Chapter 3. Results

Diagnostic Reliability of McDonald Criteria

Introduction

This section addresses results for 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 multiple sclerosis (MS) according to the Poser criteria?*

The diagnosis of MS has traditionally been based on clinical history and examination, and on the absence of an alternative diagnosis. Paraclinical tests such as imaging techniques,[9] evoked potentials,[10] and CSF analyses have been used to aid diagnosis, but their precise role in formal diagnostic criteria for MS has been a source of debate. With the advent of MRI, which provides much more detailed images of the brain and spinal cord than were possible with computed tomography (CT), debate over the role of paraclinical tests in the diagnosis of MS has been renewed.

Of the various diagnostic criteria proposed for MS, the most widely used have been those by Poser et al. (1983).[4] The Poser criteria focus on objective examination evidence of abnormalities in at least two separate areas of the central nervous system, with historical information to suggest at least two periods of MS involvement over time. The criteria may be summarized as follows:[4]

- Clinically definite MS:
  - 2 attacks and clinical evidence of 2 separate lesions; *or*
  - 2 attacks; clinical evidence of 1 lesion and paraclinical evidence of another, separate lesion.
- Laboratory-supported definite MS:
  - 2 attacks; clinical or paraclinical evidence of 1 lesion; and CSF immunologic abnormalities; *or*
  - 1 attack; clinical evidence of 2 separate lesions; and CSF immunologic abnormalities; *or*
  - 1 attack; clinical evidence of 1 lesion and paraclinical evidence of another, separate lesion; and CSF immunologic abnormalities.
- Clinically probable MS:
  - 2 attacks and clinical evidence of 1 lesion; *or*
  - 1 attack and clinical evidence of 2 separate lesions; *or*
  - 1 attack; clinical evidence of 1 lesion and paraclinical evidence of another, separate lesion.
- Laboratory-supported probable MS:
  - 2 attacks and CSF immunologic abnormalities.

In light of the growing potential of paraclinical tests – particularly MRI – to contribute to the diagnosis of MS, the International Panel on the Diagnosis of Multiple Sclerosis met in July 2000 to reassess existing diagnostic criteria. The panel recommended changes to the criteria that (1) integrated MRI into the overall diagnostic scheme and (2) provided for the diagnosis of primary progressive disease.[5] The new criteria (here referred to as the McDonald criteria)[5] were

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

designed to be used by practicing physicians and adapted, as necessary, for clinical trials. They dropped the Poser categories of "clinically definite," "laboratory-supported definite," "clinically probable," and "laboratory-supported probable" MS and proposed that the disease be diagnosed using the following criteria:[5]

- 2 or more attacks; objective clinical evidence of 2 or more lesions
  - No additional data needed
- 2 or more attacks; objective clinical evidence of 1 lesion; *plus*
  - Dissemination in space, demonstrated by:
    - MRI, *or*
    - 2 or more MRI-detected lesions consistent with MS plus positive CSF, *or*
    - Further clinical attack implicating a different site.
- 1 attack; objective clinical evidence of 2 or more lesions; *plus*
  - Dissemination in time (demonstrated by MRI).
- 1 attack; objective clinical evidence of 1 lesion (monosymptomatic presentation, clinically isolated syndrome); *plus*
  - Dissemination in space, demonstrated by:
    - MRI, *or*
    - 2 or more MRI-detected lesions consistent with MS plus positive CSF, *and*
  - Dissemination in time, demonstrated by:
    - MRI *or*
    - Second clinical attack.
- Insidious neurological progression suggestive of MS; *plus*
  - Positive CSF, *and*
  - Dissemination in space, demonstrated by:
    - 9 or more T2 lesions in brain, *or*
    - 2 or more lesions in spinal cord, *or*
    - 4-8 brain lesions plus 1 spinal cord lesion, *or*
    - abnormal VEP associated with 4-8 brain lesions, *or*
    - abnormal VEP with fewer than 4 brain lesions plus 1 spinal cord lesion; *and*
  - Dissemination in time, demonstrated by:
    - MRI, *or*
    - Continued progression for 1 year.

The publication of the McDonald criteria has renewed the discussion of the value of incorporating paraclinical testing, particularly MRI, in the diagnosis of MS. The Social Security Administration (SSA), aware of the difficulties brought on by improper or inadequate diagnosis, has sought evidence from the literature regarding the value of the various diagnostic criteria. Question 1a seeks specific information regarding the reliability of the McDonald criteria.

In this review, we sought to identify specifically those studies that evaluated whether patients diagnosed with MS according to the McDonald criteria would later meet the Poser criteria for clinically definite MS. We also sought articles that, although not specifically using the McDonald criteria, may have evaluated components of the McDonald criteria against later diagnosis of MS according to the Poser criteria.

There exists a significant literature regarding the diagnostic utility of newer MRI techniques, such as magnetization transfer or various measures of atrophy,[11-21] but we did not include such articles as they are not a part of currently used diagnostic criteria in general or the McDonald criteria specifically.

Results

Thirteen articles[22-34] describing 11 different study populations met the inclusion criteria for this question (see Evidence Table 1a in Appendix F). Of the included articles, two[27,34] specifically addressed the comparative performance of the Poser and McDonald criteria, nine[22-25,28,30-33] examined the performance of various standard MRI techniques in addition to clinical diagnosis, one[29] evaluated the performance of other paraclinical testing in addition to Poser criteria, and one[26] contained data regarding the performance of the McDonald criteria, although it was not specifically designed to answer this question.

Those studies that did not specifically address the relative performance of the Poser and McDonald criteria nevertheless provided background information regarding the improved diagnostic yield achieved by adding MRI and paraclinical testing to the clinical diagnosis. These studies generally did not utilize MRI in the same manner as the McDonald criteria, and therefore it is difficult to apply these results directly. For example, several studies examined the performance of MRI abnormalities at baseline with subsequent diagnosis of MS. As the data summarized in Table 3 indicate, the results of these studies document the significant predictive value of baseline MRI findings for the subsequent development of MS. The McDonald criteria utilize serial MRI studies to add to the diagnostic performance. Therefore, although these studies provide background information documenting the utility of MRI, they do not specifically address the McDonald criteria use of MRI.

The two studies that did specifically address the performance of the McDonald criteria in comparison with the Poser criteria[27,34] reported remarkably similar results (Table 4). The first study[27] evaluated 119 patients with clinically isolated syndromes at baseline and reevaluated 50 of these patients at 3 years. MRI studies from 3 months, 1 year, and 3 years were retrospectively analyzed in a blinded fashion using the McDonald criteria. In this study, the diagnosis of MS had been made clinically (using the Poser criteria) prior to the retrospective application of the McDonald criteria. The second study[34] followed 139 patients with clinically isolated syndromes prospectively for a mean of 39 months, with 86 patients followed for at least 3 years. In this study, the diagnosis of MS was made prospectively according to the Poser criteria and compared with the prospective application of the McDonald criteria. The two studies showed sensitivity of the McDonald criteria ranging from 0.73 to 0.94, with specificity ranging from 0.83 to 0.87, when compared to diagnosis using the Poser criteria.

Using the Poser criteria as the diagnostic gold standard in this way is problematic, mainly because the limited duration of any clinical study means that some patients may be classified as non-diseased at the conclusion of a study who later go on to meet the Poser criteria for MS. Once the Poser criteria are defined as the gold standard, other tested criteria can at best correlate highly with Poser diagnosis. It is possible that the McDonald criteria might perform better than the Poser criteria (in terms of sensitivity, ability to diagnose MS early, and/or association with prognosis) if both could be compared to a (hypothetical) true gold standard. The high level of agreement of McDonald criteria with later Poser criteria diagnosis observed in the studies reviewed here is not inconsistent with this possibility.

Inter-rater Reliability of Diagnosis with McDonald and Poser Criteria

Introduction

The preceding discussion of Question 1a was concerned with the validity of new diagnostic criteria for MS; this section examines the inter-rater reliability of the application of these criteria as posed by 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?* High inter-rater reliability of the McDonald criteria would suggest that their use in clinical practice should achieve similar validity to the findings of the literature as reviewed in Question 1a.

Our primary goal was to evaluate studies that provide direct analysis of inter-rater reliability of the McDonald or Poser criteria. We also sought studies that evaluated the inter-rater reliability of components of the McDonald criteria, such as MRI studies of T2 lesions or gadolinium-enhancing lesions. We excluded studies of inter-rater reliability of MRI techniques that are not utilized in the McDonald criteria, such as measurements of cerebral atrophy, magnetic resonance spectroscopy, and measurements of T2 lesion volume.

Results

Two studies met our inclusion criteria (see Evidence Table 1b in Appendix F).[37,38] Both examined inter-rater reliability of neurologist-physicians in diagnosing MS according to the Poser criteria. One of the studies[38] also examined inter-rater reliability for diagnosing MS according to the McDonald criteria. We found no data examining the inter-rater reliability among other clinicians.

Ford et al. (1996)[37] examined the inter-rater reliability of the Poser criteria. Overall, there was substantial agreement between the two observers in classifying MS (kappa = 0.65; 95 percent confidence interval [CI], 0.52 to 0.78). There was disagreement as to whether a patient had one or more "attacks" of MS. Agreement was substantial for clinical evidence of separate lesions and was almost perfect for both paraclinical evidence and laboratory support (all kappa values > 0.90). The primary disagreement was in defining "attacks" of MS, and the authors appropriately noted that this might be due to the retrospective nature of the evaluation.

The primary shortcomings of this study were that only two evaluators were compared, and diagnoses were based on retrospective analysis of data from clinical records. This process is significantly different from prospective diagnosis. Prospective diagnosis allows the clinician access to detailed information that may be inadequately recorded in the medical record, such as lesion location, intensity of enhancement, and overall appearance of MRI changes. Likewise the details of clinical exacerbations may be communicated to an examining physician making prospective diagnoses, but may be inadequately or incompletely recorded in the medical record. In general, retrospective diagnosis is frequently based on relatively less complete information as recorded in the medical record.

Zipoli et al. (2003)[38] specifically addressed Question 1b by examining the inter-rater reliability of both the Poser and McDonald criteria (see Table 5). In this study, four neurologists assessed all patients consecutively admitted for diagnosis to a university department of

neurology from September 2001 through June 2002 who had been followed for at least 6 months. The mean follow up was 12.7 months. The study evaluated 41 MS patients, of whom 15 had relapsing-remitting MS, two had secondary progressive MS, five had primary progressive MS, and 19 had clinically isolated syndromes. Three additional patients who were not diagnosed with MS were included in the evaluation. Data were abstracted from medical records onto standardized forms by a non-evaluating neurologist. No diagnoses were recorded. The four evaluating neurologists were stated to have similar clinical experience in MS diagnosis and management. The study does not report any specific training of these neurologists with regard to the McDonald criteria or MRI analysis. The four evaluating neurologists reviewed the standardized forms and all MRI scans independently without discussing their findings. Eighteen patients had follow-up MRI scans in addition to baseline scans.

In this study, the primary difficulty in the McDonald criteria appeared to be decreased agreement in MRI interpretation, specifically in those patients with high lesion loads. The authors commented that this study utilized neurologist evaluators rather than neuroradiologists. Previous studies have correlated level of radiographic training with agreement in interpretation.[39,40] Judging dissemination in time was of particular difficulty in those patients with clinically isolated symptoms. The authors suggested that neuroradiologists be encouraged to interpret scans in MS patients with the McDonald MRI criteria in mind, providing specific information regarding lesion location and timing.

Predictors of Physical and Mental Impairments at 12 Months

Introduction

This section examines the evidence pertaining to Question 2: *What clinical indicators, including particularly time-course of impairments, predict physical or mental impairment at 12 months?* The notion here is that an individual diagnosed with MS but not yet work disabled could, within the follow-up period of disability evaluation, become work disabled. In formulating a response to this question, we considered three crucial study features: the assessment of possible prognostic features, the period of follow up, and the assessment of a valid measure of physical or mental impairment. The assessment of possible prognostic features had to occur between the time of diagnosis of MS and the occurrence of work disability; in most studies this assessment occurred at or near study entry, but this was not necessary. The 12-month timeframe specified in this question was the major limiting criterion in the number of studies that could be included for evaluation; to expand the pool of studies, we included any evaluations in which follow up occurred within a window of 9 to 24 months from study entry (or measurement of the purported prognostic feature). Thus, a study that related prognostic data available at one time with outcome data at another time 9 to 24 months later would be eligible; furthermore, a study might provide such data over multiple different time frames (see Figure 1). In all cases, the available outcome measure was the Disability Status Scale (DSS) or the Expanded Disability Status Scale (EDSS). For purposes of judging a level approximately associated with work disability, we focused on the transition from an EDSS score of less than 6 to greater than 6. Since deficits associated with MS unrelated to acute exacerbations are not expected to regress, we did not include studies in which patients were already severely disabled.

Results

Twelve publications[41-52] describing 11 separate studies met the inclusion criteria (see Evidence Table 2 in Appendix F). Four main categories of predictors were analyzed in the included articles: (1) clinical characteristics; (2) imaging study results; (3) laboratory test results; and (4) self-reported status. The following discussion is organized by these categories.

Clinical characteristics. Clinical characteristics were described in four reports[42,43,45,49] describing three studies. In a 1989 study by Goodkin et al.,[45] exacerbation rates and adherence to disease types were analyzed. A non-population-based cohort of 425 patients was followed; 254 patients with definite MS completed the evaluations, including the following patterns of disease: stable (n = 80), relapsing-remitting stable (n = 155), relapsing-remitting progressive (n = 48), and chronic progressive (n = 142). After a follow up of 1 to 5 years (mean, 2.6 years), disease pattern was not found to be a stable characteristic of patients, with 44 percent of patients changing from stable or relapsing-remitting to progressive disease, while 40 percent of patients with progressive disease stabilized. Further, disease pattern did not predict change in EDSS scores at 2 years.

Runmarker et al. (1994)[49] followed patients with relapsing-remitting course at onset for up to 25 years. Of 308 patients identified at the onset, 200 had sufficient data for analysis. Multivariate survival models were developed to predict time from onset to start of progressive disease and time from onset to a DSS score of 6; similar models were developed for predicting events from the end of the fifth year of disease onward. Several factors appeared to be predictive in one or more of the models, including age at onset, sex, degree of remission after relapse, mono- or poly-regional symptoms, type of affected nerve fibers, and number of affected neurological systems. Patients with early onset had a low initial risk of progression, rising to a maximum over about 15 years; patients with a late onset had a higher initial risk, which rose for several years, then fell. The predictions were not validated internally or externally.

The most recent included study, by Cottrell et al. (1999a[42] and 1999b[43]), evaluated a prospective, population-based cohort of patients with primary progressive MS. The original cohort was followed for a mean of 23 years; this was supplemented by an additional cohort, with a follow-up time not reported. The probability of progressing from one DSS level to the next was highest initially (87 percent probability of progressing from level 1 to 2 in 1 year), relatively constant for patients in DSS levels 2 to 5 (probability of progression of 1 level ranging from 26 percent to 40 percent), and lower for patients in DSS levels 6 to 9 (ranging from 2 percent to 10 percent). Using a multivariate analysis of potential predictors of progression to DSS of 8, several factors were significant, including sex, age at onset, years from onset to reaching a DSS of 3, and number of systems involved. The model was not validated, nor was its discriminative ability examined. However, in univariate survival curves, the discrimination of these individual factors appears modest at best.

Imaging study results. Three studies of MRI satisfied inclusion criteria, with the majority of exclusions attributable to lack of comparison of MRI with subsequent course. Two studies[46,48] evaluated monthly MRIs on patients with relapsing-remitting disease. Koziol et al. (2001)[46] studied 50 patients in the context of a trial of cladribine in an effort to identify predictors of exacerbation. Three potential predictors were considered: enhancing lesions in three consecutive monthly MRIs, new enhancing lesions in three consecutive months, and new hypointense lesions in three consecutive monthly images. In all cases, the sequential MRIs had poor sensitivity (36 percent, 31 percent, and 31 percent, respectively), but higher specificity (85

percent, 89 percent, and 89 percent, respectively). As a result, the positive predictive value was low (approximately 25 percent for an exacerbation in the next month), but the negative predictive value was fairly high (89 percent for each predictor).

In a substudy of patients in a trial of glatiramer acetate, Rovaris et al. (2003)[48] examined the univariate correlations of baseline T1 and T2 lesion volume versus change in EDSS score. Without adjusting for baseline EDSS (which was correlated with baseline lesion volume), modest Spearman rank correlation coefficients were seen in the comparison of baseline lesion volume to EDSS at 9 months. Absolute changes in EDSS were not reported, and the investigators acknowledged that the associations were not clinically strong.

In one imaging study,[50] 28 patients with various disease patterns and 13 healthy control subjects had spinal cord images performed at the second cervical level. While changes in spinal cord size were noted, there was no statistically significant association noted between cord area and change in EDSS.

Laboratory test results. Laboratory tests were evaluated in four included studies: one of apolipoprotein E (APOE) $\in$4 allele,[41] two of immunological markers (one in blood,[51] one in cerebrospinal fluid[52]), and one of evoked potentials.[44]

APOE $\in$4 allele, associated with impaired neuronal repair, was assessed by Chapman and colleagues for its ability to predict clinical progression of MS.[41] Forty-seven patients with relapsing-remitting MS were evaluated for the presence of APOE $\in$4 allele. Results indicated that the presence of this genotype was associated with more rapid disease progression, with mean increase in EDSS of 4.0 during a 2-year follow up in patients with the allele compared to a mean increase of 2.7 for individuals without the genotype. Notably, the APOE $\in$4 allele was not found to be associated with other factors possibly associated with relapse, including baseline EDSS, prior rate of relapse, or exacerbation rate in the prior 2 years.

Villar et al. (2002)[52] examined intrathecal immunoglobulin-M (IgM) synthesis in 22 patients with relapsing-remitting disease to determine if its presence during early stages of the disease correlated with a worse prognosis. Patients were evaluated as to their time to conversion to clinically definite MS, number of relapses, and changes in EDSS. While half of the patients with intrathecal IgM synthesis progressed at least one EDSS unit after 1 year, none of the patients without this marker progressed in that period. Notably, relapse rate was similar in patients with and those without intrathecal IgM synthesis.

Trotter et al. (1989)[51] measured interleukin-2 (IL-2) levels in serum of 10 patients with chronic progressive MS and 12 normal controls, as well as concavalin A suppressor, mitogen stimulation, and phenotyping of peripheral blood mononuclear cells. Only IL-2 levels at baseline correlated with disability over 18 months. Though no cutoff value for elevation was selected a priori, a value of 40 U/mL corresponded to a sensitivity of 67 percent and specificity of 100 percent for the outcome of worsening one or more units in EDSS. Notably, these results were not examined after adjusting for other predictive features, particularly those diagnostic of MS.

One study was identified that compared results from motor and visual evoked potentials at baseline with subsequent EDSS. Fuhr et al. (2001)[44] studied 30 patients with relapsing-remitting (n = 25) or secondary progressive (n = 5) patterns using measures of motor and visual evoked potentials, such as the sum of Z scores of central motor conduction time. Results of these tests at baseline were moderately correlated with EDSS score at 24 months (0.43, $p = 0.03$). No cutoff was defined a priori for a positive evoked potential study. However, from the data provided, a Z score at baseline exceeding 0 was associated with a sensitivity of 53 percent and specificity of 70

percent for any worsening of EDSS at 24 months. For the population studied, the overall likelihood of progressing was 17/27 (63 percent); of 11 patients with a positive study, nine (82 percent) progressed, and of 15 with a negative study, eight (53 percent) progressed.

Self-reported status. Self-reported health status was compared to subsequent course in one study.[47] Of 97 patients with relapsing-remitting MS, six were lost to follow up before 12 months. Quality of life (QOL) was assessed using the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36), with results dichotomized into poor/fair versus good/very good/excellent. The relative risk for any worsening in EDSS at one year was 1.9 (95 percent CI, 1.0 to 3.5). In absolute terms, having a poor or fair QOL assessment at baseline was associated with a 42 percent likelihood of progression to any degree, while a better QOL response was associated with a 23 percent chance of worsening. This association persisted with multivariate adjustment for baseline clinical characteristics. The association did not extend to other dimensions of the SF-36. Moreover, no assessment was made regarding the presence of a disability claim.

Disease-modifying Therapies and Long-term Improvement

Introduction

In this section we report results for 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?* In attempting to answer this question, we were looking for data indicating whether individual patients *improve* on current disease-modifying therapy. Specifically, we were looking for data suggesting that disease-modifying treatments might result in enough clinical improvement to decrease an individual's level of impairment to the point where he or she might not be disabled at some future time.

As readers familiar with the relevant literature will recognize, most studies of the efficacy of disease-modifying agents assume that individual patient improvement is unlikely and do not consider data on improvement as part of the outcome assessment. In general, the literature assumes that while disease-modifying agents may be expected to reduce progression, they are not likely to result in improvement. This "disconnect" between the question we were asking and the assumptions of the clinical trial literature led us to conduct a broad search of the literature to examine indicators of efficacy, with a more specific examination of any available data delineating individual patient improvement.

Our search strategy focused on reports that described data from randomized controlled trials of disease-modifying agents in patients with any degree of impairment. For the data to be meaningful with regard to improvement, we selected studies that included the outcome measures of change in physical function, cognitive function, work/employment status, and reduction in relapse rates where outcomes were reported over at least 12 months. We excluded studies evaluating therapies not currently in use. Ultimately, our focus was on those studies that reported data on individual patient outcomes – particularly the *improvement* of individual patients. We excluded studies that did not contain placebo controls because the demonstration of individual patient improvement without a comparison to control patients might only reflect

improvement that is part of the natural history of MS rather than a treatment-specific improvement.

Results

Fifty-one publications[53-103] describing 34 separate trials, met our inclusion criteria (see Evidence Table 3a in Appendix F). The treatments evaluated included interferon beta, glatiramer acetate, mitoxantrone, glucocorticoids, intravenous (IV) immunoglobulin-G (IgG), azathioprine, cyclophosphamide, plasma exchange, methotrexate, cladribine, cyclosporine, and combinations of these therapies. Outcome measures were primarily group or mean changes in rates of relapses, changes in rates of progression on the EDSS, and occasionally changes in some measures of cognition or quality of life. These studies are adequate to answer issues regarding the efficacy of therapy, but do not provide information regarding individual patient improvement.

Relapse rates. Data on relapse rate outcome measures are summarized in Table 6.[53,54,59,62-65,68,72-78,82-84,86,89,92,93,95-97,100,103] In most of these studies the baseline mean annual relapse rate was approximately one relapse per year. The effect of most of the therapies studied was a mean reduction in the range of 0.3 relapses per year; higher reductions (up to one relapse per year) were reported after treatment with IV IgG when baseline relapse rates were in the range of 1.5 to 2 relapses per year. We found only one study that reported reductions in individual patient relapse rates. Kappos et al. (2001)[79] reported that treatment with interferon beta increased the proportion of patients who were relapse-free or had decreased relapse rates from 45 percent (placebo) to 53.1 percent (treatment). In summary, the effect of treatment with any of the currently used therapies is a mean reduction in relapses of less than one relapse per year. The benefit may be higher in patients with higher baseline relapse rates, but we did not find data stratified in this manner.

Physical function (EDSS). Studies reporting data on individual patient improvement in EDSS scores are summarized in Table 7.[53,55,58,59,63-65,67,68,75-78,80,83,90,96,97,102,103] Most of these studies provided a definition (and in some cases more than one definition) of improvement, usually corresponding to at least a one-point decrease in EDSS score. Changes of half a point (0.5) in EDSS may not be reliable due to uncertainty regarding this degree of change and uncertain clinical significance. The proportion of patients meeting varied definitions of improvement in placebo groups ranged from 0 to 22 percent; the differences in proportions of patients improving ranged from -1 to 49 percent, although most fell into the range of 0 to 12 percent. Higher EDSS change thresholds (-2 points or -3 points) resulted in lower proportions of patients meeting improvement definitions in one study.[55]

Improvement in cognitive function. Two included studies evaluated the efficacy of therapy with regard to cognitive function.[76,78] Both studies demonstrated benefit of therapy over placebo but did not present data with regard to the quantitative improvement seen in individual patients.

Improvement in quality-of-life measures. Two included studies (described in three publications) evaluated the efficacy of therapy with regard to quality-of-life measures.[59,66,79] Both studies demonstrated benefit of therapy over placebo in the groups studied, but did not present data with regard to the quantitative improvement seen in individual patients.

Symptom Management and Improvement

Introduction

This section of the report considers Question 3b: *Among patients with MS, do treatments aimed at symptom management result in improvements in physical or mental outcomes compared to usual care?* Several symptomatic complaints are common among patients with MS. Among the most common are fatigue and voiding dysfunction. Less well recognized are cognitive dysfunction and pain. We assessed the effectiveness of therapies aimed at particular symptoms including spasticity, voiding dysfunction, fatigue, depression, and cognitive impairment, as well as more comprehensive treatment aimed at multiple symptoms or multiple areas of functional status, i.e., rehabilitation. We excluded trials of drugs that are not currently commercially available in the US.

We sought randomized controlled trials reporting physical and mental health outcomes. We excluded studies that reported data on patient preferences for treatments as the only outcome. We examined outcomes 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. In contrast to our treatment of the previous question (Question 3a), we did not require data to be reported as the proportion of patients meeting a definition of symptom "improvement." We report all data comparing treatment and control groups on the above outcomes regardless of the format in which they were presented or analyzed.

Results

We included a total of 68 articles describing 65 separate studies (see Evidence Table 3b in Appendix F). By topic, these included: for spasticity, 35 articles[104-138] describing 32 separate studies; for rehabilitation, 10 articles/studies;[139-148] for depression, eight articles/studies;[149-156] for fatigue seven articles/studies;[157-163] for voiding dysfunction, seven articles/studies;[164-170] and for cognitive problems, one article/study.[171]

Spasticity. Among the 32 studies of specific treatments aimed at management of spasticity were randomized controlled trials (RCTs) of oral medications, intramuscular medications, and intrathecal medications, as well as physical treatments (magnetic stimulation and electrical neuromuscular stimulation). Trials of drug treatments were most common and considered the following agents: tizanidine, baclofen, diazepam, gabapentin, dantrolene, threonine, botulinum toxin, cannabinoids/delta-9-tetrahydrocannabinol (delta-9-THC), progabide, and amantadine. Baclofen was the subject of six comparisons with placebo,[105,109,117,124,129,130] six comparisons with tizanidine,[104,108,113,122,126,134,136] two comparisons with diazepam,[106,110] and one with stretching exercises.[105] In addition, one trial compared intrathecal baclofen with placebo.[125]

Tizanidine was studied in two placebo comparisons,[133,137] six comparisons with baclofen (see above), and one with diazepam.[127]

Dantrolene was studied in one placebo comparison[111] and one comparison with diazepam.[131,132] We found two studies each of gabapentin,[107,121] progabide,[119,120,128] botulinum toxin,[114,135] and threonine[112,116] versus placebo. Delta-9-THC was tested in two trials[115,138]

against both placebo and cannabis sativa plant extract. Electrical[118] and magnetic[123] neuromuscular stimulation were each compared to sham stimulation.

Study quality. All 32 trials randomly allocated patients to treatment groups, but the methods of randomization were clearly described for only 5/32 (16 percent). Most failed to provide enough information to determine whether there had been adequate concealment of the allocation schedule (25/32; 78 percent). In the seven studies that reported enough information for evaluation, there was adequate concealment in four, and inadequate concealment in three.

Thirty of 32 trials (94 percent) were described as double-blind; patients were described as blinded in 29 trials, and in one trial[118] it was unclear whether patients were blinded. Treating investigators were blinded in 28 trials; outcome assessors were blinded in 30 trials, but it was unclear whether they were blinded in another study.[105] Most studies (27/32; 84 percent) reported the number of withdrawals in each treatment group.

Twenty of the 32 studies were of crossover design; the remaining trials were of parallel-group design. Among the crossover trials, we assessed several dimensions of quality unique to the crossover design. The presence of period effects or carry-over effects was not discussed in over half the studies (11/20; 55 percent). One trial reported having a period or carry-over effect and appropriately analyzed first-period data only.[121] Most crossover trials (17/20; 85 percent) included a washout period. Thirteen (65 percent) failed to report clearly the number of patients assigned to each treatment sequence. In six studies it was unclear whether patients who did not complete both treatment periods were excluded; two studies appeared to use an inappropriate unpaired analysis retaining patients who did not complete the crossover in the final analysis.

Study populations. Only 11 of the studies describe the study population in terms of EDSS score at baseline; mean baseline EDSS ranged from 4.7 to 7.4 in studies reporting means. The two studies of botulinum toxin[114,135] appeared to include the most severely impaired patients, with median EDSS scores above 7.5. Other study populations included mostly subjects with EDSS scores in the range of 5.0 to 7.0.

Studies used a variety of outcome measures, many of which were not validated. Of the placebo-controlled trials, 12 of 20 (60 percent) used the Ashworth Scale or Modified Ashworth Scale. However, this scale was often combined with other measures[135] or summed across sides and muscle groups to create outcome variables and statistical testing that have not been validated.

Findings. Trial design and results are summarized in Table 8. Assessments of muscle tone were not consistently positive for most of the drug treatments, with the exception of two studies each of gabapentin[107,121] and progabide.[119,120,128] Results of functional assessments were largely inconclusive; the only positive findings on functional assessments were associated with delta-9-THC; deterioration in 9-Hole Peg Test (9-HPT) and Multiple Sclerosis Functional Composite (MSFC) was associated with delta-9-THC in one small study[115] while a much larger study found improvements in 10-meter walk time[138] with delta-9-THC compared to placebo. Both studies reported these changes in the absence of a significant effect on objective measures of muscle tone. However, beneficial effects in the latter study for mobility are not easily ascribable to the drug's psychoactive properties alone.

In contrast to oral treatments, which were of uncertain effectiveness, intrathecal baclofen treatment showed a profound effect on muscle tone and spasms in one trial;[125] however, the patients in this study were selected based on response to a pre-trial intrathecal dose of baclofen.

None of the trials comparing active drug treatments showed statistically significant differences in efficacy between agents, although there were some differences in tolerability.

Rehabilitation. Among the 10 studies we classified as rehabilitation were RCTs of a variety of interventions delivered in a variety of settings. Probably the most comprehensive was a study of home-based management, which included not only traditional rehabilitation services but also nursing, education, psychological support, and social services.[144] Several studies examined more typical rehabilitation interventions delivered in inpatient or outpatient settings.[139,140,142,145] One study examined supervised exercise,[143] and one physiotherapy.[148] Three studies used interventions that were probably less intense than traditional rehabilitations described as professionally guided self-care,[141] wellness intervention (education in health behaviors and lifestyle change),[146] and symptom management and adjustment (education and behavioral therapy).[147]

Only one study used an active treatment comparison group, comparing inpatient rehabilitation to outpatient rehabilitation.[139] Two other studies used control interventions presumed to be ineffective: a home exercise program[145] and self-exercise treatment.[142]

Study quality. All 10 trials randomly allocated patients to treatment groups, but the methods of randomization were clearly described for only five (50 percent). Most failed to provide enough information to determine whether there had been adequate concealment of the allocation schedule (six; 60 percent). In the four studies that reported enough information for evaluation, there was adequate concealment in three and inadequate concealment in one.

Because of the nature of the interventions, none of these studies was able to blind patients or therapists; however, four (40 percent) masked outcome assessors. Most studies (seven; 70 percent) reported the number of withdrawals in each treatment group. All but one of the studies was parallel-group in design; one study of home physiotherapy used a crossover design.[148]

Study populations. Most of the studies (seven; 70 percent) described the study population in terms of EDSS score at baseline. Two studies had populations with notably lower mean EDSS scores of 3.4: the single study of supervised exercise[143] and one of the more educational/behavioral intervention studies,[147] which notably had a broad range of EDSS scores, from 0 to 9. In other studies, the populations were more impaired on average, with mean EDSS scores of 5.9 to 6.2 and medians of 5.5 to 6.5.

Outcome measures used in these studies included measures of impairment (e.g., EDSS), disability (e.g., Functional Independence Measure [FIM]), and handicap (e.g., SF-36), and not only related to individual functions, but often used comprehensive measures of overall disability and handicap.

Findings. Physiotherapy interventions failed to influence impairments as measured by EDSS (see Table 9). These interventions were, however, associated with measurable changes in functional status. Improvements in health (handicap) were observed in the SF-36 and several other measures.

Depression. Among the eight studies of treatments aimed at depression were RCTs of a variety of interventions delivered in a variety of settings. Two placebo-controlled studies used drug treatments either alone[155] or with a psychotherapy co-intervention.[156] One study tested psychotherapy,[149] and five trials used behavioral therapy, described as cognitive-behavioral therapy[150,151,153,154] or cognitive remediation.[152]

The two drug studies were placebo-controlled and were the only studies to employ blinding of patients and investigators. The remaining studies used control groups that were passive (wait-list[150,151] or no treatment[152,154]) or active (current events discussion group,[149] support group[153]); one study used both an active and passive control group.[149]

Study quality. All eight trials randomly allocated patients to treatment groups, but the methods of randomization were clearly described for only two (25 percent). Six trials (75 percent) failed to provide enough information to determine whether there had been adequate concealment of the allocation schedule; the other two trials did not have adequate concealment of allocation. Because of the nature of the interventions, none of the behavioral therapy studies was able to blind patients or therapists; however, neither did any of these studies mask outcome assessors. Both studies of drug treatments were double-blind. Half of the studies failed to report the number of withdrawals in each treatment group. All but one of the studies were parallel-group in design; one study of amitriptyline used a crossover design.[155] Although this study did not report the number of patients in each sequence, it did employ a 1-week washout period and stated that there was no period or carry-over effect.

Study populations. Among the eight trials, the populations studied were quite variable with regard to the presence of depression or depressive symptoms. Some studies selected patients with MS without regard to the presence of a diagnosis of depression or depressive symptoms;[149,150,152] however, most studies either required a formal diagnosis of a depressive disorder[153,156] or required a certain score on a depression scale.[151,154] One study specifically included patients based on the particular complaint of episodes of involuntary laughing or weeping.[155]

Findings. The included trials are summarized in Table 10. All but one of the studies reported statistically significant improvement in mood outcomes; Mendoza et al. (2001)[152] failed to achieve statistical significance for changes in the Beck Depression Inventory (BDI), but the results did show a strong trend in favor of cognitive remediation. Treatment effects were similarly significant for studies of patients with depression diagnoses or those meeting minimum depression scale scores.

The studies were consistent in finding that untreated control groups[149-152,154] showed no improvement over time. In contrast, a control group of patients receiving psychotherapy as a co-intervention did show improvement.[156] Two studies that used active controls (current events discussion group[149] or support group[152]) found marginal improvements in some outcomes, but still allowed detection of treatment effects.

Fatigue. Among the seven studies of specific treatments aimed at management of fatigue associated with MS were RCTs of oral and topical medications. Amantadine was the subject of four placebo comparisons.[157-159,161] Pemoline was evaluated in three placebo comparisons,[159,161,163] two of which were three-arm studies including both amantadine and pemoline.[159,161] Two more placebo comparisons evaluated 4-aminopyridine[162] and transdermal histamine/caffeine.[160]

Study quality. All seven trials randomly allocated patients to treatment groups, but the methods of randomization were clearly described for only one.[157] None provided enough information to determine whether there had been adequate concealment of the allocation schedule. All of the studies were described as double-blind, and all described the number of dropouts or withdrawals from each treatment group.

Four of the trials used a crossover design.[157,158,162,163] One reported a period effect and appropriately analyzed first-period data;[157] the other three crossover studies tested for and reported no period or carry-over effects. One of the trials did not use a washout period.[162] Two reported the number of patients in each sequence.

Study populations. All studies selected patients based on complaints of persistent fatigue; some studies required either a threshold score on a fatigue measurement scale or a run-in period

during which fatigue symptoms were demonstrated to be stable. Mean EDSS scores at baseline were reported for six of the seven studies (86 percent). Two studies had populations with higher EDSS scores than the others: Rossini et al. (2001),[162] with a mean EDSS score of 6.2; and Gillson et al. (2002),[160] with entry criteria requiring EDSS between 5.0 and 6.5. The other studies had less impaired populations on average, with mean EDSS scores ranging from 2.6 to 4.3.

Findings. Included trials are summarized in Table 11. Four studies comparing amantadine with placebo found statistically significant results on a few outcome measures, and statistical trends on others, consistent with a treatment effect. The two largest studies[157,161] found statistically significant treatment effects on the MS-Specific Fatigue Scale (MS-FS) or visual analog fatigue scores. Two of three studies comparing pemoline to placebo also reported statistically significant treatment effects.[161,163] One small study reported a treatment effect on the Modified Fatigue Impact Scale (MFIS) associated with transdermal histamine/caffeine, but no effects on functional measures (25-foot timed walk, 9-HPT) or cognitive performance.[160] The single trial of 4-aminopyridine included in our review failed to show a treatment effect.[162]

Voiding dysfunction. Among the seven studies of treatments for voiding dysfunction were five placebo-controlled trials of desmopressin nasal spray,[164-167,170] one trial of pelvic floor rehabilitation,[169] and one trial of abdominal vibration or pressure.[168]

Study quality. All seven trials randomly allocated patients to treatment groups, but none clearly described the methods of randomization. None provided enough information to determine whether there had been adequate concealment of the allocation schedule. All of the desmopressin studies were described as double-blind. Blinding was not possible for the vibration and rehabilitation studies. All but one[167] of the studies described the number of dropouts or withdrawals from each treatment group.

Six of the seven studies were of crossover design; one study of amitriptyline used a parallel-group design.[169] Only one of the crossover trials incorporated a washout period,[168] and only one tested for carry-over or period effect.[170]

Study populations. Only two studies characterized populations in terms of EDSS, with means of 4.4 and 6.7. All studies recruited patients with MS and some kind of voiding dysfunction such as nocturia, urinary frequency, or incontinence.

Findings. Included studies are summarized in Table 12. The five studies of desmopressin nasal spray all demonstrated statistically significant reductions in short term urinary frequency (number of voidings) during 6-hour periods after dosing. Most studies dosed at bedtime to control nocturia, but one study used daytime dosing to control daytime urinary frequency.[166] At longer time intervals (24 hours) the treatments failed to show significant effects. The number of episodes of incontinence showed a statistical trend in one study,[170] but this outcome was not reported in the other studies of desmopressin.

Among the two nonpharmacological treatments, a program of pelvic floor rehabilitation (biofeedback and exercise) was effective in reducing incontinence episodes and nocturia;[169] this study also showed a measurable reduction in urinary symptom-related handicap ($p < 0.05$). A small study of the use of a handheld abdominal vibrator at time of voiding succeeded in reducing post-void residual urine volumes and also appeared to reduce incontinence, but not urinary frequency.[168] Interpretation of the results of this study was limited by poor reporting of the data and statistical analysis (no statistical tests were reported for several outcomes).

Cognition. Studies aimed at treatment of cognitive deficits in MS are challenging to characterize. We did not identify a single study that (1) selected a study population with

demonstrable cognitive deficits, (2) delivered a treatment (drug, behavioral, or psychological) aimed at improving cognitive performance, and (3) measured effects of treatment using tests of cognitive performance. However, one trial met at least two of these criteria and merits discussion here.[171] In addition, some of the studies aimed at treatment of depression assessed cognitive performance at baseline or as secondary outcomes.[152] Some of the rehabilitation trials evaluated mood-related outcomes in addition to functional outcomes, but none evaluated cognitive performance as an outcome measure. Finally, two of the studies aimed at treatment of fatigue also evaluated effects on cognitive function.[159,160]

Study quality. Lincoln et al. (2002)[171] describe both the method of randomization and adequate concealment of allocation. Because of the nature of the intervention, blinding of patients and therapists was not possible; however, outcome assessors were masked. The study reported the number of withdrawals and dropouts in each treatment group.

Study populations. No studies selected patients based on screening for cognitive deficits at study entry. Lincoln et al. (2002)[171] enrolled MS patients without regard to cognitive performance for their study of a cognitive intervention; however, 78 percent of the study population enrolled scored greater than 1 (the recommended cut off) on the mental disability question from Guy's Neurological Disability Scale (GNDS), and 95 percent either reported cognitive problems or showed significant impairment (score below cut off) on the Brief Repeatable Battery (BRB-N). Thus, while the proportion of patients with cognitive deficits may be elevated in this study due to selection biases, the prevalence of measurable cognitive deficits in other MS trials is likely also elevated.

Findings. Lincoln et al. (2002)[171] found no differences between the cognitive rehabilitation, assessment, and screening groups in any of the outcomes measured at 4 months (shortly after the end of the 6-week intervention) or at 8 months. Although the prevalence of cognitive deficits at baseline was relatively high, a large number of patients had self-reported cognitive deficits that were not detectible on GNDS or BRB-N, which may have hampered the ability to detect treatment effects. Furthermore, the population was heterogeneous with regard to their cognitive performance, which at best results in increased variance, making it more difficult to show statistical significance to differences between groups.

Geisler et al.(1996)[159] evaluated the effect of amantadine and pemoline on fatigue and cognition. Although the Symbol Digit Modalities Test (SDMT), a measure of attention and visuomotor search, showed a drug treatment effect (with the amantadine-treated group showing the greatest improvement), all other neurobehavioral outcomes failed to demonstrate a treatment effect. The authors acknowledge that the study did not have sufficient statistical power to detect small- to medium-sized treatment effects. But also, the study found no statistically significant correlation between fatigue symptoms and cognitive deficits either at baseline or post-treatment.

Gillson et al. (2002)[160] reported results from the Paced Auditory Serial Additions Test (PASAT) in a trial of transdermal histamine and caffeine in a population of MS patients with fatigue. No significant treatment effect was detected.

Association of Clinical Findings with Work Ability

Introduction

This section summarizes findings related to Question 4: *Among individuals with MS, what physical, mental, laboratory, or radiographic findings have been associated with inability to work?*

Aspects surrounding ability to work are multifactorial in nature involving individual skill level, education level, interest in working, financial needs, domestic responsibilities, transportation needs, and other factors such as social norms and expectations. Functional limitations (physical and/or mental) associated with MS are presumably the primary determinants of work capacity. Adverse, intermittent, and poorly measurable symptoms such as fatigue and pain compound the task of determining if an individual with MS is able to work. Employers' willingness or ability to provide workplace accommodations for workers who have limitations also impacts whether or not an individual with MS can work.

Several different types of physical and cognitive function measures (e.g., EDSS, disease subtype, mobility aids) were employed within and across studies to determine work ability among individuals with MS. For the purpose of this review we have grouped the summary of study findings according to these types of measures, which are listed in Table 13. Some studies used various types of measures and therefore are listed and discussed in more than one section. Categorizing studies or specific tests into physical function versus mental/cognitive function for this review was somewhat challenging since some tools measure both types of function. For example, the EDSS focuses on the functioning of numerous systems, including mental functioning, but is weighted more heavily towards testing ambulation. For the purpose of this review, tools like this are categorized according to the dominant area of function tested, resulting in the categorization of EDSS under "physical function." Tools testing multiple types of function are discussed in greater detail below.

Results

A total of 22 articles[172-193] representing 20 research studies were included in this review (see Evidence Table 4 in Appendix F). The majority of studies (n = 15) employed tests to measure physical function, while fewer studies (n = 5) measured cognitive or mental function (Table 13). To date, no studies have been ascertained that used laboratory or radiographic measures to determine work ability among individuals with MS.

Most of the studies (n = 18) were cross-sectional in design, where work status or work ability was measured at a single point in time, primarily at the time of study enrollment. The other two studies were retrospective in design, including one case-control study[193] and one retrospective cohort study.[181] No prospective studies, where changes in physical and/or cognitive function over time were considered in relation to changes in work status or ability, were identified for this review.

Twenty papers describing 18 studies used work status (yes/no or full-time, part-time, unemployed, retired, housewife) as a proxy measure for work ability among individuals with MS.[172-189,192,193] The remaining two[190,191] attempted to incorporate several aspects of work ability as a study outcome: the Hyllested criteria. The remainder of this section is organized by

types of measurement tools used to examine inability to work: Hyllested criteria, EDSS/DSS, cognitive measures, use of mobility aids, MS disease subtype, job type/characteristics, and self-report.

Hyllested criteria. As indicated above, only two[190,191] of the 20 included studies sought to determine ability to work among individuals with MS beyond the measurement of work status. Findings from these studies are reported separately (Table 14) from other studies that employed similar methods of measuring physical and mental functioning. Two cross-sectional studies conducted in Israel by Rozin et al. (1975[191] and 1982[190]) used similar methods for determining work ability among selected groups of individuals with MS ages 17 to 50 years. Interviews were conducted by social workers in the study participants' homes where demographic and occupational information was collected, as well as information about desire to be trained and employed. Subjects were evaluated by neurologists to determine degree of disability using a scale similar to the EDSS (described below) called the Hyllested criteria, which ranges from 0 (no functional disability, no residual signs) to 6 (bedridden, incontinent, requires constant supervision), with the mid-level score of 3 defined as moderate disability with work impairment sufficient to require a lighter job. Using these data, study participants were categorized into one of three groups: (1) Group A - completely handicapped with no rehabilitation potential; (2) Group B - potential for vocational rehabilitation, but unemployed or currently employed, but needs rehabilitation services for continuation of employment; or (3) Group C - currently working without need of rehabilitation intervention. All Group B patients underwent additional tests to evaluate their functioning potential by a rehabilitation physician, occupational therapist, and psychologist. Study participants were also categorized into types of disability (physical, mental, or both); however, it is unclear how researchers measured mental disability. Furthermore, it is unclear if physical disability was determined strictly on the basis of the Hyllested criteria or if additional information was used. (The earlier study by Rozin et al. [1975][191] addresses additional aspects of employment by examining disability type and level by job type, discussed in greater detail under "Job type/characteristics," below.)

Group A participants who were handicapped without rehabilitation potential were more likely to be disabled due to a decrease in physical or physical and mental function, with few being disabled strictly due to a decrease in mental function. Both studies[190,191] observed that about half of Group C (fully employed without need of vocational rehabilitation) were physically disabled and ranked as being either mild to moderately disabled on the Hyllested scale. Subjects in Group B (those who would benefit from vocational rehabilitation) were more likely to have moderate to severe physical disability. Although these patients had significant physical limitations, they were still considered to be individuals who would benefit from vocational rehabilitation and capable of working.

The obvious advantages of these studies include the consideration of work ability beyond work status, as well as the examination of both physical and mental function among the same sample population with regard to work ability, which most other studies did not report. However, a limitation of these methods is that researchers used current work status to determine ability to work, which could possibly bias the outcome of their evaluation. If a study participant was not working and expressed no desire to work, but was actually capable of performing a job, they may have been classified as someone who is not a candidate for vocational rehabilitation. Additional limitations include the omission of detailed information about how mental function was measured. These studies were also limited by small sample sizes, and data collected about education and marital status were not included as possible confounders in multivariate analyses.

Because these were cross-sectional studies, the levels of physical and mental functioning were measured at the time of the study and not at the time when study participants ceased employment. The timing between impaired function and inability to work was not established.

EDSS/DSS. Three cross-sectional studies (Table 15)[174,180,188] and one case-control study[193] used the EDSS, and two cross-sectional studies[183,187] used the DSS to assess ability to work among individuals with MS. The EDSS (and its earlier version, the DSS) is a clinical tool commonly used for rating neurological impairment in individuals with MS.[194] Clinicians determine a patient's EDSS level by first assigning a separate grade for eight functional systems including pyramidal, cerebellar, bowel and bladder, cerebral, brain stem, sensory, visual, and other functions. A composite of grades is then used to determine an individual's EDSS score ranging from 0 (normal neurological exam) to 10.0 (death due to MS).[194] The level of function for each of the eight systems is considered for EDSS score; however, assignment of a level is superseded by an individual's ability to ambulate (e.g., free from mobility aids vs. need for mobility aids), possibly giving more weight to ambulation than the other seven functional systems. For example, individuals with MS who are able to walk without ambulatory aid would receive a score of 0 through 4.5, whereas a need for constant bilateral assistant (e.g., canes, crutches or braces) would predetermine an individual to receive a score of 6.5. Although mental function is factored into the EDSS scoring system as one of eight systems, it is not considered independently of ambulation at any EDSS level.

As detailed in Table 15, a lower frequency of employment was consistently observed in groups with higher EDSS levels in all three cross-sectional studies. Unemployment among study participants with an EDSS ranging from 3 to 6 was reported to be approximately 42 percent,[174] 52 percent,[180] and 72 percent,[188] respectively, while employment among lower EDSS levels (≤ 2.5) was 37 percent, 42 percent, and 51 percent, respectively. Unemployment was most common among individuals with higher levels of EDSS (≥ 6.5). A case-control study[193] observed that mean EDSS levels were significantly different between unemployed cases (0 = 5.4; standard deviation [SD] = 0.1) and employed controls (0 = 4.5; SD = 0.1) with MS (p = 0.01). This is only a 1-point difference on the EDSS scale, but 4.5 and 5.4 straddle the scale's demarcation of work ability, with 4.5 defined as "able to work a full day" and 5.5 defined as "disability severe enough to preclude full daily activities."[194] These findings may reflect the timing of the neurological exam to assess EDSS, which was conducted at the outset of the study and not at the point when employment ceased. Physicians may not have been blinded to study participants' work status at the time of the exam, possibly biasing their evaluation.

Hammond et al. (1996)[183] conducted a large (n = 2099) cross-sectional study in Australia and reported that after adjusting for age, men with moderate DSS levels (4-6) were almost three times more likely to be unemployed (prevalence ratio [PR], 2.7; 95 percent CI, 2.1 to 3.6), and women were four times more likely to be unemployed (PR, 4.0; 95 percent CI, 2.7 to 5.8) compared to men and women (respectively) with lower DSS levels (0-3). Men and women with severe DSS (7-9) were also more likely to be unemployed (men PR, 17.9; 95 percent CI, 7.5 to 41.5; women PR, 24.6; 95 percent CI, 8.0 to 76.1) when compared to this same group. The second study[187] observed that a 1-point increase in DSS was associated with a seven percent decrease in the likelihood of being employed, and being male increased the probability of employment by 11 percent after controlling for numerous factors such as age, sex, education, marital status, and parenthood.

Findings from these studies suggest that individuals with higher EDSS/DSS levels are more likely to report not working. The three cross-sectional studies that examined EDSS had small

sample sizes such that adjustment of prevalence ratios for other aspects associated with work ability was not possible. Extrapolation of these findings is limited because they focus only on a single dimension of work ability.

Studies by Larocca et al. (1982)[187] and Hammond et al. (1996)[183] included multivariate analyses where adjusted estimates were reported; however, no measures of cognitive impairment or job characteristics and responsibilities were considered. Again, these studies were cross-sectional, and the assessment of EDSS during enrollment in the study failed to establish the timing between impaired physical function and inability to work.

From a more global perspective, a semantic issue with EDSS deserves mention. Level 5.5 denotes disability severe enough to preclude full daily activities. It is not clear to what degree clinicians equate this EDSS-based activity preclusion with being incapable of working, without exploring other aspects of work ability such as cognitive function and employer accommodations. Conversely, cognitive impairment sufficient to impair work capacity would not typically be reflected in the EDSS score.[189]

Cognitive measures. Three studies primarily examined cognitive function and work status among patients with MS (Table 16). Two of these[172,189] administered a battery of cognitive tests, while the third[177] collected data on cognitive function (attention/concentration, planning/organizing, retrospective and prospective memory) and ambulatory assistance through a self-report survey. The former studies examined a broad spectrum of function including verbal skills, memory, visuospatial perception, problem solving, and attention and concentration. In addition to these tests, Beatty et al. (1995)[172] also administered the Ambulation Index (which is highly correlated with the EDSS, r = 0.96). Ambulation, short-term memory, delay recall, age, and verbal ability were found to explain 49 percent of the variance in employment status. Patients who were still working attained significantly higher scores on most of the individual measures of cognitive performance and were impaired on significantly fewer cognitive domains. Rao et al. (1991)[189] reported that cognitively impaired patients were also less likely to be employed compared to individuals who were cognitively intact, but information on which specific cognitive tests (or impairments) were associated with employment was not reported, and level of physical function was not considered in the analyses. Self-perceived cognitive deficit and need for mobility assistance were also associated with unemployment, as were fewer years of education and age.[177] However, occupational level (socioeconomic index), number of people living at home, and duration of illness did not impact employment status. A study by Genevie et al. (1987)[179] also considered self-reported cognitive function in combination with physical function and other symptoms, but details of the types of cognitive limitations were not described (see Evidence Table 4 in Appendix F for details of study limitations).

One additional study focused on self-reported expressive communication disorder, and study participants were asked if the communication disorder interfered with employment.[173] Employment status among those with self-reported communication problems was less compared to the entire study sample. Methodological problems (described in Evidence Table 4 in Appendix F) prevented further interpretation of this study.

Common sense suggests that impaired cognitive function has the potential to seriously impact work ability. However, the three studies described in Table 16 do not provide the evidence needed to determine the type and/or level of cognitive impairment when an individual with MS is no longer able to work. Unlike the studies by Rao et al. (1991)[189] and Edgely et al. (1991),[177] Beatty et al. (1995)[172] provided far more detail about the specific cognitive tests that were associated with not working. In addition, these researchers also considered level of

ambulation in combination with cognitive function and demographic characteristics. Since only a global measure of variance was provided it is difficult to interpret the strength of association between each of these domains (cognitive function, level of ambulation, demographic characteristics) and work ability. The 1991 study by Rao et al.[189] did provide details about the types of tests that were administered, but used a global measure of "intact versus not intact" to examine work ability. The method of self-report of cognitive function used by Edgley et al. (1991)[177] has limitations in that someone with impairment may not be able to objectively measure their own level of cognitive function. Finally, the temporal relationship between cognitive impairment and cessation of work among study participants was not captured in these cross-sectional studies.

Mobility aids. The number and type of mobility aids study participants used was measured in two studies[176,186] as a proxy measure for degree of disability. Kornblith et al. (1986)[186] developed a three-level Mobility Dysfunction Index (MDI) ranging from no assistance needed (Level 1), to any combination of cane, walker, leg brace, etc. (Level 2), to use of a wheel chair for more than half the time in- or outdoors (Level 3). A 1-point increase in MDI decreased the probability of males working by 24.3 percent, while it decreased the likelihood employment for females by 15.4 percent, leading investigators to conclude that mobility was a major determinant of employment, while age and duration of disease were minor. Dyke et al. (2000)[176] considered the number of mobility aids used and reported that only 20 percent of the variance in employment among a sample of women was accounted for by the number of mobility aids used, age, and education.

Use of certain mobility aids (e.g., wheelchair) can certainly provide a measure of degree of physical disability, as well as indicate the possible level to which the disease progression has hindered physical function, but Dyke et al. (2000)[176] considered only the number of aids used. A limitation of using the number of mobility aids to measure degree of disability is that it most likely is not sensitive enough to detect changes in other aspects of disease status and mental function, and it certainly does not capture job requirements. Desk jobs that require only sitting may enable someone who uses a wheelchair, but is not cognitively impaired, to continue working.

Disease subtype. Although there is great variability in the course of MS, three subtypes of disease are generally recognized: (1) relapsing-remitting; (2) primary progressive; and (3) secondary progressive.[195] The terms of these subtypes have changed over time due to refinements made within each classification, which are reflected in the different terms used in the following studies. One cross-sectional study[184] and one case-control study[193] compared MS patients' current work status with disease subtype. Both studies reported a higher frequency of employment among study participants with relapsing-remitting compared to primary progressive[184] and relapsing-progressive MS.[193] These findings are consistent with the greater degree of disability typically noted among individuals with progressive MS;[196] however, the analyses for both of these studies were crude and did not consider other factors associated with ability to work, except for Jacobs et al. (1999),[184] which attempted to control for age by restricting analyses to individuals less than 60 years of age. Furthermore, disease subtype was not measured until enrollment into the study. The disadvantage of using disease subtype for determining work ability is that the range of cognitive and/or physical function within each classification can vary tremendously. Furthermore, these studies do not provide the needed information for the measurement of physical and/or cognitive function that results in cessation of employment.

Job type/characteristics. Six studies (Table 17) either focused primarily on job type or work characteristics[181,192,193] or included information about occupation as a secondary aim in their study.[183,187,191] Although the purpose of this review was to summarize information about measurements of physical and/or mental function among individuals with MS associated with inability to work, examining job requirements provides an indirect measure of the physical and/or mental levels of function needed to sustain employment.

In a case-control study conducted by Verdier-Talliefer et al. (1995)[193] several job characteristics were examined for their relationship with unemployment among MS patients. After adjusting for age, sex, type of disease, and level of education, an elevated odds of unemployment was observed among study participants whose jobs required physical strength (odds ratio [OR], 7.6; 95 percent CI, 3.2 to 18.2), manual precision (OR 3.1; 95 percent CI, 1.6 to 6.3), and frequent moves (OR, 2.5; 95 percent CI, 1.3 to 4.9). Furthermore, the odds of unemployment decreased when the job was a "desk job" (OR, 0.3; 95 percent CI, 0.1 to 0.5), or one that required sitting (OR, 0.3; 95 percent CI, 0.1 to 0.7). When all of the demographic and job characteristics were considered together in a multivariate model, work in the public sector was protective against unemployment (OR, 0.4; $p < 0.05$), and work requiring physical strength increased the odds of unemployment (OR, 4.5; $p < 0.001$). Analyses stratified by sex revealed that factors associated with unemployment for men involved a rigid work schedule (OR, 17.1; $p < 0.01$), while for women unemployment was strongly associated with work requiring physical strength (OR, 4.5; $p < 0.05$). These findings are consistent with those of Scheinberg et al. (1980),[192] who observed that currently employed ($n = 51$) study participants were more likely to hold jobs that were clerical (35.3 percent) or professional (37.2 percent) as opposed to skilled (13.7 percent) or unskilled (2.0 percent) labor. An early study by Rozin et al. (1975)[191] categorized study participants into groups according to their level of function and ability to work, described in greater detail above. This study crudely assessed changes in employment from the time of diagnosis to the time of the study. Study participants included in Group B (unemployed, but had the potential for vocational rehabilitation or employed, but needed rehabilitation services to continue employment) were more likely to remain in clerical and professional type jobs, compared to those in skilled and unskilled labor. Furthermore, those in Group C (currently working without need of rehabilitation intervention) were able to remain in the workplace, although they shifted employment from labor-intensive jobs to clerical or professional. Authors note that of the 131 clients with working potential, only 18 percent indicated that they stopped work because of MS, but provided no additional information about why study participants left work. Again, the limitations of this study are its small sample size and the lack of consideration of additional factors that influence ability to work. Grønning et al. (1990)[181] reported consistent findings that heavy work was predictive of early unemployment; however, this study had serious limitations with regard to how jobs were categorized into heavy versus light work (see Evidence Table 4 in Appendix F).

Two studies did not focus primarily on job characteristics in the analyses, but did provide commentary about it. LaRocca et al. (1982)[187] reported that a significant portion of the variance in employment status was unexplained by typical demographic characteristics such as age, education, and occupation, but was explained by more subjective measures of workplace characteristics, social support, and coping style. Unfortunately, additional information about workplace characteristics was not provided. Dyck et al. 2000[176] commented that 17 percent of the women in their study reported that they quit work because they were unable to negotiate reduced work hours with their managers.

From what we know about the possible physical and cognitive limitations associated with MS, as well as the resulting fatigue and other symptoms, it is not surprising that unemployment is more common among individuals whose jobs required physical exertion. The strength of the case-control study by Verdier-Talliefer et al. (1995)[193] is that it considered numerous working conditions that increased or decreased the odds of unemployment. The remaining studies provided descriptive information that was parallel with findings reported by Verdier-Talliefer et al. (1995).[193] Unfortunately, none of these studies systematically examined whether employers' willingness or ability to provide workplace accommodations or flexible work schedules fostered continued employment.

Self-report. Several studies provided descriptive information about conditions or situations that influenced individuals with MS to cease employment (Table 18). Physical difficulty, ambulation problems, visual difficulties, emotional problems, and fatigue were reasons for ceasing employment among participants in two studies[177,192] who indicated that they left work because of MS. In addition, both studies reported that a significant percentage of women (37.4 percent and 26 percent, respectively) indicated leaving for reasons other than MS, including marriage and/or pregnancy. Among individuals who remained at work, fatigue was reported as the most common symptom impeding work performance or restricting the work that could be done in two cross-sectional studies.[176,182] An additional study not included in Table 18[178] reported that 10 percent (n = 30) of study participants indicated that they quit work because of fatigue. Although these findings do not involve specific clinical tests to determine the presence or absence of cognitive or physical impairment, qualitative data like these are useful for shaping quantitative data analyses, as well as shaping future research.

Environmental Factors and Work Ability

Introduction

This section describes the evidence pertaining to Question 5: *Among individuals with MS, how does elevated temperature or other environmental factors impair the capacity to work?* The precise scope of this question was defined in collaboration with the project's technical advisory panel and representatives of SSA. This process led to consensus that temperature was the sole environmental factor that warranted investigation. We therefore focused specifically on evidence regarding the associations between thermal (ambient or climatic temperature) conditions in the work environment and the work capacity, work status, or disability status of patients with MS. It was recognized that occupational physical activity might be a modifier of the effect of environmental temperature.

The importance assigned to temperature effect in MS is based on longstanding clinical impressions that excessive or high heat conditions may be associated with transient worsening of symptoms and/or function in some patients with MS.[197,198] We did not attempt to assess the quality of the evidence underlying such clinical impressions, nor did we investigate possible mechanisms whereby ambient and/or body temperature might act physiologically to affect MS.[199-201] We also did not examine treatment modalities based on thermal sensitivity in relation to Question 5; rather, randomized controlled trials of cooling garments and other temperature-lowering interventions were considered for inclusion under Question 3b (symptomatic treatments).

A key issue for SSA disability determination in MS would seem to be assessment of reported functional or activity limitations due to worsening symptoms or decreases in functional capacity that are temporally associated with particular activity demands or the physical environments attendant to certain jobs. If such worsening were consistent and had significant adverse impact, the associated demand or situation might be considered a "critical job demand" for some jobs (i.e., the limiting point of a potentially significant mismatch between a person's work capacity and the job requirements). Another research question of interest to SSA might be whether environmental temperature conditions characteristic of particular workplaces or jobs are associated with different rates of MS work disability.

Results

A single research study met the inclusion criteria for this question (see Evidence Table 5 in Appendix F).[182] Other candidate articles were typically excluded due to a lack of data regarding work capacity, work status, or disability status variables. In considering some of the excluded interventional studies aimed at short-term reduction of individual body temperature,[199,201-203] it was clear that an unacceptable degree of extrapolation would be required to relate the conclusions of these studies to work demands or circumstances of likely relevance to SSA.

The single included article,[182] a questionnaire survey, also has significant limitations in its applicability to SSA-relevant issues. It does, however, suggest that some MS patients perceive temperature as a factor that can either impede (high temperature) or enhance (cool temperature) work performance. This cross-sectional study divided 508 respondents with MS into four self-reported work status categories. Table 19 summarizes subjects' responses to open-ended questions regarding impediments to and enhancers of work/chore performance. Independent raters coded the actual responses into several categories. Table 19 includes only temperature-related conditions/situations and those endorsed by all four work groups. Overall, 53 work-impeding categories were identified, with 22 of those being endorsed by at least five percent of respondents. For work-enhancement, 27 total categories were identified, with 17 being endorsed by at least five percent of respondents.