Psychological Medicine, 2005, **35**, 1337–1348. © 2005 Cambridge University Press
doi:10.1017/S0033291705005210 Printed in the United Kingdom

Chronic fatigue in a population sample: definitions and heterogeneity

PATRICK F. SULLIVAN[1]*, NANCY L. PEDERSEN[2,3], ANDREAS JACKS[4] AND BIRGITTA EVENGÅRD[4]

[1] *Departments of Genetics, Psychiatry and Epidemiology, University of North Carolina at Chapel Hill, NC, USA*; [2] *Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden*; [3] *Department of Psychology, University of Southern California, Los Angeles, CA, USA*; [4] *Department of Laboratory Medicine, Karolinska Institute, Stockholm, Sweden*

ABSTRACT

Background. Numerous nosological decisions are made when moving from the common human symptom of unusual fatigue to the rare chronic fatigue syndrome (CFS). These decisions have infrequently been subjected to rigorous evaluation.

Method. We obtained telephone interview data on fatiguing symptoms from 31 406 individuals twins in the Swedish Twin Registry aged 42–64 years; 5330 subjects who endorsed fatigue and possessed no exclusionary condition formed the analytic group. We evaluated the definition and classification of CFS-like illness using graphical methods, regression models, and latent class analysis.

Results. Our results raise fundamental questions about the 1994 Centers for Disease Control criteria as (1) there was no empirical support for the requirement of four of eight cardinal CFS symptoms; (2) these eight symptoms were not equivalent in their capacity to predict fatigue; and (3) no combination of symptoms was markedly more heritable. Critically, latent class analysis identified a syndrome strongly resembling CFS-like illness.

Conclusions. Our data are consistent with the 'existence' of CFS-like illness although the dominant nosological approach captures population-level variation poorly. We suggest that studying a more parsimonious case definition – impairing chronic fatigue not due to a known cause – would represent a way forward.

INTRODUCTION

This one of a series of papers that describes the epidemiology and genetic epidemiology of fatigue-related illness in the large, population-based Swedish Twin Registry (STR) (Evengard *et al*. 2005). In this paper, the focus is on definitional issues and heterogeneity in individuals aged 42–64 years.

The prevalence of chronically fatiguing illness ranges by over an order of magnitude, from the rare (0·235%) chronic fatigue syndrome (CFS) in population-based studies (Reyes *et al*. 2003) to a common constellation of symptoms in primary care (9·0%) (Skapinakis *et al*. 2003). Fatiguing illnesses are widely assumed to possess considerable phenotypic and etiological heterogeneity. These assumptions are relatively strongly supported by clinical studies using diverse multivariate statistical approaches (Hickie *et al*. 1995; Kirk *et al*. 1999; Nimnuan *et al*. 2001; Sullivan *et al*. 2003) and by the prevailing consensus criteria for CFS (Fukuda *et al*. 1994; Reeves *et al*. 2003). Not accounting for both types of heterogeneity has multiple adverse consequences for investigations of the causes and treatments for CFS.

* Address for correspondence: Dr P. F. Sullivan, Department of Genetics, CB#7264, 4109D Neurosciences Research Building, University of North Carolina, Chapel Hill, NC 27599-7264, USA.
(Email: pfsulliv@med.unc.edu)

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

The Centers for Disease Control (CDC-94) criteria for CFS (Fukuda *et al*. 1994; Reeves *et al*. 2003) have been particularly influential and have provided a unifying approach that has fostered communication and diagnostic precision, particularly in the research setting. Despite these clear advantages, the CDC-94 criteria for CFS rely heavily on expert opinion, and have been repeatedly criticized as lacking in empirical support (Lewis & Wessely, 1992; Buchwald, 1996; Levine, 1997; Komaroff & Buchwald, 1998). For example, the combinations of eight ancillary symptoms (e.g. muscle pain, post-exertional malaise) have seldom been examined rigorously. The CDC-94 criteria implicitly assume that the ancillary symptoms are equivalent, act additively, and that subjects with from zero to three total criteria are qualitatively different from subjects with four to eight criteria.

The present report had three broad goals that addressed the definition and heterogeneity of chronic fatigue in a large, population-based sample of twins. First, we sought to evaluate the distributional and predictive properties of the basic CDC-94 criteria for CFS. Second, we explored the phenotypic heterogeneity of these criteria using latent class analysis. Finally, using heritability (Neale & Cardon, 1992; Plomin *et al*. 2003) as an external validator (Robins & Guze, 1970), we attempted to discover the item combinations that maximized heritability, reasoning that a markedly higher heritability of some item combination might identify an etiologically important illness subtype. All three goals serve the overarching aim of attempting to delineate different aspects of heterogeneity in this complex, intriguing, and costly human syndrome.

METHOD

Swedish Twin Registry

The Swedish Twin Registry (STR, 2005) is the largest population-based registry of twin births in the world (Lichtenstein *et al*. 2002; Pedersen *et al*. 2002). Data on current vital status, marital status, address, and place of birth are available on $>99\%$ of the 180 000 individuals in the registry along with record linkage to Swedish national cancer and inpatient hospitalization registries.

Interview procedures

Over a 4-year period ending in December 2002, we screened all living, contactable, interviewable, and consenting twins born in Sweden on or before 31 December 1958 for a range of disorders that included chronic fatigue. Due to the confounding influence of aging, questions about chronic fatigue were asked only of twins aged 42–64 years (i.e. born between 1 January 1935 and 31 December 1958). The interview procedure had the following steps. Each month, the master twin register was matched to national records to update vital status and current address information, $\sim$1000 pairs were randomly selected for interviews and were sent introductory letters describing the study, and, about 2 weeks later, we then attempted to contact and interview these twins via telephone. All interviews were conducted by trained lay interviewers using a computer-based data collection system.

Assessment of chronic fatigue

When the telephone interview for chronic fatigue was designed in 1996–1997, no generally recognized assessment instrument was available. Therefore, we designed a screening module for chronic fatigue that emulated closely the CDC-94 consensus criteria for chronic fatigue syndrome (Fukuda *et al*. 1994). The following data were collected. The stem question was 'have you felt abnormally tired during the last 6 months?' and was used to define 'fatigue'. The time-frame was the 6 months prior to interview, as assessment of lifetime fatigue was believed to be considerably less reliable. Only subjects who endorsed this item were asked further questions. Subjects were then asked about the continuousness of fatigue in the prior 6 months and about the duration of continuous fatigue. Impairment was considered present if subjects considered themselves 'too tired to live a normal life', that fatigue had caused social problems, or that fatigue had caused $\geqslant 25\%$ work incapacity. Finally, subjects were asked about eight ancillary symptoms during the period of abnormal tiredness: substantial impairment in short-term memory or concentration; sore throat; tender lymph nodes; muscle pain; multi-joint pain without swelling or redness; headaches of a new type, pattern, or severity; unrefreshing sleep;

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

and post-exertional malaise lasting more than 24 hours. The presence of four or more of these ancillary symptoms are a component of the definition of chronic fatigue syndrome (Fukuda *et al*. 1994).

Fatigue-related definitions

We defined fatigue as the presence of self-reported abnormal tiredness in the absence of an exclusionary condition. Exclusionary conditions were determined from multiple sources.

(1) Exclusions from the telephone interview: morbid obesity (self-reported current Quetelet's body mass index $\geqslant 40$ m/kg^2); lifetime history of anorexia nervosa or bulimia nervosa (APA, 1994) via the Structured Clinical Interview for DSM-III-R (SCID; Spitzer *et al*. 1992); 'alcohol abuse' via the SCID (Spitzer *et al*. 1992) [operationalized as lifetime DSM-IV alcohol dependence (APA, 1994) including both tolerance and withdrawal and with the subject endorsing alcohol consumption in the past year]; and self-reported history of congestive heart failure, stroke, emphysema, Parkinson's disease, multiple sclerosis, polio, systemic lupus erythematosis, Crohn's disease, or ulcerative colitis.

(2) Exclusions from Swedish national registers: inclusion in the Swedish National Cancer Registry (registration only for cervical cancer *in situ* and non-melanoma skin cancers did not lead to exclusion) and any hospitalization with a discharge diagnosis corresponding to narrow definitions of schizophrenia, schizoaffective disorder, and bipolar disorder.

(3) Exclusions via physician review of all available medical records that revealed the presence of any other exclusionary diagnosis. These included: drug or alcohol dependence; a sleep disorder with adequate work-up; infection with hepatitis B or C or HIV; neurological disorders including multiple sclerosis and myasthenia gravis; significant rheumatological disorders (e.g. rheumatoid arthritis or systemic lupus erythematosis); chronic obstructive pulmonary disease; significant endocrine disorder; or inflammatory bowel disease (see Evengard *et al*. 2005, for details).

Prolonged fatigue was the presence of fatigue with no exclusionary condition plus duration of $\geqslant 1$ month. Definition A of chronic fatigue (CF-A) was the presence of fatigue (and no exclusions) plus duration $\geqslant 6$ months, CF-B was defined as the prior definition plus impairment, and CF-C was the prior definition plus the presence of $\geqslant 4$ or 8 specific symptoms.

Assessment of additional conditions

Lifetime DSM-III-R major depression (APA, 1987) was assessed during the telephone interview via an adaptation of the Composite International Diagnostic Interview (WHO, 1990). Chronic widespread pain (akin to fibromyalgia but without the required physical exam) was assessed during the telephone interview via an adapted version of the American College of Rheumatology criteria for fibromyalgia (Wolfe *et al*. 1990).

Reliability

To establish test–retest reliability, 105 individual twins who completed the telephone screening were selected at random and reinterviewed within 2 weeks. Cohen's κ (Cohen, 1960) for the presence of fatigue was 0·69 [95% confidence interval (CI) 0·49–0·89], 0·75 (95% CI 0·56–0·94) for prolonged fatigue, and 0·76 for chronic fatigue (95% CI 0·53–0·99). These κ values suggest very good to excellent test–retest reliability (Schlesselman, 1982).

Statistical analyses

Descriptive analyses were conducted with SAS (SAS Institute Inc., 1999) and S-Plus (MathSoft, 2000). Prediction of classification stringency was conducted using logistic regression with generalized estimating equations (GEE) to account for sample clustering implicit in ascertaining twin pairs (Zeger & Liang, 1986; Zeger *et al*. 1988; Zeger & Liang, 1992).

We used latent class analysis to investigate heterogeneity. Latent class analysis is unsupervised (i.e. it is given no prior knowledge about how symptoms might interrelate) and can be considered akin to cluster analysis (do the observed data arise from a single homogeneous cluster or from multiple different clusters?). To conduct latent class analysis (McCutcheon, 1987; Yang & Becker, 1997; Sullivan *et al*. 2003) we used a FORTRAN program (Eaves *et al*. 1993; Bucholz *et al*. 1996) that employed an efficient estimization-maximization algorithm (Dempster *et al*. 1977). To determine the optimal number of latent classes, we fit 1, 2, 3, … , 10 latent class models to the data (with 50 runs within each

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

class, using randomized starting values) and used the Bayesian Information Criterion to select the number of latent classes. Finally, we calculated all possible linear combinations of the central CFS criteria, computed the tetrachoric or polychoric correlations separately for monozygotic and dizygotic twin pairs, and estimated heritability as $a^2 = 2(r_{MZ} - r_{DZ})$. For the univariate discrete case, this simple approach functions well and has improved accuracy in comparison with parsimony-based modeling (Sullivan & Eaves, 2002).

RESULTS

Sample screening

Of all eligible twins ($n = 41\,499$), 31 406 individual twins responded, giving an individual response rate of 75·68 %. Data were obtained from both members of 12 407 pairs and from one member of 6592 pairs. Of the complete pairs, 3269 pairs were monozygotic, 9010 pairs dizygotic, and 128 pairs of unknown zygosity (Lichtenstein *et al.* 2002).

Predictors of participation

We had access to data on all eligible twins independent of participation and were thus able to examine the predictors of participation. Significant predictors of participation in a multivariate logistic regression model were: being a monozygotic twin [odds ratio (OR) 1·68, $p < 0·0001$], having a co-twin who was also eligible for the interview (OR 1·44, $p < 0·0001$), female gender (OR 1·39, $p < 0·0001$), fewer total hospitalizations (dichotomized at the 75th percentile of four hospitalizations, OR 1·18, $p < 0·0001$), age (OR 1·007, $p = 0·0001$) but not registration in the Swedish Cancer Registry (OR 1·01, $p = 0·85$). Given the large sample size, some strongly significant variables are not of singular import; for example, the R^2 for age was only 0·02 % and the mean difference in age between participants and non-participants was only about 86 days.

Descriptive analyses

Abnormal tiredness in the prior 6 months was endorsed by 6571 individual twins (20·92 %), denied by 24 467 twins (77·91 %), and 368 twins provided no usable answer (1·17 %). Of the individuals endorsing fatigue, 5330 (81·11 %) had none of the exclusions defined above and 1241 subjects (18·89 %) had one or more exclusionary condition as determined by self-reported medical history, findings in national registers, or physician evaluation of medical records.

The 5330 subjects who endorsed fatigue and who possessed no exclusionary condition are the analytic group for the remainder of this report. Our analytic goals required data on fatiguing symptoms which were available only for these 5330 subjects. These individuals were predominantly female (64·17 %). The duration of fatigue was <1 month for 28·07 %, 1 to <6 months for 23·68 %, and $\geqslant 6$ months for 48·11 % of fatigued subjects. Significant impairment was reported by 60·54 % of fatigued subjects. The mean number of ancillary symptoms was 2·40 (S.D. 1·61) and the median was 2·0 (interquartile range 1–3). The four most common ancillary symptoms were unrefreshing sleep (79·87 %), memory/concentration impairment (48·48 %), muscle pain (37·26 %), and joint pain (28·26 %). The four least common symptoms were tender lymph nodes (6·89 %), headaches of a new type (11·11 %), sore throat (11·99 %), and post-exertional malaise (15·76 %).

Fig. 1 depicts the influence of the total number of ancillary symptoms on a number of symptomatic characteristics. The distributions lack any consistent or notably sharp inflections as implied by the CDC-94 criteria. Panel (*a*) shows the proportion of subjects with 0, 1, …, 8 ancillary symptoms. Panel (*b*) shows the mean number of months of fatigue for individuals with zero through eight ancillary symptoms. With the exception of a minor dip for individuals with seven symptoms, the mean months of fatigue increases nearly linearly. Similarly, in panel (*c*), the proportion of subjects with impairment related to fatigue increases approximately linearly with the number of ancillary symptoms, as does the proportion of women [panel (*d*)]. Panels (*e–l*) depict the proportion of subjects with each specific ancillary symptom by the total number of ancillary symptoms. By definition, each of these panels is anchored at 0 % for subjects with zero total symptoms and at 100 % for subjects with eight total symptoms. The relationship with total number of symptoms is approximately linear. The possible exception is the most common symptom [unrefreshing sleep, panel (*k*)] which has a sharp increase from

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

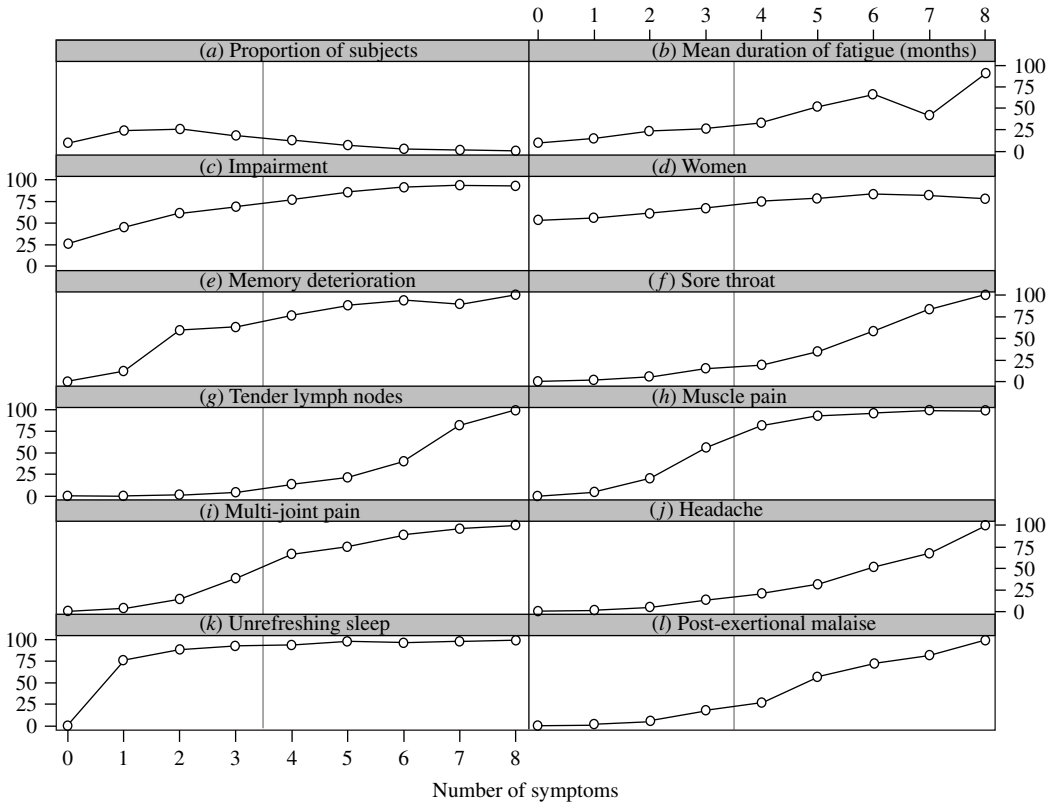


FIG. 1. Trellis graph depicting the relationships with the 0, 1, 2, …, 8 cardinal chronic fatigue syndrome symptoms. A vertical reference line is shown at the Centers for Disease Control (CDC) cut-point between three and four total symptoms. Panels (*a*)–(*d*) are essentially simply histograms [e.g. panel (*c*) shows the proportion of subjects with no symptoms, two symptoms, etc.]. Panel (*a*) shows the proportion of subjects. Panel (*b*) depicts the mean duration of fatigue in months. Panel (*c*) shows the proportion of females. Panels (*e*)–(*l*) depict the proportions of subjects with each of the individual symptoms and are more akin to cumulative distributions. By definition, Panels (*e*)–(*l*) are fixed at 0% for zero total symptoms (as no subject with zero symptoms can have, for example, memory deterioration) and 100% for eight total symptoms (as all subjects with eight symptoms must have, for example, sore throat).

zero to one total ancillary symptoms and approximately linear changes thereafter.

The CDC criteria for chronic fatigue syndrome (Fukuda *et al.* 1994) specify that four or more of eight ancillary symptoms are required for a positive diagnosis. This cut-point implies the existence of a 'point of rarity' (Kendell & Brockington, 1980; Kendell, 1989) between three and four total ancillary symptoms (indicated by the vertical gray reference line in Fig. 1) that separates normalcy from pathology. However, there is no evident abrupt increase or discontinuity in any of the panels in Fig. 1 at any cut-point. We obtained a similar pattern of results when we limited the descriptive analyses only to subjects with definition CF-A (unexplained fatigue of $\geqslant$6 months duration, data not shown).

Multivariate prediction

We next considered whether we could identify predictors of increased classification stringency in a logistic regression model (Table 1). The rows show the four dependent variables (the presence or absence of prolonged fatigue, CF-A, CF-B, or CF-C). The columns list the eight ancillary symptoms in the CDC definition of CFS (Fukuda *et al.* 1994). Note that the CF-C definition requires a certain number of ancillary symptoms which are considered independent variables in the lower row; thus, these odds ratios are artifactually inflated.

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

Table 1. *Prediction of increased classification stringency using logistic regression (with GEE)*

Dependent variable	Memory deterioration	Muscle pain	Multi-joint pain	Unrefreshing sleep	Post-exertional malaise	Tender lymph nodes	Headache	Sore throat
Prolonged fatigue (fatigue of ⩾1 month duration)	1·76 (1·55–2·01)	1·38 (1·19–1·61)	1·64 (1·39–1·94)	1·37 (1·18–1·59)	1·39 (1·14–1·70)	0·87 (0·67–1·14)	1·19 (0·96–1·48)	0·70 (0·57–0·85)
CF-A (fatigue of ⩾6 months' duration)	1·90 (1·69–2·14)	1·67 (1·46–1·91)	1·75 (1·52–2·02)	1·40 (1·21–1·62)	1·26 (1·07–1·48)	1·02 (0·80–1·29)	1·04 (0·86–1·25)	0·65 (0·54–0·78)
CF-B (add impairment)	2·56 (2·25–2·91)	1·77 (1·53–2·04)	1·83 (1·58–2·12)	1·78 (1·49–2·12)	1·80 (1·52–2·12)	1·12 (0·87–1·44)	1·25 (1·03–1·51)	0·71 (0·58–0·87)
CF-C (add ⩾4 of 8 symptoms	11·3 (8·30–15·2)	8·92 (6·80–11·7)	7·54 (5·85–9·71)	8·77 (5·10–15·09)	5·61 (4·24–7·44)	2·26 (1·40–3·62)	3·61 (2·58–5·03)	1·47 (1·00–2·16)

Shown above are the odds ratios (95% confidence intervals) from logistic regression. Generalized estimating equations (GEE) were used to control for the non-independence of twin pairs. Four regression models were run corresponding to the rows in the table. Each model had eight independent variables corresponding to the ancillary symptoms included in the 1994 CDC definition of chronic fatigue syndrome.

The ancillary symptoms of memory deterioration, muscle pain, and multi-joint pain performed well. The OR estimates for each increased with classification stringency and never included unity, suggesting their utility. The pattern for unrefreshing sleep and post-exertional malaise was similar although the ORs did not evidence an increasing trend over all definitions of fatiguing illness. In contrast, tender lymph nodes and headache did not robustly predict fatigue classification. Finally, the presence of the sore throat ancillary symptom was associated with the *absence* of three of the four definitions of fatiguing illness and was not a significant predictor for CF-C.

Latent class analysis

We entered 11 dichotomous items into the latent class analysis (gender, impairment due to fatigue, duration ⩾6 months, and the eight ancillary symptoms). In the data, we observed 679 different patterns out of a theoretical maximum of 2048 patterns (2^{11}, 33·2%). The data suggested five latent classes (via the Schwarz Bayesian Criterion). The observed class-item frequencies are shown in the upper portion of Table 2. These values describe a particular class in terms of the 11 items entered into latent class analysis. Class-item probabilities that deviate from those of the entire sample by ⩾15% are noted by arrows. The middle and lower portions of Table 2 provide additional data about the latent classes through variables not used in the latent class analysis.

Class 1

Class 1 (14% of the fatigued sample) consisted principally of women who had relatively high endorsement rates for the eight ancillary symptoms (mean of 5·03 symptoms) with near-universal impairment and long duration of fatigue. Most individuals in Class 1 met the CF-C definition of illness and many met criteria for a lifetime major depressive episode and chronic widespread pain. We labeled this class 'CFS-like' given its resemblance to clinical depictions of individuals with chronic fatigue syndrome.

Class 2

Class 2 (29% of the sample) had a predominance of males and relatively few symptoms (mean 0·86). The main symptom was unrefreshing sleep, and the prevalence of major depression and chronic widespread pain was relatively low. We termed this the 'Residual' class.

Class 3

Class 3 (17% of the sample) contained individuals with greater muscle and multi-joint pain of long duration. Relatively few of these individuals met criteria for CF-C (11%) although this class had the second-highest proportion with chronic widespread pain. We labeled this class 'Rheumatic'.

Class 4

Class 4 (35% of the sample) consisted of individuals with memory deterioration, impairment,

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

Table 2. *Latent class analysis of 11 fatigue-related characteristics*

	Fatigued sample	Class 1	Class 2	Class 3	Class 4	Class 5
Class probability	1·00	0·14	0·29	0·17	0·35	0·05
Class label/mnemonic	—	CFS-like	Residual	Rheumatic	Depressive	APS
Female gender	0·62	0·87 ↑	0·44 ↓	0·65	0·65	0·61
Memory deterioration	0·49	0·98 ↑	0·09 ↓	0·33 ↓	0·74 ↑	0·33 ↓
Sore throat	0·12	0·27 ↑	0·04	0·03	0·05	0·94 ↑
Tender lymph nodes	0·07	0·22 ↑	0·00	0·03	0·02	0·51 ↑
Muscle pain	0·38	0·95 ↑	0·09 ↓	0·79 ↑	0·14 ↓	0·69 ↑
Multi-joint pain	0·29	0·75 ↑	0·05 ↓	0·79 ↑	0·04 ↓	0·40
Headache	0·11	0·35 ↑	0·02	0·07	0·11	0·19
Unrefreshing sleep	0·80	0·97 ↑	0·56 ↓	0·85	0·91	0·81
Post-exertional malaise	0·16	0·55 ↑	0·02	0·14	0·12	0·26
Impairment	0·61	0·99 ↑	0·16 ↓	0·64	0·84 ↑	0·57
Duration ⩾6 months	0·49	0·87 ↑	0·25 ↓	0·67 ↑	0·51	0·17 ↓
Fatigue classification						
Prolonged fatigue	0·73	0·94	0·54	0·83	0·78	0·51
CF-A	0·49	0·87	0·25	0·67	0·51	0·17
CF-B	0·36	0·86	0·02	0·44	0·44	0·08
CF-C	0·14	0·86	0·00	0·11	0·01	0·04
Mean number of symptoms	2·4 (1·6)	5·03 (1·01)	0·86 (0·69)	3·03 (0·89)	2·14 (0·82)	4·13 (1·20)
Mean months of fatigue (S.D.)	27·3 (60·5)	62·1 (84·8)	10·2 (36·6)	38·8 (69·8)	24·2 (55·3)	7·65 (29·1)
Validators						
Mean age at interview (S.D.)	52·5 (5·6)	52·2 (5·35)	52·9 (5·81)	53·4 (5·44)	51·9 (5·53)	52·7 (5·43)
Major depressive episode in lifetime	0·37	0·57	0·20	0·35	0·44	0·31
Chronic widespread pain	0·08	0·30	0·01	0·12	0·03	0·04

Shown in the top panel are the class-item probabilities from a five-latent-class solution. The lower two panels show potential validators of the latent class solution relating to fatigue (middle panel) and characteristics of relevance to chronic fatigue (lower panel) where the values shown are proportions of the class or mean (S.D.) as appropriate. APS, Acute physical syndrome.

and lesser probabilities of muscle and multi-joint pain. Few met the CF-C criteria, although a substantial proportion met criteria for a lifetime major depressive episode (44%). We termed this class 'Depressive'.

Class 5

Class 5 (5% of the sample) had higher probabilities of sore throat, tender lymph nodes, and muscle pain and lower probabilities of memory deterioration and long duration. In some respects, these individuals resemble an acute viral illness leading to the label of 'Acute Physical Syndrome'.

Item heritability

To investigate the possibility that some subset of the eight ancillary symptoms possessed notably higher heritability, we calculated all possible linear combinations of the eight ancillary symptoms (total 2^8-1 or 255). There were eight one-item combinations, 28 two-item combinations, 58 three-item combinations, 70 four-item combinations, 58 five-item combinations, 28 six-item combinations, 8 seven-item combinations, and 1 eight-item combination. In Fig. 2, we depict the heritability by the number of items for female-female and male-male twin pairs. Each point shows the heritability of a unique combination of items.

The Spearman correlation between the heritabilities of female and male twin pairs was modest (Spearman $\rho=0{\cdot}25$, $p<0{\cdot}0001$). No particular combination of items had markedly higher heritabilities. The possible exceptions were the single items of multi-joint pain in females (heritability 49%) and males (67%), the combination of sore throat-new headaches in males (80%), and sore throat-tender lymph nodes-new headaches in males (68%). The magnitude of these point estimates has to be considered against the wide confidence intervals on each estimate (not shown).

DISCUSSION

Unusual fatigue is a common symptom in twins aged 42–64 years from the STR. This finding is

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

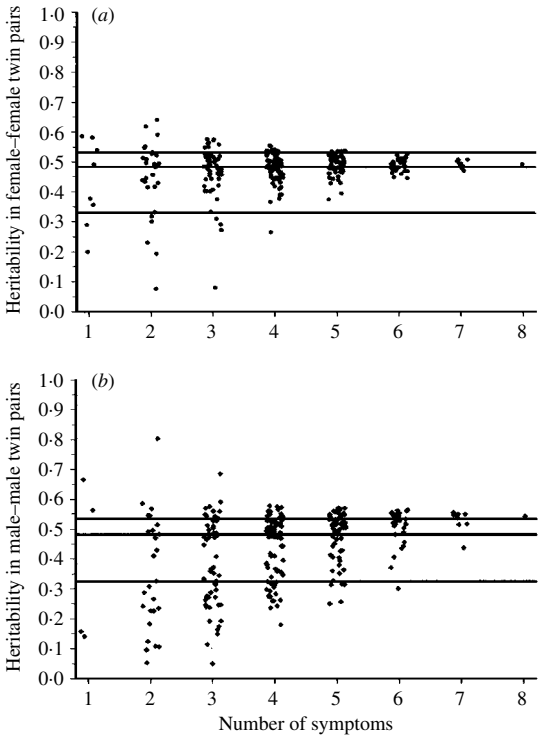


FIG. 2. Depicted are the estimated heritabilities for all possible combinations of items. (*a*) Data for female–female twin pairs; (*b*) male–male twin pairs. Horizontal reference lines show the overall 75th, 50th, and 25th percentiles for each gender.

similar to that found in other samples worldwide (David *et al.* 1990; Lewis & Wessely, 1992; Pawlikowska *et al.* 1994; Lawrie & Pelois, 1995; Chester, 1997; Jason *et al.* 1999). However, in moving from a common symptom (unusual fatigue) to an uncommon syndrome (CFS-like illness or CFS itself), numerous decisions are required. The overarching question addressed by this report examines the performance of the CDC-94 criteria for CFS (Fukuda *et al.* 1994; Reeves *et al.* 2003) in delineating normal from abnormal states of fatigue.

In brief, our results suggest that the CDC-94 criteria have fundamental flaws for the classification of fatigue in a population-based sample. On the other hand, a critically important result was in the identification of a syndrome resembling CFS-like illness using a relatively unbiased classification strategy. The latter result, in our view, is an important finding that validates the existence of CFS-like illness. The good news is that CFS-like illness appears to 'exist' (and only partially overlaps with major depression and chronic widespread pain). The bad news is that the predominant international classification criteria do not function well: (1) there was no empirical support for the requirement of four of eight cardinal CFS symptoms; (2) the eight symptoms were not equivalent in their capacity to predict fatigue; and (3) no combination of symptoms was markedly more heritable.

Definitional issues

In order to delineate the common symptom of fatigue from uncommon CFS, a number of decisions are required and a key feature of the dominant CDC-94 criteria (Fukuda *et al.* 1994; Reeves *et al.* 2003) is the required endorsement of four of eight ancillary symptoms whose validity rests on a number of implicit assumptions.

First, the four or more of eight criterion assumes that there is an important difference between individuals with unexplained chronic fatigue with zero to three *versus* four to eight of the ancillary criteria. The criteria imply that there is a 'point of rarity' (Kendell & Brockington, 1980; Kendell, 1989) between a total of three and four ancillary criteria that 'carves nature at the joint'. This assumption was not supported by our data. If this assumption were true, discontinuities should have been evident in several panels in Fig. 1. For the proportion of subjects [panel (*a*)], the proportion of subjects with impairment [panel (*c*)], and essentially all of the specific symptoms, there were no marked differences between subjects with three and four total symptoms. Moreover, the data did not suggest any reasonable alternative cut-point.

Second, the criteria assume that the ancillary symptoms are fundamentally interchangeable and thus similar in their impact on the probability of CFS. Simple prevalence data do not support this assumption as endorsement proportions for the ancillary symptoms ranged from ~7% (tender lymph nodes) to over 80% (unrefreshing sleep). More direct assessment of the adequacy of the CFS ancillary symptoms to predict fatigue classification is shown in Table 1. Three symptoms performed well (memory deterioration, muscle pain, and multi-joint pain), two performed adequately (unrefreshing sleep and post-exertional malaise), and three performed poorly (tender lymph nodes, headache, and sore throat). These data are consistent with

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

prior studies (Wessely *et al.* 1996) and suggest that the latter three ancillary symptoms do not strongly assist in the delineation of more extreme types of fatiguing illness. Notably, the sore throat ancillary symptom significantly predicted the *absence* of several definitions of chronic fatiguing illness.

Latent class analysis of heterogeneity

We next investigated the heterogeneity of chronic fatiguing illness using latent class analysis (Table 2). We believe that these results are of particular interest.

First, latent class analysis identified a class (class 1) in this epidemiological sample that strongly resembles clinical samples with CFS. This statistical technique is not influenced by nosological assumptions; we emphasize that the identification of this class was not strongly influenced by any prior beliefs about the nature of CFS or its definition. Individuals in this CFS-like class – comprising 2·4% of the entire sample – endorsed a large number of ancillary symptoms (mean of 5), nearly all had some degree of impairment, and a majority had longer duration of fatigue. In addition, there was a female predominance and substantial but not universal co-morbidity with major depression and chronic widespread pain. The ancillary symptoms identified as performing poorly in the analyses above were not highly prevalent in this class. The CFS-like class has features highly similar to clinical samples with CFS and its identification in a population-based sample supports its existence as a syndromic entity.

Second, the remaining classes provide some insight into the heterogeneity of fatigue in a population sample. Class 5 ('Acute physical syndrome') endorsed symptoms reminiscent of a time-limited acute viral syndrome (fatigue, sore throat, muscle pain, lymphadenopathy, and briefer duration). The presence of this class was at least part of the reason why the sore throat criterion predicted the absence of various definitions of fatigue. These individuals could presumably be found in samples ascertained from medical primary care settings but would be less likely in a tertiary CFS clinic.

In latent class analyses, it is not uncommon to find a class characterized by relatively few symptoms such as Class 2 ('Residual'). These individuals endorsed the unusual fatigue item but then had few commonalities other than a low number of ancillary symptoms, male gender, and briefer duration of fatigue. It is possible that this class represents individuals impacted by time-limited but diverse stressors that prompted endorsement of unusual fatigue. The remaining two classes seemed to represent common complaints in middle-aged population samples. Class 3 ('Rheumatic', i.e. syndromic stiffness or pain of the extremities or back) endorsed muscle and multi-joint pain of relatively long duration. Few met the criteria for the most extreme definition of fatigue although a minority met criteria for chronic widespread pain. It is likely that this class consists of individuals with diverse musculoskeletal complaints such as osteoarthritis, rheumatoid arthritis, fibromyalgia, and even non-pathological aches and pains associated with middle age. Finally, Class 4 ('Depressive') had a relatively high prevalence of lifetime major depression, memory deterioration, and impairment. If these individuals were older, one might expect the impact of a dementing illness but this is less likely in this cohort.

A number of prior studies have used factor analysis, latent class analysis, and latent profile analysis to attempt to delineate heterogeneity. Direct comparison of the results of these studies to that presented here is difficult owing to the different samples and item data entered into the multivariate statistical technique. One of the four studies that used latent class analysis found a three class solution in a CFS clinical sample that resembled the CFS-like, Rheumatic, and Residual classes presented here (Hadzi-Pavlovic *et al.* 2000). The items used in two prior latent class analyses were not comparable with those presented here (Sullivan *et al.* 2002; Sullivan *et al.* 2003). Insufficient data were presented to evaluate confidently the two class solutions from a latent class analysis (Hickie *et al.* 1995) and a latent profile analysis (Wilson *et al.* 2001). Factor analysis is rather different conceptually from latent class analysis but one study from primary care identified factors reminiscent of fibromyalgia, chronic fatigue, somatic depression, somatic anxiety, and irritable bowel syndrome (Robbins *et al.* 1997). Other studies identified factors akin to the acute physical syndrome (Nisenbaum *et al.* 1998) and the Rheumatic class (Bourdette *et al.* 2001) found here.

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

Heritability as a validator

Heritability is often considered as a critical external validator (Robins & Guze, 1970). Indeed, clarifying the heritable essence of a phenotype is often a strong motivation of modern twin and family studies (provided one is willing to assume the unimportance of shared familial environmental effects). We approached this issue via a 'brute force' technique whereby we computed all possible linear combinations of the eight ancillary items to determine the item combinations that maximized the difference of the correlations of monozygotic and dizygotic twin pairs. No combination of items stood out as markedly more heritable. Therefore, using these eight items, we were not successful in employing heritability as a phenotypic validator. A future report will describe the twin modeling results for various definitions of fatiguing illness.

Suggestions for the definition of CFS-like illness

The broad conclusions from these results dovetail to suggest an alternative approach to the definition of CFS-like illness. First, it is clear that empirical evaluation of the criteria to be used for the definition of a complex trait should be conducted prior to its publication and widespread adoption. We do not mean to take the framers of the CDC-94 criteria specifically to task on this issue; expert consensus is a complex process with numerous hazards and advantages and we note that certain of the ICD-10 (WHO, 1993) and DSM-IV criteria (APA, 1994) are open to the same criticism.

Second, the assumptions behind the four or more of eight ancillary symptoms criterion appear to be incorrect. The items themselves are not equivalent. Several items do not perform as expected. Most important, there appears to be no inherent difference between people with between zero and three criteria and between four and eight criteria.

Finally, there is an obvious alternative definitional strategy that could be used. If the ancillary criteria were dropped entirely, the definition of CFS-like illness could simply focus on impairing chronic fatigue not due to a known physical cause. This definition is more parsimonious and isolates the most salient public health issue. Such an alternative definition would also be far easier for practitioners to remember and apply more precisely. For researchers, there are two key advantages: the ability to target larger numbers of subjects and the avoidance of the assumption that imperfectly reported clinical symptoms (i.e. the ancillary features) actually correspond to biological human differences. The impact of such criteria could always be the subject of *post hoc* analyses.

Strengths and limitations

The results reported here had several distinctive strengths. First, it is one of the largest population screenings conducted for fatiguing illness. Second, we could assess non-response bias. The significant predictors of participation (monozygosity, female gender, age, having a co-twin who was also eligible, and fewer hospitalizations) are not unexpected and reassuringly small in magnitude.

The results of this study should be viewed in light of two limitations. First, at present, we only possess approximations of the full definition of CFS. The complete criteria require clinical evaluation which has not yet been conducted on this sample. Second, there is a tension between research based in epidemiological and clinical settings. Population-based work such as that presented here is less likely to be influenced by the numerous referral biases that can impact on clinical samples; alternatively, the study of presumptive cases in the population may possess limited relevance to clinicians.

CONCLUSIONS

In this investigation of a large population sample of twins, we found: little support for the CDC-94 criterion of four or more of eight ancillary symptoms; several ancillary symptoms (tender lymph nodes, headache, and sore throat) discriminated poorly; most notably, a CFS-like class using latent class analysis; and that heritability estimates did not identify a more essential definition of CFS-like illness.

ACKNOWLEDGEMENTS

This work was supported by NS-041483 (NLP) and AI-056014 (PFS). The Swedish Twin Registry is supported by grants from the

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

Swedish Department of Higher Education, the Swedish Scientific Council, and an unrestricted grant from AstraZeneca.

DECLARATION OF INTEREST

None.

REFERENCES

APA (1987). *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised) (DSM-III-R). American Psychiatric Association: Washington, DC.

APA (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th edn) (DSM-IV). American Psychiatric Association: Washington, DC.

Bourdette, D. N., McCauley, L. A., Barkhuizen, A., Johnston, W., Wynn, M., Joos, S. K., Storzbach, D., Shuell, T. & Sticker, D. (2001). Symptom factor analysis, clinical findings, and functional status in a population-based case control study of Gulf War unexplained illness. *Journal of Occupational and Environmental Medicine* **43**, 1026–1040.

Bucholz, K. K., Heath, A. C., Reich, T., Hesselbrock, V. M., Kramer, J. R., Nurnberger, Jr., J. I. & Schuckit, M. A. (1996). Can we subtype alcoholism? A latent class analysis of data from relatives of alcoholics in a multicenter family study of alcoholism. *Alcoholism, Clinical and Experimental Research* **20**, 1462–1471.

Buchwald, D. (1996). Fibromyalgia and chronic fatigue syndrome. *Rheumatic Disease Clinics of North America* **22**, 219–243.

Chester, A. C. (1997). Chronic fatigue syndrome criteria in patients with other forms of unexplained chronic fatigue. *Journal of Psychiatric Research* **31**, 45–50.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Education and Psychological Measurement* **20**, 37–46.

David, A., Pelosi, A., McDonald, E., Stephens, D., Ledger, D., Rathbone, R. & Mann, A. (1990). Tired, weak, or in need of rest: fatigue among general practice attenders. *British Medical Journal* **301**, 1199–1202.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* **39**, 1–38.

Eaves, L. J., Silberg, J. L., Hewitt, J. K., Rutter, M., Meyer, J. M., Neale, M. C. & Pickles, A. (1993). Analyzing twin resemblance in multisymptom data: genetic applications of a latent class model for symptoms of conduct disorder in juvenile boys. *Behavior Genetics* **23**, 5–19.

Evengard, B., Jacks, A., Pedersen, N. L. & Sullivan, P. F. (2005). The epidemiology of chronic fatigue in the Swedish Twin Registry. *Psychological Medicine*. doi:10.1017/S0033291705005052.

Fukuda, K., Strauss, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G. & Komaroff, A. (1994). The chronic fatigue syndrome: a comprehensive approach to its definition and study. *Annals of Internal Medicine* **121**, 953–959.

Hadzi-Pavlovic, D., Hickie, I. B., Wilson, A. J., Davenport, T. A., Lloyd, A. R. & Wakefield, D. (2000). Screening for prolonged fatigue syndromes: validation of the SOFA scale. *Social Psychiatry and Psychiatric Epidemiology* **35**, 471–479.

Hickie, I., Lloyd, A., Hadzi-Pavlovic, D., Parker, G., Bird, K. & Wakefield, D. (1995). Can the chronic fatigue syndrome be defined by distinct clinical features? *Psychological Medicine* **25**, 925–935.

Jason, L. A., Richman, J. A., Rademaker, A. W., Jordan, K. M., Plioplys, A. V., Taylor, R. R., McCready, W., Huang, C. F. & Plioplys, S. (1999). A community-based study of chronic fatigue syndrome. *Archives of Internal Medicine* **159**, 2129–2137.

Kendell, R. E. (1989). Clinical validity. *British Journal of Psychiatry* **19**, 45–55.

Kendell, R. E. & Brockington, I. F. (1980). The identification of disease entities and the relationship between schizophrenic and affective psychoses. *British Journal of Psychiatry* **137**, 324–331.

Kirk, K. M., Hickie, I. B. & Martin, N. G. (1999). Fatigue as related to anxiety and depression in a community-based sample of twins aged over 50. *Social Psychiatry and Psychiatric Epidemiology* **34**, 85–90.

Komaroff, A. L. & Buchwald, D. S. (1998). Chronic fatigue syndrome: an update. *Annual Review of Medicine* **49**, 1–13.

Lawrie, S. M. & Pelois, A. J. (1995). Chronic fatigue syndrome in the community: prevalence and associations. *British Journal of Psychiatry* **166**, 793–797.

Levine, P. H. (1997). Epidemiologic advances in chronic fatigue syndrome. *Journal of Psychiatric Research* **31**, 7–18.

Lewis, G. & Wessely, S. (1992). The epidemiology of fatigue: more questions than answers. *Journal of Epidemiology and Community Health* **46**, 92–97.

Lichtenstein, P., De Faire, U., Floderus, B., Svartengren, M., Svedberg, P. & Pedersen, N. L. (2002). The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *Journal of Internal Medicine* **252**, 184–205.

MathSoft (2000). *S-Plus 5 for UNIX Guide to Statistics*. MathSoft: Seattle, WA.

McCutcheon, A. L. (1987). *Latent Class Analysis*. Sage Publications: Beverly Hills.

Neale, M. C. & Cardon, L. R. (1992). *Methodology for the Study of Twins and Families*. Kluwer Academic: Dordrecht, The Netherlands.

Nimnuan, C., Rabe-Hesketh, S., Wessely, S. & Hotopf, M. (2001). How many functional somatic syndromes? *Journal of Psychosomatic Research* **51**, 549–557.

Nisenbaum, R., Reyes, M., Mawle, A. C. & Reeves, W. C. (1998). Factor analysis of unexplained severe fatigue and interrelated symptoms. *American Journal of Epidemiology* **148**, 72–77.

Pawlikowska, T., Chalder, T., Hirsch, S. R., Wallace, P., Wright, D. J. M. & Wessely, S. C. (1994). Population based study of fatigue and psychological distress. *British Medical Journal* **308**, 763–766.

Pedersen, N. L., Lichtenstein, P. & Svedberg, P. (2002). The Swedish Twin Registry in the Third Millenium. *Twin Research* **5**, 427–432.

Plomin, R., DeFries, J. C., Craig, I. W. & McGuffin, P. (2003). *Behavioral Genetics in the Postgenomic Era*. APA Books: Washington, DC.

Reeves, W. C., Lloyd, A., Vernon, S. D., Klimas, N., Jason, L. A., Bleijenberg, G., Evengard, B., White, P. D., Nisenbaum, R. & Unger, E. R. (2003). Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. *BMC Health Services Research* **3**, 25.

Reyes, M., Nisenbaum, R., Hoaglin, D. C., Unger, E. R., Emmons, C. Randall, B., Stewart, J. A., Abbey, S., Jones, J. F., Gantz, N., Minden, S. & Reeves, W. C. (2003). Prevalence and incidence of chronic fatigue syndrome in Wichita, Kansas. *Archives of Internal Medicine* **163**, 1530–1536.

Robbins, J. M., Kirmayer, L. J. & Hemami, S. (1997). Latent variable models of functional somatic distress. *Journal of Nervous and Mental Diseases* **185**, 606–615.

Robins, E. & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *American Journal of Psychiatry* **126**, 107–111.

SAS Institute Inc. (1999). *SAS/STAT® Software, Version 8*. SAS Institute, Inc.: Cary, NC.

Schlesselman, J. J. (1982). *Case-Control Studies: Design, Conduct, Analysis*. Oxford University Press: New York.

Skapinakis, P., Lewis, G. & Meltzer, H. (2003). Clarifying the relationship between unexplained chronic fatigue and psychiatric morbidity: results from a community survey in Great Britain. *International Reviews in Psychiatry* **15**, 57–64.

Spitzer, R. L., Williams, J. B. W., Gibbon, M. & First, M. B. (1992). The Structured Clinical Interview for DSM-III-R (SCID) I:

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press

history, rationale, and description. *Archives of General Psychiatry* **49**, 624–629.

STR (2005). Swedish Twin Registry (http://www.mep.ki.se/twinreg/index_en.html). Department of Medical Epidemiology and Biostatistics. Karolinska Institute, Sweden. Accessed 1 March 2005.

Sullivan, P. F. & Eaves, L. J. (2002). Evaluation of analyses of univariate discrete twin data. *Behavior Genetics* **32**, 221–227.

Sullivan, P. F., Kovalenko, P., York, T. P., Prescott, C. A. & Kendler, K. S. (2003). Fatigue in a community sample of twins. *Psychological Medicine* **33**, 263–281.

Sullivan, P. F., Smith, W. & Buchwald, D. (2002). Latent class analysis of symptoms associated with chronic fatigue syndrome and fibromyalgia. *Psychological Medicine* **32**, 881–888.

Wessely, S., Chalder, T., Hirsch, S., Wallace, P. & Wright, D. (1996). Psychological symptoms, somatic symptoms, and psychiatric disorder in chronic fatigue and chronic fatigue syndrome: a prospective study in the primary care setting. *American Journal of Psychiatry* **153**, 1050–1059.

Wilson, A., Hickie, I., Hadzi-Pavlovic, D., Wakefield, D., Parker, G., Straus, S. E., Dale, J., McCluskey, D., Hinds, G., Brickman, A., Goldenberg, D., Demitrack, M., Blakely, T., Wessely, S., Sharpe, M. & Lloyd, A. (2001). What is chronic fatigue syndrome? Heterogeneity within an international multicentre study. *Australian and New Zealand Journal of Psychiatry* **35**, 520–527.

WHO (1990). *Composite International Diagnostic Interview* (*CIDI*). World Health Organization: Geneva, Switzerland.

WHO (1993). *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. World Health Organization: Geneva.

Wolfe, F., Smythe, H. A., Yunus, M. B., Bennett, R. M., Bombardier, C., Goldenberg, D. L., Tugwell, P., Campbell, S. M., Abeles, M., Clark, P., Fam, A. G., Farber, S. J., Fiechtner, J. L., Franklin, C. M., Gatter, R. A., Hamaty, D., Lessard, J., Lichtbroun, A. S., Masi, A. T., McCain, G. A., Reynolds, W. J., Romano, T. J., Russell, I. J. & Sheon, R. P. (1990). The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Report of the Multicenter Criteria Committee. *Arthritis and Rheumatism* **33**, 160–172.

Yang, I. & Becker, M. P. (1997). Latent variable modeling of diagnostic accuracy. *Biometrics* **53**, 948–958.

Zeger, S. L. & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121–130.

Zeger, S. L. & Liang, K. Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine* **11**, 1825–1839.

Zeger, S. L., Liang, K. Y. & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049–1060 [Erratum (1989), **45**, 347].

https://doi.org/10.1017/S0033291705005210 Published online by Cambridge University Press