raises the question about the COVID-19 risk in patients with ME/CFS.

As basic information, ACE2 is encoded by the X-linked ACE2 gene whose expression is predominant in the lungs, heart, skin, and kidneys [15, 16, 17, 18]. Its expression can also be detected in monocytes [19] and activated macrophages [20]. However, the percentage of ACE2-expressing cells is below 5% in the main immune-cell populations [20]. Accordingly, current RNA-Seq studies suggest a residual ACE2 expression in PBMCs from healthy controls [18]. ACE2 has an amino-acid sequence identity of 41% with its homologous angiotensin-converting enzyme (ACE) [21]. This sequence similarity increases to 61% at the nucleotide level [21]. The enzymes ACE and ACE2 are members of the renin-angiotensin-aldosterone system (RAAS), which regulates blood pressure and vascular resistance [22]. In particular, ACE and ACE2 have vasoconstriction and vasodilation effects, respectively. Given this counteracting effect, high ACE:ACE2 ratios are possible indicators of severe COVID-19 outcomes, linked to increased reactive oxygen species (ROS) production, vasoconstriction, and inflammation [23].

To answer our research question, we performed a meta-analysis of public DNA methylation and gene expression data of ACE2 and ACE in PBMCs. Similar study was conducted on the DNA methylation pattern of ACE2 in the same cell type from patients with systemic lupus erythematosus [24], an autoimmune disease whose symptoms overlap with the ones from ME/CFS [25]. To complement our findings, we also compared the mRNA levels of these two genes in PBMCs from a new cohort of female patients with ME/CFS and healthy women.

2. Materials and methods

2.1. Eligible diagnostic criteria of ME/CFS

In our meta-analysis, we selected public data from studies using either the 1994 Centre for Diseases Control criteria (1994 CDC/Fukuda) [1] or the 2003 Canadian Consensus Criteria (2003 CCC) [2] for the disease diagnosis. These criteria are defined by the presence of several key symptoms while excluding known medical conditions (e.g., multiple sclerosis or lupus) that can also explain fatigue. The choice of using these two criteria for study selection complies with the research standards set by the European Network on ME/CFS [26].

2.2. Analysis of published DNA methylation association studies

Our meta-analysis was based on six genome-wide DNA methylation association studies (Table 1), four of which [27, 28, 29, 30] were previously reviewed [31], and other two published after this review [32, 33]. Briefly, these studies aimed at identifying differentially methylated CpG dinucleotide sites between patients and healthy controls. Illumina methylation arrays were used to measure the respective DNA methylation levels with the exception of a single study (Table 1). In this study, the measurements were made by the reduced representation bisulfite sequencing [33].

With respect to the exclusion criteria, one study excluded individuals who were taking beta-blockers or ACE inhibitors [30]. Three studies excluded participants who were treated with immunomodulatory effects or affecting the underlying DNA methylation levels at the time of data collection [28, 29, 32].

In four of the published DNA methylation studies, patients and healthy controls were matched for age, gender, and body mass index (Table 1) [28, 29, 30]. In two other studies, the matching was only based on age and gender [27, 33]. Ethnicity was also used for further matching [30, 32] or the same matching could be assumed in studies that only recruited white females [28, 29]. The DNA methylation levels were quantified in CD4+ T cells [27], PBMCs [28, 29, 30, 33], and T lymphocytes [32].

We conducted a joint analysis of the four array-based studies which made the data available [28, 29, 30, 32]. We first retrieved the data from all the CpG probes located in the coding regions and the transcription starting sites (TSS) of ACE and ACE2, respectively. We then restricted our data analysis to the 27 probes shared between the Infinium

Table 1. Summary of the six DNA methylation studies under analysis.

Reference	Sample type	ME/CFS patients			Healthy controls, n	Technology (manufacturer)	NCBI GEO Accession number
		n	Sample characteristics	Case definition			
[27]	CD4+ T cells	25	Female/male adults Mean age: 50 years old Mean BMI: not reported	1994 CDC/Fukuda	18	Infinium HumanMethylation450K Array (Illumina)	NA
[28]	PBMC	12	Female adults Mean age: 41 years old Mean BMI: 23 kg/m ²	1994 CDC/Fukuda & 2003 CCC	12	Infinium HumanMethylation450K Array (Illumina)	GSE59489
[29]	PBMC	49	Female adults Mean age: 50 years old; Mean BMI: 23 kg/m ²	1994 CDC/Fukuda & 2003 CCC	25	Infinium HumanMethylation450 Array (Illumina)	GSE93266
[30]	PBMC	13	Female adults Mean age: 50 years old Mean BMI: 26 kg/m ²	1994 CDC/Fukuda & 2003 CCC	12	Methylation EPIC Array (Illumina)	GSE111183
[32]	T lymphocytes	61	Female/male adults Mean age: 32 years old Mean BMI: 27 kg/m ²	1994 CDC/Fukuda & 2003 CCC	48	Infinium HumanMethylation450K Array (Illumina)	GSE156792
[33]	PBMC	10	Female/male adults Mean age: Not reported Mean BMI: not reported	2003 CCC	10	Reduced representation Bisulfite sequencing	GSE153667

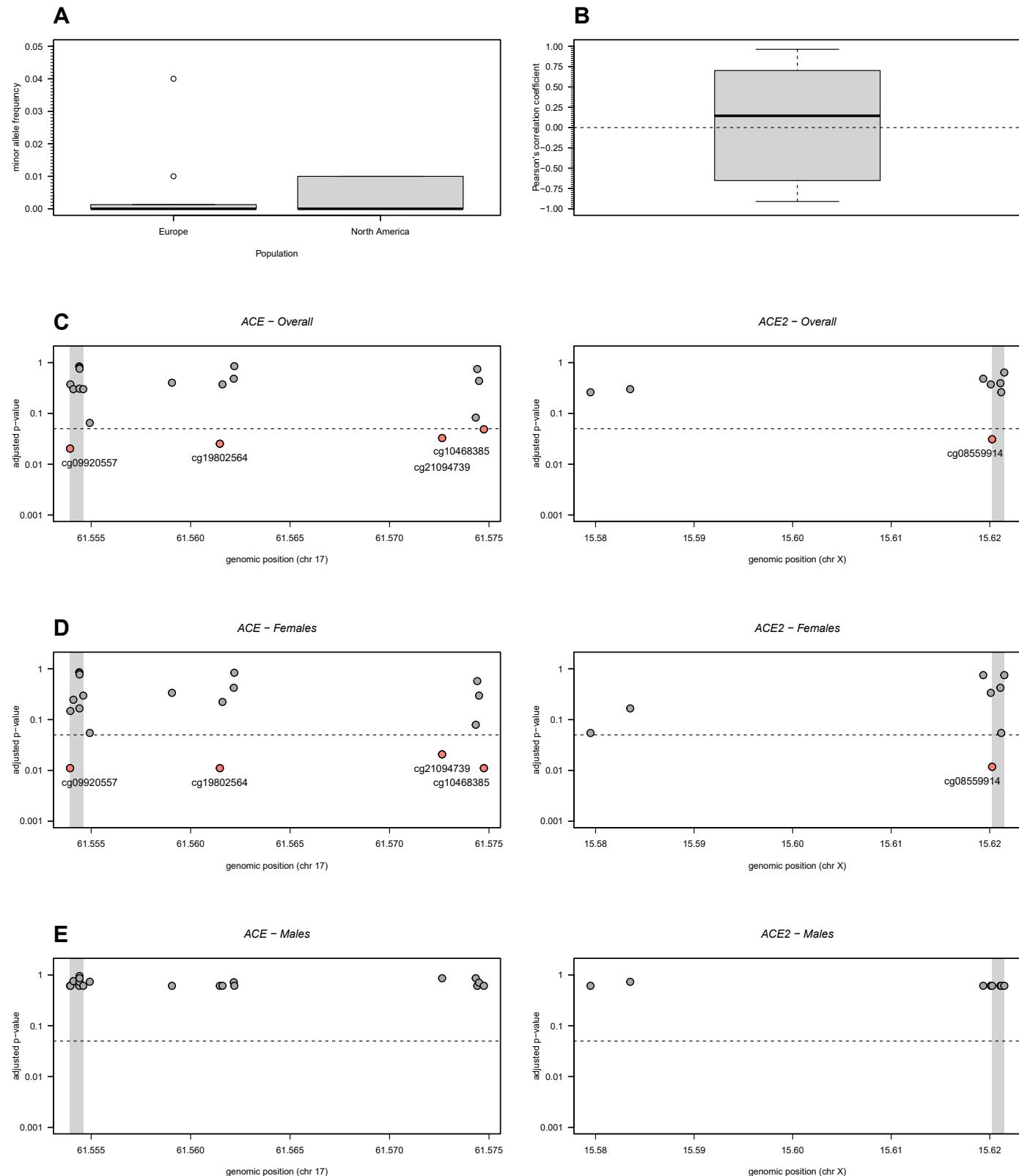


Figure 1. DNA methylation analysis of 19 and 8 CpG probes located in the *ACE* and *ACE2* genes, respectively. (A) Minor allele frequency in European and North American populations of SNPs located in the probes under analysis (see the respective data in Supplementary Table 2). (B) Boxplot of all possible Pearson's correlation coefficients (y axis) between the M-values of the probes under analysis. Horizontal dashed line represents the situation of lack of correlation. (C) Adjusted p-values for the overall association between each probe and ME/CFS. Adjusted p-values were calculated according to the Benjamini-Hochberg procedure with a false discovery rate of 5% (dashed line). Grey areas in the plots represent the TSS of the genes. (D) and (E) The same analyses as shown in C but for women and men separately.

HumanMethylation450K and the Infinium HumanMethylationEPIC arrays (Supplementary Table 1).

Before conducting the statistical analysis itself, we checked whether (1) the selected probes showed a high probability of detection, (2) they were not cross-reactive with other genomic regions, and (3) they were not affected by single nucleotide polymorphisms (SNPs) with high minor allele frequencies [34]. In the latter criterion, the SNPs included in the selected probes had a minor allele frequency less than 5% in Europeans and North Americans (Figure 1A; Supplementary Table 2) referring to the sampled populations of the studies. All probes passed the remaining basic quality control checks.

We analyzed the M-values of a given probe instead of the respective β -values to ensure a good approximation of the Normal distribution to the data [35]. Briefly, the β -values were calculated as the proportion of the methylation signal relative to the total signal for a given probe. The M-values were finally obtained by applying a logit transformation to the β -values.

To analyze the M-values of each probe, we initially estimated a linear regression model where the respective covariates were the study indicator and the disease status of the participants. In this model, we included the main effects of the covariates and the interaction. The model parameters were then estimated by the maximum likelihood method. Note that the main effect of the disease status is usually seen as the pooled effect of this covariate across all studies, as done in meta-analysis.

We then simplified the model using a backward stepwise procedure based on Akaike's information criteria. Since the effect of the study indicator was significant for the data of each probe, we tested the association between ME/CFS and a given probe using a likelihood ratio test. In this test, we compared the model including the study indicator only with the best model including that covariate and the one associated with disease status (i.e., either the model only including the main effects or the model including both main effects and the interaction term).

To control for multiple testing, we adjusted the raw p-values using the Benjamini-Hochberg procedure [36]. This adjustment ensured a false discovery rate of 5% under the assumption of independent tests.

Pearson's correlation coefficient was used to check the validity of this assumption (Figure 1B).

We also repeated the same association analysis for women and men separately. Note that three studies only recruited women [28, 29, 30] while the remaining study recruited both men and women [32]. In the latter study, there was no information available about the gender of each participant. In this case, we estimated this missing information using the function `getSex` of the R package `minfi` applied to the genome-wide DNA methylation data [37]. The resulting frequencies of men and women matched with those reported in the original study.

In the women-specific analysis, we performed the same association analysis as described above. In the men-specific analysis, we compared a linear regression model with the disease status as the single covariate against another model without that covariate, when analyzing data from each probe. The comparison was done by the likelihood ratio test whose p-values were then adjusted for multiple testing in the same way as described above.

Finally, for the study which did not share the respective data [27], we checked whether the reported differentially methylated CpG probes were located in either ACE or ACE2 (see Table 1 from this study). We did the same for the study based on the reduced representation bisulfite sequencing technology [33] (see Additional File 1 from this study).

2.3. Analysis of gene expression studies

Our meta-analysis of gene expression studies was focused on eight reports using microarray technology (Table 2) [11,38–44]. These studies complied with the Minimum Information about a Microarray Experiment (MIAME) standard [45] and, therefore, they were considered to have sufficient quality for their inclusion in the meta-analysis. In particular, these studies normalized the data which ensured comparability between different samples and between different measurements of the same genes.

Gene expression of these studies was performed in PBMCs (5 studies), whole blood (2 studies) and muscle biopsies (one study). One study excluded participants who were taking any regular medication [43].

Table 2. Summary of the 8 microarray-based gene expression studies under analysis, ordered by the year of publication.

Reference	Sample type	ME/CFS patients		Healthy controls, n	Technology (Manufacturer)	ACE/ACE2 available	Data availability (NCBI GEO Accession number)	
		n	Sample characteristics	Case definition				
[38]	PBMCs	5	Female adults Mean age: 42 years old Mean BMI: not reported	1994 CDC/Fukuda	5	Atlas Glass Human 3.8 I Microarray (BD Biosciences Clontech)	No/No	No (NA)
[39]	PBMCs	25	Female/male adults Mean age: 41 years old Mean BMI: not reported	1994 CDC/Fukuda	25	Custom microarray (Nimblegen)	Unclear	No (NA)
[40]	Whole blood	25	Female/male adults Mean age: 43 years old Mean BMI: not reported	1994 CDC/Fukuda	50	GeneChip Human Genome U133 Plus 2.0 (Affymetrix)	Yes/Yes	No (NA)
[41]	Whole blood	11	Female/male adults Mean age: 34 years old Mean BMI: 20.3 kg/m ²	1994 CDC/Fukuda	11	Custom microarray (NA)	Yes/No	Yes (NA) ^a
[42]	Muscle biopsies	4	Female/male adults Mean age: 45/37 years old Mean BMI: not reported	1994 CDC/Fukuda	5	Operon V2.0 (CRIBI University of Padova)	Yes/Yes	No (NA)
[43]	PBMCs	8	Male adults Median age: 36 years old Mean BMI: not reported	1994 CDC/Fukuda	7	GeneChip Human Genome U133 (Affymetrix)	Yes/Yes	Yes (GSE14577)
[11]	PBMCs	37	Female/male adults Mean age: 51 years old Mean BMI: 29.4 kg/m ²	1994 CDC/Fukuda	25	MWG 20K human Array (Biotech MWG)	Yes/Yes	No (NA)
[44]	PBMCs	33	Female/male adults Mean age: not reported Mean BMI: not reported	1994 CDC/Fukuda	21	GeneChip Human Gene ST (Affymetrix)	Yes/No	No (NA)

^a Data shared as a supplementary file in the online version of the study.

Another study reviewed the medications taken by the participants [11]. However, it was unclear which medications were considered as a part of the exclusion criteria. A third study reported that healthy controls were free from any medication at the time of sampling [41].

Three additional studies using microarray technology [46, 47, 48] were excluded from our meta-analysis due to unclear or ineligible case definitions of ME/CFS. We also excluded four RNA-seq studies [49, 50, 51, 52], because of insufficient reporting on the basic quality control checks. In particular, these studies did not report the percentage of reads that could be mapped onto the reference transcriptome, the percentage of the transcriptome covered, the average number of mapped reads per transcript, the relationship between the GC content and the mapped read distribution, as recommended elsewhere [53]. More importantly, given the high sequence homology between ACE and ACE2, these studies did

not explain how their mapping algorithms dealt with reads that could be ambiguously mapped onto different locations in the transcriptome.

The selected studies were conducted in small cohorts of patients with ME/CFS (mean sample size = 18.5; range = 4–37) and healthy controls (mean sample size = 18.6; range = 5–50 individuals) (Table 2). In these studies, the patients and healthy controls were matched for age and gender. Different commercial and custom microarray technologies were used for the respective gene expression quantification. There was only one study in which the microarray did not include any probe in the genes of interest [38]. Another study used a custom array based on 9,522 genes from the RefSeq database, as available in August 2002 [39]. However, this study did not provide the list of genes included in the respective microarray. In terms of data sharing, one study made the data available in the GEO database [43] and another one within the respective publication

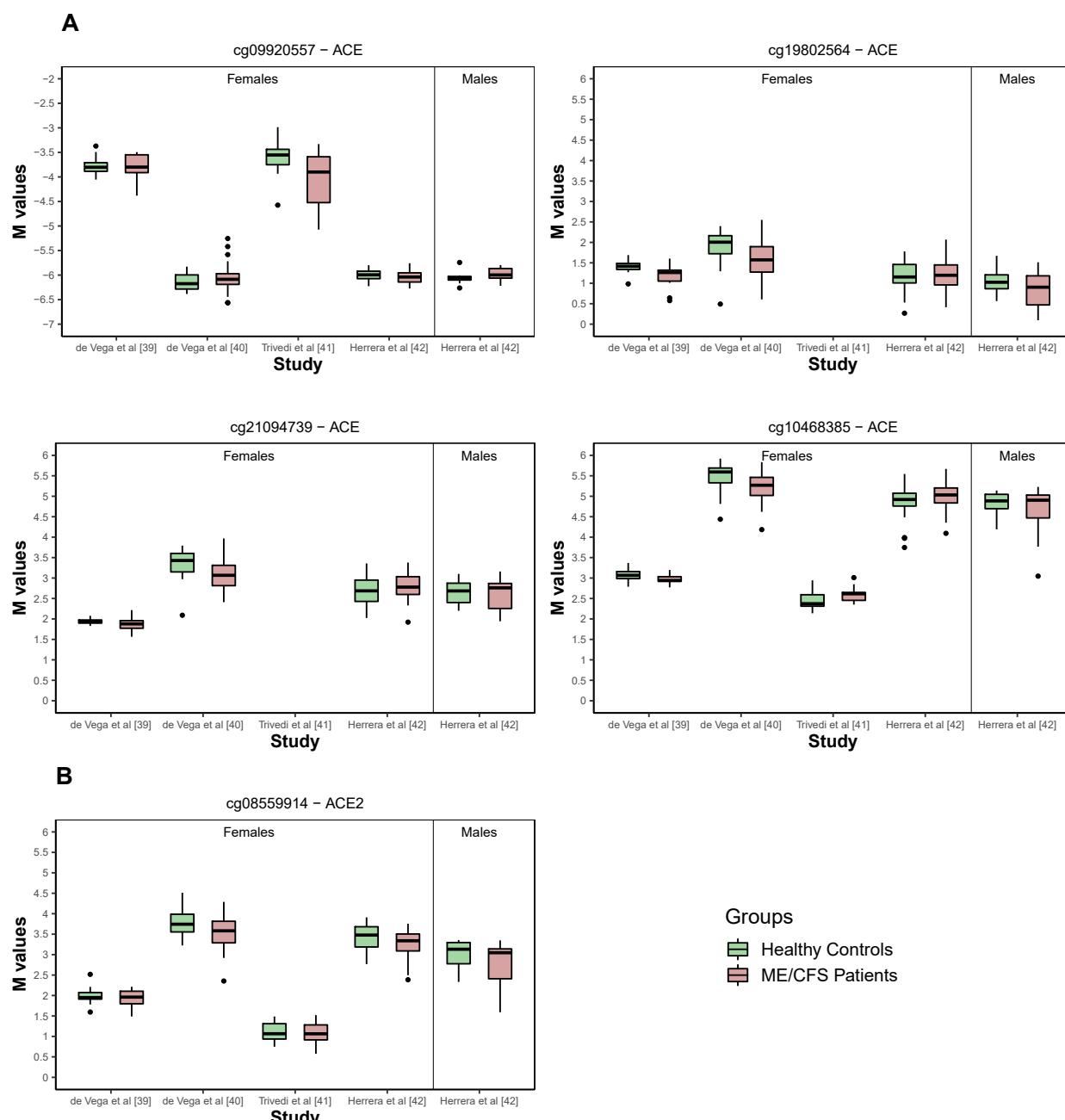


Figure 2. Boxplots per study, group and gender of the M-values referring to probes identified in Figures 1C and 1D. (A) Significant probes located in ACE. (B) Significant probe located in ACE2.

[41]. The latter study used a custom microarray that measured the expression of stress-related genes including *ACE* but excluding *ACE2*.

Before conducting a meta-analysis of the available data, we first re-analyzed two studies where the normalized data were available [41, 43]. In the first study [41], we calculated the mean of the \log_2 (fold-change) for *ACE* and the respective standard error. Note that the microarray used in this study did not include any probe in *ACE2*. In the second study [43], we initially calculated the mean and the respective standard error of the \log_2 (fold-change) for each probe located in *ACE* and *ACE2*. We then pooled each pair of means for the same gene using the inverse-variance weighting method [54]. A third study reported the mean of the \log_2 (fold-change) for *ACE2* and the respective p-value using a two-tailed Student's test [11]. In this case, we determine the quantile of the t-distribution associated with half of the reported p-value, equated it to the test statistic, and solved the resulting equation as a function of the standard error. No information was available from this study concerning the expression levels of *ACE*.

Finally, we pooled the different estimates for the same gene from different studies using the inverse-variance weighting method [54].

2.4. Analysis of new RNA data on the *ACE*/*ACE2* gene expression in ME/CFS

2.4.1. Study participants

Thirty-seven women with ME/CFS were recruited in 2020 from the outpatient clinic for immunodeficiencies at the Institute for Medical Immunology at the Charité-Universitätsmedizin Berlin, Germany. These patients were diagnosed according to the 2003 CCC while excluding other medical or neurological diseases which could explain fatigue [2]. Thirty-four women with self-reported healthy status were recruited from staff.

2.4.2. Experimental procedure for RNA isolation and expression

Consistently with previous studies of ME/CFS, the gene expression quantification was performed in PBMCs. These cells were isolated from heparinized whole blood by density gradient centrifugation using Biocoll Separating Solution (Merck Millipore). Total RNA was isolated and extracted from 2×10^6 PBMCs according to the manufacturer's instructions (NucleoSpin RNA Kit, Macherey-Nagel, cat. nr. 740955.50). Afterwards cDNA was prepared by reverse transcription (High-Capacity cDNA Reverse Transcription Kit, Applied Biosystems, cat. nr. 4368814) and real-time PCR was performed using TaqMan® Universal PCR Master Mix (cat. nr. 4305719) and TaqMan® Gene Expression Assays (cat. nr. 4331182) for *ACE* (Hs00174179_m1), *ACE2* (Hs01085333_m1) and the housekeeping gene *HPRT1* (Hs02800695_m1) (Applied Biosystems). The amplification of *ACE* and *HPRT1* was based on 20 ng template cDNA. For the amplification of *ACE2*, this quantity was increased to 100 ng. All measurements were performed with the ABI7200 and software Step One Plus as absolute quantification according to manufacturer's instruction. Relative gene expression was analysed using the ΔCT method.

2.4.3. Statistical analysis

We first tested whether patients and healthy controls were matched for age using the Kolmogorov-Smirnov test for two independent samples. For statistical convenience, gene expression values were independently transformed for *ACE* and *ACE2* using a Box-Cox transformation [55]. The parameter estimates of this transformation were 0.303 and 0.225 for *ACE* and *ACE2*, respectively. The transformed values for each gene were then analyzed as the outcome variable of a linear regression model specifying age and disease status of the participants as the respective covariates. The linear regression model was estimated using the maximum likelihood method. After estimating the models, we tested the normal distribution in the resulting residuals using the Shapiro-Wilk test. We also visually inspected the assumption of constant variance of the same residuals as a function of the covariates.

Note that we were unable to quantify the *ACE2* expression in 11 patients due to cDNA material below the limit of detection. These problematic samples could be due to a lower expression of *ACE2* in ME/CFS patients than in healthy controls. To test this hypothesis, we compared the respective proportion of samples below the limit of detection using the Pearson's χ^2 test for two-way frequency tables.

The significance level of the statistical analysis was set at 5%.

2.4.4. Ethical approval

The protocol of this study was approved by the Ethics Committee of Charité-Universitätsmedizin Berlin in accordance with the 1964 Declaration of Helsinki and its later amendments (reference number EA2/067/20). All patients and healthy controls gave written informed consent to participate in the study.

2.5. Statistical software

We performed our statistical analysis in the R software version 4.0.3. In this analysis, we used the following Bioconductor packages: *hgu133a.db*, *hgu133plus2.db*, *IlluminaHumanMethylation450kanno.ilmn12*, *hg19*, and *IlluminaHumanMethylationEPICanno.ilmn10b2.hg19* to retrieve the annotation of the GeneChip HG-U133A, GeneChip U133 + 2, Infinium HumanMethylation450K Array and HumanMethylationEPIC arrays, respectively; *minfi* to estimate the sex of each individual from DNA methylation data [37]. The R scripts are freely available from the first and last authors upon request.

3. Results

3.1. Meta-analysis of *ACE*/*ACE2* DNA methylation in ME/CFS patients

The oldest DNA methylation study [27] did not make the data available and hence, we screened the list of 120 differentially methylated probes (see table 1 from this study). Although located in 70 genes, these probes were neither located in *ACE* nor *ACE2*. We also screened the list of differentially methylated probes reported by the study based on the reduced representation bisulfite sequencing technology (see Additional File 1 from ref. [33]). Again, none of these probes was in the *ACE* or *ACE2* loci.

For the four array-based studies [28, 29, 30, 32], we conducted a joint analysis of the respective data in accordance with a meta-analysis. We first observed that the M-values of the 27 probes under investigation tended to be uncorrelated with each other (Figure 1B). This observation supported the use of the Benjamini-Hochberg procedure to adjust the raw p-values under a multiple testing scenario.

The subsequent analysis suggested four CpG probes in *ACE* to be associated with ME/CFS (Figure 1C). The probe cg09920557 belongs to the TSS region of the gene while the remaining probes (cg19802564, cg21094739, and cg10468385) are located in the gene body. The best linear regression models for each probe included both the main effects of the study indicator and of the disease status and the respective interaction term (Supplementary Table 3). The statistical interaction between these two covariates could be seen when plotting the whole data set (Figure 2A). Although not significant, the estimated main effect of the disease status was negative for each of the significantly associated probes.

Concerning the probes in *ACE2*, the only significant association with ME/CFS was obtained for cg08559914 located in the TSS region of the gene (Figure 1C). According to the best linear regression model for this probe, there was a negative association between the respective M-values and ME/CFS (coefficient estimate = -0.141 with a standard error of 0.048; Figure 2B and Supplementary Table 3). Given that a hypomethylated promoter region is typically indicative of an increased expression of the respective gene, this finding suggested an increased *ACE2* expression in patients with ME/CFS.

We then repeated the same analysis for women and men separately. For women, we obtained the same disease associations, as described

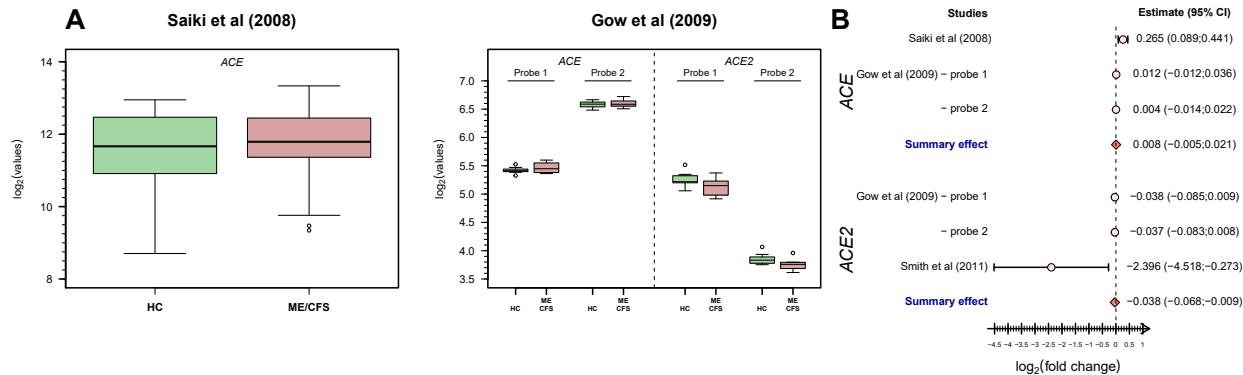


Figure 3. Analysis of ACE/ACE2-related data from eligible microarray-based gene expression studies. (A) Boxplots of the data from these studies (Saiki et al (2008), ref. [41]; Gow et al (2009); ref. [43]). (B) Forest plot for the study-specific and pooled estimate of the mean of the $\log_2(\text{fold-change})$ between patients with ME/CFS and healthy controls using data shown in A.

above (Figure 1D and Supplementary Table 3). For men, we did not find any significant associations, probably due to data from a single study [32] (Figure 1E).

3.2. Meta-analysis of ACE/ACE2 gene expression in ME/CFS patients

We first conducted a re-analysis of the two studies in which the expression levels of ACE or ACE2 were available for each participant (Figure 3A) [41,43]. In the first study [41], there was evidence for an increased expression of ACE in patients with ME/CFS (mean of the $\log_2(\text{fold-change}) = 0.265$; 95% CI=(0.089; 0.441)). In the second study [43], the means of the $\log_2(\text{fold-change})$ were estimated at 0.012 (95% CI=(-0.012; 0.036)) and 0.004 (95% CI=(-0.014; 0.022)) for the two probes in ACE. The corresponding estimates for the two probes in ACE2 were -0.038 (95% CI=(-0.085; 0.009)) and -0.037 (95% CI=(-0.083; 0.008)) (Figure 3A). The pooled estimates for this study were 0.007 (95% CI=(-0.006; 0.020)) and -0.038 (95% CI=(-0.067; -0.008)) for ACE and ACE2, respectively.

Although not sharing the data, there was a study [11] that reported a significant negative association between ME/CFS and ACE2 expression (see online Supplementary Table 2 of this study). In this case, we obtained the following mean of the $\log_2(\text{fold-change}) = -2.396$ and 95% CI=(-4.518; -0.273).

We then pooled the estimates from different studies for the same gene: 0.008 (95% CI=(-0.005; 0.021)) and -0.038 (95% CI=(-0.068; -0.009)) for ACE and ACE2, respectively (Figure 3B). Therefore, our meta-analysis suggested a reduced expression of ACE2 but not of ACE in patients with ME/CFS when comparing to healthy controls.

Finally, the remaining gene expression studies neither shared the respective data nor reported any differential ACE/ACE2 expression between patients and healthy controls.

3.3. Analysis of ACE/ACE2 gene expression from a new female cohort

To complement our findings from the above meta-analysis, we measured the ACE and ACE2 mRNA levels in PBMCs from 37 women with ME/CFS (mean age = 41.1 years old) and 34 healthy women (mean age = 37.4 years old) (Table 3). Patients and healthy participants were matched for age (Kolmogorov-Smirnov test, $p = 0.38$). There was no information about the disease duration for 4 patients. The average disease duration for the remaining patients was 5.4 months in relation to the time of diagnosis (range = 0–24 months).

We observed higher mRNA levels of ACE than of ACE2 (Table 4, Figure 4A). There was no evidence for a significant correlation between ACE and ACE2 expression levels (Spearman's correlation coefficient = -0.120) (Figure 4B). In contrast to the above meta-analysis, we could not find a reduced expression of ACE2 in patients with ME/CFS using the

complete case scenario (Table 4). However, there were 11 (29.7%) of the 37 samples from patients in which the expression level of ACE2 was below the limit of detection. This proportion of samples was significantly higher than that for healthy controls given that the expression of ACE2 could be quantified in all the samples (29.7% versus 0%; Pearson's χ^2 test, $p = 0.002$). Consequently, we could not rule out that the patients with ME/CFS from this cohort have a decreased expression of ACE2 when compared to healthy controls. Finally, in accordance with our meta-analysis, there was no evidence of differential expression of ACE between patients and healthy controls from this cohort.

4. Discussion

In this work, we investigated potential differences in ACE/ACE2 DNA methylation and expression levels between patients with ME/CFS and

Table 3. Summary statistics for the gene expression of ACE and ACE2 from the German female study participants where data of ACE2 were only available for 26 affected patients.

Summary statistic	Healthy controls	ME/CFS patients
N	34	37
Mean age (range), years	37.4 (23; 65)	41.1 (19; 60)
Mean disease duration since diagnostic (range), months	NA	5.4 (0; 24)
ACE		
Geometric mean	0.153	0.144
Interquartile range	0.087	0.073
ACE2		
Geometric mean	0.002	0.001
Interquartile range	0.005	0.004

Table 4. Analysis of the linear regression models for the Box-Cox-transformed ACE and ACE2 mRNA levels where data were only available for 26 ME/CFS patients.

Analysis	Estimate (SE)	P-value
Box-Cox transformed ACE		
(Intercept)	0.541 (0.032)	<0.001
Age	0.001 (0.001)	0.328
Disease Status (ME/CFS)	-0.013 (0.018)	0.481
Box-Cox transformed ACE2		
(Intercept)	0.307 (0.038)	<0.001
Age	-0.001 (0.001)	0.137
Disease Status (ME/CFS)	-0.006 (0.021)	0.789

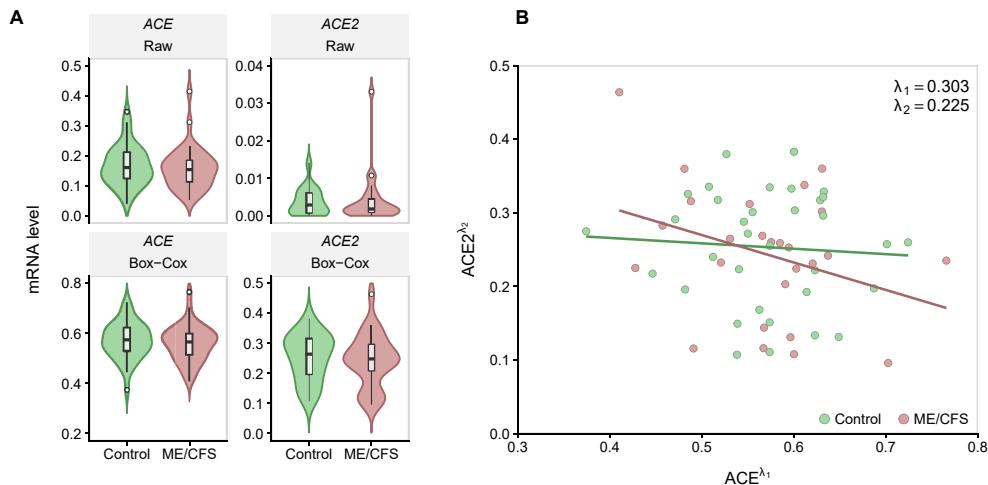


Figure 4. Analysis of *ACE* and *ACE2* expression levels from the German study. (A) Violin plots of *ACE* (left side) and *ACE2* (right side) mRNA raw data (upper row) and transformed data using a Box-Cox transformation (lower row). (B) Scatterplot between the transformed *ACE* and *ACE2* expression levels (Spearman's correlation coefficient = -0.120).

healthy controls. With the identification of these differences, we expected to determine the health risk of patients with ME/CFS if infected by SARS-CoV-2. However, we stumbled upon hurdles related to (i) data unavailability for a possible re-analysis, (ii) availability of data derived from PBMCs and related subsets in which *ACE2* is not particularly expressed, (iii) studies with unclear data quality, and (iv) studies using disease case definitions that are not recommended for research. As a consequence, we could not provide a more definite answer to our main research question.

Notwithstanding these difficulties, we could identify four CpG probes on *ACE* and another one on *ACE2* with decreased DNA methylation levels in patients with ME/CFS. This finding suggested an increased expression of the respective genes. However, our meta-analysis of public data suggested the opposite. Such decrease in *ACE2* expression was partially confirmed by new data in which there was a significant higher proportion of samples below the limit of detection in patients with ME/CFS than in healthy controls. Nonetheless, it was clear that *ACE2* is not particularly expressed in PBMCs from both patients with ME/CFS and healthy controls, as mentioned in the introduction.

In general, *ACE2* downregulation is known to occur after host-cell entry by SARS-CoV-2 [56]. This downregulation is particularly problematic in individuals affected by cardiovascular diseases, diabetes, and other medical conditions, due to their low *ACE2* levels before the infection [57]. SARS-CoV-2 infection is then expected to further increase the *ACE*:*ACE2* ratio, thus, promoting vasoconstriction, increased production of ROS and inflammation in patients with these co-morbidities [23]. In this scenario, a putative reduction of the *ACE2* expression makes patients with ME/CFS similar to these patients with a high risk for COVID-19. As a consequence, patients with ME/CFS could be considered a priority group for vaccination by public health authorities. The fundamental question is then to know whether our findings based on PBMCs could recreate what occurs in pulmonary epithelial and endothelial cells, the main targets of SARS-CoV-2. Future research should be conducted to answer this question, as similarly done in past studies aiming at understanding how the gene expression profiles from PBMCs could mimick those present in other tissues affecting by a given disease [58, 59, 60].

Given the residual *ACE2* expression in PBMCs under normal conditions, one is tempted to say that SARS-CoV-2 does not infect these cells. However, earlier studies on SARS-CoV-1 found this virus within T lymphocytes, macrophages, and dendritic cells [61]. More recently, an *in vitro* study was able to infect PBMCs with SARS-CoV-2 [62]. Monocytes are particularly susceptible to such infections. In this context, one cannot

rule out that SARS-CoV-2 might use alternative receptors when infecting PBMCs.

Among the alternative receptors for SARS-CoV-2, the human transmembrane protease serine 2 (TMPRSS2) was suggested as a strong candidate [63] due to its role on SARS-CoV-1 infection [64, 65]. This protease seems to induce SARS-CoV-2 cell entry through endocytosis via a mechanism of *ACE2* cleavage [14]. Another candidate receptor is the A disintegrin and metalloproteinase domain 17 protein (ADAM17) recognized by the immune system as a stress-response signal [66]. Like TMPRSS2, ADAM17 can also cleave *ACE2* but with a reduced viral invasion efficiency [67].

With respect to the role of these proteases in ME/CFS, a targeted gene expression study analyzed ADAM17 and other stress-response proteins [41]. This study did not report any differential expression of this protease between patients with ME/CFS and healthy controls. However, this study is likely to be affected by a low statistical power due to small sample sizes for both groups. In addition, one of the selected DNA methylation studies suggested a decrease in the DNA methylation levels of one ADAM17-related CpG probe in patients with ME/CFS [30].

Dipeptidyl peptidase-4 (DPP4), also known as the lymphocyte cell surface protein CD26, was found to be the main receptor for the Middle East respiratory syndrome-related coronavirus [68, 69]. In contrast to *ACE2*, this surface protein is highly abundant in PBMCs including CD4+ and CD8+ T cells [18]. Bioinformatic analysis also suggested a strong interaction potential between this protein and SARS-CoV-2 [70, 71]. Finally, DPP4 inhibitors were found to be protective against severe COVID-19 in patients with diabetes mellitus when compared to RAAS blockers [72]. After initial concerns, this finding combined with others suggested an interesting therapeutic avenue against COVID-19 using DPP4 blockers [73].

Interestingly, there is evidence for an increased proportion of natural killer cells and T cells expressing DPP4/CD26+ in patients with ME/CFS [7, 74]. However, the number of DPP4/CD26 molecules was significantly reduced in T lymphocytes and natural killer cells of these patients [74]. If DPP4 is indeed a relevant receptor for immune-cell invasion by SARS-CoV-2, research about this receptor should be prioritized when analyzing PBMCs from patients with ME/CFS.

Sialic acids were also hypothesized as binding receptors used by SARS-CoV-2, as reported for other human coronaviruses [75]. These acids are highly expressed in the epithelium cells of the lungs and oral cavity [76]. *In vitro* and *in silico* studies demonstrated the same binding potential for SARS-CoV-2 [77, 78, 79]. However, the *ACE2* glycosylation inhibition studies suggested that sialic acids on *ACE2* receptor prevent

ACE2–virus interaction [80, 81]. Again, detailed research on these putative receptors could help to determine the health risk of patients with ME/CFS when infected by SARS-CoV-2.

It was suggested that the arousal state experienced by patients with ME/CFS protects them against microbial infections [82]. This suggestion came from a clinical trial where patients were treated with clonidine to decrease such a state. Treated patients got their symptoms worsened and had their inflammation markers increased during the trial. In contrast, basic epidemiological studies reported many patients with frequent viral infections and flu-like symptoms [3, 4, 83]. The question is how an infection by SARS-CoV-2 lies in this contrasting evidence. A possible answer can be given with the assistance of the so-called sustained arousal model of ME/CFS [84]. According to this model, a sustained arousal state promotes in the long-run deleterious alterations of different body systems, including the immune system. Similar prediction was made by a recent study discussing the natural history of ME/CFS [85]. If so, patients with longer disease durations are more likely to show these immunological alterations than patients at the early stages of the disease. However, we could not analyze the effect of disease duration on our results, because this variable was not available in the public data sets included in our meta-analyses.

Finally, our original idea was also to include a meta-analysis of ACE/ACE2 data from published genome-wide association studies on ME/CFS [11, 32, 86, 87, 88]. However, we could not materialize this idea, because such studies did not make their data publicly available. Nevertheless, evidence is scarce for a putative role of ACE/ACE2 polymorphisms on ME/CFS. Two studies reported many candidate SNPs for such association, but none was located in ACE or ACE2 [11, 86]. Two other studies did not find any significant SNPs associated with ME/CFS [32, 88]. The most optimistic study reported thousands of SNPs related to the disease [87]. However, this study did not perform all the basic quality control checks [89].

5. Conclusions

Notwithstanding the low expression of ACE2 in PBMCs in general, there is evidence for a decreased expression of the gene in these cells from patients with ME/CFS. If PBMCs can qualitatively recreate what is occurring in the main cellular targets of SARS-CoV-2, then patients with this disease could be at a higher COVID-19 risk. In this regard, a recent preliminary report suggested that patients with ME/CFS got their symptoms worsened upon SARS-CoV-2 infection [91]. Altogether, these patients could be considered a priority group for vaccination against COVID-19, even though vaccines could trigger ME/CFS [92, 93] or even exacerbate ME/CFS symptoms as the case of the natural immunization by SARS-CoV-2. To further consolidate the existing evidence, future research should prioritize the collection of data from the main cellular targets in patients with ME/CFS. Further investigation should be also conducted on alternative SARS-CoV-2 receptors (i.e., DPP4 and sialic acids). At last, future research should also consider investigating putative sex differences in patients with ME/CFS given that, in general, men are more affected by COVID-19 than women [90].

Declarations

Author contribution statement

João Malato, André Fonseca, Anna D Grabowska, Luís Graça, Clara Cordeiro, Luís Nacul, Eliana M Lacerda, Jesus Castro-Marrero, Francisco Westermeier: Analyzed and interpreted the data; Wrote the paper.

Franziska Sotzny: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Sandra Bauer, Helma Freitag: Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Carmen Scheibenbogen: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Nuno Sepúlveda: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

João Malato and André Fonseca were fully funded by FCT – Fundação para a Ciência e Tecnologia, Portugal (ref.grant: SFRH/BD/149758/2019 and SFRH/BD/147629/2019, respectively). Nuno Sepúlveda and Clara Cordeiro were partially funded by FCT – Fundação para a Ciência e a Tecnologia, Portugal (ref. grant: UIDB/00006/2020). Luís Nacul and Eliana M Lacerda acknowledge the funding from the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH -Award Number: R01AI103629), and from the ME Association (Award number: PF8947) for their studies on ME/CFS.

Data availability statement

Data are available from the GEO database under the accession number GSE59489, GSE93266, GSE111183, GSE156792, GSE153667, GSE14577.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2021.e07665>.

References

- [1] K. Fukuda, S.E. Straus, I. Hickie, M.C. Sharpe, J.G. Dobbins, A. Komaroff, The chronic fatigue syndrome: a comprehensive approach to its definition and study, *Ann. Intern. Med.* 121 (1994) 953–959.
- [2] B.M. Carruthers, A.K. Jain, K.L. De Meirlier, D.L. Peterson, N.G. Klimas, A.M. Lemer, A.C. Bested, P. Flor-Henry, P. Joshi, A.C.P. Powles, J.A. Sherkey, M.I. Van de Sande, Myalgic encephalomyelitis/chronic fatigue syndrome: clinical working case definition, diagnostic and treatment protocols, *J. Chronic Fatigue Syndrome* 11 (2003) 7–115.
- [3] S.C. Johnston, D.R. Staines, S.M. Marshall-Gradisnik, Epidemiological characteristics of chronic fatigue syndrome/myalgic encephalomyelitis in Australian patients, *Clin. Epidemiol.* 8 (2016) 97–107.
- [4] L. Chu, I.J. Valencia, D.W. Garvert, J.G. Montoya, Onset patterns and course of myalgic encephalomyelitis/chronic fatigue syndrome, *Front. Pediatr.* 7 (2019) 12.
- [5] S. Rasa, Z. Nora-Krukle, N. Henning, E. Eliassen, E. Shikova, T. Harrer, C. Scheibenbogen, M. Murovska, B.K. Prusty, Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), *J. Transl. Med.* 16 (2018), 268.
- [6] M.E. Ariza, Myalgic encephalomyelitis/chronic fatigue syndrome: the human herpesviruses are back!, *Biomolecules* 11 (2021) 1–17.
- [7] N.G. Klimas, F.R. Salvato, R. Morgan, M.A. Fletcher, Immunologic abnormalities in chronic fatigue syndrome, *J. Clin. Microbiol.* 28 (1990) 1403–1410.
- [8] L. Lorusso, S.V. Mikhaylova, E. Capelli, D. Ferrari, G.K. Ngonga, G. Ricevuti, Immunological aspects of chronic fatigue syndrome, *Autoimmun. Rev.* 8 (2009) 287–291.
- [9] E.W. Brenu, M.L. van Driel, D.R. Staines, K.J. Ashton, S.B. Ramos, J. Keane, N.G. Klimas, S.M. Marshall-Gradisnik, Immunological abnormalities as potential biomarkers in chronic fatigue syndrome/myalgic encephalomyelitis, *J. Transl. Med.* 9 (2011), 81.
- [10] J.R. Kerr, Gene profiling of patients with chronic fatigue syndrome/myalgic encephalomyelitis, *Curr. Rheumatol. Rep.* 10 (2008) 482–491.
- [11] A.K. Smith, H. Fang, T. Whistler, E.R. Unger, M.S. Rajeevan, Convergent genomic studies identify association of GRIK2 and NPAS2 with chronic fatigue syndrome, *Neuropsychobiology* 64 (2011) 183–194.
- [12] W. Li, M.J. Moore, N. Vasiliu, J. Sui, S.K. Wong, M.A. Berne, M. Somasundaran, J.L. Sullivan, K. Luzuriaga, T.C. Greeneugh, H. Choe, M. Farzan, Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus, *Nature* 426 (2003) 450–454.
- [13] X.Y. Ge, J.L. Li, X. Lou Yang, A.A. Chmura, G. Zhu, J.H. Epstein, J.K. Mazet, B. Hu, W. Zhang, C. Peng, Y.J. Zhang, C.M. Luo, B. Tan, N. Wang, Y. Zhu, G. Cramer, S.Y. Zhang, L.F. Wang, P. Daszak, Z.L. Shi, Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor, *Nature* 503 (2013) 535–538.

- [14] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.H. Wu, A. Nitsche, M.A. Müller, C. Drosten, S. Pöhlmann, SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor, *Cell* 181 (2020) 271–280.e8.
- [15] I. Hamming, W. Timens, M.L.C. Bulthuis, A.T. Lely, G.J. Navis, H. van Goor, Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis, *J. Pathol.* 203 (2004) 631–637.
- [16] K.F. To, A.W.I. Lo, Exploring the pathogenesis of severe acute respiratory syndrome (SARS): the tissue distribution of the coronavirus (SARS-CoV) and its putative receptor, angiotensin-converting enzyme 2 (ACE2), *J. Pathol.* 203 (2004) 740–743.
- [17] M.Y. Li, L. Li, Y. Zhang, X.S. Wang, Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues, *Infect. Dis. Poverty* 9 (2020), 45.
- [18] U. Radzikowska, M. Ding, G. Tan, D. Zhalparov, Y. Peng, P. Wawrzyniak, M. Wang, S. Li, H. Morita, C. Altunbulakli, M. Reiger, A.U. Neumann, N. Lunjani, C. Traidl-Hoffmann, K.C. Nadeau, L. O'Mahony, C. Akdis, M. Sokolowska, Distribution of ACE2, CD147, CD26, and other SARS-CoV-2 associated molecules in tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and COVID-19 risk factors, *Allergy* 75 (2020) 2829–2845.
- [19] M. Rutkowska-Zapala, M. Suski, R. Szatanek, M. Lenart, K. Weglarczyk, R. Olszanecki, T. Grodzicki, M. Strach, J. Gasowski, M. Siedlar, Human monocyte subsets exhibit divergent angiotensin I-converting activity, *Clin. Exp. Immunol.* 181 (2015) 126–132.
- [20] X. Song, W. Hu, H. Yu, L. Zhao, Y. Zhao, X. Zhao, H.H. Xue, Y. Zhao, Little to no expression of angiotensin-converting enzyme-2 on most human peripheral blood immune cells but highly expressed on tissue macrophages, *Cytometry* (2020) 1–10.
- [21] S.R. Tipnis, N.M. Hooper, R. Hyde, E. Karran, G. Christie, A.J. Turner, A human homolog of angiotensin-converting enzyme: cloning and functional expression as a captopril-insensitive carboxypeptidase, *J. Biol. Chem.* 275 (2000) 33238–33243.
- [22] F. Westermeier, M. Bustamante, M. Pavelz, L. García, M. Chiong, M.P. Ocaranza, S. Lavandero, Novel players in cardioprotection: insulin like growth factor-1, angiotensin-(1-7) and angiotensin-(1-9), *Pharmacol. Res.* 101 (2015) 41–55.
- [23] P. Pagliaro, C. Penna, ACE/ACE2 ratio: a key also in 2019 coronavirus disease (Covid-19)? *Front. Med.* 7 (2020) 335.
- [24] A.H. Sawalha, M. Zhao, P. Coit, Q. Lu, Epigenetic dysregulation of ACE2 and interferon-regulated genes might suggest increased COVID-19 susceptibility and severity in lupus patients, *Clin. Immunol.* 215 (2020) 108410.
- [25] J. Blomberg, C.G. Gottfries, A. Elftaoui, M. Rizwan, A. Rosén, Infection elicited autoimmunity and Myalgic encephalomyelitis/chronic fatigue syndrome: an explanatory model, *Front. Immunol.* 9 (2018), 229.
- [26] D.F.H. Pheby, D. Araja, U. Berkis, E. Brenna, J. Cullinan, J.-D. de Korwin, L. Gitto, D.A. Hughes, R.M. Hunter, D. Trepel, X. Wang-Steverding, The development of a consistent Europe-wide approach to investigating the economic impact of myalgic encephalomyelitis (ME/CFS): a report from the European Network on ME/CFS (EUROMENE), *Healthcare* 8 (2020) 88.
- [27] E.W. Bremu, D.R. Staines, S.M. Marshall-Gradisnik, Methylation profile of CD4+ T cells in chronic fatigue syndrome/myalgic encephalomyelitis, *J. Clin. Cell. Immunol.* 5 (2014) 1–14.
- [28] W.C. De Vega, S.D. Vernon, P.O. McGowan, DNA methylation modifications associated with Chronic Fatigue Syndrome, *PLoS One* 9 (2014), e104757.
- [29] W.C. De Vega, S. Herrera, S.D. Vernon, P.O. McGowan, Epigenetic modifications and glucocorticoid sensitivity in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), *BMC Med. Genomics* 10 (2017), 11.
- [30] M.S. Trivedi, E. Oltra, L. Sarria, N. Rose, V. Beljanski, M.A. Fletcher, N.G. Klimas, L. Nathanson, Identification of myalgic encephalomyelitis/chronic fatigue syndrome-associated DNA methylation patterns, *PLoS One* 13 (2018), e0201066.
- [31] E. Almenar-Pérez, T. Ovejero, T. Sánchez-Fito, J.A. Espejo, L. Nathanson, E. Oltra, Epigenetic components of myalgic encephalomyelitis/chronic fatigue syndrome uncover potential transposable element activation, *Clin. Ther* 41 (2019) 675–698.
- [32] S. Herrera, W.C. de Vega, D. Ashbrook, S.D. Vernon, P.O. McGowan, Genome-epigenome interactions associated with myalgic encephalomyelitis/chronic fatigue syndrome, *Epigenetics* 13 (2018) 1174–1190.
- [33] A.M. Helliwell, E.C. Sweetman, P.A. Stockwell, C.D. Edgar, A. Chatterjee, W.P. Tate, Changes in DNA methylation profiles of myalgic encephalomyelitis/chronic fatigue syndrome patients reflect systemic dysfunctions, *Clin. Epigenet.* 12 (2020) 167.
- [34] S. Dedeurwaerder, M. Defrance, M. Bizez, E. Calonne, G. Bontempi, F.F. Fuks, A comprehensive overview of Infinium HumanMethylation 450 data processing, *Brief. Bioinform.* 15 (2013) 929–941.
- [35] P. Du, X. Zhang, C.C. Huang, N. Jafari, W.A. Kibbe, L. Hou, S.M. Lin, Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis, *BMC Bioinf.* 11 (2010), 587.
- [36] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B* 57 (1995) 289–300.
- [37] M.J. Aryee, A.E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A.P. Feinberg, K.D. Hansen, R.A. Irizarry, Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays, *Bioinformatics* 30 (2014) 1363–1369.
- [38] T. Whistler, J.F. Jones, E.R. Unger, S.D. Vernon, Exercise responsive genes measured in peripheral blood of women with Chronic Fatigue Syndrome and matched control subjects, *BMC Physiol.* 5 (2005), 5.
- [39] N. Kaushik, D. Fear, S.C.M. Richards, C.R. McDermott, E.F. Nuwaysir, P. Kellam, T.J. Harrison, R.J. Wilkinson, D.A.J. Tyrrell, S.T. Holgate, J.R. Kerr, Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome, *J. Clin. Pathol.* 58 (2005) 826–832.
- [40] J.R. Kerr, R. Petty, B. Burke, J. Gough, D. Fear, L.I. Sinclair, D.L. Mattey, S.C.M. Richards, J. Montgomery, D.A. Baldwin, P. Kellam, T.J. Harrison, G.E. Griffin, J. Main, D. Enlander, D.J. Nutt, S.T. Holgate, Gene expression subtypes in patients with chronic fatigue syndrome/myalgic encephalomyelitis, *J. Infect. Dis.* 197 (2008) 1171–1184.
- [41] T. Saiki, T. Kawai, K. Morita, M. Ohta, T. Saito, K. Rokutan, N. Ban, Identification of marker genes for differential diagnosis of chronic fatigue syndrome, *Mol. Med.* 14 (2008) 599–607.
- [42] T. Pietrangelo, R. Mancinelli, L. Toniolo, G. Montanari, J. Vecchiet, G. Fanò, S. Fulle, Transcription profile analysis of vastus lateralis muscle from patients with chronic fatigue syndrome, *Int. J. Immunopathol. Pharmacol.* 22 (2009) 795–807.
- [43] J.W. Gow, S. Hagan, P. Herzyk, C. Cannon, P.O. Behan, A. Chaudhuri, A gene signature for post-infectious chronic fatigue syndrome, *BMC Med. Genomics* 2 (2009), 38.
- [44] M.G. Jeffrey, L. Nathanson, K. Aenlle, Z.M. Barnes, M. Baig, G. Broderick, N.G. Klimas, M.A. Fletcher, T.J.A. Craddock, Treatment avenues in myalgic encephalomyelitis/chronic fatigue syndrome: a split-gender pharmacogenomic study of gene-expression modules, *Clin. Ther.* 41 (2019) 815–835.e6.
- [45] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton, T. Gaasterland, P. Glenisson, F.C.P. Holstege, I.F. Kim, V. Markowitz, J.C. Mateescu, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, M. Vingron, Minimum information about a microarray experiment (MIAME) - toward standards for microarray data, *Nat. Genet.* 29 (2001) 365–371.
- [46] S. Galbraith, B. Cameron, H. Li, D. Lau, U. Vollmer-Conna, A.R. Lloyd, Peripheral blood gene expression in postinfective fatigue syndrome following from three different triggering infections, *J. Infect. Dis.* 204 (2011) 1632–1640.
- [47] S.D. Vernon, E.R. Unger, I.M. Dimulescu, M. Rajeevan, W.C. Reeves, Utility of the blood for gene expression profiling and biomarker discovery in chronic fatigue syndrome, *Dis. Markers* 18 (2002) 193–199.
- [48] C.B. Nguyen, L. Alsøe, J.M. Lindvall, D. Sulheim, E. Fagermoen, A. Winger, M. Kaarbo, H. Nilsen, V.B. Wyller, Whole blood gene expression in adolescent chronic fatigue syndrome: an exploratory cross-sectional study suggesting altered B cell differentiation and survival, *J. Transl. Med.* 15 (2017), 102.
- [49] J. Bouquet, J.L. Gardy, S. Brown, J. Pfeil, R.R. Miller, M. Morshed, A. Avina-Zubieta, K. Shojania, M. McCabe, S. Parker, M. Uyaguri, S. Federman, P. Tang, T. Steiner, M. Otterstater, R. Holt, R. Moore, C.Y. Chiu, D.M. Patrick, RNA-seq analysis of gene expression, viral pathogen, and B-cell/T-cell receptor signatures in complex chronic disease, *Clin. Infect. Dis.* 64 (2017) 476–481.
- [50] J. Bouquet, T. Li, J.L. Gardy, X. Kang, S. Stevens, J. Stevens, M. VanNess, C. Snell, J. Potts, R.R. Miller, M. Morshed, M. McCabe, S. Parker, M. Uyaguri, P. Tang, T. Steiner, W.S. Chan, A.M. De Souza, A. Mattman, D.M. Patrick, C.Y. Chiu, Whole blood human transcriptome and virome analysis of ME/CFS patients experiencing post-exertional malaise following cardiopulmonary exercise testing, *PloS One* 14 (2019), e0212193.
- [51] E. Sweetman, M. Ryan, C. Edgar, A. Mackay, R. Vallings, W. Tate, Changes in the transcriptome of circulating immune cells of a New Zealand cohort with myalgic encephalomyelitis/chronic fatigue syndrome, *Int. J. Immunopathol. Pharmacol.* 33 (2019) 1–8.
- [52] R.P.H. Rajmakers, A.F.M. Jansen, S.P. Keijmel, R. Ter Horst, M.E. Roerink, B. Novakovic, L.A.B. Joosten, J.W.M. Van Der Meer, M.G. Netea, C.P. Bleeker-Rovers, A possible role for mitochondrial-derived peptides humanin and MOTS-c in patients with Q fever fatigue syndrome and chronic fatigue syndrome, *J. Transl. Med.* 17 (2019), 157.
- [53] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M.W. Szcześniak, D.J. Gaffney, L.L. Elo, X. Zhang, A. Mortazavi, A survey of best practices for RNA-seq data analysis, *Genome Biol.* 17 (2016), 13.
- [54] J. Hartung, G. Knapp, B.K. Sinha, *Statistical Meta-Analysis with Applications*, first ed., John Wiley & Sons, New Jersey, 2008.
- [55] Ö. Asar, O. Ilk, O. Dag, Estimating Box-Cox power transformation parameter via goodness-of-fit tests, *Commun. Stat. Simul. Comput.* 46 (2017) 91–105.
- [56] P.K. Datta, F. Liu, T. Fischer, J. Rappaport, X. Qin, SARS-CoV-2 pandemic and research gaps: understanding SARS-CoV-2 interaction with the ACE2 receptor and implications for therapy, *Theranostics* 10 (2020) 7448–7464.
- [57] P. Verdecchia, C. Cavallini, A. Spanevello, F. Angeli, The pivotal link between ACE2 deficiency and SARS-CoV-2 infection, *Eur. J. Intern. Med.* 76 (2020) 14–20.
- [58] I.C. Gerling, R.A. Ahokas, G. Kamalov, W. Zhao, S.K. Bhattacharya, Y. Sun, K.T. Weber, Gene expression profiles of peripheral blood mononuclear cells reveal transcriptional signatures as novel biomarkers of cardiac remodeling in rats with aldosteronism and hypertensive heart disease, *JACC Hear. Fail.* 1 (2013) 469–476.
- [59] T. Takamura, M. Honda, Y. Sakai, H. Ando, A. Shimizu, T. Ota, M. Sakurai, H. Misu, S. Kurita, N. Matsuzawa-Nagata, M. Uchikata, S. Nakamura, R. Matoba, M. Tamino, Kichi Matsubara, S. Kaneko, Gene expression profiles in peripheral blood mononuclear cells reflect the pathophysiology of type 2 diabetes, *Biochem. Biophys. Res. Commun.* 361 (2007) 379–384.
- [60] F.S. Manoel-Caetano, D.J. Xavier, A.F. Evangelista, P. Takahashi, C.V. Collares, D. Puthier, M.C. Foss-Freitas, M.C. Foss, E.A. Donadi, G.A. Passos, E.T. Sakamoto-Hojo, Gene expression profiles displayed by peripheral blood mononuclear cells from patients with type 2 diabetes mellitus focusing on biological processes implicated on the pathogenesis of the disease, *Gene* 511 (2012) 151–160.
- [61] M.Z. Tay, C.M. Poh, L. Rénia, P.A. Macary, L.F.P. Ng, The trinity of COVID-19: immunity, inflammation and intervention, *Nat. Rev. Immunol.* 20 (2020) 363–374.
- [62] A.C. Codo, G.G. Davanzo, L.B. de Monteiro, G.F. de Souza, S.P. Muraro,

- J.V. Virgilio-da-Silva, J.S. Prodonoff, V.C. Carregari, C.A.O. de Biagi Junior, F. Crunfli, J.L. Jimenez Restrepo, P.H. Vendramini, G. Reis-de-Oliveira, K. Bispo dos Santos, D.A. Toledo-Teixeira, P.L. Parise, M.C. Martini, R.E. Marques, H.R. Carmo, A. Borin, L.D. Coimbra, V.O. Boldrini, N.S. Brunetti, A.S. Vieira, E. Mansour, R.G. Ulaf, A.F. Bernardes, T.A. Nunes, L.C. Ribeiro, A.C. Palma, M.V. Agrela, M.L. Moretti, A.C. Sposito, F.B. Pereira, L.A. Velloso, M.A.R. Vinolo, A. Damasio, J.L. Proenca-Módena, R.F. Carvalho, M.A. Mori, D. Martins-de-Souza, H.I. Nakaya, A.S. Farias, P.M. Moraes-Vieira, Elevated glucose levels favor SARS-CoV-2 infection and monocyte response through a HIF-1 α /Glycolysis-Dependent Axis, *Cell Metabol.* 32 (2020) 437–446.e5.
- [63] W. Sungnak, N. Huang, C. Bécavin, M. Berg, R. Queen, M. Litvinukova, C. Talavera-López, H. Maatz, D. Reichart, F. Sampaziotis, K.B. Worlock, M. Yoshida, J.L. Barnes, N.E. Banovich, P. Barbry, A. Brahma, J. Collin, T.J. Desai, T.E. Duong, O. Eickelberg, C. Falk, M. Farzan, I. Glass, R.K. Gupta, M. Haniffa, P. Horvath, N. Hubner, D. Hung, N. Kaminski, M. Krasnow, J.A. Kropski, M. Kuhnemund, M. Lako, H. Lee, S. Leroy, S. Linnarson, J. Lundberg, K.B. Meyer, Z. Miao, A.V. Misharin, M.C. Nawijn, M.Z. Nikolic, M. Noseda, J. Ordovas-Montanes, G.Y. Oudit, D. Pe'er, J. Powell, S. Quake, J. Rajagopal, P.R. Tata, E.L. Rawlins, A. Regev, P.A. Reyfman, O. Rozenblatt-Rosen, K. Saeb-Parsy, C. Samakovlis, H.B. Schiller, J.L. Schultz, M.A. Seibold, C.E. Seidman, J.G. Seidman, A.K. Shalek, D. Shepherd, J. Spence, A. Spira, X. Sun, S.A. Teichmann, F.J. Theis, A.M. Tsankov, L. Vallier, M. van den Berge, J. Whitsett, R. Xavier, Y. Xu, L.E. Zaragozi, D. Zerti, H. Zhang, K. Zhang, M. Rojas, F. Figueiredo, SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes, *Nat. Med.* 26 (2020) 681–687.
- [64] I. Glowacka, S. Bertram, M.A. Muller, P. Allen, E. Soilleux, S. Pfefferle, I. Steffen, T.S. Tsegaye, Y. He, K. Giriss, D. Niemeyer, H. Schneider, C. Drosten, S. Pohlmann, Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response, *J. Virol.* 85 (2011) 4122–4134.
- [65] S. Matsuyama, N. Nagata, K. Shirato, M. Kawase, M. Takeda, F. Taguchi, Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2, *J. Virol.* 84 (2010) 12658–12664.
- [66] S. Düsterhöft, J. Lokau, C. Garbers, The metalloprotease ADAM17 in inflammation and cancer, *Pathol. Res. Pract.* 215 (2019), 152410.
- [67] A. Heurich, H. Hofmann-Winkler, S. Gierer, T. Liepold, O. Jahn, S. Pohlmann, TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein, *J. Virol.* 88 (2014) 1293–1307.
- [68] N. van Doremalen, K.L. Miazgowicz, S. Milne-Price, T. Bushmaker, S. Robertson, D. Scott, J. Kinne, J.S. McLellan, J. Zhu, V.J. Munster, Host species restriction of Middle East respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4, *J. Virol.* 88 (2014) 9220–9232.
- [69] W. Widagdo, N.M.A. Okba, W. Li, A. de Jong, R.L. de Swart, L. Begeman, J.M.A. van den Brand, B.-J. Bosch, B.L. Haagmans, Species-specific colocalization of Middle East respiratory syndrome coronavirus attachment and entry receptors, *J. Virol.* 93 (2019) e00107–19.
- [70] Y. Li, Z. Zhang, L. Yang, X. Lian, Y. Xie, S. Li, S. Xin, P. Cao, J. Lu, The MERS-CoV receptor DPP4 as a candidate binding target of the SARS-CoV-2 spike, *IScience* 23 (2020), 101160.
- [71] N. Vankadari, J.A. Wilce, Emerging WuHan (COVID-19) coronavirus: glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26, *Emerg. Microb. Infect.* 9 (2020) 601–604.
- [72] S.Y. Rhee, J. Lee, H. Nam, D.S. Kyung, D.W. Shin, D.J. Kim, Effects of a DPP-4 inhibitor and RAS blockade on clinical outcomes of patients with diabetes and COVID-19, *Diabetes Metab. J.* 45 (2021) 251–259.
- [73] A.J. Scheen, DPP-4 inhibition and COVID-19: from initial concerns to recent expectations, *Diabetes Metab.* 47 (2021), 101213.
- [74] M.A. Fletcher, X.R. Zeng, K. Maher, S. Lewis, B. Hurwitz, M. Antoni, G. Broderick, N.G. Klimas, Biomarkers in chronic fatigue syndrome: evaluation of natural killer cell function and dipeptidyl peptidase IV/CD26, *PloS One* 5 (2010), e10817.
- [75] X.-L. Sun, The role of cell surface sialic acids for SARS-CoV-2 infection, *Glycobiology* (2021), cwab032.
- [76] B.W. Cross, S. Ruhl, Glycan recognition at the saliva – oral microbiome interface, *Cell. Immunol.* 333 (2018) 19–33.
- [77] E. Milanetti, M. Miotti, L. Di Renzo, M. Monti, G. Gosti, G. Ruocco, In-Silico evidence for two receptors based strategy of SARS-CoV-2, *BioRxiv* (2020) 6197, 2020.03.24.
- [78] M. Awasthi, S. Gulati, D.P. Sarkar, S. Tiwari, S. Kateriya, P. Ranjan, S.K. Verma, The sialoside-binding pocket of SARS-CoV-2 spike glycoprotein structurally resembles MERS-CoV, *Viruses* 12 (2020), 909.
- [79] A.N. Baker, S.J. Richards, C.S. Guy, T.R. Congdon, M. Hasan, A.J. Zwetsloot, A. Gallo, J.R. Lewandowski, P.J. Stansfeld, A. Straube, M. Walker, S. Chessa, G. Pergolizzi, S. Dedola, R.A. Field, M.I. Gibson, The SARS-CoV-2 spike protein binds sialic acids and enables rapid detection in a lateral flow point of care diagnostic device, *ACS Cent. Sci.* 6 (2020) 2046–2052.
- [80] H. Chu, B. Hu, X. Huang, Y. Chai, D. Zhou, Y. Wang, H. Shuai, D. Yang, Y. Hou, X. Zhang, T.T. Yuen, J.P. Cai, A.J. Zhang, J. Zhou, S. Yuan, K.K.W. To, I.H.Y. Chan, K.Y. Sit, D.C.C. Foo, I.Y.H. Wong, A.T.L. Ng, T.T. Cheung, S.Y.K. Law, W.K. Au, M.A. Brindley, Z. Chen, K.H. Kok, J.F.W. Chan, K.Y. Yuen, Host and viral determinants for efficient SARS-CoV-2 infection of the human lung, *Nat. Commun.* 12 (2021), 134.
- [81] Q. Yang, T.A. Hughes, A. Kelkar, X. Yu, K. Cheng, S.J. Park, W.C. Huang, J.F. Lovell, S. Neelamegham, Inhibition of SARS-CoV-2 viral entry upon blocking N-and O-glycan elaboration, *Elife* 9 (2020) 1–44.
- [82] D. Sulheim, E. Fagermoen, A. Winger, A.M. Andersen, K. Godang, F. Muller, P.C. Rowe, J.P. Saul, E. Skovlund, M.G. Oie, V.W. Bruun, Disease mechanisms and clonidine treatment in adolescent chronic fatigue syndrome a combined cross-sectional and randomized clinical trial, *JAMA Pediatr.* 168 (2014) 351–360.
- [83] J. Slomko, J.L. Newton, S. Kujawski, M. Tafil-Klawe, J. Klawe, D. Staines, S. Marshall-Gradisnik, P. Zalewski, Prevalence and characteristics of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME) in Poland: a cross-sectional study, *BMJ Open* 9 (2019), e023955.
- [84] V.B. Wyller, H.R. Eriksen, K. Malterud, Can sustained arousal explain the chronic fatigue syndrome? *Behav. Brain Funct.* 5 (2009) 1–10.
- [85] L. Naclu, S. O’Boyle, L. Palla, F.E. Naclu, K. Mudie, C.C. Kingdon, J.M. Cliff, T.G. Clark, H.M. Dockrell, E.M. Lacerda, How myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) progresses: the natural history of ME/CFS, *Front. Neurosci.* 11 (2020), 826.
- [86] K.A. Schlauch, S.F. Khaiboullina, K.L. De Meirleir, S. Rawat, J. Petereit, A.A. Rizvanov, N. Blatt, T. Mijatovic, D. Kulick, A. Palotás, V.C. Lombardi, Genome-wide association analysis identifies genetic variations in subjects with myalgic encephalomyelitis/chronic fatigue syndrome, *Transl. Psychiatry* 6 (2016), e730.
- [87] M. Perez, R. Jaundo, K. Hilton, A. Del Alamo, K. Gemayel, N.G. Klimas, T.J.A. Craddock, L. Nathanson, Genetic predisposition for immune system, hormone, and metabolic dysfunction in myalgic encephalomyelitis/chronic fatigue syndrome: a pilot study, *Front. Pediatr.* 7 (2019), 206.
- [88] J.J. Dibble, S.J. McGrath, C.P. Ponting, Genetic risk factors of ME/CFS: a critical review, *Hum. Mol. Genet.* 29 (2020) R118–R125.
- [89] A.D. Grabowska, E.M. Lacerda, L. Naclu, N. Sepúlveda, Review of the quality control checks performed by current genome-wide and targeted-genome association studies on myalgic encephalomyelitis/chronic fatigue syndrome, *Front. Pediatr.* 8 (2020), 293.
- [90] N. Gadi, S.C. Wu, A.P. Spihlman, V.R. Moulton, What’s sex got to do with COVID-19? Gender-based differences in the host immune response to coronaviruses, *Front. Immunol.* 11 (2020), 2147.
- [91] <http://www.meaction.net/wp-content/uploads/2021/04/Report-on-the-impact-of-Covid-19-on-ME.pdf>. (Accessed 1 August 2021).
- [92] R.K. Gherardi, G. Crépeaux, F.J. Authier, Myalgia and chronic fatigue syndrome following immunization: macrophagic myofasciitis and animal studies support linkage to aluminum adjuvant persistency and diffusion in the immune system, *Autoimmun. Rev.* 18 (2019) 691–705.
- [93] J. Phelan, A.D. Grabowska, N. Sepúlveda, A potential antigenic mimicry between viral and human proteins linking Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) with autoimmunity: the case of HPV immunization, *Autoimmun. Rev.* 19 (2020), 102487.

F.7 Chapter 9

CORRESPONDENCE



Risk of BA.5 Infection among Persons Exposed to Previous SARS-CoV-2 Variants

TO THE EDITOR: In recent months, omicron (B.1.1.529) became the dominant variant of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), displaying some degree of immune evasion.¹ The initial omicron subvariants, BA.1 and BA.2, are being progressively displaced by BA.5 in many countries, possibly owing to greater transmissibility and partial evasion of BA.1- and BA.2-induced immunity.^{2,3} The protection afforded by BA.1 against infection by the BA.5 subvariant is critical because adapted vaccines under clinical trials are based on BA.1.

Portugal was one of the first countries affected by a BA.5 predominance. We used the national coronavirus disease 2019 (Covid-19) registry (SINAVE) to calculate the risk of BA.5 infection among persons with documented infection with past variants, including BA.1 and BA.2. The registry includes all reported cases in the country, regardless of clinical presentation.

The national SARS-CoV-2 genetic surveillance identified periods when different variants represented more than 90% of the isolates.⁴ We identified all persons who had a first infection in periods of dominance of each variant, to calculate their infection risk during the period of BA.5 dominance (Fig. 1A). We pooled BA.1 and BA.2 because of the slow transition between the two subvariants in the population. Finally, we calculated the risk of BA.5 infection for the population that did not have any documented infection before BA.5 dominance (June 1, 2022).

We found that previous SARS-CoV-2 infection had a protective effect against BA.5 infection (Fig. 1B and Table S1 in the Supplementary Appendix, available with the full text of this letter

at NEJM.org), and this protection was maximal for previous infection with BA.1 or BA.2. These data should be considered in the context of breakthrough infections in a highly vaccinated population, given that in Portugal more than 98% of the study population completed the primary vaccination series before 2022.

The study design cannot eliminate all confounders (see the Discussion section in the Supplementary Appendix). In addition, one limitation is the putative effect of immune waning in a population with hybrid immunity (previous infection and vaccination). We found that BA.1 or BA.2 infection in vaccinated persons provided higher protection against BA.5 than infection with pre-omicron variants, in line with a recent report with a test-negative design.⁵ However, BA.1 or BA.2 infections occurred closer to the period of BA.5 dominance than infections with previous variants. There is a perception that the protection afforded by previous BA.1 or BA.2 infection is very low, given the high number of BA.5 infections among persons with previous BA.1 or BA.2 infection. Our data indicate that this perception is probably a consequence of the larger pool of persons with BA.1 or BA.2 infection than with infection by other subvariants, and it is not supported by the data.

Overall, we found that breakthrough infections with the BA.5 subvariant were less likely

THIS WEEK'S LETTERS

- 953 Risk of BA.5 Infection among Persons Exposed to Previous SARS-CoV-2 Variants**
- 955 Convalescent Plasma for Covid-19**

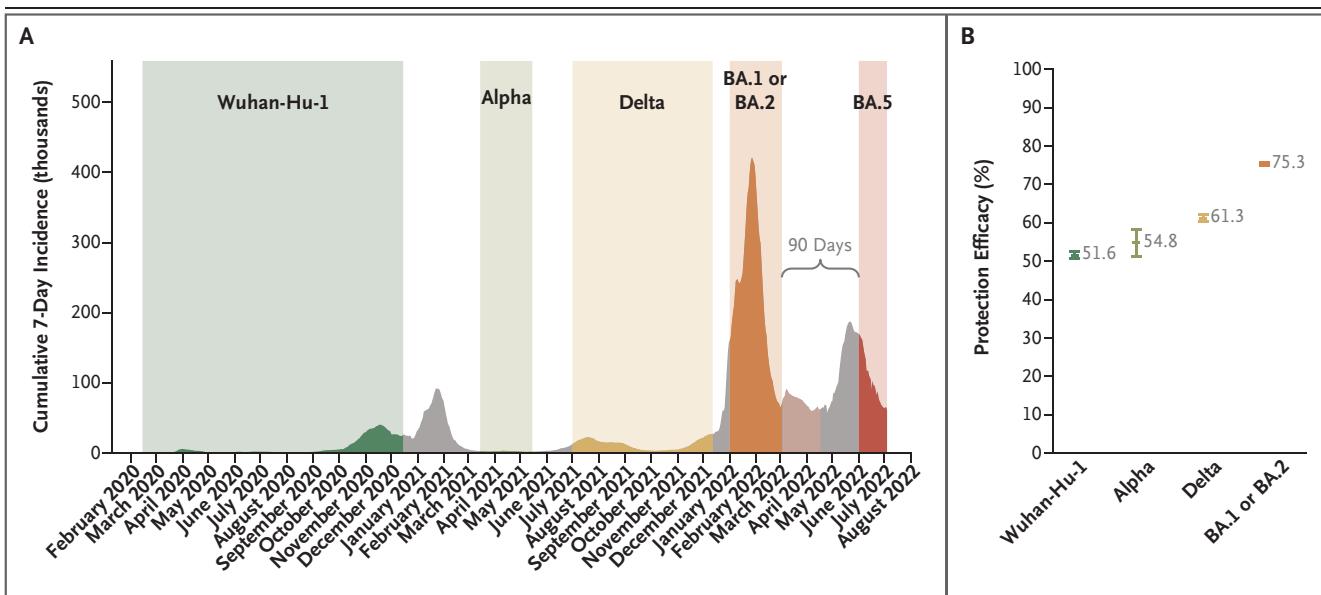


Figure 1. Protective Effect of Previous SARS-CoV-2 Infection on Infection with the Omicron BA.5 Subvariant.

As shown in Panel A, we identified the periods (in different colors) when one variant was represented in more than 90% of sample isolates (data from the national severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2] genetic diversity surveillance⁴). The periods in gray represent times when more than one variant was in circulation. Given the relatively slow transition between dominance by the omicron BA.1 subvariant and dominance by the omicron BA.2 subvariant, we pooled BA.1 and BA.2 in the analysis. We did not include anyone infected in the 90 days before dominance by the omicron BA.5 subvariant. Panel B shows protection efficacy against infection during the period of BA.5 dominance (from June 1, 2022) among persons with one infection in the periods of dominance of different variants, as represented in Panel A, as compared with persons without any documented infection until June 1. Persons with two infections before June 1 were not included in the study. I bars represent 95% confidence intervals.

among persons with a previous SARS-CoV-2 infection history in a highly vaccinated population, especially for previous BA.1 or BA.2 infection, than among uninfected persons.

João Malato, M.Sc.

Instituto de Medicina Molecular João Lobo Antunes
Lisbon, Portugal

Ruy M. Ribeiro, D.Phil.

Los Alamos National Laboratory
Los Alamos, NM

Pedro P. Leite, M.D.

Pedro Casaca, M.D.

Eugénia Fernandes, Ph.D.

Direção Geral da Saúde
Lisbon, Portugal

Carlos Antunes, Ph.D.

Universidade de Lisboa
Lisbon, Portugal

Válter R. Fonseca, M.D., Ph.D.

Direção Geral da Saúde
Lisbon, Portugal

Manuel C. Gomes, Ph.D.

Universidade de Lisboa
Lisbon, Portugal

Luis Graca, M.D., D.Phil.

Instituto de Medicina Molecular João Lobo Antunes
Lisbon, Portugal
lgraca@medicina.ulisboa.pt

Drs. Gomes and Graca contributed equally to this letter.

Supported by the European Union Horizon 2020 research and innovation program (ERA project number, 952377-iSTARS) and by Fundação para a Ciência e a Tecnologia through 081_596653860 and PTDC/MAT-APL/31602/2017 and through National Institutes of Health grant R01-AI116868.

Disclosure forms provided by the authors are available with the full text of this letter at NEJM.org.

This letter was published on August 31, 2022, at NEJM.org.

1. Qu P, Faraone J, Evans JP, et al. Neutralization of the SARS-CoV-2 omicron BA.4/5 and BA.2.12.1 subvariants. *N Engl J Med* 2022;386:2526-8.

2. Yu J, Collier AY, Rowe M, et al. Neutralization of the SARS-CoV-2 omicron BA.1 and BA.2 variants. *N Engl J Med* 2022;386:1579-80.

3. Cao Y, Yisimayi A, Jian F, et al. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by omicron infection. *Nature* 2022;608:593-602.

4. Instituto Nacional de Saúde Doutor Ricardo Jorge. Genetic diversity of the novel coronavirus SARS-CoV-2 (COVID-19) in Portugal. (In Portuguese) 2022 (<https://insaflu.insa.pt/covid19>).

5. Altarawneh HN, Chemaitlely H, Ayoub HH, et al. Protection of SARS-CoV-2 natural infection against reinfection with the omicron BA.4 or BA.5 subvariants. July 12, 2022 (<https://www.medrxiv.org/content/10.1101/2022.07.11.22277448v1>). preprint.

DOI: 10.1056/NEJMc2209479

F.8 Chapter 10



Published Online
January 5, 2023

[https://doi.org/10.1016/
S1473-3099\(22\)00833-7](https://doi.org/10.1016/S1473-3099(22)00833-7)

neutralisation was reduced by more than 10 times compared with the neutralisation of B.1_{pp}.

Finally, we assessed the sensitivity of XBB.1_{pp} to neutralisation by antibodies induced by vaccination or vaccination plus breakthrough infection (figure B; appendix pp 1–2). Plasma of triple vaccinated individuals had almost no detectable neutralising activity against XBB.1_{pp} (neutralising titre 50 [NT₅₀] 2), whereas the neutralising activity against B.1_{pp} was high (NT₅₀ 1165) and against BA.5_{pp} was moderate (NT₅₀ 127). Next, we measured the plasma of triple vaccinated individuals with breakthrough infection during the BA.5 wave in Germany (June to November, 2022). The plasma samples showed high neutralising activity against B.1_{pp} (NT₅₀ 1779), moderate neutralising activity against BA.5_{pp} (NT₅₀ 538), and low neutralising activity against XBB.1_{pp} (NT₅₀ 14). Similar findings were made for plasma from triple vaccinated individuals who received either monovalent or bivalent (ie, B.1 or B.1 plus BA.5) booster vaccination: B.1_{pp} NT₅₀ 1806 for B.1 or 1939 for B.1 plus BA.5; BA.5_{pp} NT₅₀ 206 for B.1 or 525 for B.1 plus BA.5; and XBB.1_{pp} NT₅₀ 8 for B.1 or 5 for B.1 plus BA.5.

Collectively, our data suggest that the SARS-CoV-2 XBB.1 lineage exhibits an extraordinarily strong ability for antibody evasion, which makes XBB.1 similar to BQ.1 and BQ.1.1;⁹ two highly neutralisation-resistant sublineages of omicron that are currently increasing in incidence in several countries worldwide. The finding that most mAbs do not neutralise XBB.1_{pp} highlights that novel mAbs are needed for the treatment of COVID-19 and that other or additional treatment options (eg, paxlovid, molnupiravir, or remdesivir) should be considered in areas with high incidence of the XBB sublineages. The observation that host-cell entry of XBB.1_{pp} is reduced as compared with BA.5_{pp} suggests that the increased ability of XBB.1 to evade antibody-mediated

neutralisation might have come at the cost of a moderately reduced efficiency of host-cell entry.

SP and MH do contract research on the testing of vaccinee serum samples for neutralising activity against SARS-CoV-2 for Valneva, unrelated to this work. GMNB served as an advisor for Moderna and SP served as an advisor for BioNTech, unrelated to this work. All other authors declare no competing interests. SP acknowledges funding for this project by the German Federal Ministry of Education and Research (01K12006D), the EU project UNDINE (grant agreement number 101057100), the Ministry for Science and Culture of Lower Saxony (14-76103-184, MWK HZI COVID-19), and the German Research Foundation (PO 716/11-1 and PO 716/14-1). H-MJ received funding from the German Federal Ministry of Education and Research (01K12043, NaFoUniMedCovid19-COVIM 01KX2021), Bavarian State Ministry for Science and the Arts; and DFG through the research training groups RTG1660 and TRR130, the Bayerische Forschungsstiftung (Project CORAd), and the Kastner Foundation. GMNB acknowledges funding by the German Center for Infection Research (grant number 80018019238) and a European Regional Development Fund (Defeat Corona, ZW7-8515131). The funding sources had no role in study design, data collection, data analysis, data interpretation, writing of the Correspondence, or the decision to submit the manuscript for publication. We did not receive payment by a pharmaceutical company or other agency to write this Correspondence. We were not precluded from accessing data in the study and we accept responsibility to submit for publication.

Prerna Arora, Anne Cossmann,
Sebastian R Schulz,
Gema Morillas Ramos,
Metodi V Stankov, Hans-Martin Jäck,
Georg M N Behrens, Stefan Pöhlmann,
*Markus Hoffmann
mhoffmann@dpz.eu

Infection Biology Unit, German Primate Center, Leibniz Institute for Primate Research, Göttingen 37077, Germany (PA, SP, MH); Faculty of Biology and Psychology, Georg-August University Göttingen, Göttingen, Germany (PA, SP, MH); Department for Rheumatology and Immunology, Hannover Medical School, Hannover, Germany (AC, GMR, MVS, GMNB); Division of Molecular Immunology, Department of Internal Medicine 3, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany (SRS, H-MJ); German Centre for Infection Research, Hannover-Braunschweig, Hannover, Germany (GMR, GMNB); Centre for Individualized Infection Medicine, Hannover, Germany (GMNB)

- 1 Focosi D, Maggi F. Recombination in coronaviruses, with a focus on SARS-CoV-2. *Viruses* 2022; **14**: 1239.
- 2 Taghizadeh P, Salehi S, Heshmati A, et al. Study on SARS-CoV-2 strains in Iran reveals potential contribution of co-infection with and recombination between different strains to the emergence of new strains. *Virology* 2021; **562**: 63–73.

- 3 Turakhia Y, Thornlow B, Hinrichs A, et al. Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape. *Nature* 2022; **609**: 994–97.
- 4 Wertheim JO, Wang JC, Leelawong M, et al. Detection of SARS-CoV-2 intra-host recombination during superinfection with alpha and epsilon variants in New York city. *Nat Commun* 2022; **13**: 3645.
- 5 Chen C, Nadeau S, Yared M, et al. CoV-spectrum: analysis of globally shared SARS-CoV-2 data to identify and characterize new variants. *Bioinformatics* 2021; **38**: 1735–37.
- 6 WHO. Tracking SARS-CoV-2 variants. <https://www.who.int/activities/tracking-SARS-CoV-2-variants> (accessed Nov 22, 2022).
- 7 Schmidt F, Weisblum Y, Muecksch F, et al. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J Exp Med* 2020; **217**: e20201181.
- 8 Arora P, Zhang L, Nehlmeier I, et al. The effect of cilgavimab and neutralisation by vaccine-induced antibodies in emerging SARS-CoV-2 BA.4 and BA.5 sublineages. *Lancet Infect Dis* 2022; **22**: 1665–66.
- 9 Qu P, Evans JP, Faraone J, et al. Enhanced neutralization resistance of SARS-CoV-2 omicron subvariants BQ.1, BQ.1.1, BA.4.6, BF.7, and BA.2.75.2. *Cell Host Microbe* 2022; published online Nov 22. <https://doi.org/10.1016/j.chom.2022.11.012>.

Stability of hybrid versus vaccine immunity against BA.5 infection over 8 months

The coverage of SARS-CoV-2 vaccination in large parts of the world, together with the high number of breakthrough infections, especially following the emergence of Omicron subvariants, makes hybrid immunity (resulting from vaccine and infection) common. Hybrid immunity, particularly after BA.1 or BA.2 infection, confers substantial protection against the BA.5 infection.^{1–3} However, although the waning of protection afforded by natural infection in non-vaccinated individuals or by vaccination has been well documented,^{4,5} the stability of hybrid immunity, specifically against the BA.5 subvariant, now dominant in many countries, has not been thoroughly addressed.

We used the Portuguese COVID-19 registry (SINAVE), which includes all notified cases of infection in the

country on the basis of an official positive test and irrespective of clinical presentation, to investigate the risk of reinfection with BA.5 in a highly vaccinated population previously infected with BA.1 or BA.2 subvariants. We included the population aged 12 years or older, for whom the vaccination coverage was greater than 98% at the end of 2021 (appendix pp 4–5). The registry is very comprehensive due to legal requirements for compensation payment during mandatory isolation. We include infection data from the start of the pandemic until Sept 14, 2022.

We identified the periods of dominance (over 90% of the isolates) of BA.1 and BA.2 (Jan 1–Apr 17, 2022) and BA.5 infections (June 1–Sept 14, 2022) using the national SARS-CoV-2 genetic surveillance data and divided those periods into 15 day intervals (figure A). We then calculated the relative risk (RR) of BA.5 infection in each interval for individuals that had the first infection during each BA.1 and BA.2 dominance subinterval, compared with individuals also vaccinated but without any previous documented infection. Reinfection was defined as two positive tests in the same individual, at least 90 days apart. We found that the RR increased from around 0·06 to around 0·35 between 3 months and 8 months post BA.1 or BA.2 infection (figure B, appendix p 12). Indeed, the RR initially increases rapidly, then more slowly, stabilising at around 0·37.

The present authors previously assessed the effect of unreported infections in the calculation of RR.¹ Here, we mitigate this effect by calculating the RR for the same interval of BA.5 infection for individuals infected by BA.1 or BA.2 in distinct periods, thus with a constant frequency of unreported infections. In any case, our findings are consistent throughout the entire dataset (appendix p 12). Our

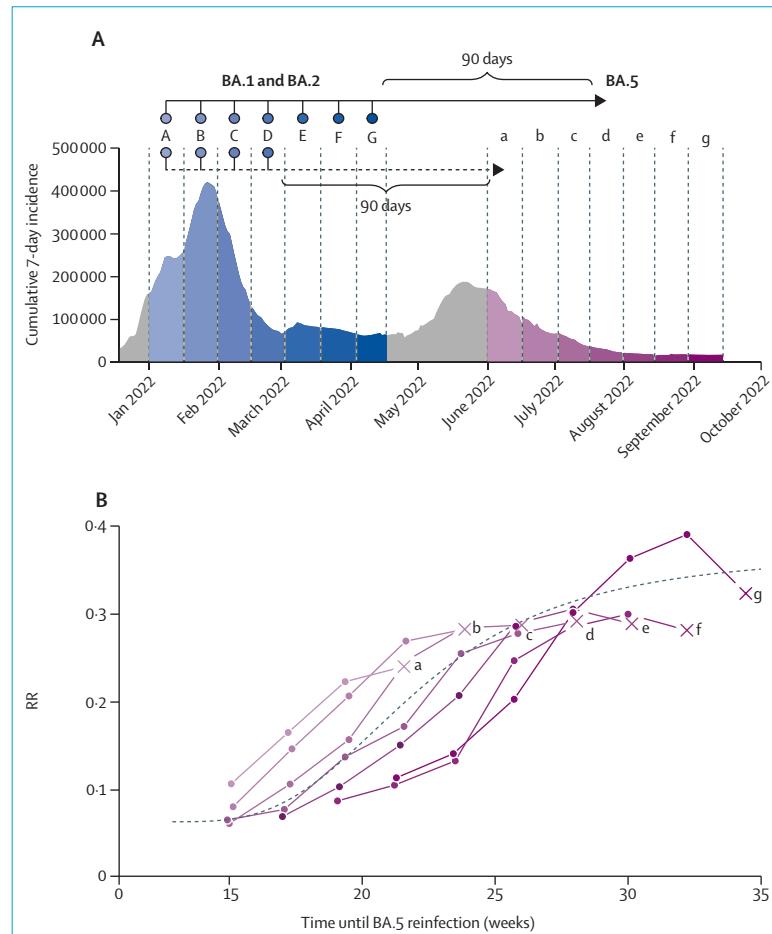


Figure: Stability of hybrid immunity protection against BA.5 infection following infection with BA.1 or BA.2 subvariants

(A) Incidence of documented SARS-CoV-2 infection overlaid with the period of dominance of the BA.1 and BA.2 variants, Jan 1–Apr 14, 2022, divided into 15-day sub-intervals (shades of blue), and the period of dominance of the BA.5 variant, Jun 1–Sep 14, 2022, also divided into 15-day sub-intervals (shades of purple). Two illustrative comparisons are represented. In period d (BA.5 dominance), the risk of infection was compared between individuals with a first documented infection in one of the seven subintervals of BA.1 and BA.2 dominance (A–G), represented with the solid arrow. In the second example with the dashed arrow, in period a of BA.5 dominance, the risk of infection was compared between individuals with a first documented infection in the first four periods of BA.1 and BA.2 dominance (A–D), as reinfections were only considered 90 days following the first infection. (B) RR of reinfection versus first infection in each subinterval of the period of BA.5 dominance (curves a–g, corresponding with the periods of the same letter as in (A) over time since the first infection. The increase in risk is well described by a saturating function (appendix pp 5, 9) as represented by the fitted line (dashed, black). RR=relative risk.

For how reinfection was defined see <https://www.who.int/publications/item/WHO-2019-nCoV-Surveillance-Guidance-2022>.

registry-based dataset includes data on essentially the whole population, but only includes data on positive tests. This feature precludes using a test-negative study design, which has been successfully used in other studies of RR.^{2,6} However, previous reports indicate that the estimates of protection efficacy using the national registry are well aligned with studies that used a test-negative design, albeit in a different population.^{1,2}

Studies since 2021 have made clear the potential for immune imprinting, with one study⁷ suggesting that protection against infection waned after the booster (relative to primary series). In our study, essentially the whole population is vaccinated with the booster dose, and therefore we cannot distinguish effects of booster versus primary series. However, our results of increased protection with hybrid immunity versus vaccine

immunity, agrees with the overall conclusion of that study that "imprinting effects are unlikely to negate the overall public health value of booster vaccinations".⁷

This study shows that hybrid immunity following infection with Omicron BA.1 or BA.2 when compared with vaccine-only immunity leads to substantially increased protection against BA.5 reinfection for up to 8 months.

JM reports funding by Fundação para a Ciência e a Tecnologia, Portugal (grant numbers SFRH/BD/149758/2019 and UIDB/00006/2020).

RMR reports funding by Fundação para a Ciência e a Tecnologia, Portugal (grant number PTDC/MAT-APL/31602/2017), and National Institutes of Health, USA (grant numbers R01-AI116868 and 5TUS54ACTHL143541). CA reports funding by the Fundação para a Ciência e a Tecnologia, Portugal (PIDDAC; grant number UIDB/50019/2020).

LG reports funding by Fundação para a Ciência e a Tecnologia, Portugal (grant number 081_596653860) and Fundación la Caixa, Spain (grant number HR22-00741). All other authors declare no competing interests.

João Malato, Ruy M Ribeiro, Eugénia Fernandes, Pedro Pinto Leite, Pedro Casaca, Carlos Antunes, Válder R Fonseca, †Manuel Carmo Gomes, *†Luis Graca Igrraca@medicina.ulisboa.pt

*Joint senior authors

Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Centro Académico de Medicina de Lisboa, Universidade de Lisboa, Lisboa, 1649-028, Portugal (JM, RMR, VRF, LG); CEAUL - Centro de Estatística e Aplicações da Universidade de Lisboa, Lisboa, Portugal (JM); Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM, USA (RMR); Direção de Serviços de Informação e Análise, Direção Geral da Saúde, Lisboa, Portugal (EF, PPL, PC); Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal (CA, MCG); Instituto Dom Luiz, Universidade de Lisboa, Portugal (CA); Comissão Técnica de Vacinação contra a COVID-19, Direção Geral da Saúde, Lisboa, Portugal (VRF, MCG, LG)

See Online for appendix

- Malato J, Ribeiro RM, Leite PP, et al. Risk of BA.5 Infection among persons exposed to previous SARS-CoV-2 Variants. *N Engl J Med* 2022; **387**: 953–54.
- Altarawneh HN, Chemaitelly H, Ayoub HH, et al. Protective effect of previous SARS-CoV-2 infection against omicron BA.4 and BA.5 subvariants. *N Engl J Med* 2022; **387**: 1620–22.
- Hansen CH, Friis NU, Bager P, et al. Risk of reinfection, vaccine protection, and severity of infection with the BA.5 omicron subvariant: a nation-wide population-based study in Denmark. *Lancet Infect Dis* 2022; published online Oct 18. [https://doi.org/10.1016/S1473-3099\(22\)00595-3](https://doi.org/10.1016/S1473-3099(22)00595-3).

- Chemaitelly H, Tang P, Hasan MR, et al. Waning of BNT162b2 vaccine protection against SARS-CoV-2 infection in Qatar. *N Engl J Med* 2021; **385**: e83.
- Goldberg Y, Mandel M, Bar-On YM, et al. Protection and waning of natural and hybrid immunity to SARS-CoV-2. *N Engl J Med* 2022; **386**: 2201–12.
- Ayoub HH, Tomy M, Chemaitelly H, et al. Estimating protection afforded by prior infection in preventing reinfection: applying the test-negative study design. *Epidemiology* 2022; published online Jan 3. <https://doi.org/10.1101/2022.01.02.22268622> (preprint).
- Chemaitelly H, Tang P, Coyle P, et al. Protection against reinfection with SARS-CoV-2 omicron BA.2.75* sublineage. *Epidemiology* 2022; published online Oct 30. <https://doi.org/10.1101/2022.10.29.22281606> (preprint).

following a third vaccine dose in all study participants (appendix p 10), concurrent with previous data.³ However, concentrations of nasal M-IgA in participants with previous SARS-CoV-2 infection, but without omicron breakthrough infection, remained above the amount associated to 65% protection² over the 8-month study period (figure C). This finding suggests a long-lasting mucosal immunity evoked by SARS-CoV-2 infection.

We next followed systemic and mucosal immune responses in participants that had a BA.1 or BA.2 breakthrough infection during the screening study. 7 months following breakthrough infection, S-IgG concentrations waned to be lower than at baseline (appendix p 10). As previously shown,⁴ serological responses were lower among participants with a history of SARS-CoV-2 infection before breakthrough infection compared with those without and the difference remained over the 7-months follow-up (appendix p 10). Whether these findings reflect immune imprinting after previous infection⁵ or a hampered systemic viral replication due to stronger and more rapid mucosal immune responses² needs further investigation. Interestingly, although nasal M-IgA concentrations waned, they remained above the protective threshold² in 94% of participants with previous SARS-CoV-2 WT or delta infection and in 58% of previously SARS-CoV-2-naïve participants (figure B). In line with this, and in agreement with recent population-based data,^{6,7} BA.1 and BA.2 infections were strongly protective against subsequent BA.5 infection in this cohort, with a HR of 0.13 (95% CI 0.04–0.44; figure D).

To assess whether M-IgA in nasal samples originated in the mucosa, we correlated M-IgA to mucosal spike-specific secretory IgA in nasal samples, and M-IgA to spike-specific IgA in serum. Concentrations of M-IgA correlated stronger to mucosal