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.

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

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

$$\theta_1 = \gamma\theta_0 + (1-\gamma)\theta_1^* \,, \tag{1}$$

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	
5	0.24	0.43	0.50	0.42	
3	–	0.02	0.25	0.23	100
2	–	–	–	–	
1.5	–	–	–	–	
1.25	–	–	–	–	
10	0.77	0.79	0.77	0.70	
5	0.56	0.66	0.69	0.63	
3	0.20	0.41	0.53	0.50	250
2	–	–	0.23	0.26	
1.5	–	–	–	–	
1.25	–	–	–	–	
10	0.84	0.86	0.84	0.78	
5	0.70	0.76	0.78	0.73	
3	0.47	0.60	0.67	0.65	500
2	–	0.27	0.46	0.47	
1.5	–	–	0.04	0.13	
1.25	–	–	–	–	
10	0.89	0.90	0.89	0.84	
5	0.80	0.84	0.85	0.81	
3	0.64	0.72	0.77	0.75	1000
2	0.32	0.50	0.62	0.62	
1.5	–	0.05	0.32	0.38	
1.25	–	–	–	–	

Table 2. *Cont.*

| Δ_T \\ θ_0	0.05	0.1	0.25	0.5	n (per Group)
10	0.93	0.94	0.93	0.90	
5	0.88	0.90	0.90	0.88	
3	0.78	0.83	0.85	0.84	2500
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	
5	0.91	0.93	0.93	0.91	
3	0.84	0.88	0.90	0.88	5000
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).
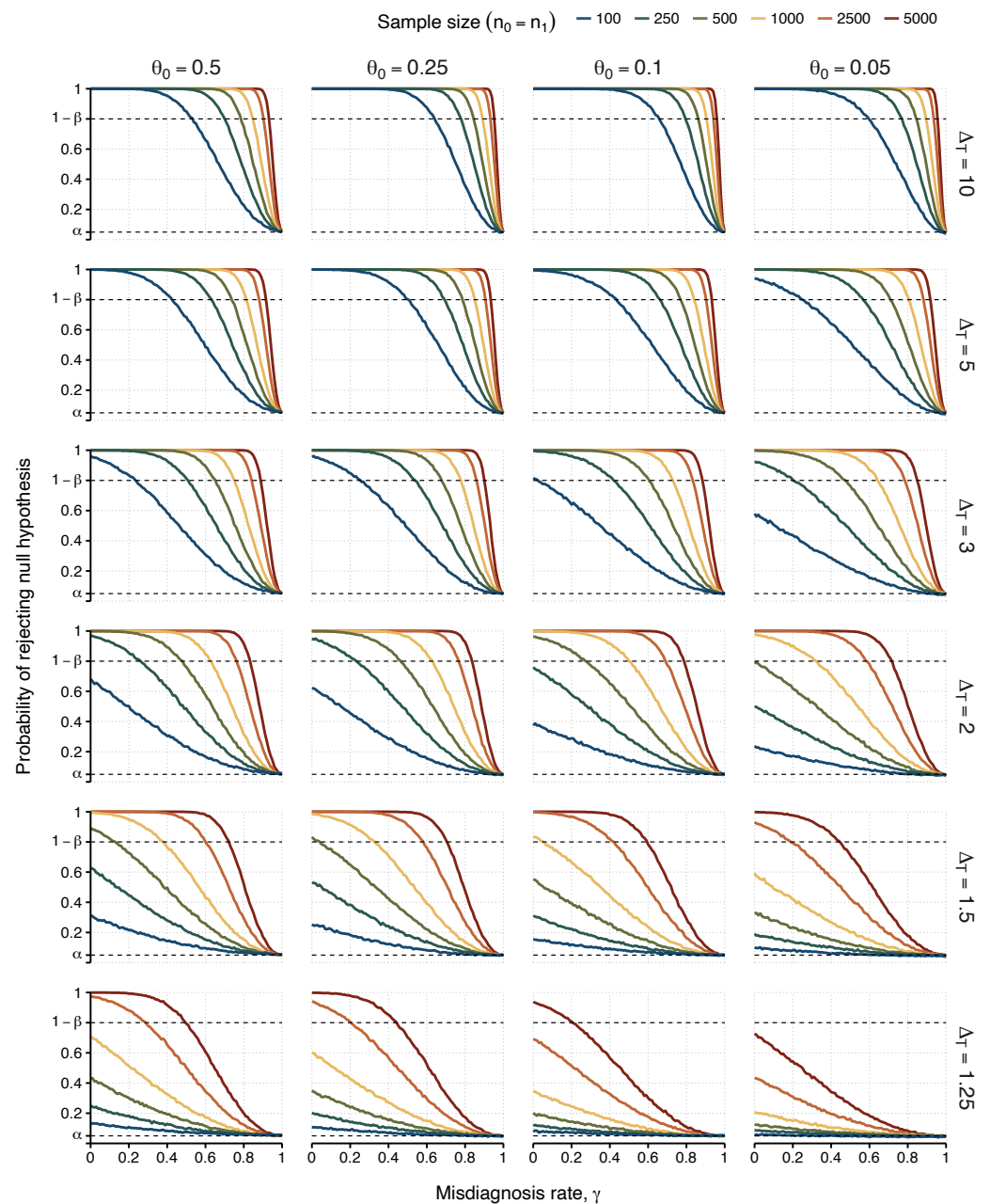
Figure 1. Probabilities of detecting an association (i.e., rejecting H_0) as a function of the misdiagnosis rate. Each column represents the values attributed to the risk allele frequency found in matched healthy controls and false positive ME/CFS cases ($\theta_0 \in \{0.05, 0.1, 0.25, 0.5\}$). Each 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.
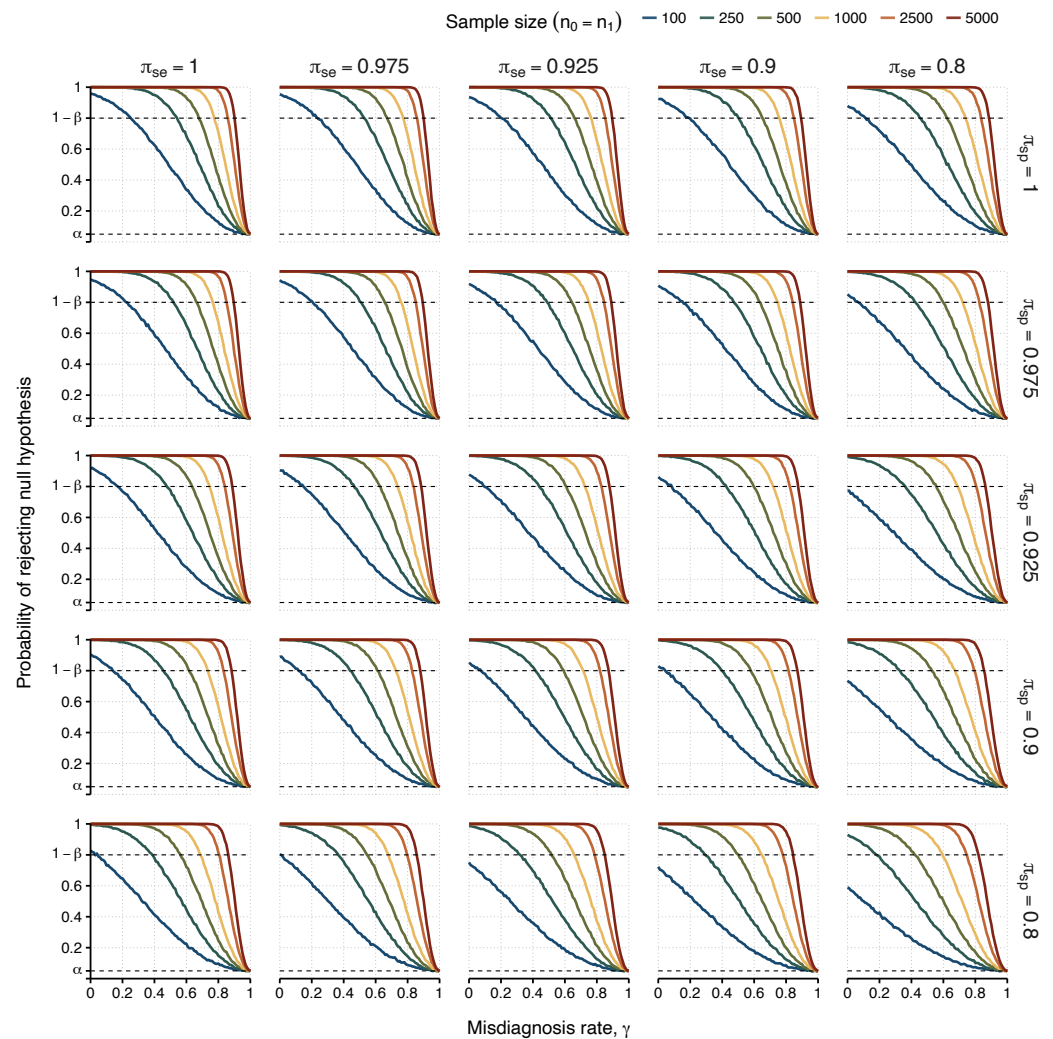
Figure 2. Probabilities of detecting an association (i.e., rejecting H_0) as a function of the misdiagnosis rate. Each scenario represents simulated results with a different combination of sensitivity (π_{se}) and specificity π_{sp} for the serological test for columns and rows, respectively. Power was 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	0.25	0.23	0.20	0.19	0.11	
0.975	0.22	0.20	0.17	0.15	0.06	
0.925	0.17	0.14	0.09	0.08	—	100
0.9	0.13	0.11	0.07	0.04	—	
0.8	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	
1	0.67	0.67	0.66	0.65	0.62	
0.975	0.66	0.65	0.64	0.63	0.59	
0.925	0.63	0.62	0.60	0.59	0.55	500
0.9	0.61	0.61	0.59	0.57	0.52	
0.8	0.56	0.54	0.51	0.50	0.42	
1	0.77	0.77	0.76	0.75	0.73	
0.975	0.76	0.75	0.74	0.74	0.72	
0.925	0.74	0.73	0.72	0.71	0.68	1000
0.9	0.73	0.72	0.71	0.70	0.67	
0.8	0.68	0.67	0.65	0.64	0.59	
1	0.85	0.85	0.85	0.84	0.83	
0.975	0.85	0.85	0.84	0.84	0.82	
0.925	0.84	0.83	0.82	0.82	0.80	2500
0.9	0.83	0.82	0.81	0.81	0.79	
0.8	0.80	0.79	0.78	0.78	0.74	
1	0.90	0.90	0.89	0.89	0.88	
0.975	0.89	0.89	0.89	0.88	0.87	
0.925	0.88	0.88	0.87	0.87	0.86	5000
0.9	0.88	0.87	0.87	0.87	0.85	
0.8	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 ($\hat{\Delta}_T$)	*p*-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.
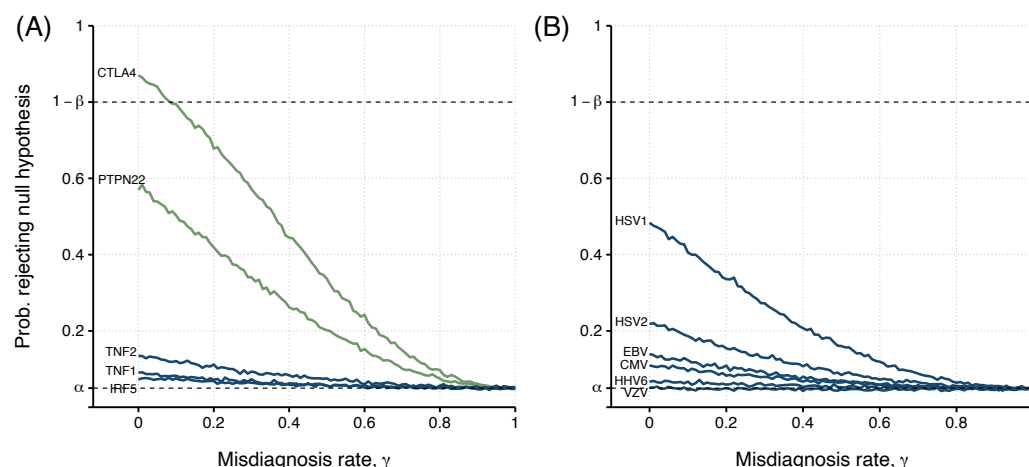
Figure 3. The relationship between the misdiagnosis probability (or rate) and the probability of detecting an association (i.e., rejecting the H_0) estimated from simulated data from two previously published studies: (**A**). Data from five different SNPs (genes *PTPN22*, *CTLA4*, *TNF* (TNF1-rs1799724 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 ($\hat{\Delta}_T$)	*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

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.

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,

$$f(x_i|n_i;\theta_i) = \prod_{i=0,1} \binom{n_i}{x_i} \theta_i^{x_i}(1-\theta_i)^{n_i-x_i} , \tag{A1}$$

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^*

$$\theta_1^* = \frac{\theta_0 \Delta_T}{1 + \theta_0 (\Delta_T - 1)} \tag{A2}$$

References

1. Fukuda, K.; Strausm, 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]
2. Carruthers, B.M.; Jain, A.K.; De Meirleir, 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]
3. 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]
4. 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 ß2-adrenergic receptors. *Autoimmun. Rev.* **2020**, *19*, 102527. [CrossRef] [PubMed]
5. 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]
6. 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]
7. Rasa, S.; Nora-Krukle, 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]
8. 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]
9. 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]
10. 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]
11. 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.
12. 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.
13. 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]
14. 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]
15. 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]
16. 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]
17. 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.
18. 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]
19. 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]
20. 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]
21. Steiner, S.; Becker, S.C.; Hartwig, J.; Sotzny, F.; Lorenz, S.; Bauer, S.; Löbel, M.; Stittrich, 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, Switzderland, 2022; pp. 215–225.

28. Domingues, T.D.; Mouriño, 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.; Mouriño, 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. 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*, 644548. [CrossRef] [PubMed]

45. Cella, 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.; Mattey, 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.; Mattey, 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]