Introduction

The Social Security Administration (SSA) operates the world's largest and most stringent disability program, processing more than 3.5 million claims each year, with multiple sclerosis (MS) representing the third most common neurological diagnosis cited as the cause for disability.1 The purpose of this project, nominated by SSA and contracted through the Agency for Healthcare Research and Quality (AHRQ), is to determine whether current medical knowledge supports the SSA's stated policies regarding MS. In January 2003, the Duke Evidence-based Practice Center began work on this 13-month task to review evidence from the medical literature for use in updating SSA's listing of impairments for multiple sclerosis (MS) and for revising its disability policy (if indicated).

Research Questions

The seven major research questions addressed during this review are as follows:

Question 1a: What is the reliability of new McDonald criteria (incorporating supplementary information from 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 MS according to the Poser criteria?

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?

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

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?

Question 3b: Among patients with MS, do treatments aimed at symptom management result in improvements in physical or mental outcomes compared to usual care?

Question 4: Among individuals with MS, what clinical indicators, including particularly time-course of impairments, predict physical or mental impairment at 12 months?

Key Terms and Definitions

Knowledge of the terms used in the SSA disability evaluation process, components of that process, and Medical Listing criteria related to MS is critical to the reader's understanding of this report. To assist in the preparation of the report, SSA provided explanations of terms and processes as currently defined by SSA regulations and rulings. The terms cited below, as well as other terms and processes used by SSA for disability determination, are defined and described in the SSA publication, Disability Evaluation Under Social Security 2003.2

The statutory definition of “Disability” is “the inability to engage in any substantial gainful activity by reason of a medically determinable physical or mental impairment(s) which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months.” This definition differs from the clinically used definition of the World Health Organization's International Classification of Impairments, Disabilities and Handicaps,3 which defines disability as “any restriction or lack of
ability to perform an activity in a manner or within the range considered normal for a human being.” While much of the medical literature uses the latter, broader definition, the reader must be aware that the goals of this report relate to the statutory definition.

The following terms are defined by current (2003) SSA regulations:
- “Claimant” is anyone who has filed a disability claim.
- “Substantial Gainful Activity” is the ability to earn an average of $800 per month.
- “Medically Determinable Impairment” is a physical or mental impairment that results from anatomical, physiological, or psychological abnormalities which can be shown by medically acceptable clinical and laboratory diagnostic techniques.
- “Evidentiary Requirements” for disability determination are described by SSA regulation. An acceptable medical source must report signs, symptoms, and laboratory findings diagnostic of an impairment. Although a claimant’s reported signs and symptoms are not sufficient to meet the evidentiary requirements for establishing the presence of a medically determinable impairment, all available evidence including the claimant’s report of symptoms is used to evaluate the impact of any documented impairment(s) on the claimant’s ability to carry out work tasks.
- “Severe Impairment” is defined by the agency as any “impairment that more than minimally limits the claimant’s ability to do basic work activities.”

The regulations include a Listing of Impairments for each body system that define disability. Often referred to as the “medical listings,” this list allows quick disability determinations to be made on the basis of medical criteria alone. The SSA publication, Disability Evaluation Under Social Security 2003, under the neurological category of impairments, includes Listing 11.09.

11.09 Multiple Sclerosis with:
A. Disorganization of motor function as described in 11.04; or
B. Visual or mental impairment as described under the criteria in 2.02, 2.03, 2.04, or 12.02; or
C. Significant, reproducible fatigue of motor function with substantial muscle weakness on repetitive activity, demonstrated on physical examination, resulting from neurological dysfunction in areas of the central nervous system known to be pathologically involved by the multiple sclerosis process.

Full details on the Medical Listing for multiple sclerosis, including the imbedded references to sections 2, 11, and 12, are available in the above-cited SSA publication.

“Residual Functional Capacity” is assessed when a claimant is determined to have a “severe” impairment that does not meet or equal the intent of the medical listings. Physical capacity (lifting, carrying, walking, standing, sitting, pedaling, and so on) and mental capacity (cognitive and behavioral, thought processing, concentration, pace, behavior) are assessed in determining residual functional capacity.

To adjudicate claims by individuals with MS for disability benefits, SSA must determine whether the claims file includes information from an acceptable medical source that documents the signs, symptoms, and laboratory findings that are diagnostic of a physical or mental impairment. SSA adjudicators also determine whether the impairment would be expected to more than minimally interfere with the claimant’s capacity to carry out basic work activities for at least 12 consecutive months or end in death. If a severe impairment is identified, the adjudicator determines whether the medical findings meet or equal an impairment in the medical listings. If the documented impairment does not meet or equal a listed impairment, the adjudicator must determine the claimant’s residual functional capacity and consider vocational factors prior to making a final disability determination.

**Methods**

A systematic and comprehensive search of the medical literature was conducted and was followed by a thorough review and evaluation of the literature determined to be relevant to the major research questions.

**Literature 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**

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. In all, there were five major searches:

1. **Search 1** was a general search targeting MS and employment issues that merged search terms for multiple sclerosis, transverse myelitis, and optic neuritis with employment terms such as disability evaluation, work capacity evaluation, employment, and activities of daily living. No study designs were excluded.

2. **Search 2** was targeted to studies on the reliability of diagnostic criteria for MS. Major search terms employed were multiple sclerosis (exploded), multiple sclerosis/employment (limited to diagnostic articles), text word options for poser and mcdonald, and exploded terms reproducibility of results/
or observer variation/ or psychometrics, along with the text word reliability. No study designs were excluded.

3. Search 3 focused on treatment of fatigue for MS and specified several drugs used in the treatment of MS-related fatigue. No study designs were excluded.

4. Search 4 looked for a wide range of symptomatic therapies (other than fatigue) and disease-modifying therapies. A wide selection of treatments was specified, and the search was limited to randomized controlled trial designs.

5. Search 5 was focused on the predictive value of the McDonald diagnostic criteria, specifically on their use of additional paraclinical diagnostic methods (MRI, VEP, and CSF) and on studies reporting sensitivity, specificity, and reproducibility.

All searches, including narrowly focused sub-searches, yielded 1,487 potentially relevant citations.

Abstract and Full-text Screening Criteria

For each question, we developed fairly detailed instructions and decision rules for the screeners' reference. There were very broad inclusion requirements for abstracts: MS study subjects and potential relevance to any of the five questions. For the full-text screening, screeners were asked to record their include/exclude decision, research question assignment, and, if appropriate, exclusion criteria that detailed insufficiencies in study design and clinical substance requirements.

The titles and abstracts of the 1,487 articles were reviewed against the inclusion/exclusion criteria by at least two of five clinical investigators. The full text of each article passing the title-and-abstract screening was retrieved from the library 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.

Approximately 50 percent of the articles were included after the abstract screening and full-text article review stages.

Data Abstraction and Development of Evidence Tables

Data from articles included after full-text screening were abstracted directly into an evidence table template, which served as a data abstraction form. The study's writer/editor began the process with a partial abstraction of each included article. The partial abstraction included descriptions of the study design, interventions, number of subjects at the start of the study, and types of outcomes data to be collected. The partial abstraction form was forwarded to a clinician for completion and then returned to the writer/editor, who checked it for completeness and consistency of information and forwarded it to a second clinician for over-reading. The over-reader returned the table to the writer/editor for a final check of the completeness of the content, editing, and formatting.

At the end of the data abstraction stage and the very close scrutiny of each article, 168 articles were included.

Results and Discussion

The primary goal of this review was to examine the evidence in the medical literature for data that can guide policy in determining disability in MS patients. Although the literature in general (and certain studies in particular) suffers from limitations, reasonably strong conclusions can still be drawn for most of the seven research questions.

Reliability of Criteria for Diagnosing MS

This topic encompassed two questions:

**Question 1a: What is the reliability of new McDonald criteria (incorporating supplementary information from radiologic and laboratory studies including MRI, VEP, and CSF analyses) compared with long-term follow-up diagnosis of clinically definite MS according to the Poser criteria?**

**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?**

Analytic approach. Regarding Question 1a, the most important difference between the Poser criteria and the new McDonald criteria is the addition of MRI findings to the diagnosis of MS, in lieu of the presence of a second attack. Our approach to this question was to identify studies in two categories: (1) those that specifically compared the new McDonald criteria to long-term diagnosis of clinically definite MS according to the Poser criteria; and (2) those that provided data on the accuracy of various MRI techniques, VEP analyses, and CSF analyses as supplements to clinical diagnosis of MS.

For Question 1b, the relevant diagnostic criteria were the Poser and McDonald criteria plus any other clinical, laboratory, neurological exam, MRI, VEP, CSF, or other data supporting the MS diagnosis. Results had to describe data on agreement or disagreement on the MS diagnosis between evaluating physicians. Agreement statistics could include kappa scores, sensitivity and specificity rates, or other data of the type that could be used to complete a two-by-two table.

Results. The validity of the McDonald criteria is well-supported by two types of evidence: (1) follow-up studies of patients with clinically isolated syndrome (CIS) diagnosed according to the McDonald criteria and (2) studies that correlate specific MRI findings (components of the McDonald criteria) with clinical diagnosis. First, two studies show that between 73 and 94 percent of patients presenting with CIS who go on to develop clinically definite MS over 1 to 4 years of follow up could be diagnosed with MS according to the McDonald criteria (but would have been undiagnosed under
From the studies identified in the review, the McDonald criteria is reasonably high, ranging from 83 to 87 percent. Second, many studies support the MRI component of the McDonald criteria by showing a strong and consistent association between the number of T2 lesions on MRI and the subsequent development of clinically definite MS among patients with CIS or optic neuritis.

Two studies examined the inter-rater reliability of neurologist-physicians in diagnosing MS according to the Poser criteria; one of these also examined inter-rater reliability in diagnosing MS according to the McDonald criteria. We found no data examining inter-rater reliability among non-neurologist clinicians. Overall, there was substantial agreement between observers in classifying MS. Poorer agreement was observed in determining whether a patient had one or more “attacks” of MS and in interpretation of MRI.

Discussion. From the studies identified in the review, the McDonald criteria appear to have substantial evidence for validity and offer the obvious potential advantage of resulting in an earlier diagnosis of MS than the Poser criteria permit. The McDonald criteria have been criticized for their complexity in comparison with previous criteria; however, we found data that demonstrate that these criteria yield a good overall diagnostic reliability, at least as good as the previous Poser criteria. However, data about reliability are available only for neurologists specializing in MS; adoption of the new criteria by clinicians with less expertise could result in deterioration of reliability. Further research on the inter-rater reliability of these criteria in broader clinical settings would be helpful to determine the quality of MS diagnosis.

Prediction of Physical or Mental Impairment at 12 Months

The research question for this topic was What clinical indicators, including particularly time-course of impairments, predict physical or mental impairment at 12 months?

Analytic approach. There were four main categories of clinical predictors of particular interest to the analysis: (1) clinical characteristics such as exacerbation rates, disease type, age at disease onset, sex, degree of remission after relapse, and type and number of neurological symptoms; (2) imaging studies, particularly MRI; (3) laboratory test results such as apolipoprotein E (APOE) e4 allele and intrathecal immunoglobulin-G (IgM) synthesis; and (4) self-reported health status using validated scales.

Our evaluation was limited to those studies with a time course of 12 months (SSA’s statutory limit), a timeframe which treating physicians would not ordinarily consider an important decision point. The course of MS has typically been studied over time horizons of many years.

Results. We found relatively little data describing changes in neurological or other impairments over 9 to 24 months; however, we used the data that were available to approximate the 12-month time horizon dictated by statutory requirements. Clinical characteristics have been the best studied, with four reports providing evidence for this review. Brain and spinal cord MRI have not been shown to be promising. Suggestive evidence is available for laboratory markers and self-reported quality of life, but these indicators will need further study to establish their reliability and utility. While clinical features do not individually provide reliable guidance on prognosis, multivariate predictive models based on relatively easy-to-obtain features may have better performance; such models have not, however, been validated.

Discussion. The ability to predict the future course of MS has been an active area of research; however, most studies examining disease course do so over relatively long time periods (5 to 20 years). The limited predictive ability of some multivariate models has not been validated in populations other than those in which the models were developed; thus, their value for predicting disability has yet to be determined.

Disease-modifying Therapies and Long-term Improvement

Research Question 1a was targeted to current disease-modifying therapies: 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?

Analytic approach. Interventions of interest were all current (2003) disease-modifying immunomodulatory and immunosuppressive treatments. Outcomes of interest were absolute improvements that might result in an individual who is unable to work becoming able to work. The following domains were considered: physical functioning (primarily Expanded Disability Status Score [EDSS]), relapse frequency, cognitive functioning, quality of life, and adverse events.

Results. Most of the data suggest that few patients improve on disease-modifying therapy. Those few who do improve generally do so only in the range of 1.0 point on the EDSS. We found no data regarding improvement in work ability and no data that would correlate a 1.0-point improvement in EDSS with improvement in work ability. The significance of a 1.0-point EDSS improvement varies depending on baseline EDSS score (because the scale is non-linear), but the improvement data available are not generally stratified according to baseline EDSS score. With regard to work ability, the significance of the available data on clinical improvement is unclear. We found no data that quantified individual patient improvement with regard to cognitive function or quality-of-life measures.

Discussion. Our review does not support a conclusion that the current therapies are likely to result in substantial improvement in a significant proportion of patients with MS. This finding is consistent with expert opinion and demonstrated by the design inherent in current clinical trials, that is, the use of lack of decline in EDSS scores as the primary
outcome measure. Current therapies are generally regarded as allowing for a modest reduction in progression of MS – particularly in the relapsing-remitting patient population – but are not generally expected to result in significant long-term improvement. Recently, however, combination therapies have begun to be used in the treatment of MS; such combinations of current therapies or new therapies may have greater potential to result in improvements in neurological status.

**Symptom Management and Improvement**

Symptom management was the focus of Question 3b: *Among patients with MS, do treatments aimed at symptom management result in improvements in physical or mental outcomes compared to usual care.*

**Analytic approach.** The effectiveness of symptomatic therapies for spasticity, rehabilitation, urinary management, fatigue, depression, and cognitive impairment was examined. Relevant outcomes were analyzed within six categories: (1) symptom-specific functional status or quality-of-life outcomes; (2) physical functioning (primarily EDSS); (3) cognitive functioning; (4) work or employment outcomes; (5) generic quality-of-life outcomes; and (6) adverse events.

**Results and discussion.** Treatment aimed at alleviation of symptomatic manifestations of MS, rather than at the underlying disease, could have an important role in maximizing functioning among people with MS. Among the six areas we investigated, the degree of impairments and the effectiveness of the treatments varied. We found:

- Although drugs such as baclofen, diazepam, dantrolene and tizanidine are often used to reduce spasticity in MS, the research evidence for a beneficial therapeutic effect is inconsistent. Uncertain findings here, as with other symptoms (cognitive impairment, fatigue), may be due, in part, to measurement issues. Better measurement tools may be required in order to confirm the clinical impression that widely used anti-spasticity drugs such as baclofen, tizanidine, and dantrolene are more effective than placebo. Given current measurement techniques, it is not surprising that active-treatment comparison studies fail to show clinically important differences among these drugs.

- Physiotherapy interventions failed to influence impairments as measured by EDSS. These interventions were, however, associated with measurable changes in functional status. Improvements in health (handicap) were observed in the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) and several other measures. The interventions employed in rehabilitation studies were multifaceted, and it is difficult to attribute beneficial effects to particular components of the interventions.

- Depression treatments, including psychotherapy, behavioral therapy, and certain drug therapies, can lead to measurable improvements in mood, but the link to improved functional status and, further, to ability to work was not demonstrated in these studies per se. There are few data linking treatment of depression to improvements in other symptoms (such as fatigue or cognitive impairments) or other outcomes (such as functional status or quality of life).

- Measurement of fatigue is limited by a definition that spans several domains, leading to difficulty with validation. Amantadine appears to have some ability to alleviate fatigue in MS, as demonstrated by statistically significant differences in some outcomes in several trials; however, the clinical significance of these effects is likely small. Pemoline has been less often studied and shows results suggesting some effect. There is little support for the efficacy of 4-aminopyridine. Modafinil has shown promising results in phase-II trials, but has not yet been evaluated in a double-blind randomized controlled trial. Further research on new pharmacological therapies (such as modafinil) and development of additional data on the validity of instruments for fatigue measurement and their sensitivity to change would be helpful directions for future research.

- Studies of treatments for voiding dysfunction show clear improvements in symptoms, but provide less clear data on how improvements in urinary symptoms impact other areas of health, and no data on how these symptomatic improvements impact work ability. Desmopressin was highly effective at reducing urine volume and also consistently effective at reducing urinary frequency. This was shown to translate into improvements in uninterrupted sleep hours and in fewer episodes of incontinence. Physical treatments, including both pelvic floor rehabilitation and use of a handheld vibrator during micturition, were also shown to reduce urinary symptoms compared with control. Many interventions commonly used for urinary disorders in MS have not been studied in randomized controlled trials of MS patients. Commonly used interventions for which no randomized controlled trials have been performed among MS patients include anticholinergic and antimuscarinic drugs, behavior modification, and intermittent or indwelling urinary catheterization.

- None of the studied treatments for cognitive impairments has had a consistent measurable effect on cognitive performance in MS. Treatment of cognitive impairments has been little studied and indirectly studied, in the sense that most data on cognitive effects are inferred from studies aimed at treatment of fatigue or depression. One study suggested that fatigue symptoms do not correlate with cognitive impairment, though they do correlate with symptoms of depression. Future studies would benefit from more precise delineation of study population based on screening for cognitive performance deficits within a relatively narrow and defined range; this would likely improve the chances of finding a treatment effect and would also make
clearer the population for whom the results would be applicable.

**Association of Clinical Findings and Work Ability**

In contrast to the previous questions, Question 4 directly linked clinical results with ability to work: Among individuals with MS, what physical, mental, laboratory, or radiographic findings have been associated with inability to work? 

**Analytic approach.** The phrasing of this question predetermined the outcome of interest as ability to work. Findings reported as absolute and relative measures of physical and mental/cognitive function and laboratory and radiographic testing related to work activity were assessed.

**Results and discussion.** There is a significant gap between what is included in the literature and the type of research evidence required to link objective clinical measures (physical, mental, laboratory, and radiographic findings) with ability to work. Although objective physical and cognitive measures have been developed, their application in the occupational literature is sparse. Furthermore, assessment of how symptoms such as pain and fatigue impact work ability is essentially absent. The reported findings on work ability displayed some consistency across studies. For example, individuals who had higher EDSS levels or low cognitive function were more likely to report not working. However, the strength of association across these studies was not clearly demonstrated, as most reported frequencies or crude estimates of association. Several studies had small sample sizes, which hindered researchers from calculating risk estimates that were adjusted for potential biases such as age, education, level of employer assistance, job type, and desire to work. In addition, most studies considered only physical function or cognitive function, when both can hinder employment.

**Environmental Factors and Work Ability**

Similar to the previous question, the focus of Question 5 was the ability to work: Among individuals with MS, how does elevated temperature or other environmental factors impair the capacity to work?

**Analytic approach.** The evidence sought for this question was on the association of workplace environmental conditions and demands (ambient temperature, individual’s body temperature, heat or cold exposure) on the ability of an individual with MS to work. Relative and absolute measures of association were assessed.

**Results.** With regard to work impairment, limitation, or disability related to temperature conditions, we found remarkably little research that met our inclusion criteria; thus, this question remains mostly unanswered. The one included report confirmed that some MS patients perceive that excessive heat impedes their work capacity.

Discussion. The evidence provides no basis for generalizations such as maximum appropriate working temperature levels unique to MS patient populations. It is unlikely that medical data in Social Security Disability Insurance (SSDI) application files in the current era will include objective diagnostic test results identifying MS patients who respond adversely to heat challenges. However, subjective patient reports may describe such associations with or without clinician comment or correlation with objective clinical status measures. Although not necessarily founded on randomized controlled trial data, current clinical impression seems to hold that ambient and/or exercise-induced body temperature effects may bear a relationship to MS symptom status in some patients, perhaps more so than is thought to be the case for chronic disease states in general.

**Future Research**

Future research about work ability among individuals with MS can shed a great deal of light on factors that foster or hinder employment. Our full report, particularly the evidence reported on association of clinical findings and work ability, highlights significant evidence and information gaps concerning

- Patterns of MS patient reports regarding functional limitations.
- Information commonly collected in medical encounters with MS patients (and therefore available to SSA).
- Knowledge about the impact on performance of specific work tasks of commonly objectified parameters such as coordination, strength, and vision, and especially of factors such as fatigue or cognitive dysfunction, which are either difficult to measure or are less commonly assessed in detail.
- Effective research methods for categorizing job or task demands in such a way as to isolate those demands that are likely to be “critical” for an SSDI applicant with MS.

In the context of these gaps, it may be productive to pursue research approaches that simultaneously address four domains:

1. Subjective reports (this domain is not sufficient alone for SSDI determination purposes).
2. Objective clinical data (ideally of the sort commonly encountered in medical records).
3. In-depth objective measures (which may be available and not widely applied clinically, but which may be used with subsets of subjects to explore correlation with other domains).
4. Work status measures (ideally longitudinal, with stratifications based on work demands).

Such an approach may apply to thermal sensitivity as well, with some additional specification and focus. Parallel assessment of concomitant ambient temperature, physical
exertion, and core body temperature would address key relevant physiological exposure factors.

Outcome measures could include the domains outlined above, for example:

- Self-perceived well-being and level of symptoms such as fatigue.
- Clinical parameters such as walking speed or muscle strength.
- In-depth measures such as potentially associated biomarkers or physiological parameters.
- Work status measures, including absenteeism and disability benefits use.

**Availability of the Full Report**

The full evidence report from which this summary was taken was prepared for the Agency for Healthcare Research and Quality (AHRQ) by the Duke Evidence-based Practice Center, under Contract No. 290-02-0025. It is expected to be available in May 2004. At that time, printed copies may be obtained free of charge from the AHRQ Publications Clearinghouse by calling 800-358-9295. Requesters should ask for Evidence Report/Technology Assessment No. 100, Criteria to Determine Disability Related to Multiple Sclerosis. In addition, Internet users will be able to access the report and this summary online through AHRQ's Web site at www.ahrq.gov.

**Suggested Citation**


**References**


