frontiers

Impact of Imperfect Diagnostic Tests in ME/CFS Association Analyses

João Malato [1,*], **Luís Graça** [1] **and Nuno Sepúlveda** [2,3]

[1] *LGraca Lab, Instituto de Medicina Molecular João Lobo Antunes, Faculty of Medicine, University of Lisbon, Lisbon, Portugal*
[2] *Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland*
[3] *CEAUL–Centro de Estatística e Aplicações da Universidade de Lisboa, University of Lisbon, Lisbon, Portugal*

Correspondence*:
Corresponding Author
jmalato@medicina.ulisboa.pt

ABSTRACT

Abstract.

Keywords: misclassification, simulation, power studies, association studies, myalgic encephalomyelitis, chronic fatigue syndrome

1 INTRODUCTION

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a complex disease with uncertainty in its diagnosis (Nacul et al., 2017). ME/CFS patients usually manifest unexplained long-lasting fatigue or post-exertional malaise that arises after slight physical or mental effort and is not alleviated by rest, along with other heterogeneous accompanying symptoms (Fukuda et al., 1994; Carruthers et al., 2003). Its prevalence is estimated between 0.4% and 1%, depending on the population, affecting twice more women than men (Lim et al., 2020).

The aetiology of ME/CFS has been proven difficult to confidently determine and there are have been proposed multiple triggers for the disease onset. There is the overall acceptance that its pathophysiology arises from a combination of both environmental factors (Blomberg et al., 2018; Chu et al., 2019) and genetic factors (Steiner et al., 2020; Lande et al., 2020). Moreover, the disease has been described as having an autoimmune onset (Lorusso et al., 2009; Sotzny et al., 2018), with the immune dysregulation often occurring after exposure to an acute viral infection (Rasa et al., 2018; Blomberg et al., 2018; Chu et al., 2019). Whether ME/CFS patients are predisposed to such infections is still to be determined (Lacerda et al., 2019). Nonetheless, association studies have proposed a relation between the development of ME/CFS and viruses, such as different types of human herpesviruses (Wallace et al., 1999; Bouquet et al., 2019) like infectious mononucleosis and Epstein-Barr virus (Straus, 1985; Wallace et al., 1999; Katz et al., 2009; Katz and Jason, 2013; Harvey et al., 2016; Blomberg et al., 2019; Shikova et al., 2020) and cytomegalovirus (Cameron et al., 2010), and other viruses, such as long-term dengue infections(Seet et al., 2007). More recently, even the COVID-19 coronavirus has been related to ME/CFS due to the post-viral syndrome developed after infection by the SARS-CoV-2 virus (Komaroff and Bateman, 2021). Patients diagnosed with ME/CFS can also display associated immune perturbations (Klimas et al., 1990; Brenu et al., 2011), which shows consistency with the proposed autoimmunity origin for the disease (Sotzny et al., 2018).

These immune abnormalities, triggered by either the external viruses or other conditions such as having a stressful life-style or history of anxiety (Lacerda et al., 2019), could be an explanation for the frequent viral infections or the high rate of flu-like symptoms reported by ME/CFS patients (Lacerda et al., 2019). Ultimately, this persistent physical impairment greatly reduces the quality of life of those affected and efforts have been made to gradually increase knowledge on the disease.

Association studies have looked into possible cases for genetic predispositions (Kaushik, 2005; Gow et al., 2009; Perez et al., 2019; Steiner et al., 2020; Szklarski et al., 2021), focusing on the determination of immunologic triggers or cofactors that could prompt the disease (Lorusso et al., 2009; Loebel et al., 2017; Lacerda et al., 2019; Cliff et al., 2019; Shikova et al., 2020). Nonetheless, the lack of a specific biomarker, or combination of biomarkers, that could objectively characterise ME/CFS leave its diagnosis to be mostly based on specific symptoms and exclusion of other diseases (Smith et al., 2014). This further increases the uncertainty surrounding an unequivocal diagnosis, which has resulted in more than 20 symptom-based criteria currently used to clinically diagnose ME/CFS based on different reported aspects of the disease (Brurberg et al., 2014). Recommendations and protocols have been made for criteria standardisation (Pheby et al., 2020), however, distinct studies will inevitably define the cohorts of patients differently, potentially with conflicting results (Nacul et al., 2017; Malato et al., 2021). This inherent level of misclassification (i.e., non-ME/CFS patients being incorrectly diagnosed as such) amongst ME/CFS cohorts has already been mentioned when characterising the genome of suspected patients (Brown et al., 2021) and should be taken into account in order to minimise its negative effects on association studies (Malato et al., 2021).

While other potential error-prone effects have been mitigated through the development of sound study designs and the use of replicable laboratory-controlled scenarios, the negative effects from misclassification have yet to be accounted for. With ME/CFS diagnosis mostly based on symptoms, the disease is currently subjected to a broader and less precise characterisation when it comes to patient selection. Furthermore, as more research is done in this heterogeneous disease, results from different fields have been presenting evidences that the same clinical diagnosis may encompass possible disease subtypes (Jason et al., 2001, 2005; Kerr et al., 2007; Stoothoff, 2017), indicating that a stratification of patients into specific characteristics, such as type of trigger or severity, may help to better study the disease.

In the present paper we studied how the inclusion of misclassified individuals into a cohort of suspected ME/CFS patients might impact the statistical power of hypothetical case-control association studies. We performed simulations on multiple scenarios, assessing the effect of variables such as sample size and the measured odds ratio of risk factors for ME/CFS. Mathematical formulations of the problem were based on cohort comparison of allele frequencies and exposure to viruses, representing candidate genes and serology association studies. For the latter, hypotheses on misclassification also accounted for sensitivity and specificity as accuracy measures of the applied serology tests. For illustrative purposes, simulations were also done on real-world data sets from a two published studies (Steiner et al., 2020; Cliff et al., 2019).

2 STATISTICAL METHODOLOGY

2.1 Formulation of the Problem

To study the impacts of inherent misclassification and lack of further stratification within ME/CFS, suppose a typical case-control association study design with two well-defined sampled cohorts for suspected ME/CFS patients and healthy controls matched for age, gender, race, and any other relevant variable that might pose an influence on the analyses. Being a hypothetical study, the “perfect” outline can be delineated, with all suspected patients diagnosed by the same diagnosis criteria and all samples collected at the same

time points, in equal conditions, and handled by the same experienced entity, to reduce any of the more usual laboratory-associated errors.

The main goal of the hypothetical outlined study described is then, to assess an association between a specific risk factor and ME/CFS. Assuming that the risk factor of interest is only classified under a binary outcome for its presence/absence, the observed data can be summarised by a 2×2 frequency table, whose sampling distribution is the product of two Binomial distributions, as such

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

where x_0 and x_1 are the observed frequencies of healthy controls and suspected cases with presence of the risk factor, respectively, n_0 and n_1 are the corresponding sample sizes of each cohort, and θ_0 and θ_1 are the probabilities for presence of risk factor in healthy controls and suspected cases, respectively.

Based on the sampling distribution from Equation (1) different scenarios can be created by varying the cohort sample sizes and parameters θ_0 and θ_1. In order to assess whether or not there is an association to the risk factor of interest in each scenario, we test the null hypothesis for lack of association to ME/CFS (H_0 : odds ratio $= 1$) through the Pearson's χ^2 test for independence. On each case, H_0 is rejected if the p-value for the Pearson's χ^2 test is less than the 5% significance level. By repeating this inference multiple times under the same parametric conditions (i.e., testing the same scenario multiple times) the power of the study is estimated as the overall proportion in which H_0 is rejected.

However, these association studies do not account for the underlying misclassified or distinct population subgroups influencing the cohort of suspected patients. Irrespective of how meticulous the sampling methods are, we can assume the possibility for an inherent number of non-ME/CFS patients being incorrectly placed on the ME/CFS's cohort. This is expected due to the heterogeneity of the studied disease (Nacul et al., 2017). The lack of objectivity in ME/CFS diagnosis creates the possibility for patients with similar combination of overlapping symptoms to be diagnosed as ME/CFS while in fact having other diseases such as fibromyalgia (Abbi and Natelson, 2012) or multiple sclerosis (Gaber et al., 2014). Even other conditions in a low graded state, such as prediabetes, could make this misdiagnosis possible as patients in these states do not have the necessary conditions to be fully diagnosed (). Thus, to account for the effect of misclassification, four main assumptions can be considered:

i. due to misclassification, the sampled cohort of suspected ME/CFS cases can be divided into apparent (false positives) and true positive cases of ME/CFS;

ii. misdiagnosed apparent cases are considered healthy controls in the sense that they share the same risk factor probability, θ_0;

iii. there is an overall misclassification rate, γ, creating the two distinct possibilities of apparent and true cases within the suspected cases cohort;

iv. this misclassification rate is only dependent on the true clinical status of each of the suspected cases.

Under assumption ii. and the law of total probability, the risk factor probability on suspected cases can be written as

$$\theta_1 = \gamma\theta_0 + (1-\gamma)\theta_1^* \,, \tag{2}$$

where θ_1^* is the risk factor probability on true ME/CFS cases that becomes diluted with misclassification rate, γ, disguising its true effect and possible detection of association. This results in the augmentation of the initially observable 2×2 frequency table, with the suspected ME/CFS cases subdivided into apparent and true positives (Table 1). It is worth mention that the reverse for misclassification of truly ME/CFS-positive individuals being diagnosed as healthy participants was not accounted for in this study as it is less likely for ME/CFS individuals to pass as participants and being consistently asymptomatic for long periods of time.

Table 1. Augmented version of the observable 2×2 frequency table and the respective probabilities under a Binomial and Multinomial distribution for healthy controls and suspected cases, respectively.

Observed risk factor	Controls	Suspected cases	
		(Apparent)	(True)
Presence	θ_0	$\gamma\theta_0$	$(1-\gamma)\theta_1^*$
Absence	$1-\theta_0$	$\gamma(1-\theta_0)$	$(1-\gamma)(1-\theta_1^*)$

This structure can be used when studying the association between ME/CFS and a specific gene signature based on quantification of candidate gene and allele-specific expression levels, where θ_0 and θ_1 respectively give the single-nucleotide polymorphism (SNP) probabilities for healthy controls and suspected patients. In this case, it is safe to assume that there are no significant errors on the transcriptional profiling methods for the identification of SNPs of interest (with the outcome given as present vs. absent), with misclassification of patients being the only uncertainty associated to the study.

However, when studying the association between ME/CFS the level of exposure towards viruses (as exposed vs. non-exposed) in serological analyses, even Table 1 does not account potential accuracy errors given by the implemented serology test. If poorly calibrated, this errors can potentially return false positive or false negative on both cohorts. When considering a serological study scenario, four additional assumptions must be taken into account, with transversal effect to all data sets:

v. for each test performed, individuals can only be classified as seropositive or seronegative, in opposition to other 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 negatives for individuals poorly measured by the serological assessment;
viii. the binary exposure outcomes given by π_{se} and π_{sp} are independent from the assessed cohort.

Under these assumptions, the the probability from Equation 2 can be extended into

$$\theta_1 = \pi_{se}\gamma\theta_0 + (1-\pi_{sp})\gamma(1-\theta_0) + \pi_{se}(1-\gamma)\theta_1^* + (1-\pi_{sp})(1-\gamma)(1-\theta_1^*) \,. \tag{3}$$

Taking the eight assumptions into account, the observable 2×2 frequency table can be further augmented with sensitivity and specificity respectively defining the serology tests' overall accuracy to determine seropositive (either true of false positive) and seronegative (either true or false negative) populations on

both cohorts (Table 2). Note that Equation (3) includes parameters related to the accuracy of serology tests. Based on this formulation one can return to Equation (2) by simply assuming $\pi_{se} = \pi_{sp} = 1$.

Table 2. 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 $\overline{E}$ for non-exposed).

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

2.2 Simulation Structure

For the two hypothetical case-control association studies scenarios described above, the impact of inherent misclassification or patients stratification, as well as the effect of serology tests' accuracy was assessed through multiple simulations on different parametric values. For illustrative purposes we performed our simulation study considering cohort sample sizes of $n_0 = n_1 = \{100, 250, 500, 1000, 2500, 5000\}$. On each case, the overall value for θ_1^* was determined based on a predefined true odds ratio, Δ_t, for the levels of association between the measured risk factor and the disease and θ_0, as such

$$\theta_1^* = \frac{\theta_0 \Delta_T}{1 + \theta_0(\Delta_T - 1)} \, . \tag{4}$$

For simulations of candidate gene scenarios, this hypothetical true odds ratio parameter was determined as $\Delta_T = \{1.25, 1.5, 2, 5, 10\}$, ranging from limited, to mild, to a strong association between potential allelic polymorphisms and ME/CFS. Also in this scenarios, the SNP probabilities in the population of healthy controls and apparent cases were defined as $\theta_0 = \{0.05, 0.1, 0.25, 0.5\}$.

To demonstrate the influence of sensitivity and specificity, simulations on a second scenario focusing serology assessment fixed $\theta_0 = 0.25$ and $\Delta_T = 1.5$, with $\pi_{se} = \pi_{sp} = \{0.80, 0.90, 0.925, 0.975, 1.0\}$.

Finally, to assess the power of rejecting H_0 on each case, 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. For each parameter set, power was estimated as the proportion of simulated data sets in which H_0 was rejected.

2.3 Application to Real-world Data

As example of real-world applications, we looked at data from a candidate gene study (Steiner et al., 2020) and a serological study (Cliff et al., 2019). In the study by Steiner et al. (2020), 305 Canadian Consensus Criteria-diagnosed ME/CFS patients (Carruthers et al., 2003), excluded for other medical or neurological diseases that may cause fatigue, and 201 controls, recruited from the laboratory staff, were

genotyped for five immune gene SNPs related to autoimmune diseases: tyrosine phosphatase non-receptor type 22 (PTPN22, rs2476601), cytotoxic T-lymphocyte-associated protein 4 (CTLA4, rs3087243), tumor necrosis factor (TNF, rs1800629 and rs1799724), and interferon regulatory factor 5 (IRF5, rs3807306). The study found significant associations with PTPN22 and CTLA4 autoimmunity-risk alleles only in a particular subset of ME/CFS patients that had reported acute onset of the disease after being exposed to an infection. Under these results, the study concluded on the evidence for ME/CFS as an autoimmune disease, focusing on the importance of prior acute infections as a potential variable for subgrouping ME/CFS patients. Based on the results, we assumed the population of true ME/CFS cases to be the subset of patients with infection trigger onset and estimated the power to reject H_0 on different values of γ through simulation, with parameters for the SNP probability on controls and apparent cases and Δ_T, defined as the odds ratio of the subgroup of interest (Table 3). For consistency to the hypothetical simulation scenarios, the values from the subgroup of ME/CFS patients without infection trigger onset were not accounted and the false positive patients given the same parametric values as the cohort of healthy controls.

Table 3. Parameter values estimated by Steiner et al. (2020) where each SNP detection probability and true odds ratio for ME/CFS patients with infection trigger onset are given by θ_0 and Δ_T, respectively.

SNP	θ_0	Δ_T
PTPN22	0.08	1.63
CTLA4	0.56	1.54
TNF1	0.16	0.89
TNF2	0.13	0.84
IRF5	0.51	0.94

The study by Cliff et al. (2019) sampled ME/CFS patients in compliance with the Canadian Consensus Criteria (Carruthers et al., 2003) and/or CDC-1994 (Fukuda et al., 1994), and part of the UK ME/CFS Biobank (Lacerda et al., 2018). The cohort of 251 patients was further categorised into mild ($n = 197$) or severely affected ($n = 54$) regarding their symptoms. For this analysis we admitted that the more severe group could be considered as true cases [Why did we choose them? Justify?]. This cohort of stratified patients was compared against 107 healthy controls and 46 patients with Multiple Sclerosis with the intent being to study the effect of exposure to different viruses on the differentiation of specific T, B cell, natural killer cell and monocyte populations, as response to cytokine stimulation. To do so, the association study focused on the exposures to six different herpes viruses: human cytomegalovirus (CMV), Epstein–Barr virus (EBV), herpes simplex virus 1 and 2 (HSV1 and HSV2), varicella-zoster virus (VZV), and human herpesvirus (HHV6). For results, the study identified differences across different cell populations, but no significant differences regarding seroprevalence between ME/CFS and controls.

To assess exposure to each one of the viruses, plasma concentrations of IgGs were measured using a commercial quantitative ELISA diagnostic testkits from Demeditec Diagnostics (). Information on this testkit mentions that the 2 standard deviations from the mean (2σ rule) was applied as cutoff to determine the seropositivy diagnostic. Based on this information we determined $\pi_{se} = \pi_{sp} = 0.975$ and once again estimated the power to reject an association for different levels of misclassification, under parameters θ_0, for probability of exposure to each virus, and Δ_T, for the true odds ratio on the subgroup of severy affected ME/CFS, both estimated in the study (Table 4).

Table 4. Parameter values from Cliff et al. (2019) where the probability of exposure to each virus and true odds ratio for severely affected ME/CFS patients are given by, θ_0 and Δ_T, respectively. Legend: human CMV—cytomegalovirus; EBV—Epstein–Barr virus; HSV1—herpes simplex virus 1; HSV2—herpes simplex virus 2; VZV—varicella-zoster virus; HHV6—human herpesvirus

Herpes virus	θ_0	Δ_T
CMV	0.37	0.84
EBV	0.93	0.65
HSV1	0.42	1.60
HSV2	0.34	1.36
VZV	0.97	0.75
HHV6	0.95	1.27

3 RESULTS

As expected, the estimated power to detect the hypothetical association decreased with misclassification rate (Figure 1 and Supplemental Figure 2). Looking at the extreme cases, the estimated power was highest when no misclassification was considered and all suspected ME/CFS cases were true positives ($\gamma = 0$). Irrespective of the scenario, as misclassification increases, the overall power is reduced towards 5% at the opposite most extreme value ($\gamma = 1$). This values is simply the significance level specified for the Pearson's χ^2 test. It is worth mention that both extreme case scenarios, with $\gamma = (0, 1)$, are but theoretical controls, identifying the maximum and minimum power of each study.

3.1 Simulations on Multiple Scenarios

Looking at the first set of simulations, along with gradually increasing the misclassification rate from 0 to 1, the power of each study was estimated through the combination of values for SNP probability in healthy controls, θ_0, and hypothetically defined true odds ratio, Δ_T, used to estimate the true probability of risk factor occurrence in ME/CFS, θ_1^*, more or less dilluted, depending from γ in Equation (2) (Figure 1). Defining Δ_T with values of 10 and 5 defined a high level of affinity between the hypothetical risk factor of interest and ME/CFS. Here, a power reduction below 80 (i.e., the specified power threshold to identify what can be considered as having acceptable reproducibility level) was only observed when $\gamma \geq 0.53$ at $\Delta_T = 10$, and $\gamma \geq 0.24$ at $\Delta_T = 5$, at the smallest sample size of 100 individuals per cohort (Table 5). Indeed studies with odds ratio of such magnitude are usually found when there is a clear association, such as a defined gene or locus alteration that serve as an undisputed marker to identify a disease. This is the case of monogenic disorders. However, with no biomarkers determined thus far, one might assume that candidate associations to ME/CFS are mild, not surpassing $\Delta_T = 2$. Under this assumption, to maintain the studies' reproducibility consistently above the 80% threshold, the compared sample sizes must be increased to a minimum of $n_i \geq 250$. Nevertheless, even at the highest simulated sample size, $n_i = 5000$, when considering the mild association of $\Delta_T = 1.25$ and $\theta_0 = 0.05$ a studies were not able to mainain consistency. This results also showed how increasing θ_0 influenced the power to reject H_0. Although overall, this parameter is not as expressive as odds ratio or sample size.

Simulated results from a serological association study scenario focused on what could be considered the application of an usual serology test, with both accuracy-defining parameters sensitivity, π_{se}, and specificity, π_{sp}, never decreasing below 80% (Supplemental Figure 2). By establishing parameters $\theta_0 = 0.25$ and $\Delta_T = 1.5$ one could make a priori predictions for the impact of different sample sizes regarding the maximum expected power, based on the previous results. To contextualise, when $\pi_{se} = \pi_{sp} = 1$ the results

Table 5. Maximum values of misclassification rate, γ, that maintain power of at least 80% to reject the null hypothesis of lack of association, for different values of allele frequency on healthy controls, θ_0, true odds ratio, Δ_T, and sample sizes, n_i, $i = (0, 1)$. Cells with no value indicate the inability to reach the power threshold between cohort, even at $\gamma = 0$.

Δ_T \ θ_0	0.05	0.1	0.25	0.5	n_i
10	0.59	0.65	0.63	0.53	
5	0.24	0.42	0.50	0.42	
2	–	–	–	–	100
1.5	–	–	–	–	
1.25	–	–	–	–	
10	0.76	0.79	0.77	0.70	
5	0.56	0.66	0.69	0.62	
2	–	–	0.22	0.26	250
1.5	–	–	–	–	
1.25	–	–	–	–	
10	0.84	0.86	0.84	0.78	
5	0.71	0.76	0.78	0.74	
2	–	0.27	0.46	0.47	500
1.5	–	–	0.03	0.14	
1.25	–	–	–	–	
10	0.89	0.90	0.89	0.84	
5	0.80	0.84	0.85	0.81	
2	0.31	0.50	0.62	0.62	1000
1.5	–	0.05	0.32	0.38	
1.25	–	–	–	–	
10	0.93	0.94	0.93	0.90	
5	0.88	0.90	0.90	0.88	
2	0.59	0.69	0.76	0.76	2500
1.5	0.19	0.41	0.57	0.60	
1.25	–	–	0.20	0.28	
10	0.95	0.95	0.95	0.93	
5	0.91	0.93	0.93	0.91	
2	0.71	0.78	0.83	0.83	5000
1.5	0.44	0.59	0.70	0.72	
1.25	–	0.20	0.44	0.49	

are the same as the previous scenario. Thus, similarly to past results, this scenario also shows reduced reproducibility at sample sizes of 100 to 500 individuals per cohort (Table 6). Under these sample sizes, even at the extreme control situation of only true positive ME/CFS patients ($\gamma = 0$), the studies failed to reject H_0 above the power threshold, when the effect of sensitivity and specificity were considered. Comparatively, simulated results with $n_i \geq 1000$ were able to provide better consistency to the association studies. Ultimately, despite not being as impactful as when adjusting parameters for odds ratio or probability of exposure, sensitivity and specificity showed an expected reduction in the overall power by increasing uncertainty across both cohorts.

Table 6. Maximum values of misclassification rate, γ, that maintain a discrimination power between healthy controls and diagnosed ME/CFS patients of at least 80%, for different values of sensitivity, π_{se}, specificity, π_{sp}, and sample sizes, n_i, $i = (0, 1)$. Cells with no value indicate the inability to reach the power threshold between cohort, even at $\gamma = 0$.

π_{Sp} \ π_{Se}	1	0.975	0.925	0.9	0.8	n
1	–	–	–	–	–	100
0.975	–	–	–	–	–	
0.925	–	–	–	–	–	
0.9	–	–	–	–	–	
0.8	–	–	–	–	–	
1	–	–	–	–	–	250
0.975	–	–	–	–	–	
0.925	–	–	–	–	–	
0.9	–	–	–	–	–	
0.8	–	–	–	–	–	
1	0.03	0.02	–	–	–	500
0.975	–	–	–	–	–	
0.925	–	–	–	–	–	
0.9	–	–	–	–	–	
0.8	–	–	–	–	–	
1	0.32	0.30	0.29	0.27	0.20	1000
0.975	0.29	0.28	0.25	0.23	0.16	
0.925	0.23	0.21	0.17	0.16	0.06	
0.9	0.20	0.17	0.14	0.11	–	
0.8	0.06	0.03	–	–	–	
1	0.57	0.57	0.55	0.55	0.50	2500
0.975	0.56	0.55	0.53	0.52	0.47	
0.925	0.51	0.50	0.48	0.47	0.41	
0.9	0.50	0.49	0.45	0.44	0.37	
0.8	0.41	0.39	0.35	0.33	0.23	
1	0.70	0.69	0.68	0.68	0.65	5000
0.975	0.69	0.68	0.66	0.66	0.63	
0.925	0.66	0.65	0.63	0.63	0.58	
0.9	0.64	0.63	0.61	0.61	0.56	
0.8	0.58	0.56	0.54	0.53	0.45	

3.2 Real-world Data

Simulations with parameters defined from published studies further exemplified the inherent impacts of misclassification and stratification of patients in the reproducibility of identified results (Figure 3). Results for the case-control candidate gene study (Steiner et al., 2020), where ME/CFS patients with infection trigger onset were considered as true positive cases, showed similar decreasing curves overall, with maximum estimated power at $\gamma = 0$, declining with misclassification towards the level of significance (Figure 3A). However, none of the simulated studies was able to sustain a power above the reproducibility threshold, regardless of the SNPs used, with different allele frequencies in healthy controls and true odds ratio (Table 3).

Results considering the serological study (Cliff et al., 2019), with severely affected ME/CFS patients being considered true patients in the simulations, had similar results (Figure 3B). With an additional level

of uncertainty with respect to accuracy of the applied serology tests, the results were in agreement with the original conclusions as no associations between ME/CFS and the exposure to any virus, with distinct seroprevalences, were detected. This can be seen through the lower power to reject the hypothesis for association, with no virus surpassing the reproducibility threshold, and only HHV1 ($\theta_0 = 0.42$, $\Delta_t = 1.60$) and HHV2 ($\theta_0 = 0.34$, $\Delta_t = 1.36$) even having a power above the 20% rejection probability, indicating the difficulty, under the study conditions, to validate the viruses as a risk factors for ME/CFS.

4 DISCUSSION

Our simulation study showed how misclassification of patients poses an impact on the ability to consistently recognise true associations to a triggering polymorphism or viral exposure prior to the disease onset. Irrespective of the scenario simulated, by gradually increasing misclassification rate, the studies' power to reject the null hypothesis (i.e., each study ability to correctly and consistently detect true associations through a case-control study design) changed from an optimal situation of having no poorly classified patients and a clear distinction between the two cohorts, to scenarios with the effects of true positive ME/CFS diluted under the influence of false positives sharing traits with healthy controls. Simulations on the hypothetical case-control association studies showed how the strength between a risk factor the disease can drastically improve the reproducibility and replicability of the results (rows on Figure 1). However, while still researching for biomarkers able to discriminate the disease, affinity of risk factors for ME/CFS is expected to be $1.25 \leq \Delta_T \leq 2$. Allied to this consideration, sample size is also a major implication on the studies' power. Until now, misclassification studies mostly focused on identifying the extent of misdiagnosed patients when using distinct diagnosis criteria, not looking at sample sizes (Malato 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 250–300 patients. Numbers of this magnitude struggle for consistency when in situations of mild values for odds ratios, which allows for the identification of potential sporadic significant associations that ultimately cannot be replicated in follow-up studies. As example, our results based on Steiner et al. (2020), with different allele frequencies for five candidate genes, failed to reach the defined reproducibility threshold, even for CTLA4 ($\theta_0 = 0.56$, $\Delta_t = 1.54$) and PTPN22 ($\theta_0 = 0.08$, $\Delta_t = 1.63$), markers that showed significant in the original study (Figure 3A). Considering parameters Δ_T and θ_0 for each one of these genes, the sampled populations should at least doubled in order to not only detect the association of these SNPs and ME/CFS more confidently, but also to grant that the results were to be reproduced under similar conditions.

Simulations on the serology scenarios showcased the reproducibility effects when the accuracy of serology tests was also taken into account. Sensitivity and specificity of these tests have here shown to produce a lessened effect overall, being further corrected their accuracy keeps improving. Nonetheless, their effects can still be 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 by researchers and manufacturers of said tests (Scheibenbogen et al., 2017; Domingues et al., 2021a,b). This important but often overlooked aspect has potential to induce incorrect outcomes by producing false negatives and false positives and can add an additional layer to the misclassification of ME/CFS, with impacts on the studies' reproducibility. The tests used in Cliff et al. (2019) were done under the 2σ rule, which gives us information that the two parameters were at least 0.975 accurate, marginally allowing for type II errors.

Conclusões sobre θ_0.

Mencionar semelhanças entre os dois resultados para real-world, quando um tem resultados significativos e o outro não?

Throughout this study, misclassification was mostly used in a broad sense, with simulations defining it as the wrong diagnosis of a patient (false positive). This simple idea has been explored previously, however to a lesser extend (Nacul et al., 2017; Malato et al., 2021). It is worth mentioning that the misdiagnosed patients were not healthy individuals per se. Being used as proof of concept, they were simply assumed to have the same characteristics of interest as a healthy participant, showed through parameter θ_0. In fact, this study could be extended to include a parameter of different value, allowing to further stratify the cohort of suspected ME/CFS to include misclassified patients identified with parameters values linked to other known overlapping diseases, instead of healthy controls, thus testing the effects of misclassification admitting three (or potentially more) subgroups within the disease group. Alternatively, aside from looking into the power and consistency of a study, other approaches could look for the effects caused by estimation bias.

ME/CFS is a multisystemic and complex disease, where there is still lack of understanding on the disease's aetiology and pathophysiology. Under these uncertainties and while still researching for possible biomarkers able to better discriminate the ME/CFS through case-control association studies, accepting that the statistical power is very likely to suffer from inclusion of poorly classified individuals as suspected cases, however small the rate, might help to improve the study designs and increase the overall 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 forward.

CONFLICT OF INTEREST STATEMENT

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.

AUTHOR CONTRIBUTIONS

The Author Contributions section is mandatory for all articles, including articles by sole authors. If an appropriate statement is not provided on submission, a standard one will be inserted during the production process. The Author Contributions statement must describe the contributions of individual authors referred to by their initials and, in doing so, all authors agree to be accountable for the content of the work. Please see here for full authorship criteria.

FUNDING

JM is funded by the Fundação para a Ciência e Tecnologia, Portugal (ref. SFRH/BD/149758/2019).

ACKNOWLEDGMENTS

Thanks!

SUPPLEMENTAL DATA

Supplementary Material should be uploaded separately on submission, if there are Supplementary Figures, please include the caption in the same file as the figure. LaTeX Supplementary Material templates can be found in the Frontiers LaTeX folder.

DATA AVAILABILITY STATEMENT

All simulations and analyses were done using R statistical software, version 4.1.0 R Core Team (2020). The datasets generated for this study can be found in the GitHub public repository jtmalato/misclassification-simulations.

REFERENCES

Abbi, B. and Natelson, B. H. (2012). Is chronic fatigue syndrome the same illness as fibromyalgia: evaluating the 'single syndrome' hypothesis. *QJM* 106, 3–9. doi:10.1093/qjmed/hcs156

Blomberg, J., Gottfries, C.-G., Elfaitouri, A., Rizwan, M., and Rosén, A. (2018). Infection elicited autoimmunity and myalgic encephalomyelitis/chronic fatigue syndrome: An explanatory model. *Frontiers in Immunology* 9. doi:10.3389/fimmu.2018.00229

Blomberg, J., Rizwan, M., Böhlin-Wiener, A., Elfaitouri, A., Julin, P., Zachrisson, O., et al. (2019). Antibodies to human herpesviruses in myalgic encephalomyelitis/chronic fatigue syndrome patients. *Frontiers in Immunology* 10. doi:10.3389/fimmu.2019.01946

Bouquet, J., Li, T., Gardy, J. L., Kang, X., Stevens, S., Stevens, J., et al. (2019). Whole blood human transcriptome and virome analysis of ME/CFS patients experiencing post-exertional malaise following cardiopulmonary exercise testing. *PLOS ONE* 14, e0212193. doi:10.1371/journal.pone.0212193

Brenu, E. W., van Driel, M. L., Staines, D. R., Ashton, K. J., Ramos, S. B., Keane, J., et al. (2011). Immunological abnormalities as potential biomarkers in chronic fatigue syndrome/myalgic encephalomyelitis. *Journal of Translational Medicine* 9, 81. doi:10.1186/1479-5876-9-81

Brown, D., Birch, C., Younger, J., and Worthey, E. (2021). 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. doi:10.1016/s1096-7192(21)00388-7

Brurberg, K. G., Fønhus, M. S., Larun, L., Flottorp, S., and Malterud, K. (2014). Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *BMJ Open* 4, e003973. doi:10.1136/bmjopen-2013-003973

Cameron, B., Flamand, L., Juwana, H., Middeldorp, J., Naing, Z., Rawlinson, W., et al. (2010). Serological and virological investigation of the role of the herpesviruses EBV, CMV and HHV-6 in post-infective fatigue syndrome. *Journal of Medical Virology* 82, 1684–1688. doi:10.1002/jmv.21873

Carruthers, B. M., Jain, A. K., Meirleir, K. L. D., Peterson, D. L., Klimas, N. G., Lerner, A. M., et al. (2003). Myalgic encephalomyelitis/chronic fatigue syndrome. *Journal of Chronic Fatigue Syndrome* 11, 7–115. doi:10.1300/j092v11n01_02

Chu, L., Valencia, I. J., Garvert, D. W., and Montoya, J. G. (2019). Onset patterns and course of myalgic encephalomyelitis/chronic fatigue syndrome. *Frontiers in Pediatrics* 7. doi:10.3389/fped.2019.00012

Cliff, J. M., King, E. C., Lee, J.-S., Sepúlveda, N., Wolf, A.-S., Kingdon, C., et al. (2019). Cellular immune function in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Frontiers in Immunology* 10. doi:10.3389/fimmu.2019.00796

Dimmock, M. E., Mirin, A. A., and Jason, L. A. (2016). Estimating the disease burden of ME/CFS in the united states and its relation to research funding. *Journal of Medicine and Therapeutics* 1. doi:10.15761/jmt.1000102

Domingues, T. D., Grabowska, A. D., Lee, J.-S., Ameijeiras-Alonso, J., Westermeier, F., Scheibenbogen, C., et al. (2021a). Herpesviruses serology distinguishes different subgroups of patients from the united kingdom myalgic encephalomyelitis/chronic fatigue syndrome biobank. *Frontiers in Medicine* 8. doi:10.3389/fmed.2021.686736

Domingues, T. D., Mouriño, H., and Sepúlveda, N. (2021b). Analysis of antibody data using finite mixture models based on scale mixtures of skew-normal distributions. *medRxiv* doi:10.1101/2021.03.08.21252807

Fukuda, K., Strausm, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., and Komaroff, A. (1994). The chronic fatigue syndrome: A comprehensive approach to its definition and study. *Annals of Internal Medicine* 121, 953. doi:10.7326/0003-4819-121-12-199412150-00009

Gaber, T. A.-Z. K., Oo, W. W., and Ringrose, H. (2014). Multiple sclerosis/chronic fatigue syndrome overlap: When two common disorders collide. *NeuroRehabilitation* 35, 529–534. doi:10.3233/NRE-141146

Gow, J. W., Hagan, S., Herzyk, P., Cannon, C., Behan, P. O., and Chaudhuri, A. (2009). A gene signature for post-infectious chronic fatigue syndrome. *BMC Medical Genomics* 2. doi:10.1186/1755-8794-2-38

Harvey, J. M., Broderick, G., Bowie, A., Barnes, Z. M., Katz, B. Z., O'Gorman, M. R. G., et al. (2016). Tracking post-infectious fatigue in clinic using routine lab tests. *BMC Pediatrics* 16. doi:10.1186/s12887-016-0596-8

Jason, L. A., Corradi, K., Torres-Harding, S., Taylor, R. R., and King, C. (2005). Chronic fatigue syndrome: The need for subtypes. *Neuropsychology Review* 15, 29–58. doi:10.1007/s11065-005-3588-2

Jason, L. A., Taylor, R. R., Kennedy, C. L., Harding, S. T., Song, S., Johnson, D., et al. (2001). Subtypes of chronic fatigue syndrome. *Journal of Chronic Fatigue Syndrome* 8, 1–21. doi:10.1300/j092v08n03_02

Katz, B. Z. and Jason, L. A. (2013). Chronic fatigue syndrome following infections in adolescents. *Current Opinion in Pediatrics* 25, 95–102. doi:10.1097/mop.0b013e32835c1108

Katz, B. Z., Shiraishi, Y., Mears, C. J., Binns, H. J., and Taylor, R. (2009). Chronic fatigue syndrome after infectious mononucleosis in adolescents. *Pediatrics* 124, 189–193. doi:10.1542/peds.2008-1879

Kaushik, N. (2005). Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. *Journal of Clinical Pathology* 58, 826–832. doi:10.1136/jcp.2005.025718

Kerr, J. R., Burke, B., Petty, R., Gough, J., Fear, D., Mattey, D. L., et al. (2007). Seven genomic subtypes of chronic fatigue syndrome/myalgic encephalomyelitis: a detailed analysis of gene networks and clinical phenotypes. *Journal of Clinical Pathology* 61, 730–739. doi:10.1136/jcp.2007.053553

Klimas, N. G., Salvato, F. R., Morgan, R., and Fletcher, M. A. (1990). Immunologic abnormalities in chronic fatigue syndrome. *Journal of Clinical Microbiology* 28, 1403–1410

Komaroff, A. L. and Bateman, L. (2021). Will COVID-19 lead to myalgic encephalomyelitis/chronic fatigue syndrome? *Frontiers in Medicine* 7. doi:10.3389/fmed.2020.606824

Lacerda, E. M., Geraghty, K., Kingdon, C. C., Palla, L., and Nacul, L. (2019). A logistic regression analysis of risk factors in ME/CFS pathogenesis. *BMC Neurology* 19. doi:10.1186/s12883-019-1468-2

Lacerda, E. M., Mudie, K., Kingdon, C. C., Butterworth, J. D., O'Boyle, S., and 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. doi:10.3389/fneur.2018.01026

Lande, A., Fluge, Ø., Strand, E. B., Flåm, S. T., Sosa, D. D., Mella, O., et al. (2020). Human leukocyte antigen alleles associated with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Scientific Reports* 10. doi:10.1038/s41598-020-62157-x

Lim, E.-J., Ahn, Y.-C., Jang, E.-S., Lee, S.-W., Lee, S.-H., and Son, C.-G. (2020). Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *Journal of Translational Medicine* 18. doi:10.1186/s12967-020-02269-0

Loebel, M., Eckey, M., Sotzny, F., Hahn, E., Bauer, S., Grabowski, P., et al. (2017). Serological profiling of the EBV immune response in chronic fatigue syndrome using a peptide microarray. *PLOS ONE* 12, e0179124. doi:10.1371/journal.pone.0179124

Lorusso, L., Mikhaylova, S. V., Capelli, E., Ferrari, D., Ngonga, G. K., and Ricevuti, G. (2009). Immunological aspects of chronic fatigue syndrome. *Autoimmunity Reviews* 8, 287–291. doi:10.1016/j.autrev.2008.08.003

Malato, J., Graça, L., Nacul, L., Lacerda, E., and Sepúlveda, N. (2021). Statistical challenges of investigating a disease with a complex diagnosis. In *Estatística: Desafios Transversais às Ciências com Dados*, ed. S. P. de Estatística (Sociedade Portuguesa de Estatística), 153–167

Mirin, A. A., Dimmock, M. E., and Jason, L. A. (2020). Research update: The relation between me/cfs disease burden and research funding in the usa. *Work* 66, 277–282. doi:10.3233/WOR-203173

Nacul, L., Lacerda, E. M., Kingdon, C. C., Curran, H., and 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. doi:10.1177/1359105317695803

Perez, M., Jaundoo, R., Hilton, K., Alamo, A. D., Gemayel, K., Klimas, N. G., et al. (2019). Genetic predisposition for immune system, hormone, and metabolic dysfunction in myalgic encephalomyelitis/chronic fatigue syndrome: A pilot study. *Frontiers in Pediatrics* 7. doi:10.3389/fped.2019.00206

Pheby, D. F., Araja, D., Berkis, U., Brenna, E., Cullinan, J., de Korwin, J.-D., et al. (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. doi:10.3390/healthcare8020088

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria

Rasa, S., , Nora-Krukle, Z., Henning, N., Eliassen, E., Shikova, E., et al. (2018). Chronic viral infections in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine* 16. doi:10.1186/s12967-018-1644-y

Scheibenbogen, C., Freitag, H., Blanco, J., Capelli, E., Lacerda, E., Authier, J., et al. (2017). The european me/cfs biomarker landscape project: an initiative of the european network euromene. *Journal of translational medicine* 15, 162. doi:https://doi.org/10.1186/s12967-017-1263-z

Seet, R. C., Quek, A. M., and Lim, E. C. (2007). Post-infectious fatigue syndrome in dengue infection. *Journal of Clinical Virology* 38, 1–6. doi:10.1016/j.jcv.2006.10.011

Shikova, E., Reshkova, V., ntoniya Kumanova, Raleva, S., Alexandrova, D., Capo, N., et al. (2020). Cytomegalovirus, epstein-barr virus, and human herpesvirus-6 infections in patients with myalgic ncephalomyelitis/chronic fatigue syndrome. *Journal of Medical Virology* doi:10.1002/jmv.25744

Smith, M. E. B., Nelson, H. D., Haney, E., Pappas, M., Daeges, M., Wasson, N., et al. (2014). *Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome*. Tech. rep. doi:10.23970/ahrqepcerta219

Sotzny, F., Blanco, J., Capelli, E., Castro-Marrero, J., Steiner, S., Murovska, M., et al. (2018). Myalgic encephalomyelitis/chronic fatigue syndrome – evidence for an autoimmune disease. *Autoimmunity Reviews* 17, 601–609. doi:10.1016/j.autrev.2018.01.009

Steiner, S., Becker, S. C., Hartwig, J., Sotzny, F., Lorenz, S., Bauer, S., et al. (2020). Autoimmunity-related risk variants in PTPN22 and CTLA4 are associated with ME/CFS with infectious onset. *Frontiers in Immunology* 11. doi:10.3389/fimmu.2020.00578

Stoothoff, J. (2017). Subtyping patients with myalgic encephalomyelitis (ME) and chronic fatigue syndrome (CFS) by course of illness. *Journal of Biosensors, Biomarkers and Diagnostics* 2, 1–9. doi:10.15226/2575-6303/2/1/00113

Straus, S. E. (1985). Persisting illness and fatigue in adults with evidence of epstein-barr virus infection. *Annals of Internal Medicine* 102, 7. doi:10.7326/0003-4819-102-1-7

Szklarski, M., Freitag, H., Lorenz, S., Becker, S. C., Sotzny, F., Bauer, S., et al. (2021). 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. doi:10.3389/fimmu.2021.644548

Wallace, H. L., Natelson, B., Gause, W., and Hay, J. (1999). Human herpesviruses in chronic fatigue syndrome. *Clinical Diagnostic Laboratory Immunology* 6, 216–223. doi:10.1128/cdli.6.2.216-223.1999

FIGURE CAPTIONS

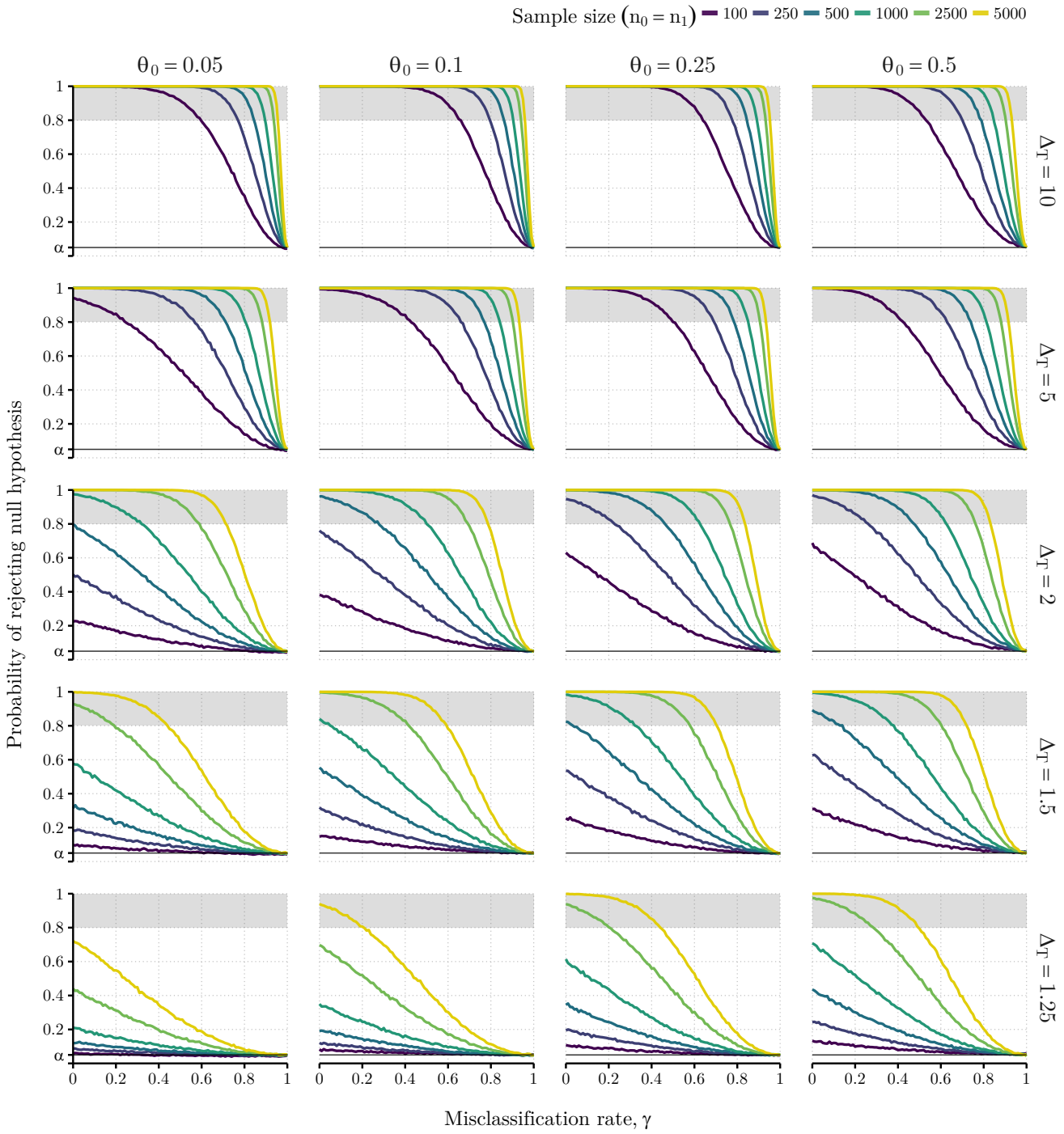

Figure 1. Probabilities of rejecting the null hypothesis, i.e., absence of association between the two populations, as function of the misclassification rate. Each column represents the values attributed to the risk allele frequency found in matched healthy controls and false positive ME/CFs cases ($\theta_0 = \{0.05, 0.1, 0.25, 0.5\}$). Each row varies the true odds ratio for the association between risk allele frequency assessed between true positive cases and healthy controls ($OR_T = \{1.25, 1.5, 2, 5, 10\}$). Power analysis was estimated for different sample sizes of 100, 250, 500, 1000, 2500, and 5000 ($n_0 = n_1$), represented by the different coloured lines on each scenario. Gray filled area indicates positive scenarios where the probability of rejecting the null hypothesis is above 80%. Dark horizontal line indicates the level of significance used, $\alpha = 0.05$.

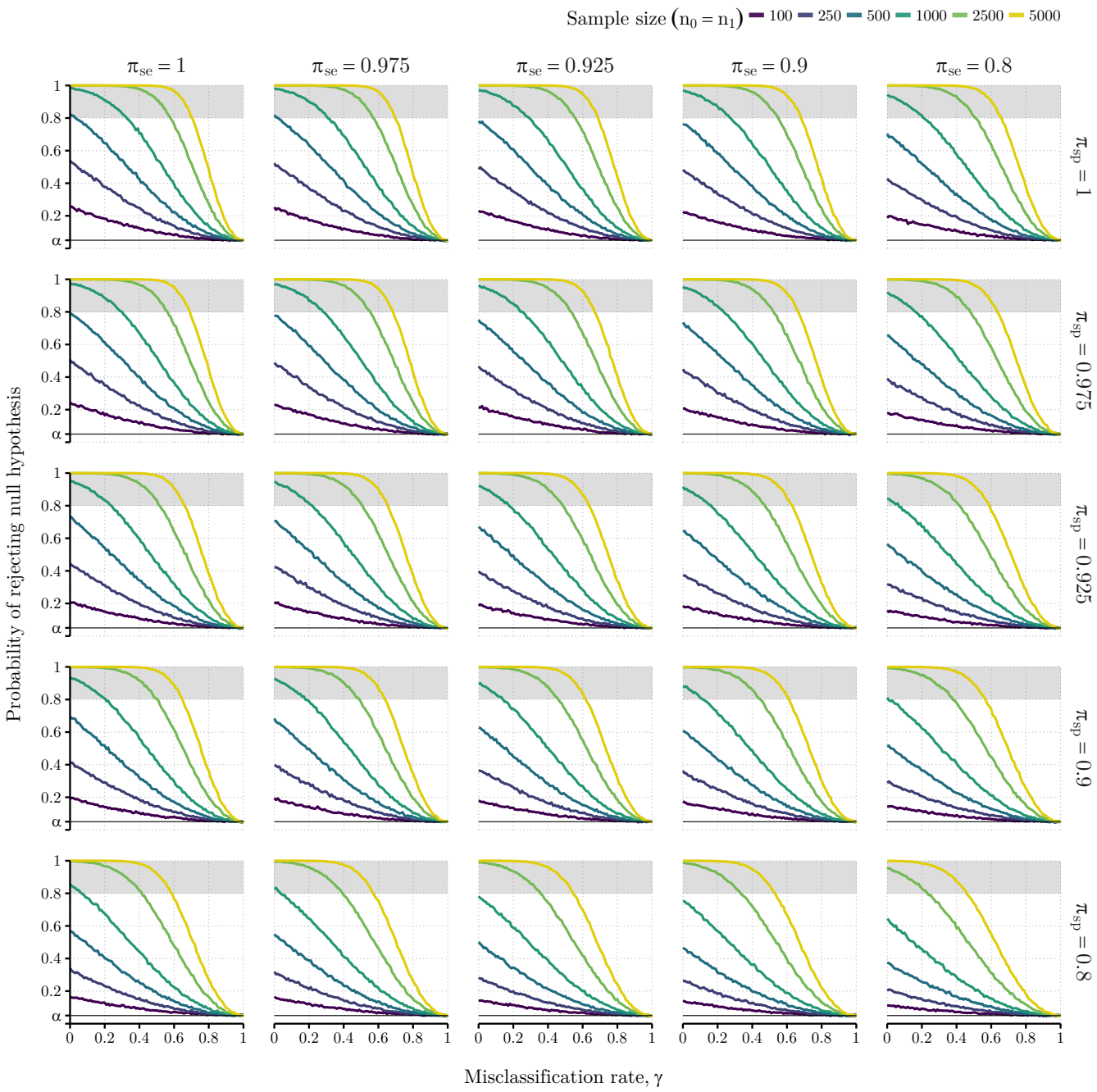

Figure 2. Probabilities of rejecting H_0 as function of the misclassification rate. Each scenario represents simulated results with combination of serology test's sensitivity, π_{se}, and specificity, π_{sp}, for columns and rows, respectively. Power analysis was estimated for different sample sizes of 100, 250, 500, 1000, 2500, and 5000 ($n_0 = n_1$), represented by the different coloured lines on each scenario, with probability of exposure in healthy controls fixed as $\theta_0 = 0.25$ and true odds ratio $\Delta_T = 1.5$. Gray filled area indicates positive scenarios where the probability of rejecting the null hypothesis is above 80%. Dark horizontal line indicates the level of significance used, $\alpha = 0.05$.

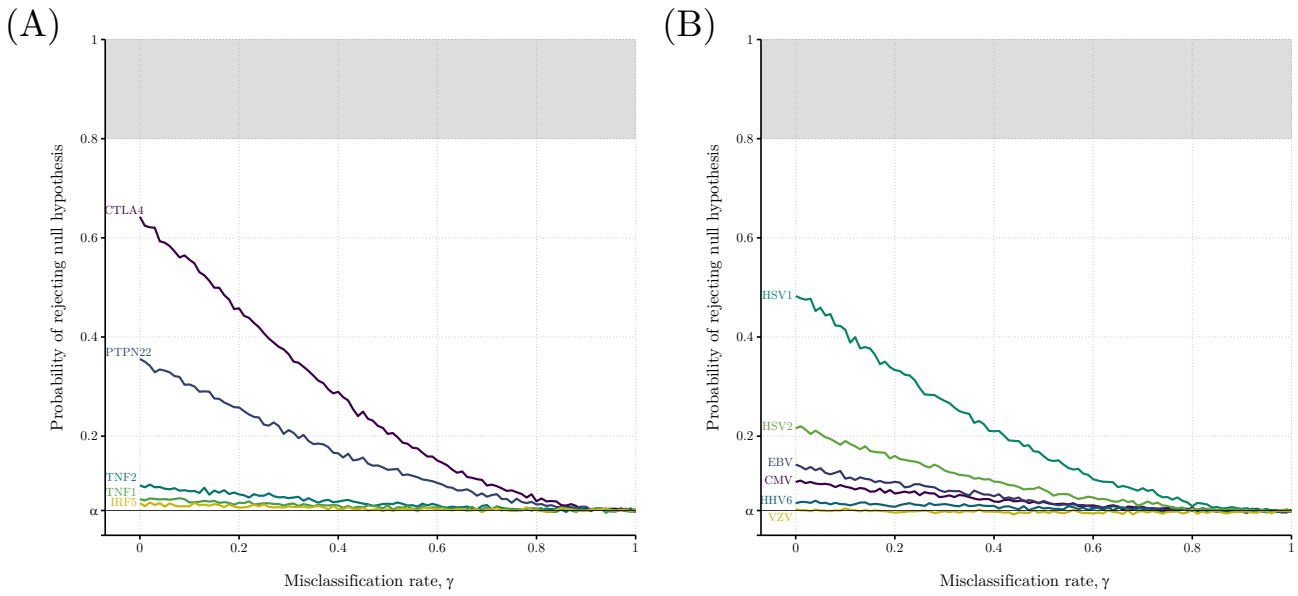

Figure 3. Probabilities of rejecting the null hypothesis using real world data in **(A)** five different SNPs studied (PTPN22, CTLA4, TNF1 and TNF2, and IRF5); or **(B)** Six different herpes viruses (CMV, EBV, HSV1 and HSV2, VZV, and HHV6), as function of misclassification rate. For each study, risk allele frequencies or probability of exposure and true patient odds ratio were determined by Steiner et al. (2020) ($n_0 = 201$; $n_1 = 305$) and Cliff et al. (2019) ($n_0 = 107$; $n_1 = 251$), as to Tables 3 and 4, respectively. Gray filled area indicated the the scenarios where the probability of rejecting the null hypothesis is above 80%. Dark horizontal line indicated the level of significance used, $\alpha = 0.05$.