

Statistical challenges of investigating a disease with a complex diagnosis

João Malato¹, Luís Graça¹, Luís Nacul², Eliana Lacerda², and Nuno Sepúlveda^{3,4}

¹Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Portugal

²Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, United Kingdom

³Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, United Kingdom

⁴Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal

Submitted: March 19th, 2021

Abstract

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

Keywords: multidimensional scaling; cluster analysis; misclassification; cohen's κ coefficient; Jaccard's similarity index.

1 Introduction

Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) is a complex disease whose patients manifest unexplained fatigue lasting for more than six months (Fukuda et al., 1994) or suffer from post-exertional malaise that is not alleviated by rest (Carruthers et al., 2003). Disease prevalence has been estimated between 0.4% and 1.0% affecting six women to one man (Morris and Maes, 2013).

16 The underlying pathological mechanisms remain poorly understood, but they are often associated
17 with environmental stressors, including severe viral infections (Rasa et al., 2018).

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

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

31 **2 The UKMEB**

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

38 analysis consists of a total of 523 participants divided into 275 suspected cases of ME/CFS, 136
39 healthy controls, and 112 patients with multiple sclerosis.

40 **3 Diagnostic agreement analysis**

41 After patients' referral for a possible integration in the biobank, suspected cases were comprehensively
42 evaluated according to four case definitions of ME/CFS: Centre for Disease Control criteria (CDC-
43 1994) (Fukuda et al., 1994), Canadian Consensus Criteria (CCC-2003) (Carruthers et al., 2003),
44 Institute of Medicine Criteria (IOM-2005) (Institute of Medicine (US), 2015), and International
45 Consensus Criteria (ICC-2011) (Carruthers et al., 2011). The CDC-1994 requires the patients to
46 have unexplained fatigue for at least 6 months and at least four out of eight fatigue-related symptoms.
47 The IOM-2005 is typically used by general practitioners and it requires the patients to show at
48 least three main symptoms such as profound fatigue, post-exertional malaise, and unrefreshing
49 sleep. The CCC-2003 requires the patients to manifest four or more fatigue specific symptoms, at
50 least two neurological or cognitive ones, and at least one autoimmune, neuroendocrine, or immune
51 symptom. Finally, the ICC-2011 is more focused on neuro-immune and cognitive symptoms, and
52 on the inability to produce sufficient energy on demand (post-exertional neuroimmune exhaustion).

53 There were 269 (97.8%), 233 (84.7%), 229 (83.3%) and 213 (77.5%) out of 275 suspected
54 cases whose symptoms agreed with CDC-1994, IOM-2005, CCC-2003, and ICC-2011, respectively
55 (Table 1). This finding suggests that the general practitioners who referred the suspected cases to a
56 possible integration in the biobank made their diagnosis based on the CDC-1994. Unsurprisingly,
57 only 62.9% of the suspected cases ($n = 173$) had a positive diagnosis across all the four case
58 definitions. Therefore, the remaining suspected cases had at least one negative diagnosis.

59 **TABLE 1.**

60 It is worth noting that there were no suspected cases who had a negative diagnosis across all
61 case definitions. There were also three individuals whose symptoms agreed with ICC-2011 only,
62 IOM-2005 only, or both criteria. These individuals were considered to be fatigued but non-ME/CFS
63 patients given that they did not agree with either the CDC-1994 or the CCC-2003 as recommended
64 for ME/CFS research (Pheby et al., 2020).

65 To better understand the agreement between diagnostic outcomes obtained from different
66 case definitions, we used the Jaccard's similarity index, J (Gower and Warrens, 2014). Note that
67 this index is usually a measure used to compare objects with shared attributes. Here we instead
68 applied this index to compare attributes themselves. For a pair of case definitions (C_i, C_j) , this
69 index was estimated as

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

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

74 The estimates of this index ranged from 0.752 (IOM-2005 versus ICC-2011) to 0.876
75 (CDC-1994 versus IOM-2005; CDC-1994 versus CCC-2003) (Table 2). The estimates showed the
76 stringency and differences in scope of each case definition. In addition, these estimates showed that,
77 even if the general practitioners applied two different case definitions of ME/CFS in their diagnosis,
78 there could still be a fraction of suspected cases where the respective diagnostic outcomes might not
79 agree with each other.

80 **TABLE 2.**

81 **4 Symptoms' similarity analysis**

82 A major advantage of using data from the UKMEB is the comprehensive symptom's characterisation
83 of all study participants. In particular, each participant had to report the severity of 57 symptoms
84 occurred a month before data collection. Severity of each symptom was categorised into absence,
85 mild, moderate, and severe. These invaluable data were then analysed to assess the similarity of
86 all participants in terms of their symptom's severity profile. With this purpose, we first computed
87 all possible 4×4 contingency tables resulting from cross-tabulating the symptom's severity data
88 for any given pair of participants (i, j) , $i, j = 1, \dots, 523$. We then calculated a similarity matrix
89 between any given pair of individuals by estimating the Cohen's κ coefficient (Agresti, 2002) in the
90 corresponding 4×4 contingency tables, that is,

$$\kappa_{ij} = \frac{\sum_{k=1}^4 p_{ij,kk} - \sum_{k=1}^4 p_{ij,k} \cdot p_{ij,\cdot k}}{1 - \sum_{k=1}^4 p_{ij,k} \cdot p_{ij,\cdot k}}, \quad (2)$$

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

96 With respect to the classical MDS, the first two components could explain 33.1% of the total
97 inertia (Figure 1A). More importantly, the first component clearly discriminated healthy controls
98 from suspected cases of ME/CFS. In the same component, patients with multiple sclerosis and the
99 three fatigued non-ME/CFS cases were located between these two groups with some overlap. As

100 expected, healthy participants were the most homogeneous cohort due to an absence or, at most, mild
101 severity of the different symptoms. In contrast, the suspected cases of ME/CFS consisted of a diverse
102 group as evidenced by their wide spread in the plot. Interestingly, a few suspected cases of ME/CFS
103 had symptom's severity profiles similar to the ones from healthy controls. In agreement with these
104 observations, the hierarchical cluster analysis revealed that some suspected cases of ME/CFS could
105 be placed in clusters together with healthy controls and patients with multiple sclerosis (Figure
106 1B); a detailed analysis on the optimal number of clusters will be done elsewhere. Therefore, it
107 was reasonable to assume that some of the suspected cases of ME/CFS, although agreeing with
108 CDC-1994 or CCC-2003, could be in fact true cases of another disease, as discussed by Nacul et al.
109 (2017).

110 *FIGURE 1.*

111 **5 Impact of misclassification on an association analysis**

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

$$f(x_0, x_1 | n_0, n_1; \theta_0, \theta_1) = \prod_{i=0,1} \binom{n_i}{x_i} \theta_i^{x_i} (1 - \theta_i)^{n_i - x_i}, \quad (3)$$

118 where x_0 and x_1 are the frequencies of exposed healthy controls and suspected cases, respectively,

119 n_0 and n_1 are the associated sample sizes, and θ_0 and θ_1 are the corresponding probabilities of
120 exposure in healthy controls and suspected cases.

121 To study the impact of a potential misclassification of suspected cases on the detection of a
122 possible association, four main assumptions were considered for the simulated data: (i) suspected
123 cases could be divided into apparent (or false positive) cases and true positive cases of ME/CFS;
124 (ii) the apparent cases were deemed equivalent to healthy controls in terms of degree of exposure,
125 i.e., the probability of exposure in these individuals was given by θ_0 ; (iii) there was an overall
126 misclassification rate, γ , for the suspected cases; and (iv) misclassification was only dependent
127 on the true clinical status of each suspected case. Under the assumption (ii) and the law of total
128 probability, the probability of exposure associated with suspected cases could be written as

$$\theta_1 = \gamma\theta_0 + (1 - \gamma)\theta_1^*, \quad (4)$$

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

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

135 We augmented the observable 2×2 frequency table where the suspected cases were subdivided
136 into apparent and true positive cases (Table 3). In this case, we simulated data from healthy controls
137 according to the Binomial distribution with a sample size of n_0 individuals and probability of
138 success θ_0 . With respect to the suspected cases, we simulated data from a Multinomial distribution
139 with a sample size of n_1 individuals and probability vector given by the probabilities shown in Table
140 3. Note that, given assumption (iv), the associated Multinomial distribution could be decomposed

141 into the following Binomial distribution

$$n_{1,m}|n_1; \gamma \rightsquigarrow \text{Bin}(n_1, \gamma), \quad (5)$$

142 referring to how many individuals were hypothetically misclassified as true positive cases, and two

143 Binomial distributions conditional to $n_{1,m}$

$$X_{1,F}|n_{1,m}; \theta_0 \rightsquigarrow \text{Bin}(n_{1,m}, \theta_0), \quad (6)$$

144 and

$$X_{1,T}|n_1 - n_{1,m}; \theta_1^* \rightsquigarrow \text{Bin}(n_1 - n_{1,m}, \theta_1^*), \quad (7)$$

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

146 and true positive cases, respectively.

147 *TABLE 3.*

148 For illustrative purposes, we performed our simulation study with $n_0 = n_1 = 100$, $\theta_0 = 0.25$,
149 and $\theta_1^* = 0.35$. According to this parameter specification, the odds ratio of true positive cases versus
150 healthy controls was 1.62, a low but reasonable value for a putative association with ME/CFS, given
151 that there is no disease-specific biomarker. To estimate the power of rejecting H_0 , we generated
152 10,000 data sets for each value of γ , ranging from 0 (no misclassification) to 1 (full misclassification)
153 with a lag of 0.01. In each data set, H_0 was rejected if the p-value of the Pearson's χ^2 test was less
154 than 0.05. For a given parameter set, power was finally estimated as the proportion of simulated
155 data sets in which H_0 was rejected.

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

165 *FIGURE 2.*

166 **6 Concluding remarks**

167 In summary, our analysis showed that suspected cases of ME/CFS from the UKMEB did not fully
168 agree with four main case definitions of the disease. In addition, some of these suspected cases
169 showed symptom's severity profiles similar to healthy controls and patients with multiple sclerosis.
170 These findings demonstrated the difficulty of diagnosing ME/CFS based on symptoms' assessment
171 alone. To overcome this and other difficulties, there are currently efforts for a stronger collaboration
172 among European researchers for accelerating the discovery of an objective disease-specific biomarker
173 (Scheibenbogen et al., 2017). However, joint efforts for biomarker discovery are very likely to
174 suffer from limited statistical power due to a possible misclassification of the suspected cases. A
175 possible solution to this problem is to take into account for misclassification in the respective
176 statistical analysis. Such a solution is also problematic because modelling misclassification leads to

177 an eventual problem of overparameterisation. From a frequentist standpoint, overparameterization
178 could be avoided by fixing the misclassification rate in a reasonable estimate for the sensitivity of
179 the diagnostic test. A more elegant way of doing so is to use Bayesian analysis where the prior
180 information about the misclassification rate takes the form of a probability distribution. However,
181 both frequentist and Bayesian solutions show a main hurdle for their implementation in the research
182 of ME/CFS. Given the lack of a disease biomarker, it is unclear which reasonable value or probability
183 distribution to choose for the sensitivity of current diagnostic tools of ME/CFS.

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

193 **Ethical approval**

194 All participants provided written informed consent for data collection (questionnaire, clinical
195 measurement and laboratory tests), and for allowing their samples to be available to any research
196 receiving ethical approval. Participants received an extensive information sheet and consent form in
197 which there was an option for participation withdraw from the study at any time. Ethical approval
198 was granted by the LSHTM Ethics Committee (Ref. 6123) and the National Research Ethics Service

199 (NRES) London-Bloomsbury Research Ethics Committee (REC ref. 11/10/1760, IRAS ID: 77765).

200 **Acknowledgements**

201 The authors acknowledge the CureME group for providing the data from the UKMEB. JM
202 acknowledges a PhD fellowship by the Fundação para a Ciência e Tecnologia, Portugal (ref.
203 SFRH/BD/149758/2019).

204 **References**

- 205 Agresti, Alan. 2002. *Categorical Data Analysis*. John Wiley & Sons, Inc.
- 206 Brurberg, Kjetil Gundro, Marita Sporstøl Fønhus, Lillebeth Larun, Signe Flottorp, and Kirsti
207 Malterud. 2014. “Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis
208 (CFS/ME): a systematic review.” *BMJ Open* 4:e003973.
- 209 Carruthers, Bruce M., Anil Kumar Jain, Kenny L. De Meirleir, Daniel L. Peterson, Nancy G. Klimas,
210 A. Martin Lerner, Alison C. Basted, Pierre Flor-Henry, Pradip Joshi, A. C. Peter Powles, Jeffrey A.
211 Sherkey, and Marjorie I. van de Sande. 2003. “Myalgic Encephalomyelitis/Chronic Fatigue
212 Syndrome.” *Journal of Chronic Fatigue Syndrome* 11:7–115.
- 213 Carruthers, B. M., M. I. van de Sande, K. L. De Meirleir, N. G. Klimas, G. Broderick, T. Mitchell,
214 D. Staines, A. C. P. Powles, N. Speight, R. Vallings, L. Bateman, B. Baumgarten-Austrheim, D. S.
215 Bell, N. Carlo-Stella, J. Chia, A. Darragh, D. Jo, D. Lewis, A. R. Light, S. Marshall-Gradisbik,
216 I. Mena, J. A. Mikovits, K. Miwa, M. Murovska, M. L. Pall, and S. Stevens. 2011. “Myalgic
217 encephalomyelitis: International Consensus Criteria.” *Journal of Internal Medicine* 270:327–338.
- 218 Fukuda, Keiji, S. E. Strausm, I. Hickie, M. C. Sharpe, J. G. Dobbins, and A. Komaroff. 1994. “The
219 Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study.” *Annals of*
220 *Internal Medicine* 121:953.
- 221 Gower, John C and Matthijs J Warrens. 2014. “Similarity, dissimilarity, and distance, measures of.”
222 *Wiley StatsRef: Statistics Reference Online* pp. 1–11.
- 223 Institute of Medicine (US), Committee on the Diagnostic Criteria for Myalgic Encephalomyeli-
224 tis/Chronic Fatigue Syndrome. 2015. *Beyond Myalgic Encephalomyelitis/Chronic Fatigue*
225 *Syndrome*. National Academies Press.
- 226 Lacerda, Eliana M., Kathleen Mudie, Caroline C. Kingdon, Jack D. Butterworth, Shennae O’Boyle,
227 and Luis Nacul. 2018. “The UK ME/CFS Biobank: A Disease-Specific Biobank for Advancing
228 Clinical Research Into Myalgic Encephalomyelitis/Chronic Fatigue Syndrome.” *Frontiers in*
229 *Neurology* 9.
- 230 Morris, Gerwyn and Michael Maes. 2013. “Myalgic encephalomyelitis/chronic fatigue syndrome
231 and encephalomyelitis disseminata/multiple sclerosis show remarkable levels of similarity in
232 phenomenology and neuroimmune characteristics.” *BMC Medicine* 11.
- 233 Nacul, Luís, Eliana M Lacerda, Caroline C Kingdon, Hayley Curran, and Erinna W Bowman.
234 2017. “How have selection bias and disease misclassification undermined the validity of
235 myalgic encephalomyelitis/chronic fatigue syndrome studies?” *Journal of Health Psychology*
236 24:1765–1769.

- 237 Paulino, Carlos Daniel, Paulo Soares, and John Neuhaus. 2003. “Binomial Regression with
238 Misclassification.” *Biometrics* 59:670–675.
- 239 Pheby, Derek F.H., Diana Araja, Uldis Berkis, Elenka Brenna, John Cullinan, Jean-Dominique
240 de Korwin, Lara Gitto, Dyfrig A Hughes, Rachael M Hunter, Dominic Trepel, and Xia Wang-
241 Steverding. 2020. “The Development of a Consistent Europe-Wide Approach to Investigating
242 the Economic Impact of Myalgic Encephalomyelitis (ME/CFS): A Report from the European
243 Network on ME/CFS (EUROMENE).” *Healthcare* 8:88.
- 244 Rasa, Santa, , Zaiga Nora-Krukle, Nina Henning, Eva Eliassen, Evelina Shikova, Thomas Harrer,
245 Carmen Scheibenbogen, Modra Murovska, and Bhupesh K. Prusty. 2018. “Chronic viral infections
246 in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS).” *Journal of Translational
247 Medicine* 16.
- 248 Scheibenbogen, Carmen, Helma Freitag, Julià Blanco, Enrica Capelli, Eliana Lacerda, Jerome
249 Authier, Mira Meeus, Jesus Castro Marrero, Zaiga Nora-Krukle, Elisa Oltra, et al. 2017.
250 “The European ME/CFS Biomarker Landscape project: an initiative of the European network
251 EUROMENE.” *Journal of translational medicine* 15:162.

252 **Figures**

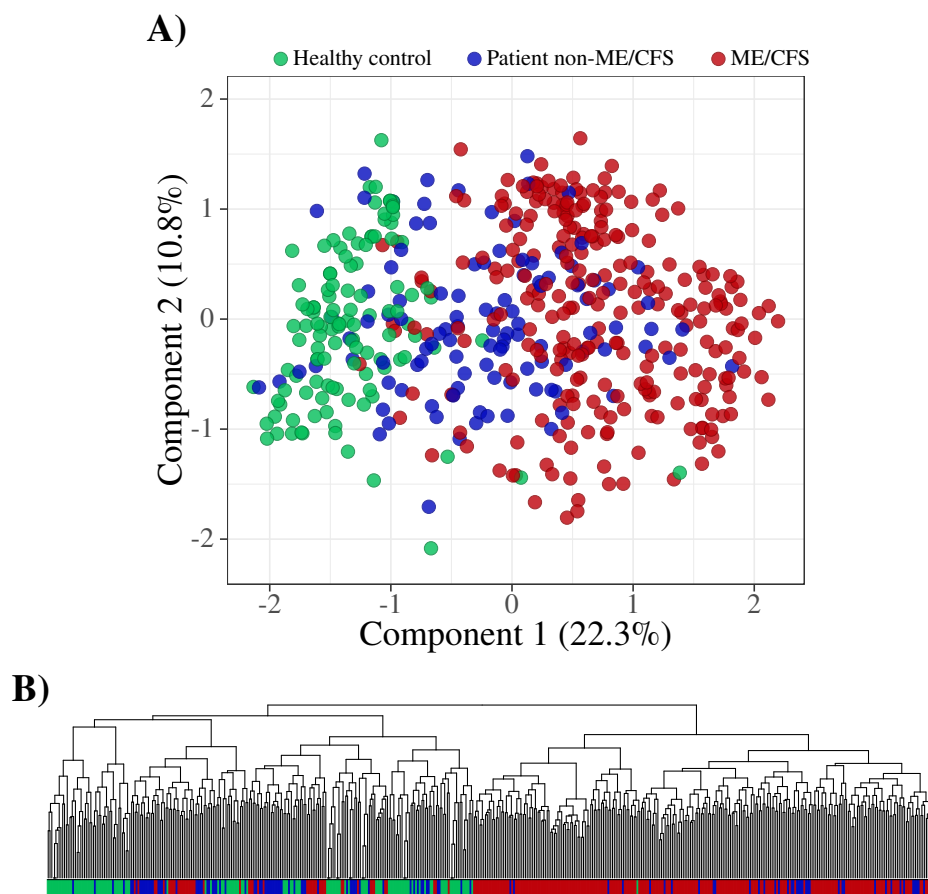


Figure 1: Symptom's similarity analysis based on the Cohen's κ coefficient: classical multidimensional scaling (A); dendrogram of hierarchical clustering analysis based on complete linkage (B) where the colour coding at the bottom is the same shown in A.

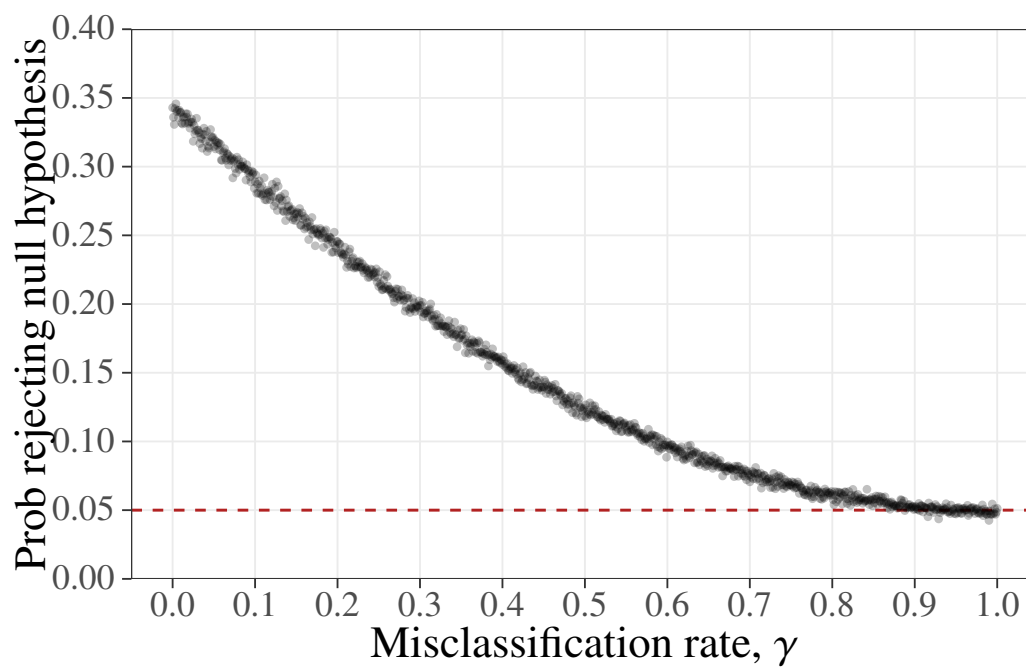


Figure 2: Estimated probability of rejecting H_0 (i.e., lack of association) as function of the misclassification rate γ .

Tables

Table 1: Frequency of suspected cases of ME/CFS according to their diagnostic outcomes using different case definitions. Percentages in the last row indicate the proportion of diagnosed cases by each case definition.

Case definition				N	% of total suspected cases
CDC-1994	IOM-2005	CCC-2003	ICC-2011		
+	+	+	+	173	62.9
+	+	+	-	32	11.6
+	-	+	+	16	5.8
+	+	-	+	16	5.8
+	-	-	-	14	5.1
+	+	-	-	10	3.6
+	-	+	-	5	1.8
+	-	-	+	3	1.1
-	-	+	+	3	1.1
-	-	-	+	1	0.4
-	+	-	-	1	0.4
-	+	-	+	1	0.4
97.8%	84.7%	83.3%	77.5%	275	100%

Table 2: Estimates of the Jaccard’s similarity index for the four case definitions of ME/CFS using data from the UKMEB.

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

Table 3: Augmented version of the observable 2×2 frequency table and the respective probabilities under a Binomial and a Multinomial distribution for healthy controls and suspected cases, respectively.

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