Chapter 2. Methods

The basis of this evidence report is a systematic comprehensive review and evaluation of the literature relevant to five topic areas proposed by the Social Security Administration.

Literature Search and Review

Sources

The primary sources of literature were MEDLINE® (1966-April 2003), CINAHL® (1983-April 2003), Cochrane Database of Systematic Reviews, and Web of Science. Searches of these databases were supplemented by reviews of reference lists contained in all included articles and in relevant review articles and meta-analyses.

Search Strategies

The basic search strategy used the National Library of Medicine's Medical Subject Headings (MeSH) key word nomenclature developed for MEDLINE® and was adapted for use in the other databases. The searches were limited to the English language and to human subjects. For efficacy of treatment topics, the searches were also limited to studies with randomized controlled trial designs. The texts of the five major search strategies are given in Appendix B. In addition, we used Web of Science (Thompson ISI, Philadelphia, PA) to identify articles that cited the recent McDonald criteria from the International Panel on the Diagnosis of Multiple Sclerosis.[5]

The searches yielded a total of 1487 citations, whose records are maintained in a ProCite (Thompson ISI ResearchSoft, Berkeley, CA) database.

Abstract and Full-text Screening Criteria

The seven specified questions spanned several topic areas and produced a considerably large and varied literature, which complicated the screening process. For each question, we developed fairly detailed inclusion/exclusion criteria. The titles and abstracts of the 1487 articles were reviewed against these criteria by at least two of five clinical investigators ("title-and-abstract screening" stage). Where no abstract was available, the title, source, and keywords were screened. At this stage, articles were included if requested by one investigator. The full text of each article passing the title-and-abstract screening was obtained for further review.

At the "full-text review" stage, each article was independently evaluated by two investigators, who forwarded their decisions to the task order manager for recording and comparison. If indicated, reviewers were asked to reconcile differences of opinion and return a reconciled final decision. If reviewers had difficulty reaching agreement, or submitted indecisive codes, the principal investigator was the arbiter.

We developed detailed screening instructions for the title-and-abstract screening (Appendix C) and additional decision rules for the full-text screening (Appendix D). For the full-text screening, we also produced a summary decision sheet, on which screeners recorded their include/exclude decision, research question assignment, and specific exclusion criterion (if appropriate) for each article.

Note: Appendixes and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/clinic/epcindex.htm.

Summaries of the results of the title-and-abstract screening and full-text review are provided in Tables 1 and 2.

Data Abstraction and Development of Evidence Tables

We determined that the data from the included articles could be abstracted directly into an evidence table template, which would serve as a "data abstraction form." To facilitate the development of the evidence tables and to use everyone's particular skills and time to their best advantage, the writer/editor began the data abstraction process with a partial abstraction of each article that was included at the full-text review stage. The partial abstraction/evidence table included descriptions of the study design, interventions, number of subjects at the start of the study, and the *types* of outcomes data to be collected; this partial abstraction was forwarded to the primary abstractor. The completed evidence table was returned to the writer/editor, who checked for completeness and consistency of information and then forwarded the table to a second investigator for over-reading. The over-reader returned the table to the writer/editor for final check of the completeness of the content, editing, and formatting. The data abstraction/evidence table templates for each research question are provided in Appendix E.

Quality Assessment Criteria

At the data abstraction stage, we evaluated each included article for factors affecting internal and external validity. The quality assessment criteria were incorporated into the last column of the data abstraction/evidence table templates (Appendix E) and varied by question. The questions and their associated criteria and range of responses follow.

Question 1a: *What is the reliability of new McDonald criteria (incorporating supplementary information form radiologic and laboratory studies including magnetic resonance imaging [MRI], visual evoked potential [VEP], and cerebrospinal fluid [CSF] analyses) compared with long-term follow-up diagnosis of clinically definite multiple sclerosis (MS) according to the Poser criteria?* The quality assessment criteria were:

1) Patients evaluated using Poser criteria regardless of results on initial tests? Yes/No/Unclear
2) Follow-up > 80%? Yes/No/Not reported (NR)/Not applicable (NA; relevant to retrospective cohort studies or case-control studies)

Question 1b: *What is the inter-rater reliability of diagnosis of MS according to Poser or McDonald criteria among neurologists or between neurologists and non-neurologist physicians?* There were two criteria:

1) Evaluating physicians blinded to one another's diagnosis? Yes/No/Unclear
2) Did study sample include an appropriate spectrum of patients (not just "difficult" cases)? Yes/No/Unclear

Question 2: *What clinical indicators, including particularly time-course of impairments, predict physical or mental impairment at 12 months?* The criteria were:

1) Study described as "population-based"? Yes/No
2) Sample of patients assembled at a *common* point in the course of their disease? Yes/No/Unclear
3) Sample of patients assembled at an *early* point in the course of their disease? Yes/No/Unclear

4) Follow up > 80%? Yes/No/NR/NA (retrospective or case-control study)
5) Outcomes assessed using a widely used scale? Yes/No
6) Outcomes assessed in a blind fashion? Yes/No/Unclear
7) If subgroups with different prognoses are identified: (a) was there adjustment for important prognostic factors? Yes/No/Unclear/NA; (b) was there independent validation? Yes/No/Unclear/NA

Question 3a: *Among patients with MS, do current disease-modifying treatments result in long-term improvements in physical or mental outcomes compared to placebo or usual care?* AND Question 3b: *Among patients with MS, do treatments aimed at symptom management result in improvements in physical or mental outcomes compared to usual care?* The criteria used for each of these questions were:

1) Described as "randomized"? Yes/No
2) Method of randomization clearly described? Yes/No
3) Concealment of allocation? Yes/No/Unclear
4) Described as "double-blind"? Yes/No
5) Patients blinded? Yes/No/Unclear
6) Investigators blinded? Yes/No/Unclear
7) Outcome assessors blinded? Yes/No/Unclear
8) Number of withdrawals in each group stated? Yes/No

For crossover trials only:

9) Period or carry-over effects? Yes/No/Not discussed
10) Washout period? Yes (give duration)/No
11) Number of patients in each sequence clearly described? Yes/No
12) Were patients who did not complete all of the period excluded from the analysis? Yes/No/Unclear

Question 4: *Among individuals with MS, what physical, mental, laboratory, or radiographic findings have been associated with inability to work?* AND Question 5: *Among individuals with MS, how does elevated temperature or other environmental factors impair the capacity to work?* The quality assessment criteria for these two questions were:

1) Study described as "population-based"? Yes/No
2) Follow up > 80%? Yes/No/NR/NA
3) Work outcomes assessed using a widely used scale? Yes/No
4) Work outcomes assessed in a blind fashion? Yes/No/Unclear
5) If subgroups with different work ability are identified: (a) was there adjustment for important prognostic factors? Yes/No/Unclear/NA; (b) was there independent validation? Yes/No/Unclear/NA

We did not sum the criteria into an overall quality assessment score, but rather we considered and reported each criterion individually. We favored this approach for several reasons:

- Previous work has shown that numeric grading systems may not discriminate well between "high-quality" and "low-quality" studies, even for randomized trials.[6,7]
- Development and use of a new quality score would require additional work for validation, for which there was no time or budget allocation in the task order.
- Identification of specific weaknesses in each study was helpful in identifying trends, which in turn assisted with our recommendations for future research.

- Describing key design components, rather than assigning a single aggregate score, is also consistent with recent recommendations from an expert panel on meta-analysis of observational studies.[8]

Peer Review Process

We employed internal and external quality-monitoring checks through every phase of the study to reduce bias, enhance consistency, and verify accuracy. Examples of internal monitoring procedures include: three progressively stricter screening opportunities for each article (title-and abstract screening, full-text article review, data abstraction review); hands-on involvement of three individuals (two clinicians) in each data abstraction; agreement of at least two clinicians on all included studies.

Our principle external quality-monitoring device was the peer-review process. Nominations for peer reviewers were solicited from several sources, including a technical advisory panel and interested federal agencies. The list of nominees was forwarded to the Agency for Healthcare Research and Quality (AHRQ) for vetting and approval.