The Quality Standards Subcommittee of the American Academy of Neurology (AAN) is charged with developing practice parameters for neurologists for diagnostic procedures, treatment modalities, and clinical disorders. The selection of topics for which practice parameters are used is based on prevalence, frequency of use, economic impact, membership involvement, controversy, urgency, external constraints, and resources required. This article addresses the usefulness of evoked potentials (EPs) in identifying clinically silent lesions in patients with suspected MS.

The diagnosis of MS remains primarily clinical, requiring evidence of white matter lesions disseminated in space and time. Some patients with suspected MS not fulfilling clinical dissemination criteria (MS suspects) have abnormal EPs that identify clinically unsuspected lesions. Current diagnostic criteria allow MS suspects to be reclassified into definite MS categories if EPs identify clinically silent lesions. The identification of clinically unsuspected lesions is one major reason clinicians use EPs in MS suspects.

Presumably, MS suspects with EP-identified clinically silent lesions are more likely to have MS and are at higher risk for developing MS-related disabilities. Accurately identifying these high-risk patients will become increasingly important if early therapies are demonstrated to be effective in preventing or delaying disability in patients with MS.

To determine the effectiveness of EP-identified silent lesions in diagnosing MS, we performed a systematic review and analysis of the literature. Based on this review we propose practice parameters for the use of EPs in patients with suspected MS.

**Process.** Confirmation of the ability of EP-identified clinically silent lesions to diagnose MS requires a comparison with an independently verifiable gold standard. Potential gold standards include pathologic confirmation of MS lesions, brain MRI, or the eventual development of clinically definite MS (CDMS). Pathologic confirmation in MS suspects is not usually practical given the invasiveness of brain biopsy. The lack of specificity of MRI limits its usefulness as a gold standard. We concluded that the eventual development of CDMS was the best available independent gold standard to evaluate the usefulness of EP-identified silent lesions in MS suspects.

To find articles describing the relationship between EP findings and the development of CDMS, we searched the National Library of Medicine’s Medline database for articles published from 1966 to January 1998, using the terms multiple sclerosis and evoked potentials. We subsequently screened the search result articles and their bibliographies for articles discussing EPs in the diagnosis of MS. We selected those studies that followed MS suspects evaluated with EPs for the development of CDMS. We eliminated articles that did not employ standard EP techniques.

From the articles fulfilling these criteria we abstracted the following methodologic characteristics: the method and setting of cohort assembly, number of patients studied, duration of follow-up, clinical and demographic features of the patients, and the criteria used to define CDMS. Based on these methodologic characteristics, we graded the studies Class I to IV using the diagnostic test strength-of-evidence rating scheme appended to this report (see below). For each EP technique studied we calculated the following parameters: sensitivity—the proportion of MS suspects developing CDMS with abnormal EPs; specificity—the proportion of MS suspects not developing CDMS with normal EPs. We also calculated the positive predictive value (PPV)—the proportion of MS suspects with abnormal EPs developing CDMS; and the complement of the negative predictive value (NPVc)—the proportion of MS suspects with normal EPs developing CDMS. Additionally, we determined the relative rate that patients with abnormal EPs developed CDMS as compared to patients with normal EPs by dividing the PPV by the NPVc. As an overall measure of the strength of the association between EP results and eventual CDMS, we calculated Goodman and Kruskal’s tau.
Table 1 Methodologic characteristics of studies following MS suspects for the development of CDMS

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Patients (n)</th>
<th>Months of FU (mean, range)</th>
<th>Patients developing CDMS, %</th>
<th>Mean age, y</th>
<th>Gender, % female</th>
<th>Spectrum of disease, n</th>
<th>CDMS criteria</th>
<th>Cohort assembly</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hume and Waxman (1988)</td>
<td>222</td>
<td>30 (12-56)</td>
<td>22.4</td>
<td>33</td>
<td>66</td>
<td>Isolated ON, 26; single lesion, 83; Poss MS, 110; Prob MS, 29</td>
<td>McDonald and Halliday</td>
<td>Consecutive referrals of MS suspects for evoked potentials</td>
<td>II</td>
</tr>
<tr>
<td>Hamburger et al. (1984)</td>
<td>96</td>
<td>36</td>
<td>51</td>
<td>Not described</td>
<td>Leg symptoms, 50; blurred vision, 45</td>
<td>Poser et al.</td>
<td>Patients referred with symptoms suggestive of MS</td>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>Deltenre et al. (1982)</td>
<td>133</td>
<td>17.6 (2-51.6)</td>
<td>33.1</td>
<td>Not described</td>
<td>Not described</td>
<td>Poss MS; Prob MS</td>
<td>McDonald and Halliday</td>
<td>Not described</td>
<td>IV</td>
</tr>
<tr>
<td>Botcher and Trojaborg (1982)</td>
<td>21</td>
<td>(19-49)</td>
<td>42.9</td>
<td>Not described</td>
<td>Not described</td>
<td>Poss MS; Prob MS</td>
<td>McDonald and Halliday</td>
<td>Not described</td>
<td>IV</td>
</tr>
<tr>
<td>Matthews et al. (1982)</td>
<td>69</td>
<td>30 (6-38)</td>
<td>40.6</td>
<td>38</td>
<td>64</td>
<td>Single lesion; poss MS; prob MS; ON excluded</td>
<td>McAlpine</td>
<td>Consecutive referrals of MS suspects for evoked potentials</td>
<td>II</td>
</tr>
</tbody>
</table>

FU = follow-up; CDMS = clinically definite MS; ON = optic neuritis; Poss = possible; Prob = probable; LSDMS = laboratory-supported definite MS.

context, tau measures the increased accuracy achieved in predicting the development of CDMS by knowing the results of EPs.)

Analysis of evidence. Study characteristics. A total of 716 articles met the Medline search criteria. Of these, 108 addressed EPs in the diagnosis of MS. Eleven articles, from nine studies, described MS suspects evaluated with EPs followed for the development of CDMS. Table 1 summarizes the methodologic characteristics of these studies. In the following discussion we refer to articles by their first author only.

The cohort assembly methods of the studies varied. Hume and Matthews both selected consecutive MS suspects from patients referred to EP laboratories. These studies introduced a potential referral-filter bias because their cohorts were assembled from subjects referred to EP laboratories. We could not determine the method Deltenre or Botcher employed to assemble patients. The study by Frederiksen appeared to be population based. His group collected all newly diagnosed patients with monosymptomatic optic neuritis (ON) from a defined geographic region.

The number of MS suspects enrolled ranged from 21 to 222. Mean ages of the MS suspects ranged from 27.5 years to 43 years. Hume excluded patients less than 15 and greater than 60 years of age. All studies describing gender reported a female preponderance. Martinelli, Deltenre, and Hamburger provided incomplete demographic information.

The spectrum of disease studied varied considerably. Frederiksen and Martinelli selected monosymptomatic ON patients only. Matthews excluded MS suspects with only ON. The remainder of the studies followed MS suspects with a broad spectrum of disease including isolated ON, “acute not diagnosed” patients with single CNS lesions common in MS patients, possible MS, and probable MS. Lee included patients fulfilling Poser’s criteria for laboratory-supported definite MS. The spectrum of MS suspects studied likely varied in ways that we could not determine.

All of the studies appeared to use standard EP techniques, although it was impossible to determine in many if they used institutionally established normal values or a 95% or 99% cutoff for abnormal. Although most authors performed both tibial and median somatosensory evoked potentials (SEPs), only Filippini described the median and tibial SEP.
The criteria used to diagnose CDMS appeared similar in most studies. Brainstem auditory evoked potentials. The mean duration of follow-up varied from 12 to 36 months. The percentage of MS suspects that developed CDMS during the follow-up period ranged from 17% to 51%. There was a trend for the studies with longer follow-up periods to have a higher frequency of CDMS. The follow-up time in these studies may have been too short to allow MS to develop in many patients. For instance, the risk for developing CDMS after an episode of isolated ON remains high for many years. Thus, it is possible that the sensitivity of EPs would increase with longer follow-up periods as more patients with abnormal EPs develop CDMS. It is also conceivable, however, that the specificity of EPs would decrease with longer follow-up times as MS suspects without abnormal EPs develop CDMS.

The criteria used to diagnose CDMS appeared similar in most studies. Although Filippini, Deltenre, and Matthews used diagnostic criteria by McDonald and Halliday, respectively, they applied only the clinical diagnostic parameters. We graded the studies by Frederiksen, Martinelli, and Hamburger Class IV because we could not determine if they used an acceptable independent gold standard for the diagnosis of MS. Frederiksen and Martinelli did not state the criteria used to diagnose CDMS. Hamburger used Poser’s criteria but did not indicate if the results of the EPs were used in the determination of definite MS. Thus he may have introduced an incorporation bias.

We were unable to calculate the strength of the EP-CDMS association and other parameters describing diagnostic accuracy in the articles by Martinelli, Deltenre, and Bottcher. Therefore, we also graded these studies Class IV. Because of serious methodologic limitations, we did not consider Class IV studies further.

We graded the studies by Filippini, Lee, Hume, and Matthews Class II. These four studies were prospective, used an acceptable independent gold standard for the diagnosis of MS, and measured EP diagnostic accuracy relative to the development of CDMS. Table 2 summarizes the abstracted statistical parameters from these Class II studies.

**Visual evoked potentials.** The studies of Lee, Hume, and Matthews demonstrated a statistically significant association between abnormal visual evoked potentials (VEPs) and an increased risk of developing CDMS. In these studies, patients with suspected MS were 2.5 to 9 times as likely to develop CDMS as patients with normal VEPs. VEP sensitivities ranged from 25% to 83%. VEPs improved the ability to predict which MS suspects will develop CDMS by as much as 29% (tau) in the studies by Hume, and as little as 8% in the study by Lee. Filippini’s study did not show that patients with abnormal VEPs were at higher risk for the development of CDMS. Of the four Class II studies, Filippini’s included the highest proportion of patients with isolated ON. Because VEPs may simply confirm the presence of a clinically suspected lesion in ON patients without identifying unsuspected lesions, it is not surprising that the CDMS predictive value of abnormal VEPs in this population would be less.

From the evidence reviewed, we conclude that an association between abnormal VEPs and an increased risk of CDMS has been established with moderate clinical certainty.

**Somatosensory evoked potentials.** There was a statistically significant association between abnormal SEPs and an increased risk of CDMS in the studies by Hume and Matthews. In these studies, patients with suspected MS with abnormal SEPs were 2.4 to 3.9 times as likely to develop CDMS as patients with normal SEPs. SEPs provided improvements in predicting CDMS ranging from 4.6% to 12.7% (tau). Reported sensitivities varied from 36% to 63%.

In contrast to those of Hume and Matthews, Lee’s study demonstrated that patients having suspected MS with abnormal SEPs were half as likely to develop CDMS than patients with normal SEPs. Additionally, Filippini’s study failed to reveal a statistically significant association between SEPs and the development of CDMS.

We conclude that the evidence describing the relationship between abnormal SEPs and the development of CDMS is inconclusive and conflicting.

**Brainstem auditory evoked potentials.** Hume observed a statistically significant, albeit weak, association between brainstem auditory evoked potentials (BAEPs) and CDMS. In Hume’s study, BAEPs increased CDMS predictive ability by only 5% (tau). The sensitivity of abnormal BAEPs was 14.6%. Filippini and Matthews studies failed to reveal a statistically significant relationship between BAEP results and the development of CDMS.

We conclude that the absence of a useful association between abnormal BAEPs and an increased risk of CDMS has been established with moderate clinical certainty.

**Multimodal evoked potentials.** Hume and Matthews provided sufficient data to allow comparison of the improvement in CDMS prediction accuracy achieved by multimodal EPs (combined VEPs, SEPs, and BAEPs) to that achieved by VEPs, SEPs, or BAEPs alone. The accuracy increase for multimodal EPs in both studies (tau 22% and 18%, respectively) was less than the accuracy increase for VEPs alone (tau 29% and 20%, respectively). The reason for this is apparent from table 2. A slight gain in sensitivity from using multimodal EPs was offset by a greater loss in specificity.


Table 2  Measures of association between EPs and the development of CDMS

<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Technique</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>PPV %</th>
<th>NPVc, %</th>
<th>Relative risk CDMS</th>
<th>95% CI relative risk</th>
<th>Tau, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filippini et al. (1994)</td>
<td>VEP</td>
<td>25</td>
<td>63</td>
<td>25.9</td>
<td>38.2</td>
<td>0.68</td>
<td>0.33-1.40</td>
<td>1.5*</td>
</tr>
<tr>
<td></td>
<td>SEP (tibial)</td>
<td>17.9</td>
<td>87</td>
<td>41.7</td>
<td>32.9</td>
<td>1.27</td>
<td>0.60-2.68</td>
<td>0.4*</td>
</tr>
<tr>
<td></td>
<td>SEP (median)</td>
<td>14.3</td>
<td>92.6</td>
<td>50</td>
<td>32.4</td>
<td>1.54</td>
<td>0.72-3.32</td>
<td>0.1*</td>
</tr>
<tr>
<td></td>
<td>BAEP</td>
<td>17.9</td>
<td>90.7</td>
<td>50</td>
<td>31.9</td>
<td>1.57</td>
<td>0.77-3.17</td>
<td>1.2*</td>
</tr>
<tr>
<td>Lee et al. (1991, 1988)</td>
<td>VEP</td>
<td>60.7</td>
<td>80.5</td>
<td>43.7</td>
<td>17.5</td>
<td>2.49</td>
<td>1.52-4.08</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>SEP</td>
<td>36.4</td>
<td>40.3</td>
<td>20.6</td>
<td>40.2</td>
<td>0.51</td>
<td>0.32-0.82</td>
<td>4.6</td>
</tr>
<tr>
<td>Hume and Waxman (1988)</td>
<td>VEP</td>
<td>83 (75.8)</td>
<td>78.5 (85.9)</td>
<td>52.7 (54.3)</td>
<td>5.9 (5.9)</td>
<td>8.96 (9.24)</td>
<td>4.42-18.2 (28.8)</td>
<td>9.9*</td>
</tr>
<tr>
<td></td>
<td>SEP</td>
<td>63 (41.4)</td>
<td>76.6 (89.7)</td>
<td>42 (44.4)</td>
<td>11.5 (11.5)</td>
<td>3.87 (3.87)</td>
<td>2.09-7.15 (12.1)</td>
<td>10.2*</td>
</tr>
<tr>
<td></td>
<td>BAEP</td>
<td>14.6 (14.6)</td>
<td>97.6 (97.6)</td>
<td>60 (60)</td>
<td>17.9 (17.9)</td>
<td>3.52 (3.34)</td>
<td>2.23-5.55 (5.1)</td>
<td>5.1*</td>
</tr>
<tr>
<td>Matthews et al. (1982)</td>
<td>Multimodal EPs</td>
<td>89.6 (86.1)</td>
<td>66.7 (77.3)</td>
<td>42.6 (47.7)</td>
<td>4.1 (4.1)</td>
<td>10.3 (11.5)</td>
<td>4.2-25.0 (21.6)</td>
<td>27.6*</td>
</tr>
<tr>
<td></td>
<td>VEP</td>
<td>53.6 (50.0)</td>
<td>87.8 (92.3)</td>
<td>75 (81.3)</td>
<td>26.5 (26.5)</td>
<td>2.83 (3.06)</td>
<td>1.66-4.80 (20)</td>
<td>23.1*</td>
</tr>
<tr>
<td></td>
<td>SEP</td>
<td>35.7 (14.3)</td>
<td>92.7 (100)</td>
<td>76.9 (100)</td>
<td>32.1 (30.5)</td>
<td>2.39 (3.28)</td>
<td>1.48-3.88 (12.9)</td>
<td>9.7*</td>
</tr>
<tr>
<td></td>
<td>BAEP</td>
<td>21.4 (21.4)</td>
<td>90.2 (92.5)</td>
<td>60 (66.7)</td>
<td>37.3 (37.3)</td>
<td>1.61 (1.79)</td>
<td>0.80-2.95 (2.7)</td>
<td>4.1*</td>
</tr>
<tr>
<td></td>
<td>Multimodal EPs</td>
<td>60.7 (56.0)</td>
<td>80.5 (87.2)</td>
<td>68 (73.7)</td>
<td>25.2 (24.4)</td>
<td>2.72 (3.01)</td>
<td>1.53-4.85 (17.7)</td>
<td>21.2*</td>
</tr>
</tbody>
</table>

Values in parentheses represent parameters from clinically silent lesions only.
Not significant at p < 0.05.
PPV = positive predictive value; NPVc = complement of the negative predictive value; relative risk of CDMS = PPV/NPVc; VEP = visual evoked potential; SEP = somatosensory evoked potential; BAEP = brainstem auditory evoked potential; EP = evoked potential.

**Limitations of available evidence.** None of the studies reviewed presented their data in sufficient detail to allow an analysis of the EP-CDMS association within MS suspect subgroups. The MS suspect populations were heterogeneous and included patients with single clinical CNS lesions, including isolated ON, as well as those meeting clinical criteria for possible MS and probable MS. Within these heterogeneous populations there may have been subgroups of patients within which specific EPs were more or less predictive of CDMS. Thus, for example, in patients presenting with isolated ON, the association between SEPs and CDMS might be stronger than the association between VEPs and CDMS. There were insufficient data to perform a subgroup analysis to make this determination.

Because EPs are commonly used in MS suspects as part of a cluster of tests (including MRI and CSF examination), it would be useful to determine the independent contribution of EPs to the diagnosis of MS. Because of the lack of detail in the data reported, we could not determine the optimal sequence of tests, nor the combination of tests that would best predict the development of CDMS. From these studies we cannot determine whether obtaining EPs and MRI provides more predictive information than either test alone. We cannot determine, for example, if MS suspects with abnormal MRIs and EP-identified silent lesions are more at risk for CDMS than MS suspects with normal MRIs and no EP-identified silent lesions. A determination of the added value of EPs in this and similar situations requires longitudinal studies, structured to allow multivariate analyses.

The impact of EPs on patient outcome remains unknown. No study demonstrated improved outcomes in MS suspects who receive EPs compared to MS suspects who do not receive EPs.

**Recommendations.** For clinicians considering evoked potentials in patients with suspected MS for the purpose of finding clinically silent lesions:

1. VEPs are recommended as probably useful to identify patients at increased risk for developing CDMS. (Guideline, Class II)
2. SEPs are recommended as possibly useful to identify patients at increased risk for developing CDMS. (Option, Class II)
3. Evidence is insufficient at this time to recommend BAEPs as a useful test to identify patients at increased risk for developing CDMS. (Guideline, Class II)

**Recommendations for future research.** The studies currently available have potential biases and do not provide sufficient information to determine the independent contribution of diagnostic modalities commonly employed in MS suspects. Future studies could include:

1. Longitudinal design with the development of CDMS, independent of paraclinical tests, as the independent gold standard of MS.

---

Lee et al. (1991, 1988) 15,16
Filippini et al. (1991, 1988) 14
Matthews et al. (1982) 17
Waxman (1988) 18
Hume and Waxman (1988) 19
Matthews et al. (1982) 20
2. Inclusion of patients with possible and probable MS by clinical criteria, as well as patients with isolated ON.
3. Follow-up periods of at least 5 years.
4. Multivariate analyses to assist in determining the optimal combination and sequence of tests that best predict the development of CDMS.
5. MS suspect subgroup analyses to determine which EPs are useful in which patients.

**Disclaimer.** This review focused on the usefulness of EPs relative to the identification of clinically unsuspected lesions in MS suspects. There are other reasons that EPs are used in evaluating the MS suspect. Evoked potentials may aid in the localization of lesions, confirm clinically ambiguous lesions, confirm the organic basis of symptoms, and suggest demyelination as the pathophysiology of a lesion. The determination of the usefulness of EPs relative to these other indications is beyond the scope of this review.

This statement is provided as an educational service of the American Academy of Neurology. It is based on an assessment of current scientific and clinical information. It is not intended to include all possible proper methods of care for a particular neurologic problem or all legitimate criteria for choosing to use a specific procedure. Neither is it intended to exclude any reasonable alternative methodologies. The AAN recognizes that specific patient care decisions are the prerogative of the patient and the physician caring for the patient, based on all of the circumstances involved.

**Acknowledgment**
The AAN QSS thanks Gary Gronseth, MD, and Eric Ashman, MD, for their service to the Academy’s membership as the authors of this practice parameter. They also thank the numerous individuals, AAN Sections, and organizations that reviewed drafts of this practice parameter, including the Neuro-Oncology Section, the Neuroimaging Section, the Sleep Section, the Spine Section, the American Clinical Neurophysiology Society, the American Association of Electrodiagnostic Medicine, and the National Multiple Sclerosis Society. In particular, they thank Douglas Goodin, MD, Barbara Vickrey, MD, and Robert Holloway, MD, for their review of the document.

**Appendix 1**
American Academy of Neurology Quality Standards Subcommittee Members: Gary Franklin, MD—Co-Chair; Catherine Zahn, MD—Co-Chair; Milton Alter, MD, PhD; Stephen Ashwal, MD; John Calverley, MD; Richard Dubinsky, MD; Jacqueline French, MD; Michael Greenberg, MD; Gary Gronseth, MD (facilitator); Deborah Hirtz, MD; Robert Miller, MD; and James Stevens, MD.

**Appendix 2**
Definitions for strength of evidence
Class I. Evidence provided by a prospective study of a broad spectrum of persons with the suspected condition. The study measures the diagnostic accuracy of the test using an acceptable independent gold standard for case definition. The test is applied in an evaluation that is masked to the persons’ clinical presentations and the gold standard is applied in an evaluation that is masked to the test result.

Class II. Evidence provided by a prospective study of a narrow spectrum of persons with the suspected condition, or by a retrospective study of a broad spectrum of persons with the condition compared to a broad spectrum of controls. The study measures the diagnostic accuracy of the test using an acceptable independent gold standard for case definition. The test is applied in an evaluation that is masked to the gold standard.

Class III. Evidence provided by a retrospective study; either the persons with the condition or the controls are of a narrow spectrum. The study measures the diagnostic accuracy of the test using an acceptable independent gold standard for case definition. The test is applied in an evaluation that is masked to the gold standard.

Class IV. Evidence provided by expert opinion or case series without controls. Any study not measuring the diagnostic accuracy of the test using an acceptable independent gold standard for case definition. Any study where the test is applied in an evaluation that is not masked to the gold standard.

Definitions for strength of recommendations
Standards. A principle for patient management that reflects a high degree of clinical certainty (usually this requires Class I evidence that directly addresses the clinical question or overwhelming Class II evidence).

Guidelines. A recommendation for patient management that reflects moderate clinical certainty (usually this requires Class II evidence or a strong consensus of Class III evidence).

Practice option. A strategy for patient management for which the clinical utility is uncertain (inconclusive or conflicting evidence or opinion).

Practice advisory. A practice recommendation for emerging or newly approved therapies or technologies based on evidence from at least one Class I study. The evidence may demonstrate only a modest statistical effect or limited (partial) clinical response, or significant cost-benefit questions may exist. Substantial (or potential) disagreement among practitioners or between payers and practitioners may exist.
References


