Invited Article:

Practice parameters and technology assessments

What they are, what they are not, and why you should care

Gary Gronseth, MD,
FAAN

Jacqueline French, MD,
FAAN

Address correspondence and reprint requests to the American Academy of Neurology, 1080 Montreal Avenue, St. Paul, MN 55116
guidelines@aan.com

A neurologist is evaluating a patient in the office after the patient reported a 5-minute episode of blindness of the right eye and clumsiness of the left arm. She orders a carotid ultrasound that reveals some carotid plaques, with a <50% stenosis of the right and left carotid artery. She consults a vascular surgeon, who recommends immediate right carotid endarterectomy. Before she proceeds, she checks the American Academy of Neurology (AAN) Web site to see if there are any relevant practice parameters on carotid endarterectomy. She discovers that endarterectomy is not recommended for her patient.¹ Armed with this definitive information, she discusses treatment options with the patient and the surgeon and prescribes antiplatelet therapy, as recommended in the guideline.

In light of her recent positive experience with guidelines, she is quick to check the AAN Web site when seeing her next patient, a child who has a new diagnosis of absence seizures. Hoping to receive guidance in selecting the proper first therapy, she finds a guideline on treatment of the patient with newly diagnosed seizures.² The only recommended therapy is lamotrigine (Level B), and she selects this treatment for her patient. The child continues to have seizures, so she asks an epileptologist for an opinion. The epileptologist questions why she did not use ethosuximide, which is considered first-line therapy for childhood absence epilepsy.³ She switches the child to ethosuximide, wondering where she and the guideline went wrong. These two scenarios illustrate the proper use and the limitations of guidelines. Guidelines are often misunderstood relative to their implications and purpose. As guidelines become an ever-increasing presence on the healthcare landscape, for neurologists as well as in other areas of medicine, it is important to explain what AAN guidelines are and what they are not.

WHAT THEY ARE An AAN practice parameter is a comprehensive synthesis of all available evidence pertinent to a specific question or set of specific questions. The Quality Standards Subcommittee (QSS) and the Technology and Therapeutics Assessment Subcommittee (TTA) have published over 100 guidelines, otherwise referred to as practice parameters and technology assessments, since the mid-1990s. These committees consist of AAN members who are appointed by the President of the AAN for 2-year renewable terms. The term “guideline” may be used for the products of these committees, but this term implies introduction of non-evidence-based consensus opinion, which does not enter into the development of the AAN’s product. However, the term “guideline” is so ingrained in the mind of the treating physician when considering or discussing this product that it is essentially used interchangeably with “practice parameter.” Guidelines are very popular products of the AAN, being rated on Academy-sponsored surveys as the second most important benefit of AAN membership, after the journal *Neurology*[®]. The published versions of the parameters are also many of the top cited articles in *Neurology*[®]. These products are similar to those created by other specialty societies, although other societies may allow more or less consensus to be used to reach conclusions. They are also similar to Cochrane reviews, a product of the Cochrane collaboration, although the Cochrane reviews rely heavily on meta-analysis, which is usually not a part of the AAN process.

AAN practice parameters and technology assessments are systematic. They follow a predefined process that is described in detail in the AAN Clinical Practice Guideline Process Manual⁴ available at www.aan.com/go/guidelines/development. The process follows a predictable sequence beginning with formulating an answerable clinical question and performing a comprehensive literature search to ensure (to the extent pos-

See page 1634

From the University of Kansas (G.G.), Kansas City; and New York University Comprehensive Epilepsy Center (J.F.), New York.

Disclosure: The authors report the following conflicts of interest: Dr. Gronseth has received speaker honoraria from Pfizer, GlaxoSmithKline, and Boehringer Ingelheim and served on the IDMC Committee of Ortho-McNeil. Dr. Gronseth estimates that <2% of his time is spent on EMG and EEG. Dr. French holds financial interests in Jazz, Eisai, Valeant, Marinus, Pfizer, and UCB. She has received research funding from the Epilepsy Therapy Development Project, FACES, UCB, Eisai, Johnson and Johnson, and Merck. Dr. French estimates that 30% of her time is spent in outpatient epilepsy practice.

sible) that all of the pertinent clinical evidence regarding the question is discovered.

AAN practice parameters and technology assessments are transparent. Transparency is assured by documenting sufficient detail regarding the process so that readers can judge for themselves the validity of the process. For example, the search terms, search strategies, and databases utilized to find all of the evidence are specifically identified so that the search could be reproduced. A broad net is cast to minimize the chance that important articles might be missed. Additionally, the reasons for classifying a specific study in a certain way are detailed in the text or in the evidence tables. Finally, the assumptions that go into formulating conclusions and recommendations are stated. Thus, a reader of an AAN clinical practice parameter should have a clear idea why the authors came to the conclusions and recommendations that they did.

Where the evidence comes from. Most physicians are familiar with the concept of evidence-based medicine (EBM). EBM describes the concept of using evidence from studies published in the medical literature to inform clinical practice. In the context of practice guidelines and EBM, evidence is narrowly defined. Evidence in this case primarily refers to studies of outcomes in a specific patient population with specified diseases who are undergoing specific interventions. Only published clinical evidence is considered. Evidence from basic science and laboratory studies is not normally cited when producing a practice parameter. Additionally, unlike some practice guidelines of other organizations, expert opinion is excluded from AAN parameters and technology assessments.

How the evidence is analyzed. Once the evidence is found, it is analyzed to determine whether the study has been designed and executed appropriately to answer the research question. The studies are given a classification score (I–IV) based upon the rigor of the study. The studies given a Class I designation have a relatively low risk of bias, whereas studies given a Class IV have a relatively high risk. The various characteristics of a study that increase the risk of bias have all been empirically validated.⁵ Different classification schemes relate to different types of clinical questions. The most well-developed and most commonly used scheme is that pertaining to therapeutic questions (see table in this issue of *Neurology*[®], page 1635). In the scenarios given above, the clinical questions important to the physician were of a therapeutic nature. Following a formal classification of evidence, the results of studies are synthesized and conclusions and recommendations are formed. Recommendations are also labeled from strong to weak

Table AAN classification of recommendations

A	= Established as effective, ineffective, or harmful (or established as useful/predictive or not useful/predictive) for the given condition in the specified population. (Level A rating requires at least two consistent Class I studies).*
B	= Probably effective, ineffective, or harmful (or probably useful/predictive or not useful/predictive) for the given condition in the specified population. (Level B rating requires at least one Class I study or two consistent Class II studies.)
C	= Possibly effective, ineffective, or harmful (or possibly useful/predictive or not useful/predictive) for the given condition in the specified population. (Level C rating requires at least one Class II study or two consistent Class III studies.)
U	= Data inadequate or conflicting; given current knowledge, treatment (test, predictor) is unproven.

*In exceptional cases, one convincing Class I study may suffice for an “A” recommendation if 1) all criteria are met, 2) the magnitude of effect is large (relative rate improved outcome > 5 and the lower limit of the confidence interval is > 2).

(Levels A–C) based on the quality of evidence available to support them (table). A separate designation is available (Level U) when evidence is insufficient to make a recommendation.

Selecting authors. AAN practice parameters and technology assessments are designed to minimize bias. Indeed, the infrastructure of the guideline development process has been specifically designed to minimize the risk that bias will affect a parameter’s conclusions or recommendations. The committees in charge of developing practice parameters and technology assessments carefully select content experts to form author panels. Scrutiny is given to potential biases of panel members. Every panel member must provide a detailed conflict of interest statement, which includes not only payments from third parties such as pharmaceutical companies, but also likelihood of bias related to income from practice activities. Since individuals with this kind of bias may also have expertise in the area, they are not excluded from participating in the process, but they may not be the chair of the writing committee, and their input will be balanced by individuals with no bias. In a recent article investigating author panel conflicts, 46 out of 50 guidelines evaluated had at least one author who had reported a conflict, the majority of which were research related.⁶

Oftentimes, the composition of an author panel is modified to ensure balance. The conclusions and recommendations are explicitly labeled to indicate the strength of evidence upon which they are based. Authors are instructed not to make conclusions or recommendations that go beyond the evidence, since even the most well-chosen panel may introduce bias through consensus and personal opinion.

Oversight. The guideline development process, including formation of the author panel, the literature

search, the classification and synthesis of evidence, and review of all guideline drafts, is supervised by several levels of the AAN including the QSS and TTA as well as the Practice Committee. This oversight as well as the strict evidence-based methodology serves to minimize the introduction of bias into the process. The resultant product represents a distillation of what is known, and the degree of certainty with which it is known, in its purest form.

The product. In the example above of the physician faced with a treatment dilemma on carotid endarterectomy, the guidelines were quickly able to summarize the most rigorous (Class I) studies, leading to a strong (Level A) recommendation pertaining to her question. If she had attempted to search the literature on her own, it would have been very difficult for her to have quickly reached the same conclusions. The amount of literature on the subject is vast and variable, including methodologically sound as well as flawed studies. Without methodologic expertise, she would have been hard pressed to differentiate between the two. In addition, the availability of a guideline hopefully raises consciousness of the clinical issue and increases awareness of strong studies in a given area.

WHAT THEY ARE NOT Practice parameters and technology assessments are not definitive. Despite their methodologic rigor, and as exemplified by the example of the child with absence seizures previously mentioned, practice parameters are sometimes incapable of answering even the simplest of clinical questions. This can happen for a number of reasons, one of the most common being that the evidence surrounding a question is incomplete, or that the studies that were performed lack generalizability. Some of the issues that physicians need to consider when seeking guidance from AAN practice parameters and technology assessments are delineated below.

Strength of the evidence. Evidence is graded as being relatively weak or strong (as indicated by its class designation). Even the strongest evidence (Class I), however, is not definitive. It is conceivable that a subsequent large, well-designed study might demonstrate a benefit of carotid endarterectomy in patients with <50% stenosis. The Level A recommendation merely indicates that this would be unlikely—but it is possible.

Generalizability. The available evidence often has limited generalizability. Patients who are included in randomized control trials, doses of medications that are used, or the outcomes that are used to measure the effect of an intervention are not necessarily the most appropriate and the most reflective

of common clinical practice. Thus, even studies judged to have a low risk of bias may have minimal applicability to practice if the patients included are somewhat atypical, the interventions given are unique, or the outcome measures are not clinically relevant. The guideline process, of course, is limited by the availability of generalizable studies with a low risk of bias. Often, clinical trials are performed at highly specialized centers that attract the most severely afflicted patients. It does not necessarily follow that a patient with less advanced disease would respond in the same manner. Moreover, each patient may have specific characteristics (gender, age, concomitant conditions) that would make one treatment preferable over another. In any study, there are responders and nonresponders who may or may not differ in important ways. These nuances cannot be studied in larger randomized trials that lead to strong recommendations because one cannot randomize enough patients in each subgroup. Oftentimes, studies of these nuances are based on observational or small studies, resulting in conclusions and recommendations that are based on low level, weak evidence which, although possibly correct, cannot be definitive. A guideline may not adequately address a specific patient with specific characteristics. For instance, an experienced clinician faced with a patient with <50% stenosis, but in the setting of recurrent TIAs and an ulcerated plaque and aspirin failure, might opt for surgery despite a high level recommendation to the contrary.

Specificity of the question. A clinical trial can only answer the question it has been designed to answer. Clinicians want therapeutic studies to answer a very specific question: Is treatment option A better than treatment option B for the patient in front of them? No clinical trial can answer that question. Rather, clinical trials are designed to answer the question: On average, is treatment option A better than treatment option B for the patients included in the trial, in the manner treatments were used in the trial, and for the outcomes assessed in the trial? For example, if the endpoints in the trial addressed 3-month outcomes, but your patient is interested in long-term outcomes, the trial results may not be applicable. Similarly, a trial of a therapeutic agent that enrolled 300 patients is unlikely to include outcomes relevant to even the most common serious idiosyncratic events, nor would it address, for example, teratogenicity as an endpoint. If these issues are important in making treatment decisions, they must be considered alongside the recommendation that the treatment has been proven to be well tolerated and effective (the usual outcome measures). These examples again demon-

strate why a clinical trial can never select “best therapy” for a specific patient and why the results of clinical trials need to be cautiously applied to individual patients.

When there is limited evidence. Absence of evidence of effectiveness is not evidence of the absence of effectiveness. In the example of the physician selecting treatment for a patient with absence epilepsy, the fact that no definitive study has been performed for most antiepileptic drugs (including ethosuximide) does not mean that those therapies are ineffective. The only evidence that should be interpreted as a recommendation not to use a therapy would derive from a well-designed and sufficiently powered study that demonstrated the treatment to be ineffective or harmful. Many standard therapies (e.g., ethosuximide for absence seizures, pyridostigmine for myasthenia gravis) have never been studied in high-quality randomized trials. For these therapies, and many others, the evidence is insufficient to support even a weak recommendation (Level U in AAN terminology) yet most neurologists would attest to their effectiveness. Under such circumstances, some insurers (and even naïve physicians) have used this lack of evidence to state that the absence of evidence is evidence that the intervention of interest is not effective. This is clearly a distortion of the guidelines and is an example of misinterpretation of incomplete evidence. The neurologist selecting therapy for her patient with absence seizures fell into this trap. When many treatments are available, and most have not been studied, the guideline will be less helpful in selecting best therapy.

Guidelines can tell you whether to use a therapy, but they are not very good at determining how to use a therapy. In many cases the titration scheme, dose selection, and other parameters of use are determined by early clinical trials, which may be performed before optimal strategies are known. For example, the definitive studies of topiramate in newly diagnosed epilepsy demonstrated the efficacy of 400 mg/day. Other doses were not tested. The conclusion can be made that 400 mg/day is an efficacious dose, but not that it is the ideal dose. Studies of other doses would be required to decide on the optimal dose. In fact, most clinicians would find this dose too high and likely to be poorly tolerated.

Guidelines are just suggestions. AAN practice parameters and technology assessments are not meant to be prescriptive. As a consequence of the fact that guidelines are never completely definitive, the recommendations that flow from the sys-

tematic review of the literature are never absolute. All practice guidelines are accompanied by a disclaimer acknowledging this fact. Thus, it is acknowledged that recommendations that are given a strength recommendation of Level A, the highest possible level, may not be appropriate for certain situations. Unfortunately, some readers of guidelines choose to interpret the recommendations as being prescriptive. Based upon the limitations of generalizability as well as the limitations of the quality of the evidence itself, this is clearly inappropriate and is so stated in the guidelines. In the past, parameters and technology assessments have been used as evidence in medical-legal cases. Again, this is an overinterpretation of “recommendations” as “rules.” Such uses are in general inappropriate and certainly not intended by the AAN.

Although the guidelines are a comprehensive review of the evidence relative to specific questions, they are not comprehensive reviews of a specific topic and are not designed to deal with all aspects of the care and treatment of patients with a particular disease. Rather, their focus is to answer with evidence very specific and sometimes controversial clinical questions. For example, Bell’s palsy is a broad topic with many therapeutic and diagnostic questions, but the practice parameter focused on therapy only.⁷ Similarly, the practice parameter on myasthenia gravis dealt solely with the effectiveness of thymectomy,⁸ and a recent epilepsy diagnostic parameter focused only on the patient with a single unprovoked seizure.⁹ As noted above, there are many other venues and publications that do comprehensively review topics. Many of these are published by the AAN in the Continuum series: Lifelong Learning in *Neurology*®.

Practice parameters/technology assessments vs consensus statements. AAN practice parameters and technology assessments are not consensus-based opinion statements of the experts forming the author panel. Although content experts are often recruited to form the author panel, the conclusions and recommendations are driven by the evidence. The inclusion of experts on the panel with knowledge of the topic at hand is often useful in formulating the questions, validating the completeness of the literature, and interpreting the results of some studies. Fundamentally, however, their content expertise does not change the nature of the conclusions or recommendations. Expert-based recommendations are explicitly prohibited from AAN practice parameters and technology assessments to clearly separate these opinions from evidence. This is not at all to say that such expert opinion is without value. Indeed, our clinician searching for guidance in selecting therapy for her absence patient would be well ad-

vised to refer to the many expert-based consensus statements on the topic.

Clinical context. In June 2007, the AAN Board of Directors approved the use of “Clinical Context” in AAN clinical practice guidelines. The first clinical practice guideline to use this is the guideline written by Thurman et al. in February 2008.¹⁰ Thus, in AAN practice parameters and technology assessments published since then, many of the limitations described above are highlighted, where appropriate, in the “Clinical Context” sections. Readers should not expect these sections to include consensus opinions that reach beyond the evidence. However, they will be warned when the data set is incomplete, or when there are issues of generalizability.

WHY YOU SHOULD CARE Given the limitations of practice guidelines enumerated above, why are they useful? In the past, a physician could rely on his or her own personal knowledge and expertise when making a treatment decision. In more recent times, the vastness of medical information has made this much more difficult. And, even if a physician was able to review the entirety of literature in a given topic area, the variable quality of studies would be a major obstacle in reaching sound conclusions. Thus, there is a need for the guideline process, by which the literature is comprehensively compiled, and each study is critically evaluated to ensure that appropriate conclusions can be drawn. One of the most important and useful aspects of guideline development is the rating of research studies. Can this research study actually answer the question it set out to ask with a high degree of certainty? The more well-designed studies that reach similar conclusions, the more likely it is that the conclusion is true.

The recent enthusiasm for evidence-based guidelines has led to the impression that they can replace physician experience and expert opinion. Evidence-based methods should not be viewed as minimizing the usefulness of the non-evidence-based means clinicians use in making decisions. Physician experience and judgment, as well as the informed opinion of experts, are all useful sources of knowledge for the practicing clinician. The fact that these are not based upon formalized evidence assessments does not mean they are not of value. However, it could be argued that the value of every opinion-based review could be improved if a formalized review of existing evidence was included.

Knowledge derived from evidence needs to be viewed as one pillar, not the sole pillar, that supports

clinical decisions. Other pillars include knowledge of established neurologic principles and informed clinical judgment. From this perspective, evidence-based guidelines can be seen as an additional tool, not the only tool, in the clinicians’ armamentarium to help them make decisions to help their patients. There can be no doubt that before making any decision, a clinician should be aware of what applicable evidence exists. Fundamentally, this is what guidelines are for.

Received July 9, 2008. Accepted in final form August 18, 2008.

REFERENCES

1. Chaturvedi S, Bruno A, Feasby T, et al. Carotid endarterectomy: an evidence-based review: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 2005;65:794–801.
2. French JA, Kanner AM, Bautista J, et al. Efficacy and tolerability of the new antiepileptic drugs I: treatment of new onset epilepsy: report of the Therapeutics and Technology Assessment Subcommittee and Quality Standards Subcommittee of the American Academy of Neurology and the American Epilepsy Society. *Neurology* 2004;62:1252–1260.
3. Wheless JW, Clarke DF, Carpenter D. Treatment of pediatric epilepsy: expert opinion, 2005. *J Child Neurol* 2005;20 suppl 1:S1–56; quiz S59–60.
4. Edlund W, Gronseth G, So Y, Franklin G. *Clinical Practice Guideline Process Manual*. St. Paul, MN: American Academy of Neurology; 2004.
5. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185–1190.
6. Holloway RG, Mooney CJ, Getchius TS, Edlund WS, Miyasaki JO. Conflicts of interest for authors of American Academy of Neurology practice guidelines. *Neurology* 2008;71:57–63.
7. Grogan PM, Gronseth GS. Practice parameter: Steroids, acyclovir, and surgery for Bell’s palsy (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2001;56:830–836.
8. Gronseth GS, Barohn RJ. Practice parameter: thymectomy for autoimmune myasthenia gravis (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2000;55:7–15.
9. Krumholz A, Wiebe S, Gronseth G, et al. Practice Parameter: evaluating an apparent unprovoked first seizure in adults (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology and the American Epilepsy Society. *Neurology* 2007;69:1996–2007.
10. Thurman DJ, Stevens JA, Rao JK. Practice parameter: assessing patients in a neurology practice for risk of falls (an evidence-based review): report of the quality standards subcommittee of the American Academy of Neurology. *Neurology* 2008;70:473–479.

Neurology[®]

Invited Article: Practice parameters and technology assessments: What they are, what they are not, and why you should care

Gary Gronseth and Jacqueline French

Neurology 2008;71;1639-1643

DOI 10.1212/01.wnl.0000336535.27773.c0

This information is current as of November 10, 2008

Updated Information & Services	including high resolution figures, can be found at: http://n.neurology.org/content/71/20/1639.full
References	This article cites 8 articles, 8 of which you can access for free at: http://n.neurology.org/content/71/20/1639.full#ref-list-1
Citations	This article has been cited by 11 HighWire-hosted articles: http://n.neurology.org/content/71/20/1639.full##otherarticles
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): All Cerebrovascular disease/Stroke http://n.neurology.org/cgi/collection/all_cerebrovascular_disease_stroke All Education http://n.neurology.org/cgi/collection/all_education Clinical neurology examination http://n.neurology.org/cgi/collection/clinical_neurology_examination MRI http://n.neurology.org/cgi/collection/mri Ocular motility http://n.neurology.org/cgi/collection/ocular_motility
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://www.neurology.org/about/about_the_journal#permissions
Reprints	Information about ordering reprints can be found online: http://n.neurology.org/subscribers/advertise

Neurology® is the official journal of the American Academy of Neurology. Published continuously since 1951, it is now a weekly with 48 issues per year. Copyright . All rights reserved. Print ISSN: 0028-3878. Online ISSN: 1526-632X.