Chapter 2. Methodology

All work included in this Task Order was carried out by MetaWorks investigators, using systematic review methods derived from the science of review research.[24, 25] These methods were generally applied according to standard operating procedures at MetaWorks and are shown in Figure 1.

The SSA submitted to AHRQ a list of questions pertinent to disability and CFS. AHRQ developed a Task Order, and presented it to MetaWorks. After MetaWorks investigators conducted a preliminary review of the literature, an Expert Panel meeting was held in Washington, DC, on November 15, 2001. The purposes of this meeting were to:

1. Establish working definition of CFS for purposes of this task order.
2. Refine key questions.
3. Get recommendations regarding breadth of literature to be reviewed, analyses to be performed, and sources of data that should be accessed to ensure the evidence report is responsive to SSA's concerns.

Key Questions

The SSA initially suggested a comprehensive list of questions to be addressed by this review. During the Expert Panel meeting, the original key questions were modified to focus more specifically on the issues of disability and impairment in CFS. The following revised questions were reviewed by the Expert Panel and representatives from SSA, and were approved by the AHRQ Task Order Officer (TOO).

1. What is the evidence that some individuals with CFS have discrete impairments that are associated with disability? (Note that impairments include both physical and mental impairments).

2. What is the evidence that in the CFS population, current neuropsychological tests reliably detect cognitive or affective impairments associated with decreased ability to work?

3. What is the evidence that in individuals with CFS, treatments are effective in restoring the ability to work?

4. What patient characteristics best define improvement in functioning or positive outcomes in the CFS population? Where it occurs, how is improvement in functioning related to the ability to engage in work activity?

Based on the Task Order, MetaWorks researchers developed a Work Plan (Appendix A) which outlined the methods to be followed for the literature search, study eligibility criteria, data elements for extraction, and methodological strategies to minimize bias and maximize precision during the process of data extraction and synthesis. The Work Plan also incorporated decisions made at the expert panel meeting held on November 15, 2001 (Appendix B), regarding the

revised key questions, CFS diagnostic criteria to be used, and recommended changes to the preliminary literature searches. The Work Plan was subsequently reviewed and accepted by AHRQ and SSA.

Causal Pathway

Based on the results of a preliminary literature review, a causal pathway was developed (Appendix A, page A-23). All of the events described in this pathway take place within the CFS universe; i.e., only patients already diagnosed with CFS are included. Patients with fibromyalgia, Gulf War Syndrome, and other related conditions are not included. To diagnose disability in the CFS universe, patients must have a medically determinable condition (defined by clinical signs and symptoms, laboratory abnormalities, or other abnormalities), leading to physical or mental impairment, that results in disability, as defined by the SSA. This causal pathway was presented at the Experts Meeting described above.

The causal pathway was not designed to function as a clinical practice guideline or algorithm for decisions regarding patient care. It was developed solely to provide guidance throughout all phases of the systematic review process specific to the project.

Literature Search

The published literature was searched from January 1, 1988 to November 15, 2001, using Medline, Current Contents®, Cochrane Library, and PsychINFO databases. In addition, the bibliographies of all accepted studies and review articles from the past two years were searched for potentially relevant citations. The retrieval cut-off date was March 15, 2002.

English language and adult population published literature only from 1988 to 2001 was sought, utilizing the following search strategy:

fatigue syndrome, chronic [MeSH] or *chronic fatigue* [*syndrome*].
limits: English language, human subjects.

The preliminary search included studies published from 1990 to 2001. Based on recommendations proposed during the expert meeting, a decision was made to extend the search window back to 1988, the year of the first operational definition of CFS published by the CDC. It was believed that many important studies may have been published immediately after publication of this definition and needed to be included. It was also recommended that the *Journal of Chronic Fatigue Syndrome*, which is not indexed by Medline, but is indexed by PsychINFO, be searched for additional relevant citations. The search was expanded to include PsychINFO database.

Exclusion Criteria

All citations and abstracts were printed and screened at MetaWorks for any mention of diagnosis and/or treatment of CFS disability or impairment (Level I screening) and reviewed for the following exclusion criteria:

- Review, meta-analysis, abstracts, letters, case reports, editorials, commentaries, and unpublished study reports.
- Studies published prior to 1988.
- Studies written in languages other than English.
- Pharmacokinetic and pharmacodynamic studies.
- Animal or *in vitro* or tissue level studies.
- Studies not *related to* or not *specific to* CFS disability or impairment.
- Studies containing < 2 patients as total sample size.
- Pediatric patient population.
- No information related to disability or impairment.
- Outcomes not extractable.
- Mixed population (unable to separate CFS from other populations).
- Studies focused on pathophysiology of CFS (lab findings/lab techniques).
- Studies not conducted in the United States, Canada, Australia or Western Europe.

The geographic limitation was imposed because the purpose of this report was to inform policy pertaining to CFS patients in the United States, and it was believed that studies pertaining to disability in CFS patients in non-Western countries would not be generalizable to CFS patients in the United States.

When it was not possible to determine the eligibility of the study from the abstract alone, full studies of abstracts lacking obvious exclusion criteria were retrieved for Level II screening, during which both inclusion and exclusion criteria were applied. Level II screening forms are shown in Appendix C.

Inclusion Criteria

The following study designs were accepted: observational (prospective, retrospective, and cross sectional), or interventional [randomized controlled trials (RCTs), non-randomized controlled trials (nRCTs), uncontrolled case series (UCS)].
Studies were required to report:

- CFS diagnosed according to one of the four accepted CFS definitions.
- CDC 1988[11] or CDC 1994[10]
- Oxford 1991[12]
- Australia 1990[13]
- Adult patients with CFS *and* disability.

- Medically determinable physical or mental impairment in CFS patients (measures of symptom severity, functional or cognitive impairment, physical activity, exercise testing, general health, or psychiatric impairment).
- At least one objective measure related to disability, per SSA guidelines.

Upon completion of Level II screening, all accepted articles were eligible for data extraction. Due to the abundance of different scales reported in each of the studies, an additional screen was performed, in which each study was reviewed. Outcomes and scales reported in each study were then extracted. From this screening process, studies that specifically reported work outcomes were selected, and pertinent data were extracted from each study.

Linked Studies

After the accepted studies were determined, linked studies were identified. These were studies in which the same patient population was reported in more than one study. Studies which contained primary data were assigned "parent" study status. "Child" studies contained supplemental information, such as follow-up data or additional analyses. Data elements were extracted from the parent studies, and supplemented by information presented in linked ("child") studies, when appropriate.

Rating the Evidence

All eligible studies were evaluated for both internal and external validity at the time of data extraction (Appendix D). One method was developed specifically for this project. Papers received 1 point for each of the following:

1. CFS is defined according to acceptable criteria, and all patients met these criteria,

2. Tests for medically determinable physical and/or mental impairment are specified and reported,

3. Control group, if present, was similar in clinically important demographic factors at the start of the study,

4. All subjects enrolled were accounted for in followup,

5. Confidence intervals or p-values were reported for numerical results,

6. Work activity or disability status was reported.

Thus, papers could receive a maximum of six points for internal validity. All studies were awarded at least two points for internal validity, because they were required to fulfill the first two criteria in order to be accepted into the database. External validity had a scale of 0-2, with zero

points awarded for a study in which the patient sample was self-selected from the CFS population, and two points if the patient sample was a random sample or all patients from a CFS cohort. Thus, the possible range of scores for each study was 2-8.

Study quality was also evaluated using a scale that graded studies based on study design (prospective longitudinal vs. cross-sectional), sufficient patient number, well-matched groups, and well-validated measurement instruments.[38] In addition, RCTs were evaluated based on a validated quality score in which points were awarded for reporting method of randomization, blinding, and withdrawals.[39]

Data Extraction

Data Extraction Forms (DEFs) were designed specifically for this project (see Appendix C), and pilot tested on a small sample of eligible studies. The pilot test allowed for necessary edits to the DEF to be made prior to implementation on all studies. Key data from each eligible study were extracted by a researcher recording data from published articles onto a DEF, and reviewed by a second researcher, checking all DEF fields against the published report. Differences were resolved prior to data entry. In all cases, at least one physician reviewed each study. Dual review of all data served to reduce error and bias in the data extraction process. The data were then entered into MetaWorks' relational database of clinical studies, MetaHub™.

Key data elements sought for extraction from each study included:

Study Characteristics:

- Citation, publication date
- Location
- Study duration, design
- Single time point or longitudinal study
- Industry sponsorship (sponsor name or not reported)
- Validity Score (see Appendix D)
- Quality Score (see Appendix D)
- Total number of patients enrolled
- CFS patients
- Healthy Controls
- Geographic location
- Institution

Treatment Arm Characteristics:

- Number of patients enrolled or randomized
- Number of patients evaluated for efficacy and safety
- Age: years (mean, median, and range)
- Gender distribution

- Duration of CFS symptoms
- Education (years)
- Employment status
- Number of patients working full-time
- Number of patients working part-time
- Number of patients unemployed
- Number of patients receiving disability benefits
- Number of patients with work limitations due to illness
- Number of patients with other medical or psychiatric diagnoses

Interventions:

- Behavioral therapy
- Psychiatric therapy
- Drug therapy
- Exercise therapy

Outcomes:

- Number of patients evaluated at followup
- Employment status
- Number of patients working full-time (including full-time students or "housewives")
- Number of patients working part-time
- Number of patients unemployed
- Number of patients receiving disability benefits
- Number of patients with work limitations due to illness
- Number of patients improved, unchanged, or worse
- Scales, by domain: baseline, outcome, or change in each score
- Cognitive
- Disease or symptom severity
- Exercise testing
- Functional
- General health
- Mental (psychiatric or affective)
- Physical Activity
- Work

The investigators categorized each scale according to one of the above domains. Some scales, such as the Checklist of Individual Strength (CIS),[26] the Sickness Impact Profile (SIP),[27] and the Medical Outcomes Study – Short Form 36 (MOS SF-36),[28] had subscales in multiple domains. Scales in the cognitive domain included the Wechsler Adult Intelligence Scale (WAIS),[29] the Hopkins Verbal Learning Scale,[30] the Everyday Attention Questionnaire (EAQ),[31] and the concentration subscales of the CIS and SIP. Scales in the disease or symptom severity domain included the Chalder Fatigue Scale,[32] and the Profile of Mood States (POMS) subscales

for fatigue, vigor, and activity.[33] The exercise testing domain included treadmill endurance tests and measures of maximum oxygen output capacity (VO_2 max). The functional domain included the total SIP scale. The general health domain included the MOS SF-36 and the Karnofsky Performance Scale (KPS).[34] The mental (psychiatric or affective) domain included the Beck Depression Inventory (BDI)[35] and the Symptom Checklist 90R (SCL 90R) subscale for depression.[36] The physical activity domain included the POMS, MOS SF-36, and SIP subscales for activity. The work domain was mainly captured as number of patients working; however the SIP work subscale was also included. This is not a complete list of scales encountered in the literature, but it encompasses the major categories. As many papers used different scales, organizing them by domain was a necessary and important first step in considering combining data from different studies. Citations for scales extracted from accepted studies are listed in Appendix E.

For each study, results from a maximum of three scales in each of the domains available were extracted. Other results available were noted as "other outcomes." Where more than three scales in a given domain of interest were reported for the same study, decisions on which scales to extract were made using the following criteria, applied sequentially:

1. Scales with a higher number of patients evaluated were extracted preferentially over those with fewer patients evaluated.

2. Scales with group means reported were preferentially extracted over those reported as group medians.

3. Scales with measures of dispersion (standard deviation or standard error) were preferentially extracted over scales where the mean or median for the group was reported, but no measure of dispersion was available.

4. Named scales, for example MOS SF-36, BDI, or Chalder fatigue scale, were preferentially extracted over study-specific or unidentified scales, on the assumption that these scales might be more amenable to pooling across studies.

5. Where, in a single domain, both total and component scale results were reported, the total was extracted preferentially.

After data extraction of all studies, decisions on which scales to analyze were based upon frequency of use.

Database Development

Data were entered from the DEFs into a relational database of clinical trials. When data entry was complete, 100 percent of the data entries were checked back against the original DEFs. In addition, a 20 percent random sample of data in the completed database was checked against the DEFs. An error rate in excess of 2 percent of this sample would have triggered a 100 percent recheck of all data elements entered into the database.

Statistical Methods

Data listings and summary data were prepared for study, patient, and treatment level characteristics, and for outcomes of interest. After the database was complete, verified, and locked, data were entered into table shells. In general, study and patient characteristics and outcomes variables were summarized using standard descriptive statistics weighted by study sample size. Given the heterogeneity of the parameters measured in different studies, the sparse reporting of common impairment measures along with similar work data, and the frequent lack of information about ranges and distributions of the instruments used, pooling of impairment scale results across studies was not possible.

Role of Consultants

The eight participants from academic and community settings who attended the multidisciplinary meeting on November 15, 2001 served as our technical expert panel (TEP), and are listed in Appendix F. All TEP members received copies of the minutes from the meeting, causal pathway, and draft report. Additionally, during the course of the project, periodic conference calls were held with the topic nominator (SSA), the Task Order Officer from AHRQ, and the external co-investigator, Dr. Nelson Gantz. During these conference calls, project updates were provided and issues of concern were addressed.

Peer Review

A group of eleven peer reviewers (Appendix F) was assembled to review a draft version of this report. The panel was composed of experts in CFS, disability, occupational medicine, family practice, and psychiatry. All reviewers were asked to complete the peer review form relative to the content of the report (Appendix G), and were encouraged to provide additional written comments as well. All responses from the TEP and peer reviewers were reviewed and, where appropriate, are incorporated into this final report.