I. Background and Objectives for the Systematic Review

Myalgic Encephalomyelitis (ME)/Chronic Fatigue Syndrome (CFS) is a condition characterized by a constellation of symptoms, with post-exertional malaise and/or chronic and disabling fatigue being the hallmark. In an attempt to describe the condition based on the proposed etiologies, several other terms have been used, including post viral fatigue syndrome and chronic fatigue immune dysfunction syndrome. The term ME was first used in the 1930s after an outbreak of neuromyasthenia and CFS was first coined in the 1980s, with both conditions having overlapping features. Over the years, there has been disagreement on the underlying etiology and whether the conditions represented by these terms reflect a single pathologically discrete syndrome, subsets of the same illness, or a nonspecific condition shared by other disease entities. The two persisting terms, ME and CFS, have more recently been used interchangeably despite the ongoing controversy.

The first set of clinical criteria defining the condition were published in 1988 and have been further developed over the years most notably with the 1994 Fukuda criteria, the 2003 Canadian clinical case definition, and in 2011 with the international consensus criteria. (see Table 1). This most recent international consensus report advocates moving away from the term CFS in favor of ME to better reflect an underlying pathophysiology involving widespread inflammation and neuropathology, and to embrace the two terms as synonymous. The syndrome that encompasses ME/CFS is characterized by various manifestations of diffuse symptoms including post-exertional malaise and/or persistent and disabling fatigue, pain, sleep disturbance, neurological and cognitive manifestations (i.e., impaired concentration, mental processing, memory), motor impairment, and altered immune and autonomic responses. The variable symptomatology and lack of an identifiable disease process with gold standard of measurement have challenged researchers and clinicians in their attempts to better understand the disease process and its effects on patients.

The prevalence of ME/CFS is also difficult to estimate given the uncertainty of the case definition. The CDC reported a U.S. prevalence of 0.3 percent with over 1,000,000 adults with the disorder in the U.S. By using different case definitions, some suggest that the rate is as high as 2.5 percent. Currently there are no U.S. Food and Drug Administration (FDA) approved medications for the treatment of ME/CFS, but many have been used to treat the associated symptoms. The burden of disease and economic impact from both medical and societal costs may be in the billions of dollars. Thus finding ways to accurately diagnose patients to optimize management has significant public health importance and consequences. Controversies surrounding ME/CFS have led to wide variations in the clinical management of patients, including uncertainty in etiology, diagnosis, and approach to treatment. Identifying treatment interventions and meaningful clinical outcomes is also challenging, given the breadth of symptoms and the uncertainty as to whether specific symptom combinations define subsets of patients that
may respond uniquely to different approaches. When patients were surveyed in April, 2013 as part of the U.S. Food and Drug Administration’s (FDS’s) patient-focused drug development initiative, treatments fell into two broad categories, those intended to treat the underlying cause of the disease and those targeting specific symptoms. The first category included immune modulators such as rintatolimod and rituximab, and antiviral and antibiotic medications. Interventions targeting symptoms included medications to treat pain, fatigue, and sleep dysfunction, and non-drug therapies which included yoga, stretching and relaxation techniques, counseling on pacing strategies, and mental exercises. An examination of the comparative effectiveness and harms of treatments for ME/CFS is important to guide clinical practice, which underscores the need for a systematic review on this topic. This report focuses on the clinical outcomes surrounding the attributes of fatigue, especially post-exertional malaise and persistent fatigue, and its impact on overall function and quality of life because these are unifying features of ME/CFS that impact patients.

This topic was nominated by the Office of Disease Prevention (ODP) at the National Institutes of Health to Inform an Evidence-based Methodology Workshop on ME/CFS. The Working Group convened by ODP assisted in refining the topic.

### Table 1. Groups With Case Definitions of ME/CFS

<table>
<thead>
<tr>
<th>ME/CFS Definition Statements</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC Fukuda et al., 1994&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Canadian Carruthers et al., 2003&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Revised Canadian Jason et al., 2010&lt;sup&gt;11&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>International Consensus Statement Carruthers et al., 2011&lt;sup&gt;7&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>London Dowsett, 1994&lt;sup&gt;12&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Oxford Sharpe, et al. 1991&lt;sup&gt;13&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

### II. The Key Questions

1. What methods are available to clinicians to diagnose ME/CFS and how do the use of these methods vary by patient subgroups?

   a. What are widely accepted diagnostic methods and what conditions are required to be ruled out or excluded before assigning a diagnosis of ME/CFS?

   b. What is the accuracy and concordance of diagnostic methods?

   c. What harms are associated with diagnosing ME/CFS?
2. What are the (a) benefits and (b) harms of therapeutic interventions for patients with ME/CFS and how do they vary by patient subgroups?

c. What are the characteristics of responders and non-responders to interventions?

PICOTS

• **Population(s)**

  Include:
  - For KQ 1: Symptomatic adults (aged 18 years or older) with fatigue
  - For KQ 2: Adults aged 18 years or older, with ME/CFS, without other underlying diagnosis

  Exclude:
  - Children and adolescents
  - Patients with other underlying diagnosis

• **Interventions**

  Include:
  - For KQ1: Case definitions: e.g., Fukada/CDC, Canadian, International and others
  - For KQ2: Forms of counseling and behavior therapy, graded exercise programs, complementary and alternative medicine (acupuncture, relaxation, massage, other), and symptom-based medication management (immune modulators, beta blockers, antidepressants, anxiolytics, stimulants, other)
• **Comparators**

  Include:

  • For KQ1: Diagnostic accuracy studies and diagnostic concordance studies with comparators
  • For KQ2: Placebo or no treatment/usual care, other active interventions (including combination therapies and head-to-head trials)

• **Outcomes**

  Include:

  • For KQ1: Sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, C statistic (AUROC), net reclassification index; concordance, any potential harm from diagnosis (such as psychological harms, labeling, risk from diagnostic test, misdiagnosis, other)
  • For KQ2: Overall function (i.e. 36-item Short Form Survey [SF-36]), quality of life, days spent at work/school, proportion working full or part time, fatigue (Multidimensional Fatigue Inventory [MFI] or similar), adverse effects of interventions, withdrawals and withdrawals due to adverse events, rates of adverse events due to interventions

• **Timing**

  Include:

  • 12 weeks or longer

• **Setting**

  Include:

  • Clinical settings
III. Analytic Framework

Figure 1. Analytic Framework for the Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

IV. Methods

A. Criteria for Inclusion/Exclusion of Studies in the Review

The criteria for inclusion and exclusion of studies will be based on the Key Questions (KQs) and are described in the previous PICOTS section.

Below are additional details on the scope of this project:

Study Designs: For all KQs we will include randomized controlled trials (RCTs). For KQ 1a we will also include derivation, validation, and cohort studies. For KQ 1c and 2b we will also include cohort studies with comparators. For all KQs we will exclude uncontrolled observational studies, case control studies, case series, and case reports. Systematic reviews will be used as sources of evidence if they address a Key Question and are assessed as being at low risk of bias, according to the AMSTAR quality assessment tool. If systematic reviews are included, we will update findings by adding primary studies identified in our searches. If multiple systematic reviews are relevant and low risk of bias, we will focus on the findings from the most recent reviews and evaluate areas of consistency across the reviews.

Non-English Language Studies: We will restrict to English language articles, but will review English language abstracts of non-English language articles to identify studies that would otherwise meet inclusion criteria, in order to assess for the likelihood of language bias.
B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions

Publication Date Range: Searches will begin in January 1988, the year the first set of clinical criteria defining CFS were published.3

Library searches will be updated while the draft report is posted for public comment and peer review to capture any new publications. Literature identified during the update search will be assessed by following the same process of dual review as all other studies considered for inclusion in the report. If any pertinent new literature is identified for inclusion in the report, it will be incorporated before the final submission of the report.

Literature Databases: Ovid MEDLINE, PsycINFO, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Health Technology Assessment, National Health Sciences Economic Evaluation Database, and Database of Abstracts of Reviews of Effects will be searched to capture both published and gray literature.

Scientific Information Packets:

Scientific information packets (SIPs) will be requested from drug and device manufacturers and a notice inviting submission of relevant scientific information will be published in the Federal Register in an effort to identify any relevant unpublished literature that may contribute to the body of evidence. All interested parties will have the opportunity to submit data for this review using the AHRQ Effective Health Care publicly accessible online SIP portal (http://effectivehealthcare.ahrq.gov/index.cfm/submit-scientific-information-packets/).

Manufacturers of currently available and FDA approved anxiolytics, antidepressants, antivirals, beta-blockers, immune modulators, non-steroidal anti-inflammatory drugs, stimulants, and triptans will be invited to provide SIPs.

Hand Searching: Reference lists of included articles and systematic reviews will be reviewed for includable literature.

Contacting Authors: In the event that information regarding methods or results appears to be omitted from the published results of a study, or if we are aware of unpublished data, we will query the authors to obtain this information.

Process for Selecting Studies: Pre-established criteria will be used to determine eligibility for inclusion and exclusion of abstracts in accordance with the AHRQ EPC Methods Guide.14 Abstracts will be included if they fulfill the population, interventions and outcomes of interest, and meets the included study design and duration. To ensure accuracy, all excluded abstracts will be dual reviewed. All citations deemed appropriate for inclusion by at least one of the reviewers will be retrieved. Each full-text article will be independently reviewed for eligibility by two team members, including any articles suggested by peer reviewers or that arise from the public posting process. Any disagreements will be resolved by consensus.

Source: www.effectivehealthcare.ahrq.gov
Published online: May 1, 2014
C. Data Abstraction and Data Management

After studies are selected for inclusion, data will be abstracted into categories that include but are not limited to: study design, year, setting, country, sample size, eligibility criteria, population and clinical characteristics, intervention characteristics, and results relevant to each key question as outlined in the previous PICOTS section. Information that will be abstracted that is relevant for assessing applicability will include the method of diagnosis, number of patients randomized relative to the number of patients enrolled, use of run-in or wash-out periods, and characteristics of the population (clinical subgroups e.g., post-exertional malaise, postural hypotension, and others), intervention, and care settings. All study data will be verified for accuracy and completeness by a second team member. A record of studies excluded at the full-text level with reasons for exclusion will be maintained.

D. Assessment of Methodological Risk of Bias of Individual Studies

Predefined criteria will be used to assess the quality of individual controlled trials, systematic reviews, and observational studies by using clearly defined templates and criteria as appropriate. Studies will be evaluated using appropriate criteria developed by the U.S. Preventive Services Task Force. Systematic reviews will be assessed using the AMSTAR quality rating instrument. Particular attention will be given to the criteria of patient population including ME/CFS case definition and spectrum of patients included in the study, comparability of groups, importance and validity of outcome measurements for the ME/CFS population, adjustment for confounders, and adherence. These criteria and methods will be used in conjunction with the approach recommended in the chapter, Assessing the Risk of Bias of Individual Studies When Comparing Medical Interventions in the AHRQ Methods Guide developed by the Agency for Healthcare Research and Quality and the chapter Options for Summarizing Medical Test Performance in the Absence of a “Gold Standard.” Studies will be rated as “good,” “fair,” or “poor,” or as specified by the particular criteria.

Studies rated “good” will be considered to have the least risk of bias, and their results will be considered valid. Good-quality studies include clear descriptions of the population, setting, interventions, and comparison groups; a valid method for allocation of patients to treatment; low dropout rates and clear reporting of dropouts; appropriate means for preventing bias; and appropriate measurement of outcomes.

Studies rated “fair” will be susceptible to some bias, though not enough to invalidate the results. These studies may not meet all the criteria for a rating of good quality, but no flaw likely to cause major bias. The study may be missing information, making it difficult to assess limitations and potential problems. The fair-quality category is broad, and studies with this rating will vary in their
strengths and weaknesses. The results of some fair-quality studies are likely to be valid, while others may be only possibly valid.

Studies rated “poor” will have significant flaws that imply biases of various types that may invalidate the results. They will have a serious or “fatal” flaw in design, analysis, or reporting; large amounts of missing information; discrepancies in reporting; or serious problems in the delivery of the intervention. The results of these studies will be least as likely to reflect flaws in the study design as the true difference between the compared interventions. We will not exclude studies rated as being poor in quality a priori, but poor-quality studies will be considered to be less reliable than higher quality studies when synthesizing the evidence, particularly if discrepancies between studies are present.

Each study evaluated will be dual-reviewed for quality by two team members. Any disagreements will be resolved by consensus.

E. Data Synthesis

We will construct evidence tables identifying the study characteristics (as discussed above), results of interest, and quality ratings for all included studies, and summary tables to highlight the main findings. We will review and highlight studies by using a hierarchy-of-evidence approach, where the best evidence is the focus of our synthesis for each key question.

Qualitative data will be summarized in summary tables as ranges, and descriptive analysis and interpretation of the results will be provided.

Meta-analyses will be conducted to summarize data and obtain more precise estimates on outcomes for which studies are homogeneous enough to provide a meaningful combined estimate. The feasibility of a quantitative synthesis will depend on the number and completeness of reported outcomes and a lack of heterogeneity among the reported results. To determine whether meta-analysis could be meaningfully performed, we will consider the quality of the studies and the heterogeneity among studies in design, patient population including method of diagnosis, interventions, and outcomes, and may conduct sensitivity analyses. Meta-regression may be conducted to explore statistical heterogeneity using additional variables on methodological or other characteristics (e.g., quality, randomization or blinding, outcome definition and ascertainment) given enough number of studies.

Results will be presented as structured by the key questions, and any prioritized outcomes will be presented first.

F. Grading the Strength of Evidence (SOE) for Individual Comparisons and Outcomes

The strength of evidence for each key question will be initially assessed by one researcher for each clinical outcome (see PICOTS) by using the approach described in the AHRQ Methods Guide. To ensure consistency and validity of the evaluation, the grades will be reviewed by the entire team of investigators for:

• Study limitations (low, medium, or high level of study limitations)
• Consistency (consistent, inconsistent, or unknown/not applicable)
• Directness (direct or indirect)
• Precision (precise or imprecise)
• Reporting bias (suspected or undetected)

The strength of evidence will be assigned an overall grade of high, moderate, low, or insufficient according to a four-level scale by evaluating and weighing the combined results of the above domains:

• High — We are very confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has few or no deficiencies. We believe that the findings are stable, i.e., another study would not change the conclusions.
• Moderate — We are moderately confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has some deficiencies. We believe that the findings are likely to be stable, but some doubt remains.
• Low — We have limited confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has major or numerous deficiencies (or both). We believe that additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.
• Insufficient — We have no evidence, we are unable to estimate an effect, or we have no confidence in the estimate of effect for this outcome. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion.

G. Assessing Applicability

Applicability will be estimated by examining the characteristics of the patient populations (e.g., demographic characteristics; type, duration or severity of symptoms; criteria used for diagnosis, presence of medical and psychiatric co-morbidities); the sample size of the studies; and clinical settings (e.g., primary care, specialty setting) and countries (e.g., patients in developing countries) in which the studies are performed. Diagnostic accuracy or concordance between case definitions may vary by subgroups, and treatment interventions may be more or less effective in specific patient subgroups such as those with post-exertional malaise or postural hypotension. Variability in the studies may limit the ability to generalize the results to other populations and settings.
V. References


VI. Definition of Terms

Not applicable.

VII. Summary of Protocol Amendments

Not applicable.

Source: www.effectivehealthcare.ahrq.gov
Published online: May 1, 2014
VIII. Review of Key Questions

Key Questions were reviewed and refined as needed by the EPC with input from the NIH Working Group and the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is being reviewed.

IX. NIH Working Group

In place of Key Informants, a NIH Working Group Planning Meeting occurred to provide input into identifying the Key Questions and guiding the scope of the report, because this topic was nominated to AHRQ from the National Institutes of Health.

X. Technical Experts

Technical Experts comprise a multi-disciplinary group of clinical, content, and methodologic experts, including the patient perspective, who provide input in defining populations, interventions, comparisons, or outcomes as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor contribute to the writing of the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified. Potential conflicts of interest are also managed by not releasing the names of Technical Experts until publication of the final report.
XI. Peer Reviewers

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodologic expertise. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. Peer reviewers do not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical briefs, be published three months after the publication of the Evidence report.

Potential Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than $10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism. Potential conflicts of interest are also managed by not releasing the names of Peer Reviewers until publication of the final report.

XII. EPC Team Disclosures

EPC core team members must disclose any financial conflicts of interest greater than $1,000 and any other relevant business or professional conflicts of interest. Related financial conflicts of interest which cumulatively total greater than $1,000 will usually disqualify EPC core team investigators.

XIII. Role of the Funder

This project was funded under Contract No. HHSA 290201200014I from the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. The Task Order Officer reviewed contract deliverables for adherence to contract requirements and quality. The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.
Appendix. Sample Search Strategy (Ovid MEDLINE)

1  exp Chronic Fatigue Syndrome/ (4528)
2  exp Encephalomyelitis/ (9563)
3  exp Fatigue/ (20358)
4  2 and 3 (373)
5  1 or 4 (4557)
6  (chronic$ adj3 fatig$ adj3 syndrom$).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier] (5194)
7  (myalg$ adj3 encephal$).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier] (419)
8  6 or 7 (5306)
9  5 or 8 (5320)
10 limit 9 to english language (4732)
11 limit 9 to abstracts (3780)
12  10 or 11 (5040)
13  exp cognitive behavior therapy/ (15929)
14  (cognit$ adj3 behav$ adj5 (therap$ or treat$ or interven$ or regimen$ or counsel$)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier] (11111)
15  13 or 14 (19599)
16  exp Exercise/ (111374)
17  ((grad$ or therap$ or treat$ or interven$ or regimen$) adj3 exercis$).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier] (36939)
18  16 or 17 (140490)
19  exp drug therapy/ (1112200)
20  exp drugs/ (641176)
21  dt.fs. (1752044)
22  19 or 20 or 21 (2743222)
23  ((drug$ or pharmac$) adj3 (therap$ or treat$ or adminis$ or prescri$ or interven$ or regimen$)).mp. (539954)
24  22 or 23 (2871189)
25  12 and 15 (334)
26  12 and 18 (365)
27  12 and 24 (563)
28  25 or 26 or 27 (1077)