BMJ Open  Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review

Kjetil Gundro Brurberg,1 Marita Sporstøl Fønhus,1 Lillebeth Larun,1 Signe Flöttorp,1,2 Kirsti Malterud3,4,5

ABSTRACT

Objective: To identify case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME), and explore how the validity of case definitions can be evaluated in the absence of a reference standard.

Design: Systematic review.

Setting: International.

Participants: A literature search, updated as of November 2013, led to the identification of 20 case definitions and inclusion of 38 validation studies.

Primary and secondary outcome measure: Validation studies were assessed for risk of bias and categorised according to three validation models: (1) independent application of several case definitions on the same population, (2) sequential application of different case definitions on patients diagnosed with CFS/ME with one set of diagnostic criteria or (3) comparison of prevalence estimates from different case definitions applied on different populations.

Results: A total of 38 studies contributed data of sufficient quality and consistency for evaluation of validity, with CDC-1994/Fukuda as the most frequently applied case definition. No study rigorously assessed the reproducibility or feasibility of case definitions. Validation studies were small with methodological weaknesses and inconsistent results. No empirical data indicated that any case definition specifically identified patients with a neuroimmunological condition.

Conclusions: Classification of patients according to severity and symptom patterns, aiming to predict prognosis or effectiveness of therapy, seems useful. Development of further case definitions of CFS/ME should be given a low priority. Consistency in research can be achieved by applying diagnostic criteria that have been subjected to systematic evaluation.

INTRODUCTION

Chronic fatigue syndrome (CFS) is a serious disorder characterised by persistent postexertional fatigue and substantial symptoms related to cognitive, immune and autonomic dysfunction.1–5 Disease mechanisms are complex,3 with no single causal factor identified. Yet there are indications that infections4–8 and immunological dysfunction9 contribute to development and maintenance of symptoms, probably interacting with genetic10 and psychosocial11–13 factors.

Studies have identified pathological patterns and structures of the central nervous system,14–15 dysregulation of body temperature and blood pressure16–17 and dysfunctional stress hormonal systems18–19 in patients with CFS compared with normal controls. None of these appears sufficiently consistent to constitute a diagnostic test, and case definitions (diagnostic criteria) are therefore used to define the CFS diagnosis. When case definitions are developed, the context of application must be considered, since different properties are needed for case definition intended for research purposes compared with case definitions used to diagnose individual patients. It is also necessary to consider whether a broad (ie, sensitive criteria ensuring that we do not miss relevant cases) or narrow (ie, specific criteria ensuring that all positive cases are definite) approach is most appropriate.

Holmes et al20 coined the term ‘chronic fatigue syndrome’ in 1988, as an alternative to ‘The chronic Epstein-Barr virus syndrome’. Since this case definition—the CDC-1988/
Holmes Criteria—was presented in 1988,20 numerous revisions have been developed, aiming for distinctive and reliable identification of individuals who represent a homogenous and consistent phenotype of the hypothesised disease entity, consistent with pathophysiological and psychosocial findings. Currently, the term ‘myalgic encephalomyelitis (ME)’ is commonly used to conceptualise a specific neuroimmunological condition, assumed to be more severe and less psychologically attributed than CFS.21 In 2003, Carruthers et al presented the Canadian-2003 Criteria for diagnosis of ME/CFS.22 A revised version was presented as International Consensus Criteria (the ICC-2011 Criteria) for ME,23 claiming to be a selective case definition for identification of patients with neuroimmune exhaustion with a pathologically low threshold of fatigability and symptom flare after exertion. The assertion that CFS and ME are different clinical entities is disputed. Below, we will pragmatically apply the term CFS/ME.

Johnston et al24 conducted a systematic review of the adoption of CFS/ME case definitions to assess the prevalence and identified eight different case definitions. There is no general agreement on a reference standard for diagnosis, and no diagnostic test is available. Bossuyt et al25,26 include case definitions in their understanding of the term ‘test’, emphasising that diagnostic tests are highly dynamic and need rigorous evaluation before they are introduced into clinical practice.

The objectives of our study were to explore strategies for evaluation of accuracy and concept validity of different case definitions for CFS/ME in the absence of a reference standard. First, we wanted to conduct a systematic review to identify and describe different case definitions (sets of diagnostic criteria) for CFS/ME. Second, we wanted to explore differences between various case definitions by identifying and reviewing validation studies.

MATERIALS AND METHODS
Protocol and registration
We developed a protocol for our study. However, we did not publish or register it.

Eligibility criteria
We included studies presenting or validating case definitions for CFS/ME for adult populations (>18 years). No language restrictions were employed.

Information sources and search
We searched Ovid MEDLINE from 1946, Ovid EMBASE from 1980, Ovid PsychINFO from 1806, Ovid AMED from 1985, The Cochrane Library from 1898, GINAHL from 1981 and PEDRO from 1929, using subject headings and text words (see online supplementary appendix 1). All searches were up to date as of 25 November 2013. We checked the reference lists of all included articles and searched for unpublished and ongoing studies by correspondence with authors and field experts.

Study selection
To select publications eligible for this review, two authors independently read all titles and abstracts in the records retrieved by the searches. We obtained publications in full text if the abstract was deemed eligible by at least one review author. At least two authors independently read the full text papers and selected studies according to the inclusion criteria. Any disagreement between review authors was resolved by discussion between the two review authors or, if necessary, by involving all authors.

Data collection process
First, we listed all the identified case definitions for CFS/ME. One author gathered information about citation from ISI and Google Scholar to indicate the impact or widespread of use, but we made no attempt to assess or rank the quality of the case definitions at this stage.

To facilitate the validity assessment, we developed a framework consisting of three different models.

Model A includes studies with independent application of different case definitions on the same population (figure 1). This model presents the interrelationship between subpopulations identified by different case definitions.

Model B includes studies where patients diagnosed with CFS/ME with one set of diagnostic criteria are diagnosed sequentially with other case definitions assumed to have increasing specificity (figure 2).

Model C includes surveys or cross-sectional studies estimating the prevalence of CFS/ME by applying different case definitions on different populations (figure 3). These studies do not directly compare different case definitions, but may be used for proxy evaluation, similar to the strategy applied by Johnston et al.24 27

Two authors reviewed all potentially relevant validation studies, and categorised them according to model A, B or C. Any disagreement between review authors at this stage was resolved by reaching consensus in the author group.

Risk of bias in individual studies
To differentiate between studies with higher and lower risk of bias, we critically appraised all included validation studies according to check lists: Studies comparing two or more case definitions directly (ie, model A or B) were appraised according to the QUADAS criteria28 (patient selection, index test, reference standard, flow and timing). For evaluation of prevalence studies (ie, model C), we used an outline for assessment of external and internal validity (11 items) of prevalence studies.

Analysis
Participation in prevalence studies, surveys and questionnaires vary across the included studies. Non-response is known to introduce bias, and methods to adjust for low response rates are available.29 In studies affected by non-
response, we have reported adjusted estimates whenever applicable. If adjusted estimates were unavailable, we have defined the proportion as the number of cases divided by the number of responders. We estimated 95% CIs for all proportions by using the Clopper-Pearson exact binomial method. We used R software V.3.0.0 and the rmeta package for statistical computations and plotting.31 32

RESULTS
Study selection
Our systematic literature search identified 1660 unique references, of which 56 articles fulfilled our inclusion criteria (figure 4). Twenty articles present different case definitions of CFS/ME for research or clinical practice20 22 23 33 34 36 37 39–41 43 44 47 (table 1). Furthermore, 38 studies were classified as validation studies, contributing data of sufficient quality and consistency for evaluation of different case definitions according to our inclusion criteria.

The degree to which the different case definitions had been applied in research and clinical guidelines varied widely, with CDC-1994/Fukuda et al39 as the most frequently cited case definition of CFS/ME.

Thirteen of the 20 identified case definitions had been assessed in one or more validation studies.20 22 23 33 34 36 37 39–41 43 44 47 For seven case definitions, no foundation for validation could be identified. We did not identify any study which rigorously assessed the reproducibility or feasibility of the different case definitions.
Independent application of several case definitions on the same population (model A)

Five studies (table 2) applied several case definitions on the same population, but only one of these reported data in a way that made it possible to compare the case definitions.50 51 Nacul et al50 used general practitioner (GP) databases and questionnaires and identified 278 patients with unexplained chronic fatigue (CF) conforming to one or more of the case definition applied, that is, CDC-1994/Fukuda et al59, Canadian-200332 or ECD-2008.34 Most of the patients who were positive according to the Canada criteria (C+) were also positive using the Fukuda criteria (F+). Forty-seven per cent of the Fukuda-positive patients were also positive according to the Canada criteria (F+/C+). In a separate publication,53 the same Fukuda-positive patients as referred in Jason et al33 were used to contrast ICC-2011. About 34% (95% CI 26% to 44%) of the Fukuda-positive patients were also ICC positive (F+/ICC+). Similar to the (F+/C+) subset, it was found that (F+/ICC+) patients experienced more functional impairments as well as more mental and cognitive problems than (F+/ICC−) patients.

None of the other four studies in this group reported data on the correlation between case definitions, patient profile and symptom burden. Application of CDC-1988/Holmes case definition was consistently associated with lower prevalence estimates than CDC-1994/Fukuda, Oxford-1991 and Australian-1990 criteria across these four studies. There was no consistent trend for the other case definitions, but the studies were heterogeneous regarding the application of different case definitions and data collection (table 2). This observation suggests that prevalence numbers obtained by different case definitions should be controlled according to diagnostic procedure, cut-off points and reasons for exclusions before concluding on differences.

Different case definitions with assumed increasing specificity applied sequentially on the same population (model B)

Twelve studies (table 3) sequentially applied different case definitions on the same population. In these studies, patients were screened by using an evaluation standard. Subsequently, test-positive individuals were screened with one or more comparators. Nine of the 12 studies applied CDC-1994/Fukuda as the evaluation standard, and then tested Fukuda-positive patients with CDC-1988/Holmes, Canadian-2003, ICC-2011, ME-2011, Empirical-2006/Reeves, London-1990/Dowsett or Neurasthenia case definitions.

We have taken the actual evaluation standard as a point of departure, and calculated the proportion of these patients still positive when applying other case definitions. Since there are no test negatives for the case definition used as point of departure, true sensitivities or specificities cannot be calculated. Results from two of the studies by Jason et al33 52 suggest that 40–70% of the Fukuda-positive patients are also Canada positives (F+/C+). One study52 concluded that there was less psychiatric comorbidity and more physical functional impairment in the subsample which was positive on both case definitions (F+/C+) than those who were negative according to the Canada criteria (F+/C−). However, the other study33 suggested a higher incidence of mental and cognitive problems among Fukuda-positive patients who were also Canada positive (F+/C+) as compared with the remaining Fukuda-positive but Canada-negative patients (F+/C−). In a separate publication,53 the same Fukuda-positive patients as referred in Jason et al33 were used to contrast ICC-2011. About 34% (95% CI 26% to 44%) of the Fukuda-positive patients were also ICC positive (F+/ICC+). Similar to the (F+/C+) subset, it was found that (F+/ICC+) patients experienced more functional impairments as well as more mental and cognitive problems and higher psychiatric comorbidity than (F+/ICC−) patients.

The comparisons presented in table 3 are associated with high risk of bias as well as random errors, and the results should be interpreted with great caution. For example, two of the included studies reported similar point prevalence according to CDC-1994/Fukuda (2.1% and 2.6%) but reported very different estimates using the Australian-1990 criteria (7.6% and 1.4%).34 35 Sometimes diagnoses were based on questionnaire responses only, sometimes following detailed clinical interviews and laboratory testing. There were also differences in the way similar case definitions were practiced in the various studies, for example, some studies applied a low threshold for exclusion of cases with psychiatric comorbidity, while others did not.

Indirect comparisons of prevalence estimates from several case definitions applied on different populations (model C)

We identified 21 studies (table 4) presenting prevalence estimates for CFS/ME (figure 3), in addition to the five studies presenting prevalence estimates following the application of multiple case definitions (table 2). Based on these studies, we extracted 17 independent estimates.
of the prevalence following application of the CDC-1994/Fukuda criteria (figure 5).

Our analysis suggests that the population prevalence of CFS/ME according to the CDC-1994/Fukuda case definition probably is less than 1% (range 0.1–6.4%; median 1%), with higher prevalence among consecutive GP attendants than from population studies. Prevalence estimates seemed higher when patients were diagnosed without a preceding medical examination. Prevalence estimates of CFS/ME according to CDC-1988/Holmes case definition seemed lower, with all the studies reporting prevalence estimates ranging from 0.0% to 0.3% (median 0.05%).

Five studies reported CFS/ME prevalence estimates according to the Oxford-1991 case definition. These estimates ranged from 0.4% to 3.7% (median 1.5%). Four studies reported prevalence estimates according to the Australian-1990 case definition ranging from 0.04% to 7.6% (median 1.2%).

**DISCUSSION**

We identified 20 studies presenting different CFS/ME case definitions, and 38 studies with data providing access to comparison and evaluation of some of these. Only a minority of existing case definitions had been submitted to comparative evaluations. The validation studies were methodologically weak and heterogeneous, making it questionable to compare the case definitions. The most cited case definition (CDC-1994/Fukuda...
et al.\(^{(50)}\) is also the most extensively validated one, whereas validation studies are few (Canadian-2003, ICC-2011\(^{(23)}\)) or missing (National Institute for Health and Care Excellence (NICE)-2007\(^{(24)}\)) for recently presented and debated case definitions. We found no empirical evidence supporting the hypothesis that some case definitions more specifically identify patients with a neuroimmunological condition.

**Strengths and weaknesses of our study**

The main strength of our study is the systematic methods used to identify and appraise articles presenting case definitions of CFS/ME and studies potentially useful to evaluate the case definitions. Furthermore, we have used systematic and transparent approaches to extract data from the validation studies, categorise the studies according to three different models and to analyse and compare the data.

The STARD initiative aims to improve the reporting on studies of diagnostic accuracy, considering any method for obtaining additional information on a patient’s health status as a test.\(^{(25)}\) Owing to the lack of a reference standard, we found this guideline less suitable for review of articles evaluating case definitions for CFS/ME. Still, issues such as study populations, test methods and rationale, technical specifications for application of the test, statistical methods for comparing measures of accuracy and uncertainty, estimates of diagnostic accuracy, variability and clinical applicability\(^{(25)}\) are relevant also for our analysis.

The validation studies we identified were small with considerable methodological weaknesses and inconsistent results. Only one study held a level of rigour where independent application of several case definitions was conducted on the same population (model A).\(^{(36)}\) Such a study should ideally be based on a population sample rather than a GP practice database, and should compare a selection of currently applied and debated case definitions, such as CDC-1994/Fukuda, Oxford-1991, Canadian-2003 and NICE-2007.

The QUADAS criteria\(^{(28)}\) demonstrate that model B is an evaluation strategy prone to several sources of bias. First, the spectrum of patients subjected to the comparator is selected and not representative of the population receiving the test if it is used alone. Second, as comparators were mostly applied subsequently to the evaluation standard, the clinical evaluations were not independent. The estimates from two of the Jason studies\(^{(33, 52)}\) suggest a comparable correspondence (40–70% of the F+ are also C+) with the results presented by Nacul et al.\(^{(50)}\) Yet, model B gives no or limited information regarding those who screened negative in the first place. We do not know whether some of those might have had a positive diagnosis if screened with one of the other case definitions.

We are even more prone to bias when exploring the consistency of different case definitions through indirect comparisons of prevalence estimates obtained from different populations (model C), and great caution is needed when such proxy comparisons are undertaken. For example, two of the included studies reported

---

**Table 2** Studies presenting prevalence estimates\(^*\) by independent application of several case definitions on the same population (model A)

<table>
<thead>
<tr>
<th>First author, year, country</th>
<th>Data collection</th>
<th>Prevalence (95% CI) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nacul,(^{50}) 2011, UK</td>
<td>609 possible cases electronically identified in databases of 29 GP practices. 70 excluded after clinical revision (explained fatigue), 135 refusals and 126 non-cases</td>
<td>ECD: 0.03 (0.02 to 0.04)</td>
</tr>
<tr>
<td>Bates,(^{56}) 1993, USA</td>
<td>995 consecutive GP visitors invited—94% screened by a questionnaire to detect major fatigue. Selected patients further evaluated by questionnaires, physical examinations and interviews</td>
<td>Canada: 0.10 (0.09 to 0.12)</td>
</tr>
<tr>
<td>Kawakami,(^{57}) 1998, Japan</td>
<td>All adults (n=508) in Town A, Kofu-city, were invited to participate in this structured psychiatric diagnostic interview survey. 137 (27%) completed the study</td>
<td>Fukuda: 0.19 (0.17 to 0.21)</td>
</tr>
<tr>
<td>Lindal,(^{55}) 2002, Iceland</td>
<td>Survey sent to 4000 randomly selected adult participants—63% responded. Questionnaire included questions on all items in the four case definitions. Diagnoses were set electronically based on received responses. No medical tests or examinations were undertaken</td>
<td>Holmes: 0.3 (0.1 to 0.9)</td>
</tr>
<tr>
<td>Wessely,(^{54, 68}) 1997, UK</td>
<td>2363 patients followed in a cohort study—84% completed. Fatigued participant subjected to detailed questionnaires, interviews and laboratory testing. Separate estimates reported for inclusion/exclusion of psychiatric comorbidity</td>
<td>Oxford: 0.4 (0.1 to 1.1)</td>
</tr>
</tbody>
</table>

\(^*\)Prevalence estimates were calculated with the number of responders in the denominator. The choice of denominator may have large implications with regard to the subsequent prevalence estimate, particularly in studies with low response rate. Hence, depending on the actual response rate, estimates presented for each study may be biased.

GP, general practitioner.
<table>
<thead>
<tr>
<th>Study recruitment</th>
<th>Case definitions</th>
<th>Conformity* (95% CI)</th>
<th>Symptom and burden profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brimacombe et al.,69 USA</td>
<td>Fukuda† (n=200) Holmes (n=171) Fukuda† (n=32)</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>(F+/H–) patients do not endorse infectious-type symptoms as often or to the same degree of severity as (F+/H+) patients (F+/H+) patients with more symptoms and functional impairment than (F+/H–). No difference in psychological comorbidity C+ patients have less psychiatric comorbidity, more physical function impairment, are more fatigued with more neurological symptoms than (F+/C–) patients (F+/C+) patients had more functional impairments, and physical, mental and cognitive problems than (F+/C–) patients. (F+/ME+) patients had more functional impairments, and more severe physical and cognitive symptoms than (F+/ME–) patients</td>
</tr>
<tr>
<td>Jason et al.,70 USA</td>
<td>Fukuda† (n=32) Holmes (n=14) Fukuda† (n=32) Canada (n=23)‡</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Jason et al.,71 USA Register</td>
<td>Fukuda† (n=24) Reeves empirical Canada (n=27) Reeves empirical (n=41)§</td>
<td>Of 24 F+ and 84 F-patients empirical criteria and Canada identified 79 and 87% correctly</td>
<td>Canada-2003 case definition appears to select more cardinal and central features of the illness than Empirical CDC-2005/Reeves case definition</td>
</tr>
<tr>
<td>Jason et al.,65 USA Register</td>
<td>Fukuda† (n=27) Reeves emp. (n=41)§</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>Empirical CDC-2005/Reeves case definition led to misclassification of major depressive disorder as CFS</td>
</tr>
<tr>
<td>Brown et al.,53 USA</td>
<td>Fukuda† (n=113) ICC (n=39)</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>ICC+ patients with more functional impairments and physical, mental and cognitive problems than (F+/ICC–) patients. The ICC+ patients also had greater rates of psychiatric comorbidity</td>
</tr>
<tr>
<td>Jason et al.,72 USA</td>
<td>Fukuda† (n=32) Dowsett (n=17)¶</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>D+ patients appear to be more symptomatic than (F+/D–) patients, especially in the neurological and neuropsychiatric areas Effect of CBT and GET similar regardless of diagnostic group affiliation</td>
</tr>
<tr>
<td>White et al.,60 UK Oxford-positive patients recruited to trial</td>
<td>Oxford† (n=641)</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Oxford-positive patients recruited to trial</td>
<td>Oxford† (n=641)</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>Effect of CBT and GET similar regardless of diagnostic group affiliation</td>
</tr>
<tr>
<td>Wearden et al.,73 UK London ME (n=329)</td>
<td>Oxford† (n=296) London ME (n=329)</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Oxford-positive patients recruited to trial</td>
<td>London ME (n=92)</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Stubhaug et al.,74 Norway Neurasthenia-positive patients recruited to trial</td>
<td>Neurasthenia† (n=72) Oxford (n=65) Fukuda (n=29)</td>
<td>1 (0.85 (0.80 to 0.90) 0.44 (0.26 to 0.62) 0.63 (0.44 to 0.79)</td>
<td>Not reported</td>
</tr>
</tbody>
</table>

*The proportion of cases relative to the evaluation standard.
†Evaluation standard.
‡Three of the 23 participants who tested positive according to the Canada criteria were negative according to Fukuda.
§14 of the 37 patients who tested positive according to Reeves were negative according to Fukuda (these 14 patients had a depression diagnosis).
¶Three of the 17 participants who tested positive according to Dowsett were negative according to Fukuda.
CFT, cognitive behavioral therapy; CFS, chronic fatigue syndrome; GET, graded exercise therapy; ICC, International Consensus Criteria; ME, myalgic encephalomyelitis.
similar point prevalence according to CDC-1994/ Fukuda (2.1% and 2.6%), but reported very different estimates following the application of the Australian-1990 criteria (7.6% and 1.4%).54 55 This inconsistency can be explained by major methodological differences seen across the included studies. Our sample includes studies in which a diagnosis of CFS/ME is made on the basis of either questionnaire responses or clinical interview. Previous studies suggest that patients who receive a standardised questionnaire report considerably more symptoms than when asked to report their symptoms spontaneously.59 There are several other sources to this between study heterogeneity, such as recruitment strategy, response rate and strategies for non-response adjustment. We were not able to identify the most important one. However, Johnston et al.57 performed an interesting subgroup analysis in their meta-analysis of 14 studies applying the CDC-1994/ Fukuda case definition, and found that the pooled prevalence for self-reporting assessment was 3.28% (95% CI 2.24% to 4.33%) compared with 0.76% (95% CI 0.23% to 1.29%) for clinical assessment. Prevalence was lower in community samples (0.87%; 0.32% to 1.42%) than in primary care samples (1.72%; 1.40% to 2.04%). The prevalence estimates based on self-reports showed high variability, while clinically assessed estimates were more consistent, especially in the community samples.

### The utility of case definitions and diagnoses

The utility of a diagnosis is linked to the potential effects of being diagnosed (eg, benefits and harms of the patient’s role, access to treatment and insurance). More importantly, a diagnosis is useful if it is linked to valid information regarding prognosis or outcomes of therapy. Reitsma et al.26 suggest clinical test validation as an alternative paradigm for evaluation of a diagnostic test when an acceptable reference standard is missing. Hence, primary studies and systematic reviews on prognosis and therapy are alternative sources to evaluate the usefulness of different case definitions of CFS/ME. We have identified only one such publication, the PACE trial.60 Here, participants were diagnosed according to the Oxford-1991 criteria, Empirical criteria-2007/Reeves and London ME-1994/National Task Force criteria, and then randomised to either standard medical treatment, graded exercise therapy, cognitive behaviour therapy or pacing. The results showed that the effectiveness of the treatments was similar across groups, irrespective of the case definition which had been used. Fluge et al.11 applied the CDC-1994/Fukuda and retrospectively added the Canada criteria in their study on the effects of rituximab in CFS with comparable results. In a recent publication, Maes et al.51 measured symptom severity, selected biomarkers and postexertional malaise in 144 patients with CFS, of whom 107 fulfilled the CDC-1994/
Fukuda criteria of CFS/ME. They claimed that CF, CFS and ME are distinct categories, although stating that patients group together in one continuum with no clear boundaries between them. Such studies would be even more useful if outcomes of specific treatment modes had also been tested.

A study comparing the prognosis of different diagnostic labels of fatigue found that patients with ME had the worst prognosis while patients with postviral fatigue syndrome had the best. This could mean that the patients destined to the worst prognosis were labelled with the ME diagnosis, or it might be explained as an adverse effect of being labelled with ME. The authors found no significant difference in recorded fatigue before the diagnosis of CFS and ME, and the data in this retrospective study supported the hypothesis of the labelling effect. Another study found that patients who attributed their fatigue to ME were more fatigued and

<table>
<thead>
<tr>
<th>Study</th>
<th>Point estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Canada</strong></td>
<td></td>
</tr>
<tr>
<td>Nacul 2011</td>
<td>0.10 (0.09 - 0.12)</td>
</tr>
<tr>
<td><strong>Fukuda</strong></td>
<td></td>
</tr>
<tr>
<td>Vincent 2012</td>
<td>0.07 (0.05 - 0.09)</td>
</tr>
<tr>
<td>Nacul 2011</td>
<td>0.19 (0.17 - 0.21)</td>
</tr>
<tr>
<td>Reyes 2003*</td>
<td>0.24 (0.14 - 0.33)</td>
</tr>
<tr>
<td>Njoku 2007</td>
<td>0.28 (0.06 - 0.80)</td>
</tr>
<tr>
<td>Steele 1998*</td>
<td>0.29 (0.20 - 0.37)</td>
</tr>
<tr>
<td>Jason 1999*</td>
<td>0.42 (0.29 - 0.56)</td>
</tr>
<tr>
<td>Bierl 2004</td>
<td>0.67 (0.50 - 0.88)</td>
</tr>
<tr>
<td>van’t Leven 2009</td>
<td>0.98 (0.79 - 1.21)</td>
</tr>
<tr>
<td>Hamaguchi 2011</td>
<td>0.98 (0.54 - 1.64)</td>
</tr>
<tr>
<td>Kawakami 1998</td>
<td>1.46 (0.18 - 5.17)</td>
</tr>
<tr>
<td>Cho 2009</td>
<td>1.64 (1.26 - 2.08)</td>
</tr>
<tr>
<td>Cho 2009</td>
<td>2.07 (1.55 - 2.72)</td>
</tr>
<tr>
<td>Lindal 2002</td>
<td>2.14 (1.61 - 2.79)</td>
</tr>
<tr>
<td>Evengård 2005</td>
<td>2.36 (2.19 - 2.53)</td>
</tr>
<tr>
<td>Reeves 2007*</td>
<td>2.55 (0.88 - 4.22)</td>
</tr>
<tr>
<td>Wessely 1997*</td>
<td>2.60 (1.70 - 3.40)</td>
</tr>
<tr>
<td><strong>Oxford</strong></td>
<td></td>
</tr>
<tr>
<td>Bates 1993</td>
<td>0.43 (0.12 - 1.09)</td>
</tr>
<tr>
<td>Lawrie 1995</td>
<td>0.58 (0.16 - 1.47)</td>
</tr>
<tr>
<td>Kawakami 1998</td>
<td>1.46 (0.18 - 5.17)</td>
</tr>
<tr>
<td>Wessely 1997*</td>
<td>2.20 (1.40 - 3.00)</td>
</tr>
<tr>
<td>Lindal 2002</td>
<td>3.73 (3.02 - 4.55)</td>
</tr>
<tr>
<td><strong>Holmes</strong></td>
<td></td>
</tr>
<tr>
<td>Lindal 2002</td>
<td>0.00 (0.00 - 0.15)</td>
</tr>
<tr>
<td>Kawakami 1998</td>
<td>0.00 (0.00 - 2.66)</td>
</tr>
<tr>
<td>Price 1992</td>
<td>0.01 (0.00 - 0.04)</td>
</tr>
<tr>
<td>Buchwald 1995</td>
<td>0.10 (0.02 - 0.29)</td>
</tr>
<tr>
<td>Versluis 1997</td>
<td>0.11 (0.07 - 0.16)</td>
</tr>
<tr>
<td>Bates 1993</td>
<td>0.32 (0.07 - 0.93)</td>
</tr>
<tr>
<td>Wessely 1997*</td>
<td>1.20 (0.50 - 1.80)</td>
</tr>
<tr>
<td><strong>Australian</strong></td>
<td></td>
</tr>
<tr>
<td>Lloyd 1990</td>
<td>0.04 (0.03 - 0.06)</td>
</tr>
<tr>
<td>Bates 1993</td>
<td>1.07 (0.51 - 1.96)</td>
</tr>
<tr>
<td>Wessely 1997*</td>
<td>1.40 (0.80 - 2.00)</td>
</tr>
<tr>
<td>Lindal 2002</td>
<td>7.62 (6.61 - 8.73)</td>
</tr>
</tbody>
</table>
more handicapped in relation to home, work, social and private leisure activities than patients who attributed their fatigue to psychological or social factors.

**Broad or narrow case definitions?**

Ideally, correspondence validity between test and target should be 100% for sensitivity (the capacity to identify patients in the target group) and specificity (the capacity to rule out patients who do not belong to the target group). More often, there is a trade-off between these measures, depending on the purpose of diagnosis. Emphasising sensitivity implies a risk of overdiagnosis, which dilutes the actual diagnostic concept, while emphasising specificity implies a risk of underdiagnosis, dismissing patients who might benefit from treatment. Development of more exclusive case definitions for CFS/ME has been proposed, claiming that existing case definitions do not select homogenous sets of patients.

More specifically, Oxford-1991, Fukuda-1994 and NICE-2007 have been criticised, especially by patient organisations, for undue overlap with psychopathology. Proponents of recent case definitions, such as Canada-2003 and ICC-2011, claim to achieve a narrow selection of patients with ME conforming to a hypothesised specific pathophysiology. Our review demonstrates, however, that these case definitions do not necessarily exclude patients with psychopathology.

A lesson could be learnt from Reeves, who tried to elaborate the CDC1994/Fukuda definition and bring methodological rigour into the diagnostic criteria by scores from standardised and validated instruments. The Empirical-2006/Reeves case definition led to a tenfold prevalence estimate as compared with the CDC1994/Fukuda definition, probably due to misclassification and inclusion of patients with major depressive disorder. The purpose of rigour had not been achieved, and the Empirical-2006/Reeves case definition was never broadly implemented. According to our review, it is uncertain whether a more homogenous subset of patients can be achieved with the Canada-2003 and ICC-2011 case definitions. The authors of the latter paper write: “Collectively, members have approximately 400 years of both clinical and teaching experience, authored hundreds of peer-reviewed publications, diagnosed or treated approximately 50 000 patients with ME and several members coauthored previous criteria.”

This declaration is no validity criterion and provides no guarantee that the case definition works according to the intentions.

**Case definitions for research or clinical practice?**

Research requires uniform and reproducible criteria, suitable for unambiguous definitions of the target population. Another concern is to compare studies across time and nations. These are arguments for an inclusive case definition, preferably one which has been in use for a while, and for which validation studies are available. In CFS/ME research, the Oxford-1991 and CDC-1994/Fukuda are the most frequently used case definitions. Our review indicates that the former might be more inclusive, with lower specificity than the latter, although the impact of this is unclear. Proponents for more restrictive case definitions dismiss findings from treatment studies documenting effects of cognitive behavioural treatment or graded exercise therapy for patients diagnosed with the Oxford-1991 or CDC-1994/Fukuda case definitions. Their claim is that for a more exclusive selection of patients with ME, defined according to specific hypothesised pathophysiology, the side effects of these treatment modalities are hazardous. So far, however, treatment studies based on the Canada-2003 or ICC-2011 case definitions are not available.

Case definitions for clinical practice should be research based, validated and manageable to provide a tool which can relieve patient’s uncertainty, indicate the most appropriate treatment and prevent adverse effects and waste of healthcare resources of unnecessary treatment and diagnostic procedures. They should be founded on available knowledge regarding the mechanisms of the actual condition, validated through credible and transparent processes and presented in a format which can be implemented in everyday practice. An argument for more inclusive case definitions for CFS/ME would be the issue of treatment, since existing evidence indicates that side effects of cognitive behavioural treatment or graded exercise therapy are negligible. For this context, the CDC-1994/Fukuda case definition appears suitable, with the NICE-2007 as a good candidate for validation studies.

**IMPLICATIONS FOR RESEARCH AND CLINICAL PRACTICE**

On the basis of our review, we argue that development of further case definitions of CFS/ME should be given low priority, as long as causal explanations for the disease are limited. It might still be useful to classify patients according to severity and symptom patterns, aiming to identify characteristics of patients that might predict differences in prognosis or expected effects of therapy.

It is likely that all CFS/ME case definitions capture conditions with different or multifactorial pathogenesis and varying prognosis. The futile dichotomy of ‘organic’ versus ‘psychic’ disorder should be abandoned. Most medical disorders have a complex aetiology. Psychological treatments are often helpful also for clear-cut somatic disorders. Unfortunately, patient groups and researchers with vested interests in the belief that ME is a distinct somatic disease seem unwilling to leave the position that ME is an organic disease only. This position has damaged the research and practice for patients suffering from CFS/ME.

**CONCLUSIONS**

Our review provided no evidence that any of the case definitions identify patients with specific or ‘organic only’ disease aetiology. Priority should be given to
Further development and testing of promising treatment options for patients with CFS/ME. Classification of patients according to severity and symptom patterns, aiming to identify characteristics of patients that might predict differences in prognosis or expected effects of therapy, might be useful. Development of further case definitions of CFS/ME should be given low priority. Consistency in research can be achieved by application of diagnostic criteria which have been systematically evaluated and compared with other case definitions.

**Contributors**
KM had the original idea, and all five authors worked together to develop an appropriate theoretical framework and design. MSF developed the search, and all authors were involved in the selection process. LL and KGB performed the statistical analysis and all authors were involved in the data interpretation. KM wrote the manuscript draft and revised the draft based on input from the other authors. All authors revised it critically for content and approved the final version.

**Funding**
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests**
None.

**Provenance and peer review**
Not commissioned; externally peer reviewed.

**Data sharing statement**
All data are extracted from the cited primary studies.

**Open Access**
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided any reuse, distribution, or modification clearly indicates that the original study is being cited.

**REFERENCES**


