



2017 Edition

Clinical Practice Guideline Process Manual

Prepared by

Gary S. Gronseth, MD, FAAN

Julie Cox, MFA

David Gloss, MD, MPH&TM, FAAN

Shannon Merillat, MLIS

Jon Dittman, MA, MFA

Melissa J. Armstrong, MD, MSc, FAAN

Thomas S. D. Getchius

On behalf of the Guideline Development, Dissemination, and Implementation Subcommittee
of the American Academy of Neurology

For the American Academy of Neurology Guideline Development, Dissemination,
and Implementation Subcommittee, the membership, and the public

2017 AAN Clinical Practice Guideline Process Manual

For more information, contact:

American Academy of Neurology
201 Chicago Avenue
Minneapolis, MN 55415
(612) 928-6000
guidelines@aan.com

The authors thank the following for their contributions:

- Jim Hopwood and Tim Streeter, MA, for their work on this edition
- Julie Cox, MFA; Erin Hagen; Siu Lee; Laura Moses Woodroffe; and Tim Streeter, MA, for their work on the 2011 edition
- Wendy Edlund; Yuen T. So, MD, PhD, FAAN; and Gary Franklin, MD, MPH, FAAN, for their work on the 2004 edition
- James C. Stevens, MD, FAAN; Michael Glantz, MD, FAAN; Richard M. Dubinsky, MD, MPH, FAAN; and Robert E. Miller, MD, for their work on the 1999 edition
- Members of the Guideline Development, Dissemination, and Implementation Subcommittee for their efforts in developing high-quality systematic reviews and evidence-based practice guidelines for the AAN membership

Guideline Development, Dissemination, and Implementation Subcommittee Members

Cynthia L. Harden, MD, Chair
Steven R. Messé, MD, FAAN, Co-Vice Chair
Sonja Potrebic, MD, PhD, FAAN, Co-Vice Chair
Stephen Ashwal, MD, FAAN, Ex-Officio
Brian Callaghan, MD
Jane Chan, MD, FAAN
Gregory S. Day, MD, MSc
Diane Donley, MD
Richard M. Dubinsky, MD, MPH, FAAN
Jeffrey Fletcher, MD, FAAN
Jacqueline French, MD, FAAN, Guideline Process Historian
David Gloss, MD, MPH&TM, FAAN
Gary S. Gronseth, MD, FAAN
Michael Haboubi, DO
John J. Halperin, MD, FAAN
Yolanda Holler-Managan, MD, FAAN
Annette M. Langer-Gould, MD, PhD
Nicole Licking, DO
David Michelson, MD
Pushpa Narayanaswami, MD, MBBS, FAAN
Maryam Oskoui, MD, FAAN
Tamara Pringsheim, MD, FAAN
Alejandro A. Rabinstein, MD, FAAN
Alexander Rae-Grant, MD
Kevin Sheth, MD, FAAN
Kelly Sullivan, PhD

This edition was approved by the AAN Guideline Development, Dissemination, and Implementation Subcommittee on July 16, 2016; by the Practice Committee on October 15, 2016; and by the AAN Institute Board of Directors on June 17, 2017.

Suggested citation: Gronseth GS, Cox J, Gloss D, et al.; on behalf of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. 2017. Clinical Practice Guideline Process Manual, 2017 ed. Minneapolis, MN: The American Academy of Neurology.

© 2017 American Academy of Neurology

Table of Contents

Preface	3	Logistics of the AAN Guideline and Case Definition Development Process	28
Introduction to Evidence-based Medicine	4	Understanding Common Uses of AAN Guidelines and Case Definitions	28
EBM Process as Applied by the AAN	5	Distinguishing Between the Types of AAN Evidence-based Documents	28
Developing the Questions	5	Identifying the Five Document Types	29
PICO Format	5	Nominating the Topic	29
Types of Clinical Questions	6	Collaborating with Other Societies	29
Development of an Analytic Framework	7	Forming the Development Panel.	30
Finding and Analyzing Evidence.	9	Implementing the AAN Conflict of Interest Policy for Guidelines and Case Definitions	30
Finding the Relevant Evidence	9	Disclosing Relationships and Determining Relevance	30
Identifying Methodologic Characteristics of the Studies	9	Identifying Relationships Considered Conflicting That Preclude Panel Involvement	31
Rating the Risk of Bias	11	Understanding Panel Composition and Responsibilities	31
Understanding Measures of Association	15	Managing Conflicts: Restrictions for Conflicted Developers and Reviewers	32
Understanding Measures of Statistical Precision	15	Disclosing Conflicts at Publication	32
Interpreting a Study	16	Identifying Violations of Conflict of Interest Policy for Guidelines	32
Synthesizing Evidence—Formulating Evidence-based Conclusions.	17	Undertaking Authorship	33
Accounting for Conflicting Evidence	18	Understanding Roles and Responsibilities	33
Knowing When to Perform a Meta-analysis	18	Completing the Project Protocol	33
Wording Conclusions for Nontherapeutic Questions	19	Developing Clinical Questions	34
Capturing Issues of Generalizability in the Conclusion	19	Developing the Search Strategy: Selecting the Search Terms and Databases	34
Making Practice Recommendations.	20	Selecting Inclusion and Exclusion Criteria	34
Developing the Recommendation, Step 1:	20	Setting the Project Timeline	35
Rating Our Overall Confidence in the Evidence from the Perspective of Supporting Practice Recommendations		Performing the Literature Search	35
Developing the Recommendation, Step 2:	21	Consulting a Research Librarian	35
Constructing the Rationale (the Logic Supporting Recommendations)		Documenting the Literature Search	35
Developing the Recommendation, Step 3:	23	Ensuring the Completeness of the Literature Search: Identifying Additional Articles	35
Crafting Actionable Recommendations		Using Data from Existing Traditional Reviews, SRs, and Meta-analyses	35
Developing the Recommendation, Step 4:	25	Minimizing Reporting Bias: Searching for Non-peer-reviewed Literature	36
Assigning Levels of Strength to the Recommendations		Selecting Articles	36
Knowing When Not to Make a Recommendation	27	Obtaining and Reviewing Titles and Abstracts	36
Making Suggestions for Future Research	27	Tracking the Article Selection Process	36
		Obtaining and Reviewing Articles	36

Table of Contents/Topic List—Continued

Extracting Study Characteristics37	Appendix 6: Steps in the AAN Development Process for Evidence-based Documents51
Developing a Data Extraction Form	37	Focused SR Development Process	51
Constructing the Evidence Tables	37	Comprehensive SR Development Process	51
Constructing Evidence Synthesis Tables	37	Practice Advisory Development Process	51
Drafting the Document37	Practice Guideline Development Process	52
Getting Ready to Write	37	Case Definition Development Process	52
Formatting the Manuscript	38	Appendix 7: Elements of AAN Evidence-based Documents53
Reviewing and Approving Guidelines and Case Definitions40	Appendix 8: Topic Nomination Priority Scoring Tool53
Stages of Review	40	American Academy of Neurology Topic Nomination Priority Score Instructions	53
Taking Next Steps (Beyond Publication)42	American Academy of Neurology Evidence-based Guideline Topic Nomination Form	54
Undertaking Dissemination	42	Appendix 9: Relationship Disclosure Form54
Responding to Correspondence	42	Appendix 10: Development Process: Roles and Responsibilities59
Updating Guidelines and Case Definitions	42	Lead Developer Responsibilities	59
Appendices	44	Codeveloper Responsibilities	59
Appendix 1: The American Academy of Neurology Institute	44	GDDI Facilitator Responsibilities	59
Appendix 2: Classification of Evidence Schemes	44	AAN Methodologist Responsibilities	59
Criteria for Rating Therapeutic Studies	44	AAN Staff Responsibilities: Development	60
Criteria for Rating Cost-effectiveness Studies	44	AAN Staff Responsibilities: Manuscript Editing	60
Criteria for Rating Causation Studies	45	Appendix 11: Project Protocol Template62
Criteria for Rating Prognostic Accuracy Studies	45	Protocol for Proposed [indicate intended final document type: focused systematic review, comprehensive systematic review, practice advisory, practice guideline] Project: Title	62
Criteria for Rating Diagnostic Accuracy Studies	46	Appendix 12: Sample Data Extraction Forms65
Criteria for Rating Population Screening Studies	46	Appendix 13: Sample Evidence Synthesis Table68
Appendix 3: Sample Evidence Table	46	Appendix 14: Manuscript Format68
Appendix 4: Tools for Building Conclusions and Recommendations	47	Appendix 15: Sample Revision Table71
Conclusions	47	References	72
Recommendations	48		
Appendix 5: Tools for Determining the Strength of the Recommendation	49		
Modified Delphi Process Questionnaire	49		
Rationale Profile Tables	50		

Preface

This manual provides instructions for developing American Academy of Neurology (AAN) focused systematic reviews (SRs), comprehensive systematic reviews (comprehensive SRs), practice advisories, practice guidelines, and case definitions. The term *guideline* or *guidelines* is used when the guidance describes these document types collectively: focused SRs, comprehensive SRs, practice advisories, and practice guidelines. Where case definitions are described, the text specifies the case definition document type.

This manual is intended for members of the AAN's Guideline Development, Dissemination, and Implementation Subcommittee (GDDI) and for developers, including development facilitators, of AAN guidelines and case definitions. The manual is also available to anyone seeking information about the AAN guideline and case definition development process, including AAN members and the public.

Practice guidelines and practice advisories are statements that include recommendations intended to optimize patient care that are informed by an SR of the evidence and an assessment of the benefits and harms of alternative care options.¹ The documents produced are intended to provide guidance to AAN members and other clinicians who evaluate, diagnose, or treat patients with neurologic disorders. AAN guidelines and case definitions cannot anticipate all clinical scenarios in which neurologic signs, symptoms, or illness may be encountered, and should not be considered as a statement of the standard of care.

All AAN guidelines are based on an SR and analysis of the literature pertinent to the specific clinical circumstance. The evidence derived from the SR informs a panel of experts who transparently develop the conclusions and recommendations of the guideline using a formal consensus development process.

This manual provides developers with specific instructions regarding the development of AAN guidelines and case definitions. However, SR and practice guideline development techniques are rapidly changing in response to growing evidence on which methodologies are effective and efficient. As a consequence, the AAN Institute Board of Directors has granted permission to the AAN GDDI to revise and expand on the processes described in this manual and to pilot those revised or expanded processes in development efforts. Any document resulting from implementation of those piloted processes that are submitted for publication before the processes are formally documented in a revised edition of this manual will include a statement describing the piloted processes used. Any change in existing processes, or introduction of new processes, will be reviewed and approved by the AAN Institute Board of Directors (see [Appendix 1](#)) in an amendment to this manual made available at AAN.com/guidelines/home/development.

This manual is divided into four sections. The first is a brief introduction to evidence-based medicine (EBM). This section closes with the rationale for the AAN's adoption of the EBM methodology for the development of its practice recommendations.

The second section is a description of the EBM process as applied by the AAN. It describes the technical aspects of each step of the process—from developing questions to formulating recommendations.

In the third section, the manual describes the logistics of AAN guideline and case definition development. It details the intricacies of the development process—from proposing a topic to formatting and writing an AAN guideline or case definition for publication.

The last section consists of appendices of supportive materials, including tools useful for the development of an AAN guideline or case definition.

It is important to note that this manual cannot describe all the situations that guideline and case definition developers may encounter. Rather, the manual provides an overview of the process, emphasizing the methodologies that are most useful in the majority of situations.

DID YOU KNOW? The Three Pillars

Evidence is only one source of knowledge that clinicians use to make decisions. The other two sources are established *principles*—for example, the neuroanatomic principles that enable neurologists to know precisely that a patient has a lesion in the lateral medulla just by examining the patient—and *judgment*, the intuitive sense clinicians rely on to help them decide what to do when there is uncertainty. One of the goals of the EBM method of analysis is to distinguish explicitly between these three sources of knowledge.

Recommendation

Judgment

Evidence

Principles

Introduction to Evidence-based Medicine

EBM concepts are best introduced with a case, such as the following example regarding ischemic stroke. A 55-year-old banker with a history of controlled hypertension is diagnosed with a small, left-hemispheric ischemic stroke. He has minimal poststroke functional deficits. The usual stroke workup does not identify the specific cause. An echocardiogram does not identify an obvious embolic source but does demonstrate a patent foramen ovale (PFO). What is the best strategy to prevent another ischemic stroke in this patient?

Note: Throughout this manual, we use many hypothetical examples. The conclusions and recommendations derived from these examples are for illustrative purposes only and are not necessarily consistent with AAN published guidelines.

Neurologists have varied and often strong opinions on the appropriate management of patients with cryptogenic stroke who have PFOs. Some would recommend closure of the PFO, as it is a potential source of paradoxical emboli. Others would consider the PFO incidental and unlikely to be causally related to the stroke.

Some would choose antiplatelet medications for secondary stroke prevention, and others would choose anticoagulation. Which treatment strategy is most likely to prevent another stroke?

Asking a question is the first step in the EBM process (see Figure 1). To answer the PFO question, the EBM method would next require looking for strong evidence. So, what is evidence?

DID YOU KNOW?

It is important to remember that in relation to AAN guidelines and case definitions, the term *evidence* refers to information from studies of clinically important outcomes in patients with specific conditions undergoing specific interventions. Although basic science studies, including animal studies, provide important information in other contexts, they are not formally considered in the development of guidelines and case definitions.

Evidence in an EBM context is information from any study of patients with the condition of interest who are treated with the intervention of interest and are followed to determine their outcomes. Evidence that would inform our question can be gained from studies of patients with cryptogenic stroke and PFO who undergo PFO closure or other therapy and are followed to determine whether they have subsequent strokes. For finding such studies, the EBM method requires comprehensive searches of online databases such as MEDLINE. The systematic literature search maximizes the chance that the reviewer will find all relevant studies.

When a study is found, the developer needs to determine the strength of the evidence it provides. For this purpose, EBM provides validated rules that determine the likelihood that a study will accurately answer the clinical question. Studies likely to be accurate provide strong evidence. Rating articles according to the strength of the evidence provided is especially necessary when different studies provide conflicting results. For example, some studies of patients with cryptogenic PFO stroke might suggest that closure lowers stroke risk, and others might suggest that medical therapy alone is as effective as PFO closure. The study providing the strongest evidence should carry more weight.

After all the relevant studies have been found and rated, the next step in the EBM process is to synthesize the evidence to answer the question. In regard to PFO, after the literature has been comprehensively searched and all the studies have been rated, one might discover that no study provides strong evidence that informs the question as to the optimal therapy. Therefore, the evidence is insufficient to support or refute the effectiveness of any of the proposed treatment strategies.

When faced with insufficient evidence to answer a clinical question, clinicians have no choice but to rely on their individual judgments. The absence of strong evidence is likely one of the reasons there is such practice variation in regard to the treatment of PFO. Importantly, relative to our PFO question, the EBM process tells us that these

treatment decisions are judgments—that is, they are merely informed opinions. No matter how strong the opinion, no one really knows which treatment strategy is more likely to prevent another stroke.

The all-too-common clinical scenario for which there is insufficient evidence to inform our questions highlights the rationale for the AAN's decision to rely on EBM methods for guideline and case definition development. In the case of insufficient evidence, an expert panel's opinion on the best course of action is sought. However, the recommendations resulting from expert deliberations in such a situation would transparently indicate that the recommendation is largely based on expert judgment. In other words, the practice advisory/guideline would highlight the uncertainty inherent in a recommendation based on insufficient evidence.

To be sure, the AAN values the opinions of experts and involves them in guideline and case definition development. However, the AAN also understands that the neurologist caring for a patient has better knowledge of that patient's values and individual circumstances. When there is uncertainty, the AAN believes decisions are best left to individual physicians and their patients after both physicians and patients have been fully informed of the limitations of the evidence.

DID YOU KNOW? Misconceptions Regarding EBM

A common, pervasive misconception of EBM is that it is “cookbook medicine” and attempts to constrain physician judgment. More often than not, EBM highlights the limitations of the evidence and emphasizes the need for individualized physician judgment in all clinical circumstances.

EBM Process as Applied by the AAN

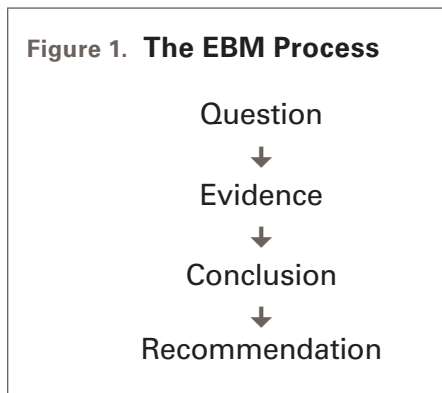
The EBM process used in the cryptogenic stroke and PFO scenario illustrates the flow of the EBM process (see Figure 1) in the development of AAN guidelines and case definitions. First, developers identify one or more clinical questions that need to be answered. The question(s) should address an area of quality concern, controversy, confusion, or practice variation.

Second, developers identify and evaluate all pertinent evidence. A comprehensive literature search is performed. The evidence uncovered in the search is evaluated and explicitly rated on the basis of content and quality.

Third, the developers draw conclusions that synthesize and summarize the evidence to answer the clinical question(s).

Finally, the developers provide guidance to clinicians by systematically translating the conclusions of the evidence to action statements in the form of practice recommendations. The recommendations are worded and rated on the basis of the quality of supporting data and other factors, including the overall magnitude of the expected risks and benefits associated with the intervention.

The subsequent sections expand on each of these steps.



PITFALL

Many guidelines have been delayed for years because of poorly formulated questions.

DID YOU KNOW?

The first three steps of the EBM process—from questions to conclusions—constitute the SR. If one stops at conclusions, one has not developed a practice advisory/guideline. Adding the additional step—from conclusions to recommendations—transforms the SR into a practice advisory/guideline.

Developing the Questions

Developing a question answerable from the evidence forms the foundation of the AAN’s EBM process. The literature search strategy, evidence-rating scheme, and format of the conclusions and recommendations all proceed directly from the question. Getting the questions right is critical.

Formulating an answerable clinical question is not a trivial step. It takes considerable thought and usually requires several iterations.

PICO Format

Clinical questions must have four components:

- Population:** The type of person (patient) involved
- Intervention:** The exposure of interest that the person experiences (e.g., therapy, positive test result, presence of a risk factor)
- Co-intervention:** An alternative type of exposure that the person could experience

(e.g., no therapy, negative test result, absence of a risk factor—sometimes referred to as the “control”)

Outcome: The outcome(s) to be addressed

Population

The population usually consists of a group of people with a disease of interest, such as patients with Bell’s palsy or patients with amyotrophic lateral sclerosis (ALS). The population of interest may also consist of patients at risk for a disease—for instance, patients with suspected multiple sclerosis (MS) or those at risk for stroke would constitute a specific population.

Often, it is important to be very specific in defining the patient population. It may be necessary, for example, to indicate that the patient population is at a certain stage of disease (e.g., patients with new-onset Bell’s palsy). Likewise, it may be necessary to indicate explicitly that the population of interest includes or excludes children.

DID YOU KNOW?—The PICO Format

In the EBM world, the necessity of formulating well-structured clinical questions is so ingrained that there is a mnemonic in common use: PICO. This helps to remind guideline and case definition developers of the need to explicitly define all four components of a clinical question.

Some EBM experts recommend adding two additional items to a clinical question: “T” for time, to explicitly indicate the time horizon one is interested in when observing the outcomes (e.g., disability at three months following a stroke); and “S” for setting, to identify the particular setting that is the focus of the question (e.g., community outpatient setting vs. tertiary hospital inpatient setting). PICO is thus sometimes expanded to PICOTS.

Intervention

The intervention defines the treatment or diagnostic procedure being considered. The question almost always asks whether this intervention should be done. An example is “Should patients with new-onset Bell’s palsy be treated with steroids?”

An example from the perspective of a diagnostic consideration would be “Should patients with new-onset Bell’s palsy routinely receive brain imaging?”

More than one intervention can be explicitly or implicitly included in the question, such as, “In patients with ALS, which interventions improve sialorrhea?” This more general question implies that developers will assess all potential interventions for treating sialorrhea.

It may be important to be highly specific in defining the intervention. For example, developers might indicate a specific dose of steroids for the Bell’s palsy treatment of interest. Likewise, developers might choose to limit the question to steroids received within the first three days of palsy onset.

The way the interventions are specifically defined in the formulation of the question will determine which articles are relevant to answering the question.

Co-intervention

The co-intervention is the alternative to the intervention of interest. For therapeutic questions, the co-intervention could be no treatment (or placebo) or an alternative treatment (e.g., L-3,4-dihydroxyphenylalanine [L-DOPA] vs. dopamine agonists for the initial treatment of Parkinson disease [PD]). For a population screening question, the alternative is not to screen.

The co-intervention is a bit more difficult to conceptualize for prognostic or diagnostic questions. Here the “intervention” is often something that cannot be actively controlled or altered. Rather it is the result of a diagnostic test (e.g., the presence or absence of 14-3-3 protein in the spinal fluid of a patient with suspected prion disease) or the presence or absence of a risk factor (e.g., the presence or absence of a pupillary light response at 72 hours in a patient post-cardiac arrest). In regard to a prognostic question, the “co-intervention” is the alternative to the presence of a risk factor—in other words, the absence of a risk

factor. Likewise, for a diagnostic test, the alternative to the “intervention”—a positive test result—is a negative test result.

Of course, there are circumstances where there may be many alternatives. The initial treatment of PD, for example, could commence with L-DOPA, a dopamine agonist, or a monoamine oxidase B inhibitor.

Finally, it is important to realize that there are times when the co-intervention is implied rather than explicitly stated in the question. The following is an example:

- In patients with Bell’s palsy, does prednisolone given within the first three days of onset of facial weakness improve the likelihood of complete facial functional recovery at six months?

Here the co-intervention is not stated but implied. The implied alternative to prednisolone in this question is no prednisolone.

Outcomes

The outcomes to be assessed should be clinically relevant to the patient. Indirect (or surrogate) outcome measures, such as laboratory or radiologic results, should be avoided—if doing so is feasible—because they often do not predict clinically important outcomes. Many treatments reduce the risk for a surrogate outcome but have no effect, or have harmful effects, on clinically relevant outcomes; some treatments have no effect on surrogate measures but improve clinical outcomes. In unusual circumstances—when surrogate outcomes are known to be strongly and causally linked to clinical outcomes—they can be used in developing a practice recommendation. (See the section on deductive inferences.)

When specifying outcomes, it is important to specify all of the outcomes that are relevant to the patient population and intervention. For example, the question might deal with the efficacy of a new antiplatelet agent in preventing subsequent ischemic strokes in patients with noncardioembolic stroke. Important outcomes needing explicit consideration include the risk of subsequent ischemic stroke (both disabling and nondisabling), death, bleeding complications (both major and minor), and other potential adverse events. Developers should attempt to include every clinically

relevant outcome. When there are multiple clinically important outcomes, it is often helpful to rate the outcomes by their importance. (The developer will need to specify the relative importance of outcomes again when assessing the strength of the recommendation.)

In addition to defining the outcomes that are to be measured, it may be helpful to state when the outcomes should be measured. The interval chosen should be clinically relevant; for chronic diseases, outcomes that are assessed after a short follow-up period may not reflect long-term outcome.

Questions should be formulated so that the four PICO elements are easily identified. The following is an example:

Population: For patients with Bell’s palsy

Intervention: Do oral steroids given within the first three days of onset

Co-intervention: Compared with no steroids

Outcome: Improve long-term facial functional outcomes?

Types of Clinical Questions

There are several distinct subtypes of clinical questions. The differences between question types relate to whether the question is primarily of a therapeutic, prognostic, or diagnostic nature. Recognizing the different types of questions is critical to guiding the process of identifying evidence and grading its quality.

Therapeutic

The easiest type of question to conceptualize is the therapeutic question. The clinician must decide whether to use a specific treatment. The relevant outcomes of interest are the effectiveness, safety, and tolerability of the treatment. The strongest study type for determining the effectiveness of a therapeutic intervention is the masked randomized controlled trial (RCT).

Diagnostic and Prognostic Accuracy

There are many important questions in medicine that do not relate directly to the effectiveness of an intervention in improving outcomes. Rather than deciding to perform an intervention to treat a disease, the clinician may need to decide whether he or she should perform an intervention to determine the

presence or prognosis of the disease. The relevant outcome for these questions is not the effectiveness of the intervention for improving patient outcomes. Rather, the outcome relates to improving the clinician's ability to predict the presence of the disease or the disease prognosis. The implication of these questions is that improving clinicians' ability to diagnose and prognosticate indirectly translates to improved patient outcomes.

For example, a question regarding prognostic accuracy could be worded, for patients with new-onset Bell's palsy, does measuring the amplitude of the facial compound motor action potential predict long-term facial outcome? The intervention of interest in this question is clearly apparent: facial nerve conduction studies. The outcome is also apparent: an improved ability to predict the patient's long-term facial functioning. Having the answer to this question would go a long way in helping clinicians to decide whether they should offer facial nerve conduction studies to their patients with Bell's palsy.

An RCT would not be the best study type for measuring the accuracy of facial nerve conduction studies for determining prognosis in Bell's palsy. Rather, the best study type would be a prospective, controlled, cohort survey of a population of patients with Bell's palsy who undergo facial nerve conduction studies early in the course of their disease and whose facial outcomes are determined in a masked fashion after a sufficiently long follow-up period.

Questions of diagnostic accuracy follow a format similar to that of prognostic accuracy questions. For example, for patients with new-onset peripheral facial palsy, does the presence of decreased taste of the anterior ipsilateral tongue accurately identify those patients with Bell's palsy? The intervention of interest is testing ipsilateral taste sensation. The outcome of interest is the presence of Bell's palsy as determined by some independent reference. (In this instance, the reference standard would most likely consist of a case definition that included imaging to exclude other causes of peripheral facial palsy.)

As with questions of prognostic accuracy, the best study type to determine the accuracy of decreased taste sensation for identifying Bell's palsy would be a prospective, controlled, cohort survey of a population of patients

presenting with peripheral facial weakness who all had taste sensation tested and who all were further studied to determine whether they in fact had Bell's palsy, using the independent reference standard. If such a study demonstrated that testing taste sensation was highly accurate in distinguishing patients with Bell's palsy from patients with other causes of peripheral facial weakness, one might recommend that clinicians routinely test taste in this clinical setting.

Population Screening

There is another common type of clinical question worth considering. These questions have a diagnostic flavor but are more concerned with diagnostic yield than with diagnostic accuracy. This type of question is applicable to the situation where a diagnostic intervention of established accuracy is employed. An example is, in patients with new-onset peripheral facial palsy, should a physician routinely obtain a head MRI to identify sinister pathology within the temporal bone causing the facial palsy? There is no concern with regard to the diagnostic accuracy of head MRI in this situation. The diagnostic accuracy of MRI in revealing temporal bone pathology is established. Rather, the clinical question of interest is whether it is useful to routinely screen patients with facial palsy with a head MRI. The outcome of interest is the yield of the procedure: the frequency with which the MRI reveals clinically relevant abnormalities in this patient population. The implication is that if the yield were high enough, clinicians would routinely order the test.

The best evidence source to answer this question would consist of a prospective study of a population-based cohort of patients with Bell's palsy who all undergo head MRI early in the course of their disease.

Causation

Occasionally, a guideline asks a question regarding the cause-and-effect relationship of an exposure and a condition. Unlike diagnostic and prognostic accuracy questions, which look merely for an association between a risk factor and an outcome, causation questions seek to determine whether an exposure causes a condition. An example is, does chronic repetitive motion cause carpal tunnel syndrome? Another example is, does natalizumab cause progressive multifocal

leukoencephalopathy? The implication is that avoidance of the exposure would reduce the risk of the condition. As in these examples, causation most often relates to questions of safety.

Theoretically, as with therapeutic questions, the best evidence source for answering causation questions is the RCT. However, in many circumstances, for practical and ethical reasons an RCT cannot be done to determine causation. The outcome may be too uncommon for an RCT to be feasible, as there may be no way to assign patients randomly to varying exposures. In these circumstances, the best evidence source for causation is a cohort survey where patients with and patients without the exposure are followed to determine whether they develop the condition. For answering the question of causation in this type of study, it is critical to strictly control for confounding differences between those exposed and those not exposed.

Determining the type of question early in guideline development is critical for directing the process. The kind of evidence needed to answer the question and the method for judging a study's risk of bias follow directly from the question type.

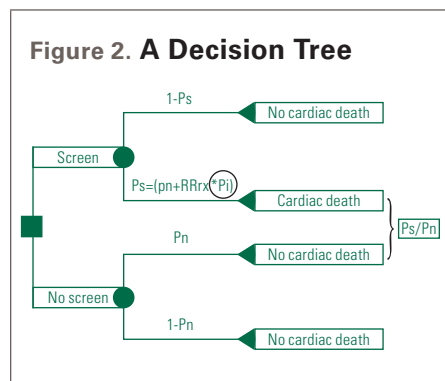
Development of an Analytic Framework

Fundamentally, all guidelines attempt to answer the question: For this patient population, does a specific intervention improve outcomes? The goal is to find evidence that directly links the intervention with a change in outcomes. When such direct evidence is found, it is often a straightforward exercise to develop conclusions and recommendations. When direct evidence linking the intervention to the outcome is not found, it may be necessary to explicitly develop an analytic framework to help define the types of evidence needed to link the intervention to patient relevant outcomes.

As a case in point, consider myotonic dystrophy (MD). Patients with MD are known to be at increased risk for cardiac conduction abnormalities. The question posed is, does routinely looking for cardiac problems in patients with MD decrease the risk that those patients will have heart-related complications such as sudden death?

One type of analytic framework that can be constructed is a decision tree.

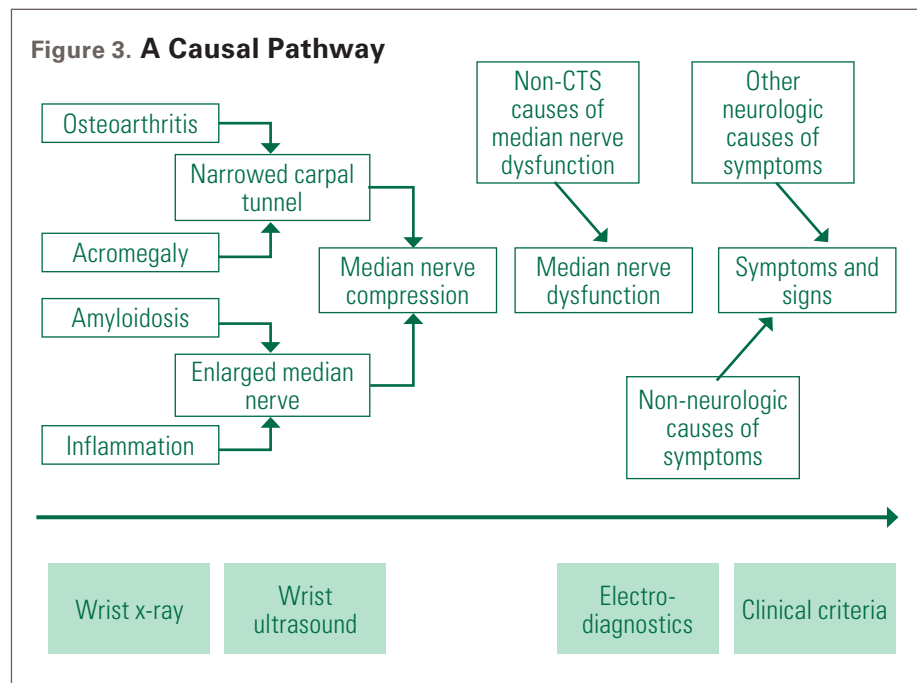
Figure 2 depicts graphically the factors that contribute to a decision that must be made (indicated by the green square—a decision node—at the base of the “sideways” tree), using the example of the risk of cardiac problems in patients with MD. If one does not screen, the patient might or might not develop a cardiac conduction problem that leads to cardiac death (this probability is depicted by the green circle labeled *Screen*—a chance node). On the other hand, even if one screens, the patient still has a chance of cardiac death (the green circle labeled *No screen*, another chance node); presumably, however, this chance would be decreased by some degree because one has identified patients at increased risk for cardiac death and treated them appropriately (perhaps placing a pacemaker after identifying heart block on a screening EKG). The probability that screening will identify an abnormality (P_i) (conduction block on an EKG) multiplied by a measure of the effectiveness of pacemaker placement for reducing cardiac death risk in patients with conduction block (RR_{rx}) should indicate how much the cardiac death risk is reduced with screening.



Direct evidence for a link between screening and reduced cardiac death would be provided by a study—ideally an RCT—that compares cardiac outcome in patients with MD who are screened with patients with MD who are not screened. If such evidence does not exist (which is probably the case), the analytic framework of the decision tree helps guideline producers identify alternative questions (and different evidence types) that might inform the decision. For example, one could find a study in which all patients with MD were

routinely screened with EKG and in which the percentage of patients with conduction block was reported. One might also find a separate study that reports the effectiveness of pacemaker placement in reducing the risk of cardiac death in patients with MD with conduction block. Using these evidence sources and the analytic framework enables one to link the intervention and outcome.

An analytic framework increases the likelihood that the guideline will identify studies with evidence that, when analyzed, will help us answer the underlying clinical question by leading us to ask related questions. In addition, the framework aids in the identification of all important outcomes.



Such analyses often suggest to guideline developers other helpful clinical questions to be asked. Rather than simply asking the therapeutic question directly linking intervention to outcome:

- For patients with MD, does routine screening with EKG (compared with not routinely screening) reduce the risk of sudden cardiac death?

guideline developers will also ask these questions:

- For patients with MD, how often does routine EKG screening (vs. no screening) identify patients with conduction block?
- For patients with MD and conduction block, does pacemaker placement (vs. no placement) reduce the risk of cardiac death?

Of course, in this example there are other potentially important outcomes to be considered, such as complications related to pacemaker screening; therefore, all important outcomes should be considered.

A decision tree is one tool that is commonly used to develop an analytic framework; a causal pathway is another. Figure 3 illustrates a causal pathway used to assist in developing questions for a guideline on the diagnostic accuracy of tests for carpal tunnel syndrome. Regardless of the tool chosen, it is worth taking the time to use an analytic framework to help define and refine the clinical questions.

Finding and Analyzing Evidence

Finding the Relevant Evidence

A comprehensive literature search distinguishes the SR that forms the basis of an AAN practice advisory/guideline from a standard nonsystematic review. The comprehensive search is performed to ensure, to the extent possible, that all relevant evidence is considered. This helps to reduce the likelihood that bias will be introduced into the process. Developers are not allowed to choose which articles they want to include (as they may select those articles that support their preconceptions). Rather, all relevant evidence is considered.

The most commonly searched database is MEDLINE. Other medical databases are also used (see the logistics section for further discussion).

DID YOU KNOW?

Studies are included even when the guideline panel members doubt the veracity of the results. A critical assumption built into the EBM process is that investigators do not lie about or fabricate data. Unless there is direct evidence of scientific misconduct (in which case the study would likely be retracted), every study is included and analyzed using the same rules.

A secondary search of the references from review articles identified in the initial search should be made to identify any relevant studies that may have been missed.

For transparency, it is important to keep a record of excluded articles and the reasons for their exclusion. After completing article selection, the developers construct a diagram depicting the flow of articles through the process, including the number excluded (see Figure 5). This diagram is included in the final (published) guideline.

Identifying Methodologic Characteristics of the Studies

After the studies are identified, it is necessary to extract essential characteristics of each study selected for inclusion. These extracted characteristics will be used to assess each study's strength.

The characteristics of each study will be included in a master (evidence) table. This table summarizes each study, including characteristics relevant to generalizability, risk of bias, and patient outcomes.

Elements Relevant to Generalizability

Developers should extract from the studies those elements that inform the judgment of each study's relevance to the clinical question and the generalizability of the results. These elements can be directly related to aspects of the clinical question.

Figure 4. Two-step Literature Review Process

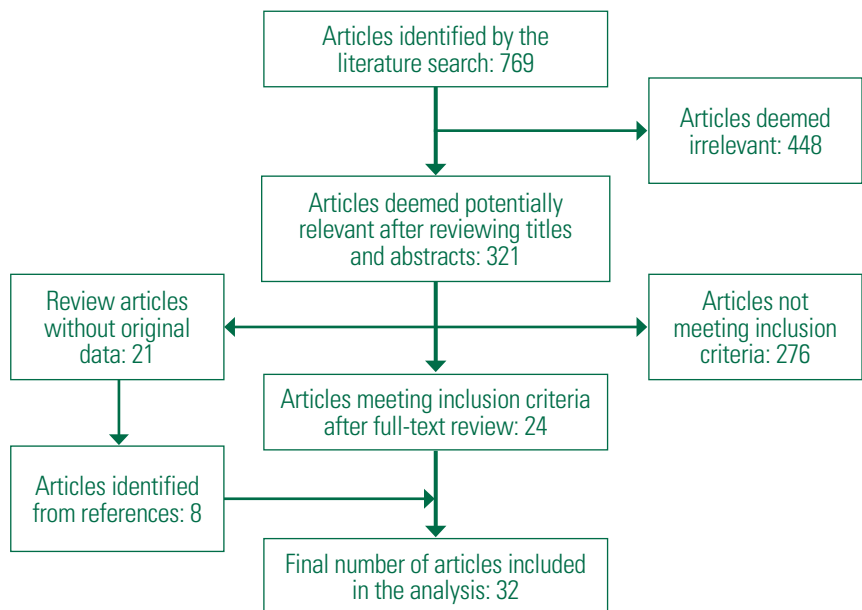


Step 1. Review abstracts
Step 2. Review full text

The initial literature search is crafted, often with the help of a research librarian, to cast a wide net to ensure that relevant articles are not missed. Content experts play an important role in this step: using their knowledge of the literature, they identify a few key articles they know are relevant to each of the clinical questions. These key articles are used to validate the search. If the key articles are missed in the search, the search strategy must be revised.

After completing a comprehensive search, developers use a two-step process (see Figure 4) to identify relevant studies. First, developers review the titles and abstracts from the comprehensive search in order to exclude citations that are obviously irrelevant to the question. Second, developers review the full text of the included titles and abstracts using prespecified inclusion criteria. The studies meeting the inclusion criteria constitute the evidence source of the guideline.

Figure 5. Flow Diagram Documenting Disposition of Articles During the Systematic Review



Elements relating to the patient population should include the following:

- Source of patients (e.g., neuromuscular referral center)
- Inclusion criterion used in the study to determine the presence of the condition of interest
- Age of the patients (e.g., mean and standard deviation)
- Sex of the included population (e.g., proportion female)

Elements relevant to the intervention and co-intervention should also be routinely extracted. These will be highly dependent on the clinical question but could include the following:

- Dose of medication used
- Timing of the intervention
- Nature of diagnostic test (e.g., CT vs. MRI)

Elements relevant to the way the study measured outcomes should also be included. These will also vary from question to question but could include the following:

- Scale used to determine the outcome (e.g., global impression of change, House-Brackman vs. Adour-Swanson scale of facial function)
- Duration of follow-up

Quality-of-Evidence Indicators

Beyond the elements pertaining to generalizability, quality-of-evidence indicators also should be extracted. The items extracted will vary according to the question type.

For therapeutic questions, critical elements include the following:

- Use of a comparison (control) group
- Method of treatment allocation (randomized vs. other)
- Method of allocation concealment
- Proportion of patients with complete follow-up
- Use of intent-to-treat methodologies
- Use of masking throughout the study (single-blind, double-blind, independent assessment)

For diagnostic or prognostic accuracy questions, important elements to be included are the following:

- Study design (case control vs. cohort survey)
- Spectrum of patients included (narrow spectrum vs. wide spectrum)

- Proportion of patients for whom both the predictor and the outcome variable are measured
- Objectiveness of the outcome variable, and whether the outcome variable is measured without knowledge of the predictor variable

For screening questions, critical elements include the following:

- Study design (data collection prospective vs. retrospective)
- Setting (population based, clinic based, or referral center based)
- Sampling method (convenience or statistical)
- Completeness (the proportion of patients in the cohort who underwent the intervention of interest)
- Masking (interpretation of the diagnostic test of interest was performed without knowledge of the patient's clinical presentation)

For causation questions, the critical elements are as follows:

- Study design (cohort or case control)
- Setting (population based, clinic based, or referral center based)
- Sampling method (selected or statistical)
- Completeness (all patients in the cohort underwent the intervention of interest)
- Masking (interpretation of the diagnostic test of interest was performed without knowledge of the patient's clinical presentation)
- The presence of confounding differences between those with and those without the putative causative factor

Patient Relevant Outcome Measures

Finally, patient relevant outcomes need to be extracted. These consist of a quantitative measure of what happened to patients within the study. For example, for a therapeutic question, how many patients improved? For a diagnostic question, how many patients had the disease?

Regardless of the question type, clinically relevant outcomes are often best measured by using discrete, categorical variables rather than continuous variables. For example, the proportion of patients with Bell's palsy who have complete facial functional recovery is a more easily interpreted measure of patient outcome than

the overall change in the median values of the House-Brackman facial function score.

Measuring patient outcomes using categorical variables involves counting patients. An example is, how many patients on drug X improved, and how many did not improve? Counting patients in this manner often enables construction of a contingency table. Table 1 is a simple two-by-two contingency table showing the numbers of patients improving on drug X vs. placebo.

	Improved	Not Improved
Drug X	13	32
Placebo	6	78

From this it is a relatively straightforward process to calculate numeric values that express the strength of association between the intervention and the outcome. Examples of measures of association are the relative risk of a poor outcome in treated patients vs. nontreated patients (the proportion of treated patients with a poor outcome divided by the proportion of nontreated patients with a poor outcome) or the poor-outcome risk difference (the proportion of treated patients with a poor outcome subtracted by the proportion of nontreated patients with a poor outcome).

Two-by-two contingency tables can also be constructed for nontherapeutic studies. For studies regarding prognosis and causation, relative risks and risk differences can also be calculated. Rather than patients grouped by whether they received treatment, patients are grouped by whether they had the risk factor of interest.

Quantitative measures of diagnostic accuracy can also be derived from a contingency table. These include sensitivities and specificities, positive and negative predictive values, and positive and negative likelihood ratios.

Finally, the quantitative measure used to describe the population screening studies is simply the yield, that is, the proportion of patients with the condition who are undergoing the test of interest.

Sometimes authors of the studies being considered might not report patient outcomes using categorical outcome variables. In such circumstances, if sufficient information is

provided, panel members themselves should attempt to construct contingency tables. If contingency tables cannot be constructed, panel members should report the quantitative outcome measure(s) as reported in the original studies. Guideline developers are encouraged to make these determinations with the help of the GDDI facilitator or methodology experts.

Rating the Risk of Bias

An important step in guideline development is to measure the risk of bias in each included study. Bias, or systematic error, is the study's tendency to measure the intervention's effect on the outcome inaccurately. It is not possible to measure the bias of a study directly. (If it were, it would imply the developer already knew the answer to the clinical question.) However, using well-established principles of good study design, the developer can estimate a study's risk of bias.

For AAN guidelines, the risk of bias in studies is measured using a four-tiered classification scheme (see [Appendix 2](#)). In this scheme, studies rated Class I are judged to have a low risk of bias; Class II, a moderate risk of bias; Class III, a moderately high risk of bias; and Class IV, a very high risk of bias. The classification rating is also known as the level of evidence.

Panel members assign each study a classification on the basis of that study's extracted quality-of-evidence characteristics. The classification scheme the AAN uses accounts only for systematic error. Random error (low study power) is dealt with separately.

A study's risk of bias can be judged only relative to a specific clinical question. The standards that are applied vary among the different question types: therapeutic, diagnostic or prognostic accuracy, screening, and causation.

The next five sections explain in more detail each study characteristic (or element) that contributes to a study's final classification for each of the five study types (therapeutic, diagnostic, prognostic, screening, and causation).

Classifying Evidence for Therapeutic Questions

Important elements for classifying the risk of bias in therapeutic articles are described next.

Comparison (Control) Group

A comparison—or control—group in a therapeutic study consists of a group of patients who did not receive the treatment of interest. Studies without a comparison group are judged to have a high risk of bias and are rated Class IV.

To be rated Class I or Class II, studies should use concurrent controls. Studies using nonconcurrent controls, such as those using external or historical controls, are rated Class III.

DID YOU KNOW?

Sometimes a study provides evidence relevant to more than one question. Often in these circumstances the study will have different ratings. For example, a study could be rated Class I for a therapeutic question and Class III for a separate, prognostic, question.

Treatment Allocation

To reduce the risk of bias, authors of a therapeutic article must ensure that treated and nontreated patient groups are similar in every way except for the intervention of interest. In other words, known and unknown baseline confounding differences between the treated and nontreated groups must be minimized.

Randomized allocation to treatment and comparison groups is the best way to minimize these confounding differences. Thus, to be rated Class I, a therapeutic study should have randomly allocated patients to the treatments being compared.

An important study characteristic that ensures patients are truly randomly allocated to different strategies is concealed allocation. Concealed allocation prevents investigators from manipulating treatment assignment. Examples of concealed allocation include the use of consecutively numbered, sealed, opaque envelopes containing a predetermined, random sequence for treatment assignment and use of an independent center that an investigator contacts to obtain the treatment assignment. In comparison, examples of nonconcealed allocation include flipping a coin (e.g., heads = treatment A, tails = treatment B) and assigning patients to treatment categories on the basis of the date (e.g., treatment A on odd-numbered days, treatment B on

even-numbered days). These nonconcealed allocation methods can be easily manipulated by investigators. For example, the coin can be flipped again, or the patient can be told to come back the next day.

DID YOU KNOW?

The effect of allocation concealment on a study's accuracy has been well established. Poor allocation concealment can introduce more bias than failure to mask outcome assessment.

In addition to a description of concealed allocation, a Class I rating requires that panel members ensure that the randomization scheme effectively balanced the treatment and comparison groups for important confounding baseline differences. In most studies the important characteristics of each treatment group are summarized in a table (usually the first table in an article describing an RCT). If important baseline differences exist, any differences in outcomes between the different treatment groups might be explained by these baseline differences rather than by any treatment effect.

Clinical studies may use methods other than random allocation to attempt to ensure that patients in different treatment groups are substantially similar in prognostically important baseline characteristics. These methods may be implemented through the study design (e.g., a matched cohort study will actively attempt to balance patients on baseline characteristics during the allocation) or during the analysis phase (e.g., stratification, multivariable analyses such as logistic regression or propensity score matching). Such studies are eligible for a Class II rating if the method used successfully matched patients in different treatment groups on all known baseline confounders demonstrated in previous studies to substantially affect prognosis. Guideline developers must use their judgment when determining which confounders substantially affect prognosis. It should be noted that the majority of studies using these techniques fail to match the patient groups on all important confounders. For this reason, these studies are typically rated no better than Class III.

Completeness of Follow-up

Patients enrolled in studies are sometimes lost to follow-up. Such losses occur for nonrandom reasons and may introduce confounding differences between the treated and nontreated groups. Thus, Class I rating requires that more than 80 percent of patients within the study have completed follow-up.

For various reasons, patients initially assigned to the treatment group might not receive the intended treatment, and patients assigned to the comparison group might receive the control treatment. If patients cross over from the treated group to the comparison group or from the comparison group to the treated group, confounding differences can be introduced. When this happens, it is important that the investigators analyze the results using intent-to-treat principles. Put simply, such principles entail analysis of the results on the basis of the group (treatment or comparison) to which each patient was originally assigned.

DID YOU KNOW?

The selection of an 80 percent completion rate is an arbitrary one. This measure of a study's quality is best understood when positioned on a continuum—the fewer patients lost to follow-up, the better. However, to fit a study into the ordinal Class I through IV system, a cutoff had to be selected.

Masking

To be rated Class I, a study addressing a therapeutic question must ensure that patients, treating providers, and persons assessing the outcome are all unaware of which treatment the patient received. Such triple-masked (blinded) studies reduce the risk of bias from the placebo effect (i.e., the biological benefit patients receive when they believe they are getting a potentially effective therapy), performance bias (e.g., treating providers might give patients perceived to be on an inferior therapy additional treatments not given to patients perceived to be on a superior therapy), and observer expectation bias (e.g., outcome assessors may incorrectly classify patients as having a good outcome if they are perceived to be on a superior therapy). Most studies

described as double-blind meet these Class I criteria.

To be rated Class II, non-triple-masked studies must ensure that persons assessing patient outcomes are unaware of the treatment assignment. Most studies labeled *single-blind* or those employing adjudicated, masked outcome assessment meet these criteria.

The requirement for masked outcome assessment can be waived (and Class II criteria met) if the outcome measure is objective. An objective outcome is one that is unlikely to be affected by observer expectation bias (e.g., patient survival or a laboratory assay). Oftentimes, determining whether an outcome is objective requires some judgment by the panel members. The AAN GDDI makes the final determination of the objectiveness of any outcome.

Class II criteria can also be met in unmasked studies using nonobjective outcomes if the study design ensures that observers' expectations regarding efficacy are substantially similar for the treatments compared. For example, a study comparing the effectiveness of one psychological treatment with another often cannot blind patients, treating providers, or outcome assessors with regard to which therapy a patient is receiving. However, such studies can control observer expectation by blinding patients and observers to the study hypothesis or by including statements in the consent process that emphasize that the investigators have no reason to expect that one treatment will be superior to another. To rate a study Class II in such circumstances, study authors must have explicitly described the methods for controlling expectations (preferably in the study protocol), and the guideline panel and GDDI members must judge that such methods were effective.

In a non-masked study using a nonobjective outcome and not controlling observer expectations, a Class III rating can be attained only if a study investigator who is not one of the treating providers has determined the outcome. Such independent outcome assessment, although not as effective as masking in reducing bias, nonetheless has been shown to be less prone to bias than having a non-masked treating clinician determine the outcome. A patient's own assessment of his or her

outcome (e.g., a seizure diary or completion of a quality-of-life questionnaire) fulfills the criteria for independent assessment.

PITFALL

It is important not to confuse allocation concealment with study masking (or blinding). Allocation concealment refers only to how investigators randomize patients to different treatments. After patients have been randomized, masking ensures that the investigators are not aware of which treatment a patient is receiving.

Primary Outcome

When designing a study, investigators are expected to designate a primary outcome, but they often do not. In addition, authors will sometimes designate multiple outcomes as primary, thereby defeating the purpose of designating a single primary outcome. Moreover, there are times when it may be impossible to determine whether study authors have reported all the outcomes they have measured.

To be rated Class I, a study must have no more than two prespecified primary outcomes. If there are three or more prespecified primary outcomes, the highest rating the study is eligible for is Class II. This rating affects all outcomes, whether primary or secondary.

Another requirement for a Class I rating is that the outcome pertinent to the guideline (regardless of whether it is a primary or secondary outcome in the original study) must have been specified a priori in the original study. This applies as well to subgroup analyses. If the article does not explicitly state that the outcome pertinent to the guideline was prespecified, then the study is eligible for no better than a Class II rating. It may be possible to confirm the selection of a prespecified primary outcome by reviewing a published protocol for the study at [ClinicalTrials.gov](https://www.clinicaltrials.gov) or similar resource.

Secondary Outcomes

When several tests are performed on the same data set, there is a chance that false-positive results (type I errors) can occur. For example, if there are five hypotheses tested, there is a 23 percent chance one of

them will be significant, even if all the tests are actually not significant. For n tests, the chance of a type I error is $1-(1-\infty)^n$. Ideally, the study authors would have adjusted for their secondary outcomes, and in this case, the guideline developers should use the study authors' reported values.

If the study authors did not adjust for multiple secondary outcomes, developers may perform a correction. The simplest and most conservative correction is the Bonferroni correction. One way to perform a Bonferroni correction is to multiply each of the observed p values by the number of comparisons measured. For articles that report five or fewer secondary outcomes, developers should perform the Bonferroni correction.

A uniformly more powerful method to correct for multiple outcomes is the Holm-Bonferroni method.³ This is the preferred method in cases where there are more than five secondary outcomes.

As an option, instead of making these corrections, the guideline panel may correct confidence intervals. In this case, the confidence intervals would be adjusted on the basis of the corrected α of the p value, and then the corrected α would be reverse-imputed.

Active Control Equivalence and Noninferiority Trials

Some therapeutic studies compare the efficacy of a new treatment with that of another standard treatment rather than placebo. Additional requirements are imposed on these trials.

To ensure that the new drug is being compared with an efficacious drug, there must be a previous Class I placebo-controlled trial establishing efficacy of the standard treatment.

In addition, the standard treatment must be used in a manner that is substantially similar to that used in previous studies (Class I placebo-controlled trial) establishing efficacy of the standard treatment (e.g., for a drug, the mode of administration, dose, and dosage adjustments are similar to those previously shown to be effective).

Furthermore, the inclusion and exclusion criteria for patient selection and the outcomes of patients receiving the standard treatment are substantially equivalent to those of a previous Class I placebo-controlled study establishing efficacy of the standard treatment.

Finally, the interpretation of the study results is based on an observed-case analysis.

Crossover Trials

A crossover trial is a type of clinical study in which the study participants are randomly assigned to receive each treatment in a different order. For example, some patients receive placebo for several weeks followed by an active drug for several weeks, whereas others receive that active drug for several weeks followed by placebo for several weeks. Often, between treatments (placebo vs. active) there is a period of no treatment called a "washout" period. With this type of study, every patient serves as his or her own control. Crossover studies are often used to increase the statistical precision of a study.

In the setting of a crossover trial, confusion can arise because the patients are serving as their own controls and so, by definition, are substantially equivalent on baseline characteristics. However, in crossover trials there can be an effect related to treatment order. Hence, it is important to ensure that the patients randomized to different treatment orders (e.g., active followed by placebo vs. placebo followed by active) are substantially equivalent on baseline characteristics.

For a crossover trial to be rated Class I, the following criteria must be met:

1. There must be a comparison of baseline characteristics across treatment order groups demonstrating substantial equivalence or with statistical adjustments for differences (if present).
2. Statistics must describe period and carryover effects, with adjustments if significant effects are present.

If the trial meets only one of these two criteria, it can be rated no higher than Class II. If it meets neither criterion, it can be rated no higher than Class III.

Classifying Evidence for Diagnostic or Prognostic Accuracy Questions

The following paragraphs present important elements to be considered when classifying evidence for a diagnostic or prognostic accuracy question.

Comparison (Control) Group

To be useful, a study of prognostic or diagnostic accuracy should include patients with and patients without the disease or

outcome of interest. Quantitative measures of accuracy cannot be calculated from studies without a comparison group. Thus, such studies are judged to have a high risk of bias and are rated Class IV.

Study Design

A Class I study of diagnostic or prognostic accuracy would be a prospective cohort survey. Investigators would start with a group of patients suspected of having a disease (the cohort). The diagnostic test would be performed on this cohort. Some patients in the cohort would have positive test results, others negative test results. The actual presence or absence of the disease in the cohort would be determined by an independent reference standard (the gold standard). Quantitative measures of the diagnostic accuracy of the test (or predictor), such as the sensitivity or specificity, could then be calculated.

In studies of diagnostic accuracy, the steps that are followed in prognostic accuracy studies are often performed in reverse order. Investigators do not start with a group of patients suspected of having the disease; rather, they select a group of patients who clearly have the disease (cases) and a group of patients who clearly do not (controls). The test is then performed on both cases and controls, and measures of diagnostic accuracy are calculated. Although such case-control studies are often easier to execute than cohort studies, this design introduces several potential biases. Thus, such studies can be rated Class II at best.

PITFALL

The term *case control* is commonly misinterpreted. Many studies have "controls." The term *case-control study*, however, is reserved specifically for studies for which investigators select patients because they have the outcome of interest (e.g., the disease) or because they do not have the outcome of interest. The former are the cases; the latter are the controls.

Data Collection

For a cohort study, data collection can be prospective or retrospective. In a prospective cohort study, both data collection and the study

itself begin before any of the patients has experienced the outcome. In a retrospective cohort study, both data collection and the study itself start after some or all of the patients have attained the outcome of interest. Retrospective data collection introduces potential bias because the investigators usually have to rely on data sources (e.g., medical records) that were not designed for the study's specific purpose. Studies with prospective data collection are eligible for a Class I rating, and those using retrospective data collection are at best Class II.

Patient Spectrum

One of the dangers of the case-control design is that such studies sometimes include either only patients who clearly have the disease or only those who clearly do not. Including such unambiguous cases can exaggerate the diagnostic accuracy of the test. To avoid this, a study employing a case-control design should include a wide spectrum of patients. A wide-spectrum study would include patients with mild forms of the disease and patients with clinical conditions that could be easily confused with the disease. A narrow-spectrum study would include only patients who clearly had the disease and normal controls. Studies employing a case-control design with a wide spectrum of patients can attain a Class II rating, and those with a narrow spectrum, Class III.

Cohort studies have a lower risk of spectrum bias than case-control studies. Occasionally, spectrum bias can be introduced into a cohort study if the study includes only patients with extreme values of the diagnostic test (or risk factor). For example, a study of the diagnostic accuracy of CSF 14-3-3 for prion disease would introduce spectrum bias if it included only patients with very high 14-3-3 levels and patients with very low 14-3-3 levels, thus excluding those with intermediate levels. The exclusion of the patients with borderline levels might exaggerate the usefulness of the test.

Reference Standard

It is essential for the usability of any study of diagnostic or prognostic accuracy that a valid reference standard be used to confirm or refute the presence of the disease or outcome. This reference standard should be independent of the diagnostic test or prognostic predictor of interest. A diagnostic test being studied can be considered

independent if the test results cannot be used in any way by the reference standard. The reference standard could consist of pathologic, laboratory, or radiologic confirmation of the presence or absence of the disease. At times, the reference standard might even consist of a consensus-based case definition. Panel members should rate as Class IV those studies that lack an independent reference standard.

Completeness

Ideally, all patients enrolled in the study should have the diagnostic test result (presence of the prognostic variable) and the true presence or absence of the disease (outcome) measured. A study can be rated no better than Class II if these variables are measured in less than 80 percent of participants.

Masking

For a study to be rated Class I or II, an investigator who is unaware of the results of the diagnostic test (presence or absence of the prognostic predictor) should apply the reference standard to determine the true presence of the disease (or determine the true outcome). In the instance of the case-control design, for the study to obtain a Class II rating, an investigator who is unaware of the presence or absence of the disease (or unaware of the outcome) should perform the diagnostic test of interest (measure the prognostic predictor of interest).

For a study to be rated Class III or better, the diagnostic test should be performed (or prognostic predictor measured) by investigators other than the investigator who determines the true presence or absence of disease (or determines the outcome).

The requirement for masked or independent assessment can be waived if the study uses an objective reference standard for determining the presence of the disease (outcome) and an objective diagnostic test (prognostic predictor) of interest. An objective measure is one that is unlikely to be affected by expectation bias.

Classifying Evidence for Population Screening Questions

For screening questions, panel members should use the study elements listed below to classify the evidence.

Data Collection

Retrospective collection of data, such as chart reviews, commonly introduces errors related to suboptimal, incomplete measurement. Thus, data collection should be prospective for a study to be classified Class I.

Setting

Studies are often performed by highly specialized centers. Because such centers tend to see more difficult and unusual cases, the patients they treat tend to be nonrepresentative of the patient population considered in the clinical question. In general, because of the potential for the study participants to be nonrepresentative, these studies from referral centers are rated Class III. Occasionally, the population of interest targeted in the screening question is primarily patients referred to specialty centers. For example, some conditions that are rare or difficult to treat may be managed only at referral centers. In these circumstances, such studies may be rated Class II.

Studies of patients recruited from nonreferral centers such as primary care clinics or general neurology clinics are more representative. These studies may be rated Class II. Population-based studies tend to be the most representative and may be rated Class I.

Sampling

The ideal methods of selecting patients for a study designed to answer a screening question are selecting all patients or selecting a statistical sample of patients. Each method ensures that the patient sample is representative. Thus, studies using a patient sample that is consecutive, random, or systematic (e.g., every other patient is included) are eligible for a Class I rating, as long as other criteria are met. Because patients may potentially be nonrepresentative, a study using a selective sample of patients (e.g., a convenience sample) can be rated at best Class III. For example, a study of the yield of MRI in patients with headache that included patients who happened to have head MRIs ordered would be at best Class III because the sample is highly selective. A study in which MRIs are performed on all consecutive patients presenting with headache is not selective and could qualify for a Class I rating (again, if other criteria are met).

Completeness

For reasons similar to those given in the sampling discussion, it is important that all patients included in the cohort undergo the test of interest. If less than 80 percent of participants receive the intervention of interest, the study cannot be rated higher than Class II.

Numerator-only Studies

Studies being used to answer a screening question are rated Class IV if patients are selected for inclusion in the study because they had the outcome of interest. For example, a study designed to determine the MRI yield in patients with migraine would be rated Class IV if only patients with abnormal MRIs were included. Such a study might be useful to describe the type of abnormalities MRI reveals in this population; however, the study cannot provide information on how often the MRI will be abnormal. In other words, such studies describe the numerator but not the denominator. A yield cannot be calculated.

Classifying Evidence for Causation Questions

In regard to patient safety in particular, it may be impractical or unethical to perform RCTs to determine whether a causal relationship exists between an exposure and a disease. A classic example of this is tobacco smoking. Because of known health risks of tobacco use, no one would advocate performing an RCT to determine whether smoking causes lung cancer. Yet the epidemiologic evidence for a causal relationship between smoking and cancer is overwhelming.

For such circumstances, the AAN has developed a causation evidence classification scheme. This enables investigators to assess the risk of bias of studies when the primary question is one of causation and the conduction of RCTs is not feasible.

The causation classification of evidence scheme is quite similar to the prognostic classification scheme. The former places additional emphasis on controlling for confounding differences between exposed and nonexposed people. Moreover, minimal thresholds for effect size are prespecified in order for studies to qualify for Class I or II designation. Finally, non-methodologic criteria centering on biological plausibility are included.

Making Modifications to the Classification of Evidence Schemes

The classification of evidence schemes described above provide general guidance for rating a study's risk of bias relative to a specific clinical question. These general schemes cannot identify all of the potential elements that contribute to bias in all situations. In specific circumstances, there can be nuances that require slight modifications to the schemes. For example, the outcome measures that are judged to be "objective" (i.e., unlikely to be affected by observer expectation bias) can vary on the basis of the clinical question asked. Those outcomes that the guideline developers deem objective, or any other modification to the classification of evidence schemes, need to be enumerated before study selection and data abstraction commence. This a priori designation of modifications is necessary to reduce the risk of bias being introduced into the review. In these circumstances, it is acceptable to modify the classification schemes slightly to correspond to the clinical questions asked. However, the schemes should not be modified to fit the evidence found. Post hoc modifications will increase the risk of bias.

Understanding Measures of Association

Interpreting the importance of the results of a study requires a quantitative measure of the strength of the association between the intervention and the outcome.

For a therapeutic question, quantitative outcomes in the treated population are usually measured relative to a nontreated population. The variables used to represent the effectiveness of an intervention quantitatively are termed *measures of effectiveness/effect* or *measures of association*. Common measures of effectiveness were introduced in the section describing study extraction and include the relative risk of an outcome and the risk difference. An example of relative risk is the proportion of patients with good facial outcomes in patients with Bell's palsy receiving steroids divided by the proportion of good facial outcomes in those not receiving steroids. An example of risk difference is the proportion of patients with good facial outcomes in patients with

Bell's palsy receiving steroids minus the proportion of good facial outcomes in those not receiving steroids.

For articles of diagnostic or predictive accuracy, the outcome variables of interest are relative risks, positive and negative predictive values, likelihood ratios, and sensitivity and specificity values.

For screening procedures, the quantitative measure of effect will be the proportion of patients with a clinically significant abnormality identified.

In general, in guideline development, absolute, categorical measures of association are preferred (e.g., risk difference) because of their greater ease of clinical interpretation; however, there are exceptions. For example, in circumstances where there is large variation in baseline risk (e.g., risk of stroke in patients with atrial fibrillation), relative measures of association may be preferable. In situations where categorical measures do not provide sufficient precision to answer the question, continuous measures may be preferred because of their greater information content.

DID YOU KNOW?

As previously mentioned, the AAN's classification of evidence scheme accounts only for the risk of bias in a study, not for the contribution of chance. Conversely, confidence intervals and *p* values do not measure a study's risk of bias. The highest-quality study has both a low risk of bias (Class I) and sufficient precision or power to measure a clinically meaningful difference.

The mathematical tools available for measuring the contribution of chance to a study's results are much more sophisticated than our ability to measure the risk of bias.

Understanding Measures of Statistical Precision

Regardless of the clinical question type or the outcome variable chosen, it is critical that some measure of random error (i.e., the statistical precision of each study) be

included in the estimate of the outcome. Random error results from chance. Some patients improve and some do not, regardless of the intervention used. In any given study, more patients may have improved with treatment than with placebo just because of chance. Statistical measures of precision (or power) gauge the potential contribution of chance to a study's results. In general, the larger the number of patients included in a study, the smaller the contribution of chance to the results.

Including 95 percent confidence intervals of the outcome measure of interest is usually the best way of gauging the contribution of chance to a study's results. A practical view of confidence intervals is that they indicate where one can expect the study results to be if the study were repeated. Most of the time, the results would fall somewhere within the upper and lower limits of the confidence interval. In other words, on the basis of chance alone, the study results can be considered to be consistent with any result within the confidence interval.

The *p* value is the next best measure of the potential for random error in a study. The *p* value indicates the probability that the difference in outcomes observed between groups could be explained by chance alone. Thus a *p* value of 0.04 indicates that there is a 4 percent probability that the differences in outcomes between patient groups in a study are related to chance alone. By convention a *p* value of < 0.05 (less than 5 percent) is usually required for a difference to be considered statistically significant.

The presence of a statistically significant association can also be determined by inspection of the upper and lower limits of the 95 percent confidence intervals. For example, if the measure of association is the relative risk or odds ratio of an outcome, and the confidence interval includes 1, the study does not show a statistically significant difference. This is equivalent to stating that the *p* value is greater than 0.05.

Relative to measures of statistical precision, 95 percent confidence intervals are preferred over *p* values. If *p* values are not provided, it is important to include measures of statistical dispersion (e.g., standard deviation, standard error, interquartile range). In most circumstances, when confidence intervals are not provided by a study's

authors, these values can be estimated from the data provided. AAN methodologists can help in this situation.

Interpreting a Study

Armed with the measure of association and its 95 percent confidence interval, one is in a position to interpret a study's results. Often the temptation here is to determine merely whether the study was positive (i.e., showed a statistically significant association between the intervention and outcome) or negative (did not show a statistically significant association). In interpreting study results, however, it is better to think of four, not two, possible outcomes. This derives from the fact that there are two kinds of differences one is interested in: whether the difference is statistically significant and whether the difference is clinically important. Henceforth, the term *significant* means *statistically significant*, and the term *important* means *clinically important*. From these two types of differences, four possible outcomes can be observed:

1. The study showed a significant and important difference between groups.

For example, an RCT of patients with cryptogenic stroke with PFO demonstrated that 10 percent of patients who had their PFO closed had strokes, whereas 20 percent of patients who did not have their PFO closed had strokes (risk difference 10 percent, 95 percent confidence interval 5–15 percent). This difference is statistically significant (the confidence interval of the risk difference does not include 0) and clinically important (no one would argue that a finding of 10 percent fewer strokes is unimportant).

2. The study showed a significant but unimportant difference between groups.

A separate RCT enrolling a large number of patients with cryptogenic stroke with PFO indicated that 10.0 percent of patients who had their PFO closed had strokes and 10.1 percent of patients who did not have their PFO closed had strokes (risk difference 0.1 percent, 95 percent confidence interval 0.05–0.015 percent). This difference is statistically significant but arguably not clinically important (there are only 1 in 1,000 fewer strokes in the patients with PFO closure).

3. The study showed no significant difference between groups, and the confidence interval was sufficiently narrow to exclude an important difference.

A third RCT enrolling a large number of patients with cryptogenic stroke with PFO showed that 5 percent of patients who had their PFO closed had strokes and 5 percent of patients who did not have their PFO closed had strokes (risk difference 0 percent, 95 percent confidence interval -0.015 percent to 0.015 percent). This difference is not statistically significant. In addition, the 95 percent confidence interval is sufficiently narrow to allow us to confidently exclude a clinically important effect of PFO closure.

4. The study showed no significant difference between groups, but the confidence interval was too wide to exclude an important difference.

Our last hypothetical RCT of patients with cryptogenic stroke with PFO demonstrates that 5 percent of patients who had their PFO closed had strokes and 5 percent of patients who did not have their PFO closed had strokes (risk difference 0 percent, 95 percent confidence interval -10 percent to 10 percent). This difference is not statistically significant. However, the 95 percent confidence interval is too wide to allow us to confidently exclude a clinically important effect of PFO closure. Because of the lack of statistical precision, the study is potentially consistent with an absolute increase or decrease in the risk of stroke of 10 percent. Most would agree that a 10 percent stroke reduction is clinically meaningful and important.

Let us consider these outcomes one at a time.

Scenario one represents the clearly positive study and scenario three the clearly negative study. A Class I study pertinent to scenario one or three would best be described as an *adequately powered* Class I study.

Scenario two usually results from a large study. The study has a very high degree of power and can show even minor differences. The minor differences may not be important. The study should be interpreted as showing no meaningful difference. A Class I study pertinent to scenario two would best be described as an *adequately powered* Class I study showing no important difference.

Scenario four results from a small study. The study is so underpowered that it is unable to show significant differences even when there might be important differences. It would be inappropriate to interpret this study as negative. A Class I study pertinent to scenario four should be described as an *inadequately powered* Class I study.

To be sure, determining what is clinically important involves some judgment.

Discussion among panel members will often resolve any uncertainty. When the clinical importance of an effect remains uncertain, it is best to stipulate explicitly in the guideline what one considers to be clinically important.

The methodologic characteristics of each informative study along with their results should be summarized in evidence tables. See *Appendix 3* for sample evidence tables.

PITFALL

In the analysis of a study that shows no significant difference between treatment groups, a common error is failing to determine whether the study had adequate power to exclude a clinically important difference. Such a study is not truly negative—rather, it is inconclusive. It lacks the precision to exclude an important difference.

Synthesizing Evidence—Formulating Evidence-based Conclusions

At this step, multiple articles pertinent to a question have been analyzed and summarized in an evidence table. These collective data must be synthesized into a conclusion. The goal at this point is to develop a succinct statement that summarizes the evidence in answer to the specific clinical question. Ideally, this summary statement will indicate the magnitude of the effect and the class of evidence on which it is based. The conclusion should be formatted in a way that clearly links it to the clinical question. In other words, the conclusion should answer the question.

Example—clinical question:

- *For patients with new-onset Bell's palsy [population]...*
- *Do oral steroids given within the first three days of onset [intervention]...*
- *Improve long-term facial outcomes [outcome]?*

Example—conclusion:

- *For patients with new-onset Bell's palsy [population]...*
- *Oral steroids given within the first three days of onset of palsy [intervention]...*
- *Are likely safe and effective for increasing the chance of complete facial functional recovery (rate difference 12 percent) (two inadequately powered Class I studies and two Class II studies) [outcome].*

Besides capturing the elements of the PICO question, four other types of information need to be considered when formulating the conclusion:

- The class of evidence
- The measure of association
- The measure of statistical precision (i.e., the random error [the power of the study as manifested by the width of the confidence intervals])
- The consistency between studies

In the example above, the class of evidence the conclusion is based on is indicated in two ways: 1) the term *likely safe and effective* indicates that the effectiveness of steroids is based on moderately strong evidence, and 2) the number and class of evidence on which the conclusion is based are clearly indicated in parentheses. To avoid confusion, one should explicitly indicate in the conclusion when studies have insufficient power to exclude a meaningful difference. *Appendix 4* provides guidance on translating evidence into conclusions.

The level of certainty directly relates to the highest class of evidence with adequate power used to develop the conclusion. Thus, conclusion language will vary on the basis of the following levels of evidence:

- Multiple Class I studies:
 - *Are highly likely to be effective...*
- Multiple Class II studies or a single Class I study:
 - *Are likely effective...*

- Multiple Class III studies or a single Class II study:
 - *Are possibly effective...*
- Multiple Class IV studies or a single Class III study:
 - *For patients with new-onset Bell's palsy, there is insufficient evidence to support or refute the effectiveness of steroids in improving facial functional outcomes.*

Analogous verbiage is used when studies demonstrate that therapy is ineffective:

- Multiple negative, adequately powered Class I studies:
 - *Are highly likely not to be effective...*
 - *Are highly likely to be ineffective...*
- Multiple negative, adequately powered Class II studies, or a single adequately powered Class I study:
 - *Are likely not effective...*
 - *Are likely ineffective...*
- Multiple negative, adequately powered Class III studies, or a single adequately powered Class II study:
 - *Are possibly not effective...*
 - *Are possibly ineffective...*

DID YOU KNOW?

When formulating evidence-based conclusions, the AAN avoids the term *proven effective* or *established as effective*. Evidence is never definitive, and therefore conclusions derived from evidence cannot be “proven” or definitively “established.”

- Multiple Class IV studies, a single adequately powered Class III study, or negative, inadequately powered Class I, II, or III studies:

- *For patients with new-onset Bell's palsy, there is insufficient evidence to support or refute the effectiveness of steroids in improving facial functional outcomes.*

Please see [Appendix 4](#) for a tool to help in constructing conclusions.

Accounting for Conflicting Evidence

When all of the studies demonstrate the same result, are of the same class, and are consistent with one another, developing the conclusion is a straightforward matter. Often, however, this is not the case. The following provides guidance on how to address inconsistent study results.

Consider a hypothetical example where the search strategy identified one Class I study, one Class II study, and one Class III study on the effectiveness of steroids in Bell's palsy. The Class I study shows a significant and important difference from placebo. The Class II and III studies show no significant or important difference from placebo. What should the development panel do? One approach would be to treat each study like a vote. Because the majority of studies (two thirds) show no benefit, the panel could conclude that steroids have no effect. This vote-counting approach is not acceptable; it ignores the sources of error within each study.

The appropriate approach to take when faced with inconsistent results in the included studies is to attempt to explain the inconsistencies. The inconsistencies can often be explained by systematic or random error.

Considering Bias First: Basing the Conclusion on the Studies with the Lowest Risk of Bias

The developers should consider systematic error first. In this example, the differences in risk of bias among the studies likely explain the inconsistencies in the results. The Class I study has a lower risk of bias than the Class II or Class III studies. Thus, the results of the Class I study are more likely to be closer to the truth. The Class II and III studies should be discounted, and, if possible, the conclusion formulated should be based solely on the Class I study.

The conclusion would be worded:

- *Oral steroids are likely effective for...*

(The “likely effective” conclusion is supported when there is a single Class I study used to formulate the recommendation. If this example is changed slightly to include two or more positive Class I studies, the conclusion would read “highly likely to be effective.”)

Considering Random Error: Are Some Studies Underpowered?

Consider another hypothetical example: that the search strategy identified three Class I studies on the effectiveness of steroids for Bell's palsy. Assume one study showed a significant and important benefit from steroids and two studies did not.

Systematic error does not obviously explain the difference, as all three studies are Class I. Therefore, the developers must consider the random error (statistical precision or power) of the studies by looking at the confidence intervals. If the confidence intervals of all three studies overlap, it is likely that random error (i.e., the lack of statistical power in some of the studies) explains the difference in the studies' results. On the basis of a single adequately powered Class I study, a “likely effective” conclusion would be justified.

Knowing When to Perform a Meta-analysis

Another solution in this circumstance would be to perform a meta-analysis. This increases the statistical precision of the conclusion by combining all of the studies. Meta-analysis is a technique that reduces

random error (but not systematic error). In this circumstance, the combined estimate of the effect of steroids would be used to develop the conclusions. For the purpose of developing conclusions for an AAN guideline, when studies are combined in a meta-analysis to increase statistical precision, the resulting pooled data are treated as though they derived from a single study.

Combining studies in a meta-analysis is often a useful way to reduce random error. However, such a practice can be inappropriate when there are differences in study design, patient populations, or outcome measures.

The strength of the conclusion (“highly likely,” “likely,” or “possibly effective”) would depend on the lowest level of evidence used in the meta-analysis. In this situation, Class I evidence from three studies would support use of the terminology *likely effective*.

Another situation in which a meta-analysis may be applicable is if all three Class I studies in the example were “negative.”

In the case of consistent negative studies, it is still important to look at the potential contribution of random error before formulating a conclusion. In this case, it might be a mistake to conclude that steroids are “highly likely not effective.” If the confidence intervals from the studies were wide—meaning that the confidence intervals included a potentially clinically important benefit of steroids because of a lack of statistical precision in the studies—the individual studies would be inconclusive. Combining the negative studies in a meta-analysis might increase the statistical power sufficiently (i.e., narrow the confidence intervals) so that a clinically important benefit of steroids is excluded. An appropriate negative conclusion could then be made.

If necessary, methodology experts on the committee can help developers perform a meta-analysis.

Considering Both Bias and Random Error

Consider a final example. Here the search strategy identifies five articles looking at the effectiveness of steroids in Bell's palsy. Two studies are Class I, two studies Class II, and one study Class III. The studies are

inconsistent in that the Class III study and Class II studies demonstrate a statistically significant difference, and the Class I studies do not.

The developers should first examine the studies with the lowest risk of bias—the Class I studies—for systematic error. They should next examine these same studies for random error. Although both Class I studies show no benefit of steroids, both studies are underpowered. They have wide confidence intervals that include potentially clinically important benefits of steroids. Combining them in a meta-analysis still shows no significant effect of steroids. However, the combined confidence interval is too wide to exclude a benefit.

Next the developers should examine the Class II studies by performing a meta-analysis that includes both the Class I and Class II studies. The meta-analysis shows a statistically significant benefit of steroids, so the developers can now formulate a conclusion.

The example conclusion used at the beginning of this section would be appropriate for this evidence. Because Class II evidence was used in the conclusion formulation, “likely effective” is used to indicate the level of certainty.

Understanding Reasons for Inconsistencies Aside from Systematic Error and Random Error

Inconsistencies between studies cannot always be explained by a systematic consideration of the level of evidence and random error. Sometimes differences between the study populations, interventions, and outcome measures are sufficient to explain inconsistencies. At times, the inconsistencies cannot be explained. In such instances, it is best to acknowledge that there is insufficient evidence to draw conclusions.

Methodology experts of the subcommittees can guide panel members in this situation.

Wording Conclusions for Nontherapeutic Questions

The examples of conclusion formulation given thus far have related to therapeutic questions. Analogous procedures are followed for questions of diagnostic or prognostic accuracy and for screening questions. The conclusions

are worded slightly differently in that the term *useful* is substituted for *effective*. Thus, the wording below might be used in a conclusion regarding the prognostic accuracy of facial compound motor action potential in identifying patients at increased risk of poor facial function:

- *For patients with new-onset Bell's palsy [population]...*
- *The measurement of facial compound motor action potentials is likely useful [intervention]...*
- *For identifying patients at increased risk for poor facial functional recovery (sensitivity 85 percent, specificity 75 percent) (three Class II studies) [outcome].*

Capturing Issues of Generalizability in the Conclusion

At times, the best evidence relevant to the question posed may be limited by issues of generalizability. In such circumstances, the evidence does not exactly answer the question posed; rather, it answers a relevant, closely related question. Limited generalizability can arise in situations that directly relate to the PICO elements of the posed question.

Population

The population may not be directly representative of the entire population of interest. This can arise when the highest-class studies pertinent to a question include only a subpopulation of patients with the disease. For example, the best studies of Bell's palsy might have been performed on women and not men.

Intervention

Limited generalizability can also result when all relevant studies determined only the efficacy of a narrow range of possible interventions encompassed by the question. For example, if all studies of patients with Bell's palsy were limited to prednisolone at 80 mg per day for three days taken within 24 hours of palsy onset (no other steroid being studied), the generalizability of this evidence to other steroids at different doses and durations is limited.

Co-intervention

Generalizability issues can also arise relative to the comparative intervention used. For example, if the literature search found only studies showing improved outcomes in patients with Bell's palsy receiving steroids compared with patients receiving thiamine (and not placebo), the applicability of this evidence to the question of steroids compared with placebo is limited.

Outcome

Finally, generalizability issues may arise relative to the measurement of the outcome. For instance, a study of steroids in patients with Bell's palsy may have determined outcome only at two months. It would be difficult to generalize this evidence to long-term outcomes.

When the generalizability of the evidence is limited, the conclusion should be worded to indicate the limited applicability of the evidence. Thus, if only high-quality (Class I) studies of patients with Bell's palsy that examined facial functional outcomes at two months in women treated with prednisolone at 80 mg per day for three days taken within 24 hours of palsy onset were compared with studies of women treated with thiamine, the conclusion should *not* read as follows:

- *For patients with Bell's palsy, it is highly likely that steroids (compared with placebo) improve facial functional outcomes (risk difference 12 percent, 95 percent confidence interval 7 to 15 percent, multiple Class I studies).*

Rather, the conclusion should be worded to capture the limited generalizability of the evidence:

- *For women with Bell's palsy, it is highly likely that prednisolone 80 mg daily for three days taken within 24 hours of palsy onset compared with thiamine improves facial functional recovery at two months (risk difference 12 percent, 95 percent confidence interval seven to 15 percent, multiple Class I studies).*

Making Practice Recommendations

The strictly evidence-based conclusions formulated using the rules discussed in the preceding section defines the end of the SR development process. The next step in the process is to develop practice recommendations.

DID YOU KNOW?

Occasionally, after completing the SR, developers will realize that the evidence base is too weak to support any meaningful practice recommendations. In these circumstances, it is appropriate to terminate the practice advisory/guideline development process rather than attempt to develop practice recommendations. The SR itself has value in informing neurologists and patients of the limitations of the evidence. In this circumstance, the AAN will seek to publish only the SR.

The first goal of the process of making a practice recommendation is to develop an actionable recommendation statement that addresses the clinical question. For example, one question with regard to patients with Bell's palsy is whether one should treat them with steroids to increase the likelihood of facial functional recovery. This includes identifying the patient population, intervention, and outcome of interest. (Here the co-intervention—no treatment—is implied.) A recommendation resulting from a review of the effectiveness of treatments for Bell's palsy might read as follows:

- *For patients with new-onset Bell's palsy [population]...*
- *Clinicians should offer oral steroids within the first three days of palsy onset [intervention]...*
- *To improve facial functional outcomes [outcome].*

The second goal is to determine and transparently indicate our confidence that adherence to the recommendation will improve outcomes—that is, the strength of the recommendation. Confidence in the strength of a recommendation in an AAN practice advisory/guideline

typically is indicated by a designation of recommendation strength of Level "A," "B," or "C." Developing actionable recommendations involves much more than a consideration of the quality of evidence on which the recommendation is based. The recommendation development process proceeds in four steps: first, rate the confidence in the ability of the evidence to support practice recommendations; second, develop a rationale to support a recommendation by explicitly considering all factors that could influence the recommendation; third, craft an actionable recommendation; and finally, assign a level for the strength of the recommendation. These steps are described next.

DID YOU KNOW?

In the AAN guideline development process, the term *class* is used to designate the risk of bias, and the term *level* is used to designate the strength of a recommendation.

Developing the Recommendation, Step 1:

Rating Our Overall Confidence in the Evidence from the Perspective of Supporting Practice Recommendations

The implicit assessment of the quality of evidence signaled by the terms possibly or likely measures our confidence that an estimate of the effect of an intervention is correct.⁴ If one's purpose was to develop only an SR, one would stop there. However, when possible, the developer would want to go further by developing actionable recommendations that provide guidance to physicians and patients. To do so requires the developer to take a second, higher-level look at the evidence. In this second look, the developer is not trying to estimate his or her confidence in the accuracy of the evidence as it relates to the effect of an intervention; rather, the developer is determining whether his or her confidence in the evidence is sufficient to support practice

recommendations. The difference is subtle but important. This second determination requires a more nuanced consideration of the evidence. For this purpose, the AAN has adopted a modified version of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) process.⁴

The modified GRADE process for evidence synthesis used by the AAN has several steps.

First, the developer anchors his or her confidence in the evidence to the risk of bias of the best informative studies.

Next, the developer considers factors that may downgrade or upgrade his or her confidence in the evidence.

Finally, the developer estimates his or her confidence in the evidence relative to the outcomes of interest on the basis of the preceding factors. In most circumstances, there will be more than one outcome of interest (e.g., a benefit of the intervention and a harm of the intervention). The process is separate for each outcome. It is possible to have different levels of confidence for different outcomes.

Anchoring the Confidence Level to the Risk of Bias

A level of confidence in the evidence relative to each outcome is then assigned. Four levels of confidence are available. Although it may differ from the final designation of confidence, the confidence in the evidence is initially anchored to the class of evidence through a method identical to that used in the development of evidence-based conclusions:

- High confidence (anchor: at least two Class I studies—corresponds to a "highly likely" conclusion from the SR)
- Moderate confidence (anchor: one Class I study or at least two Class II studies—corresponds to a "likely" conclusion from the SR)
- Low confidence (anchor: one Class II study or at least two Class III studies—corresponds to a "possibly" conclusion from the SR)
- Very low confidence (anchor: only one Class III study—corresponds to an "insufficient" conclusion from the SR)

Considering Factors That Potentially Downgrade the Confidence in the Evidence

Next, explicitly consider factors that could downgrade the confidence in the evidence to support recommendations. These factors include the number and consistency of the informative studies, the statistical precision (or power) of the studies, the directness (or generalizability) of the evidence, the biological plausibility of the intervention (e.g., the treatment efficacy of an intervention such as video-game playing for preventing colon cancer is nonsensical), and the presence of potential reporting bias. Usually, the confidence in the evidence will be downgraded only by one level for these factors even when multiple factors are present. In unusual circumstances, such as severe imprecision or severe issues with generalizability, the confidence in the evidence can be downgraded by more than one level.

Considering Factors That Potentially Upgrade the Confidence in the Evidence

Then, explicitly consider factors that could upgrade the confidence in the evidence. Such factors include a large magnitude of effect (an effect can be so large that it likely overcomes any bias in a study), the presence of a dose response relationship (this lends plausibility to the biological effect of the intervention), the direction of bias (if bias tends toward one direction but the measured effect toward the other, one can be more confident that the observed effect is real). Importantly, these factors can result in upgrading of the confidence in the evidence by only one level.

The confidence in the evidence to support recommendations derived using the modified GRADE process⁴ is one important factor to be considered when developing practice recommendations.

Conclusions Intended for an SR vs. Those Intended for a Practice Advisory/Guideline

It is important to note that the wording of the conclusions described in the “Formulating Evidence-based Conclusions Section” intended only for an SR may differ in one respect from conclusions that are used to

support practice recommendations (practice advisories/guidelines). The modified GRADE process can change the final level of confidence in the evidence from the original anchor. The conclusions in a practice advisory/guideline (in contrast to those in an SR) will be labeled with the modifiers *highly likely* (*highly probable*), *likely* (*probable*), *possibly*, or *insufficient* that correspond to the final levels of confidence in the evidence derived from the GRADE process.

- High confidence—“highly likely (highly probable)”
- Moderate confidence—“likely (probable)”
- Low confidence—“possibly”
- Very low confidence—“insufficient”

Conclusions in a practice advisory/guideline should parenthetically indicate not only the number and class of studies used to support the conclusion but also any upgrading or downgrading resulting from the modified GRADE process. Here is an example:

- For patients with Bell's palsy, it is highly likely that prednisolone 80 mg daily for three days taken within 24 hours of palsy onset compared with thiamine improves facial functional recovery at two months (risk difference 12 percent, 95 percent confidence interval 7 to 15 percent, multiple Class II studies, confidence upgraded for magnitude of effect).

Developing the Recommendation, Step 2:

Constructing the Rationale (the Logic Supporting Recommendations)

Much more than evidence must be considered when crafting practice recommendations. The evidence-based conclusions form the foundation, but other factors influence the structure of the recommendation. The impact of such factors varies among practice advisories/guidelines. Some general guidance is provided below.

Understanding the Role of Deductive Inferences from Accepted Principles

Linking the evidence to a recommendation often requires a logical inference from first principles. It is the combination of the

evidence and the inference that informs the practice recommendation. A formal analytic framework such as a decision tree or causal pathway can help in making these inferences (see section on question development). Such inferences are commonly more extensive when a nontherapeutic question is asked. For example, let us suppose that high-quality evidence rated by the screening classification of evidence scheme indicates that a large proportion of patients with new-onset trigeminal neuralgia has potentially treatable abnormalities identified by head MRI (such as a mass of lesions compressing the trigeminal nerve). This evidence alone does not indicate that patients with trigeminal neuralgia will have better outcomes if they routinely undergo MRI; rather, the evidence simply tells us that a large number of such patients will have treatable abnormalities. However, if one explicitly accepts a principle of care—that identifying treatable conditions is important—then in this example the inference logically follows that clinicians should routinely order head MRIs on patients with new-onset trigeminal neuralgia to identify potentially treatable causes. The axiomatic principle allows one to connect the dots from evidence to recommendation.

It is important to make such “dot-connecting” explicit in the published manuscript so that the process of formulating recommendations is transparent. Consider a second example. Our SR of the evidence with regard to assays of CSF 14-3-3 protein for the diagnosis of Creutzfeldt-Jakob disease (CJD) indicates that the CSF 14-3-3 protein test is 85 percent sensitive and 90 percent specific. Does the test’s moderately high accuracy indicate that the test is necessary for all patients with rapidly progressive dementia? No, there is more to consider. The principle of care just discussed—identifying treatable conditions is important—does not apply, as CJD is not treatable. It would be defensible in this situation to accept another principle of care—that reducing uncertainty even for untreatable conditions is important because it helps patients and families cope better with a devastating illness. It is likely that few clinicians would disagree with this principle, but some might. Explicitly stating the principles used in the formation of the recommendations serves to make the process transparent. If a person disagrees with the recommendation, the reason for the

disagreement will be apparent: the person does not accept this principle as axiomatic.

In the 14-3-3 example, the evidence and the explicit adoption of the principle of care do not in themselves support a recommendation to perform 14-3-3 assays routinely in patients with suspected CJD. Although of moderately high diagnostic accuracy, the 14-3-3 assay is an imperfect test. The test will not importantly change the probability of CJD in patients who are unlikely to have CJD to begin with. For example, a 72-year-old with dementia progressing over 18 months is very unlikely to have CJD, and a positive 14-3-3 test result is most likely to represent a false-positive result. Likewise, the 14-3-3 assay will provide minimal information for patients with a high likelihood of having the disease. For example, a 54-year-old with rapidly progressive dementia over three months with symmetric diffusion-weighted-imaging changes in the basal ganglia is very likely to have CJD. A negative 14-3-3 test result in this situation would most likely represent a false-negative result. These inferences are not derived from evidence as defined in the EBM context; instead, they are inferred from known principles of the course of CJD and Bayes's theorem (an important principle regarding contingent probabilities).

After examining the evidence and making several inferences from multiple explicitly stated principles (and assuming there are no other factors to consider), one might formulate the following recommendation for the 14-3-3 example:

- For patients with rapidly progressive dementia who are strongly suspected of having CJD and for whom diagnostic uncertainty remains, clinicians should order CSF 14-3-3 assays to reduce the uncertainty of the diagnosis.

When crafting recommendations, practice advisory/guideline developers must explicitly enumerate any principle-based inferences that support the recommendations.

Additionally, they must consider the strength of the inference. Not all principle-based inferences are convincing. The development panel and oversight guideline committee will determine how compelling the inferences are using a modified Delphi process.

As previously illustrated, inferences from principles are most often used in conjunction

with evidence to develop recommendations. There are unusual circumstances where compelling inferences alone can support practice recommendations without evidence. Recommendations based on compelling inferences from first principles are not often encountered in a practice advisory/guideline. Practice advisories/guidelines are typically developed for topics for which there is controversy. Compelling inferences from first principles are usually not controversial and thus are not often selected to be the topic of a practice advisory/guideline.

Although rarely needed in a practice advisory/guideline, the use of compelling inferences from first principles without evidence is illustrated by the AAN's practice recommendations regarding the determination of brain death.⁵ That practice guideline determined that there is weak evidence to support the selection of a specific observation time to ensure irreversibility of the cessation of brain function. Because of this, strong recommendations for choosing specific observation times before the declaration of brain death could not be made. Despite the absence of evidence, however, a compelling inference from first principles—in this case the requirement of irreversibility within the definition of brain death itself—supported a strong recommendation that clinicians must choose some observation period before the declaration of brain death to ensure that brain function had not returned. In the practice guideline, the selection of the specific duration of the observation period was left to physician judgment. Similar compelling inferences led to strong recommendations, despite the absence of evidence, that the clinician must know the proximate cause of the brain insult and must exclude confounding circumstances before declaring brain death.

A compelling inference from first principles alone is one circumstance wherein a strong AAN practice advisory/guideline recommendation could be developed in the absence of strong evidence.

The deductive logic justifying each recommendation must be transparently documented in a section, labeled *rationale*, that precedes the recommendation.

This section can be as brief or long as is needed to make the rationale for the recommendation explicit. The rationale will consist of one or more premises and the inferred recommendation.

Four types of premises can be used to support the recommendation:

1. One or more of the evidence-based conclusions from the SR. These are labeled *EVID* in the rationale section.
2. Generally accepted principles of care. These are labeled *PRIN* in the rationale section.
3. Strong evidence from other, related conditions (e.g., in a practice guideline on facioscapulohumeral disease [FSH], other, better-studied, muscle diseases with parallels to FSH). These are labeled *RELA* in the rationale section.
4. Deductive inferences from other premises. These are labeled *INFER* in the rationale section.

Note: *Determining what constitutes strong evidence for the RELA premises requires judgment. Some general guidance can be provided: All RELA statements need to be supported by at least one relevant citation. RELA statements directly supporting a therapeutic recommendation should be based on Class I or Class II studies. (e.g., if the SR found no studies regarding effective therapies for children with headaches following mild traumatic brain injury, it might be appropriate to make a recommendation for the use of a treatment shown to be effective in young adults with headache following mild traumatic brain injury as long as the adult evidence is Class I or Class II). In contrast, a RELA statement being used to support a counseling recommendation need not meet Class I or II criteria (e.g., a study in the psychology literature suggesting that patients understand the magnitude of risk better when the risk data are presented as fractions rather than percentages would not need to be Class I or II in order to support a related counseling recommendation).*

The rationale supporting a recommendation always requires at least one premise but does not necessarily require all types of premises. This concept is illustrated in the following example of the formulation of a hypothetical recommendation that might follow from the evidence regarding respiratory complications in FSH.

Premises:

1. Formulate an evidence-based conclusion from the SR: Patients with FSH who have severe proximal weakness, kyphoscoliosis,

and wheelchair dependence are at high risk for developing respiratory failure within one year.

2. Stipulate an assumption of known principle(s): Patients with respiratory failure from neuromuscular-related weakness often do not have symptoms such as dyspnea that precede the onset of respiratory failure. Impending respiratory failure in these patients often is only identified with pulmonary function tests.

Note: *The developer does not need to provide evidence for this—he or she can simply state that these are established principles.*

3. If needed, bring in strong evidence from related condition(s) that was not systematically reviewed here: Patients with respiratory failure secondary to muscle weakness often have improved quality of life with noninvasive pulmonary ventilation (provide a reference to a seminal article that supports this statement).

A deductive inference from these statements follows (i.e., the recommendation):

- Therefore...

For patients with FSH who have severe proximal weakness, kyphoscoliosis, and wheelchair dependence, neurologists should order pulmonary function tests every six months to identify those with impending respiratory failure from muscle weakness.

The specific part of the manuscript dealing with this recommendation might look like this:

- **Screening for Respiratory Complications**

Rationale: *Our systematic review demonstrates that patients with FSH who have severe proximal weakness, kyphoscoliosis, and wheelchair dependence are at high risk for developing respiratory failure within one year (EVID). Patients with respiratory failure from neuromuscular-related weakness often do not have symptoms such as dyspnea that precede the onset of respiratory failure (PRIN). Impending respiratory failure in these patients often is only identified with pulmonary function tests (PRIN). Patients with respiratory failure secondary to muscle weakness often have improved quality of life with noninvasive pulmonary ventilation (RELA).*

Practice Recommendation: *For patients with FSH who have severe proximal weakness, kyphoscoliosis, and wheelchair dependence, neurologists should order pulmonary function tests every six months to identify those with impending respiratory failure from muscle weakness.*

After reading this, the general neurologist should understand exactly why the recommendation was made. If he or she disagrees with the statement, he or she should be able to state the exact reason on the basis of the premises and reasoning presented in the rationale.

Note: *The premise labels EVID, PRIN, RELA, and INFER may be deleted from the final version of the manuscript body in order to improve readability; however, it is helpful to retain these labels for reference in an appendix displaying the rationale profile tables, which present the variables supporting the rationale. See Appendix 5 for sample rationale profile tables.*

Depending on how the recommendations are organized, sometimes a single rationale can support multiple recommendations. For example, the rationale just presented could support two practice recommendations as follows:

- **Practice Recommendations:**

1. *Neurologists should order pulmonary function tests on patients with FSH who have severe proximal weakness, kyphoscoliosis, and wheelchair dependence. Because of a lack of evidence, a specific monitoring frequency cannot be specified.*
2. *Neurologists should offer noninvasive pulmonary ventilation to patients with FSH and impending respiratory failure, to improve the patients' quality of life.*

Grouping recommendations in this way can make the manuscript more readable and efficiently organized.

PITFALL

At times (some would say usually), estimates of harm from an intervention cannot be made from high-quality evidence such as that provided by RCTs. Sources of evidence such as case reports or registries may be the best evidence available. Such sources of evidence of harm should not be disregarded.

DID YOU KNOW?

In order to determine the effect of a diagnostic test on patient outcomes, one must perform a utility study. Such studies involve comparing patient relevant outcomes in patients who get the test with outcomes of those who do not get the test. The utility of mammography has been tested in this way. Women were randomized either to receive routine mammography or not to receive it. In these studies, outcomes (death secondary to breast cancer) were a little better in the women getting mammography. Utility studies would be rated by the AAN's therapeutic classification of evidence system rather than by the diagnostic accuracy system and could support an actionable recommendation such as one that states "should offer."

Developing the Recommendation, Step 3:

Crafting Actionable Recommendations

AAN practice recommendations must be actionable. The most important part of a recommendation is the verb (action word) used to indicate the action that should be taken. A good verb choice for a recommendation is unambiguous and indicates a specific action that the clinician should perform. Essaihi et al.⁶ have compiled the following list of 11 suggested action verbs for practice advisory/guideline recommendation statements: *test, prescribe, perform, educate/counsel, dispose, monitor, refer/consult, prepare, document, advocate, and diagnose/conclude*. The actions advised in most practice advisory/guideline statements should

correspond to one of these 11 general action types, which should be used consistently. One of these terms (or a variant thereof) usually should be included in every AAN practice advisory/guideline recommendation statement, unless the intent of the statement would be confused by use of such terms.

DID YOU KNOW?

The step of making recommendations in the practice advisory/guideline development process necessarily requires the judgments—or opinions—of the developers. Relying on opinions has a high risk of introducing bias. To minimize this risk, the AAN has instituted the following steps:

1. Enforce a rigorous conflict of interest policy for developers.
2. Obtain consensus from developers, using a modified Delphi process. This process involves anonymous voting, facilitated discussions, group feedback, and statistical analysis of the responses. The technique minimizes biases that can be introduced by group dynamics (e.g., group reinforcing extreme opinions) or dominant personalities.
3. Transparently describe any difference of opinion.

The wording of the recommendation needs to be modified in those circumstances where the evidence indicates that the intervention is not effective or useful. For example, if

multiple adequately powered Class I studies demonstrate that an intervention is not effective, the recommendation could be worded “should not prescribe.” See [Appendix 4](#) for a more in-depth discussion of suggested wording for conclusions and recommendations.

DID YOU KNOW?

The word *consider* should not enter into an AAN practice advisory/guideline recommendation statement. Research has shown that *consider* is confusing to practice advisory/guideline users,⁷ and it is also difficult to quantify whether a person has effectively considered an action.

Table 2 presents a tool for building recommendations using AAN-suggested verbiage. A more detailed tool is presented in [Appendix 4](#).

Basing Recommendations on Surrogate Outcomes

As previously stated, developers are urged to avoid using studies where only surrogate outcomes are measured, as it is often difficult to know the relevance of such outcomes. There are situations, however, where the studies providing strong evidence relevant to a topic measure surrogate outcomes only.

For example, there is controversy with regard to the comparative effectiveness of brand name vs. generic antiepileptic drugs (AEDs).

Suppose that the only strong evidence available compares changes in serum AED levels in patients switched from brand name AEDs to generic AEDs. Serum AED levels are, of course, a surrogate outcome. It is unclear how well they correlate with clinically meaningful outcomes such as seizure control and AED-related side effects. In this situation, the AAN guideline development process permits developers to draw conclusions (and, in the case of practice advisories/guidelines, make recommendations) but only in reference to the surrogate outcome.

The conclusions and recommendations cannot imply an effect on clinically relevant outcomes. For example, assuming multiple Class I studies show the lack of pharmacologic equivalence (within some prespecified serum AED-level threshold), one might conclude the following:

- Different formulations (generic, different nongenerics) of AEDs are highly likely not to be pharmacologically equivalent (multiple Class I studies).

Note that the conclusion discusses only the surrogate outcome (consistent serum AED levels).

Crafting the recommendation becomes problematic:

- For patients with epilepsy, the same formulations of AEDs should be offered to maintain consistent serum levels of the AED (Level A).

This actionable recommendation seems inappropriate because the benefit of stable AED levels is not a clinically important outcome.

Table 2. Elements of Recommendations

Mandatory Elements	Suggested Verbiage		
When (in what circumstances and in what patient population)	(For/In) patients with condition X		
Who (the person performing the action of the recommendation statement)	Clinicians		
Level of obligation (A, B, C)	A: Must (not) prescribe, offer (Rx) Must (not) test, counsel, monitor (Scrn, Dx, Px) Must avoid (causation)	B: Should (not) offer, prescribe Should (not) test, counsel, monitor Should avoid	C: May offer, prescribe May test, counsel, monitor, educate* May avoid May choose not to offer, prescribe May choose not to test, counsel, monitor
What (do what): Intervention (co-intervention): Intervention A (as compared with intervention B)	[Describe specific intervention/test]		
To precipitate what: (outcome)	Outcome Y		
Level of evidence: (Level N)	[State recommendation level in parentheses]		

* In the special case of negative Level C recommendations, we add the word *choose* because the term may not connote a higher level of obligation than is intended. Please see [Appendix 4](#) for additional guidance for constructing recommendations.

The link between consistent levels and meaningful outcomes (seizure control, side effects) should be explicitly evaluated. If a compelling inference for this link cannot be derived from principles or strong evidence, a conditional recommendation can be made:

- *For patients with epilepsy, if consistent serum AED levels are likely to improve seizure control or decrease the risk of toxicity, the same AED formulation should be used (Level A).*

Developing the Recommendation, Step 4:

Assigning Levels of Strength to the Recommendations

When there is sufficient evidence to support an inference for the use of an intervention (i.e., the balance of benefits and harms favors the intervention), the development panel assigns one of three recommendation designations: **A**, **B**, or **C**. Each designation corresponds to a helping verb that denotes the level of strength of the recommendation. Level **A** is the strongest recommendation level and is denoted by the use of the helping verb *must*. *Must* recommendations are rare, as they are based on high confidence in the evidence and require both a high magnitude of benefit and low risk. Level **B** corresponds to the helping verb *should*. *Should* recommendations tend to be more common, as the requirements are less stringent but still based on the evidence and benefit-risk profile. Finally, Level **C** corresponds to the helping verb *may*. *May* recommendations represent the lowest allowable recommendation level the AAN considers useful within the scope of clinical practice and can accommodate the highest degree of practice variation.

Level **A** denotes a practice recommendation that *must* be done. In almost all circumstances, adherence to the recommendation will improve health-related outcomes. Almost all patients in this circumstance would desire that the recommendation be followed. A Level **B** indicates a recommendation that *should* be done. In most circumstances, adherence to the recommendation will likely improve health-related outcomes. Most patients in this circumstance would want the recommendation to be followed. A Level **C** represents a recommendation that *may* be done. In some circumstances, adherence to the recommendation might improve health-related outcomes.

When there is insufficient evidence to support an inference for the use of an intervention (i.e., the balance of benefits and harms is unknown), a Level **U** or Level **R** designation is appropriate. A Level **U** indicates that the available evidence is insufficient to support or refute the efficacy of an intervention. A Level **R** is assigned when the balance of benefits and harms is unknown and the intervention is known to be exorbitantly expensive or have important risks. This level designates that the intervention should not be used outside of a research setting.

Other, non-evidence-based factors that need to be transparently and systematically considered when formulating recommendations include the following:

- The relative value of the benefit compared with the risk; this is derived from consideration of
 - The importance to patients of the health-related outcomes (both benefits and harms)
 - The size of the intervention's effect
 - The risk of harm of the intervention (i.e., tolerability and safety)
- The feasibility of complying with the intervention (e.g., the intervention's availability)
- The cost of the intervention
- The expected variation in patient preferences relative to the risks, burdens, and benefits of the intervention

The following paragraphs explore the ways in which these non-evidence-based factors may influence a recommendation.

Identifying Other Factors Affecting the Recommendation That Potentially Change the Recommendation Level

Although compelling inferences from first principles constitute one circumstance wherein a strong recommendation can be developed in the absence of strong evidence, there are other circumstances where non-evidence-based considerations will affect the strength of the recommendation.

Clinical Importance of the Outcome

At times, high-quality evidence demonstrates a therapeutic effect on an outcome that many would judge is not very important. This scenario may apply to many surrogate outcomes, as discussed previously.

The Risk of Harm of the Intervention and the Relative Value of the Benefit Compared with the Risk

Harm includes matters of both tolerability (an unpleasant side effect that is not dangerous) and safety (a potentially dangerous side effect). Considerations of harm are dominated by matters of safety. Sometimes the evidence that is formally reviewed illuminates the frequency and magnitude of the potential harms of an intervention. When the harms are important (they occur frequently or are dangerous), it is most often useful to highlight them in the wording of the recommendation itself. For example, high-quality evidence indicates that a drug for secondary progressive MS dramatically reduces the risk of subsequent attacks (number of attacks reduced from an average of 2.3 per year to 0.6 per year) and cumulative disability but that the drug rarely (risk 1 in 1,000) causes progressive multifocal encephalopathy (PML), a usually fatal condition. In this situation, it is important that the recommendations describe both the benefit and harm, as shown in the following recommendation:

- *For patients with secondary progressive MS with an attack frequency of ≥ 1 per year despite treatment with other MS therapies, clinicians should offer drug A to reduce MS attack rates and cumulative disability. The clinician must inform the patient of the risk of PML (1 in 1,000) when discussing the potential risks and benefits of treatment (Level A).*

Note that in this example the evidence level is indicated after the sentence describing the benefit and harm. This indicates that the evidence for harm was formally reviewed and rated during SR development.

There are other situations wherein the evidence of an intervention's risk of harm was not part of the formal evidence base. The principle of care "first do no harm" justifies the inclusion of a statement regarding these harms in the recommendation even when the evidence is weak or not formally reviewed. This often happens when a rare but dangerous side effect is discovered during postmarketing surveillance. The evidence of harm may be based on weak evidence such as that provided by a case series or even isolated case reports. It is still important to include these potential harms in the recommendation itself. If we assume that this situation now applies to our example of MS

drug use and PML risk, the recommendation might read as follows:

*- For patients with secondary progressive MS whose attack frequency is ≥ 1 per year despite treatment with other MS therapies, clinicians should offer drug A to reduce MS attack rates and cumulative disability (Level **B**). The clinician should inform the patient of several isolated case reports of PML (exact risk unknown) when discussing the potential risks and benefits of treatment.*

Here the evidence level is parenthetically included only in the first sentence regarding benefit and not in the sentence describing the potential safety concern. This indicates that evidence in regard to harm was not formally assessed in the practice advisory/guideline.

Not only can developers modify a recommendation to ensure that harms are described, but they can also downgrade the recommendation strength when warranted by the relative balance of benefit and risk. In extreme circumstances where the benefit-to-risk ratio is too close to call, the recommendation can be downgraded to the point that no recommendation can be made.

There are sophisticated techniques designed to measure quantitatively the balance and tradeoffs of the risks and benefits of an intervention. These include decision analysis and cost-effectiveness analysis. Generally, such analyses are beyond the scope of a practice advisory/guideline. When concerns about cost arise, however, it is essential to incorporate them in the recommendation development process. The steps for this are described in the cost section that follows.

Variation in Patient Preferences

There may be circumstances where one can anticipate that there will be considerable variations in what patients prefer with regard to adherence to a recommendation. Such variations could result from differences in perceptions regarding the comparative importance of outcomes (e.g., the relative importance of a major gastrointestinal hemorrhage vs. an ischemic stroke) or variations in the tolerance of burdens resulting from adherence to a practice advisory/guideline recommendation (e.g., the burden of frequent laboratory testing for those taking warfarin). Such variations should influence the strength of the recommendation.

Availability and Alternative Interventions

Strong evidence might support the use of an intervention that is difficult to obtain. At a minimum, this should be discussed in the rationale sections of the practice advisory/guideline. There may be times when the strength or wording of the recommendation should be modified to convey the availability concern.

In addition, there may be alternative therapies available for the condition of interest. Ideally, comparative efficacy studies would be available to allow making recommendations pertinent to the relative merits of one drug compared with another. Often such comparative evidence is not available. Moreover, the alternative therapy might not have been studied at all.

For example, assume that amitriptyline is the only drug studied for the treatment of depression in PD and that there is one Class I study showing benefit. Because of amitriptyline's side effect profile, the potential for harm, and the availability of potentially effective alternative therapy, it would be appropriate to craft the recommendation as follows:

*- Clinicians may prescribe amitriptyline for patients with PD and depression in order to reduce depressive symptoms (Level **C**). Before prescribing amitriptyline, clinicians should assess both the patient's ability to tolerate potential anticholinergic side effects and the patient's risk of cardiac dysrhythmias.*

In addition, patients should be informed of the availability of alternative antidepressant therapies that have not been studied in PD but that have potentially better safety and tolerability profiles.

Although the evidence could support a Level **B** recommendation, the safety concerns and availability of alternative therapies have led to a downgrading of the strength of the recommendation to Level **C**. The decision to discuss the side effects and presence of alternative therapies within the recommendation itself (after the designation of evidence level) or within the rationale section needs to be made on a case-by-case basis by the developers on the basis of the potential impact of the concerns. Regardless of what is decided, the decision and its

rationale should be transparently indicated in the methods section of the manuscript.

Cost

Cost is sometimes a factor that should influence the wording or strength of a recommendation. AAN guideline developers have several options for addressing the impact of cost.

At a minimum, guideline authors will judgmentally assess the incremental cost relative to benefit of each recommendation. This is done by asking guideline panel members to explicitly indicate their assessment of cost relative to benefit during the recommendation development consensus process (see [Appendix 5](#), question 2). The recommendation strength may be downgraded if the panel judges that compliance with the recommendation will result in a high cost relative to the benefit.

In addition, in relevant instances, guideline developers should include estimates of the cost of the interventions being recommended in the guideline. Depending on the circumstances, these costs can be briefly mentioned in the rationale supporting the recommendation or presented in a short table.

At times, the cost of following a recommendation is potentially so high that more is needed than a simple enumeration of costs and judgmental assessments of the impact of cost. In these circumstances, the guideline developers can turn to cost-effectiveness analysis (CEA) to help inform their decisions. It is important to remember that most CEAs are performed from the societal perspective, whereas the typical guideline is written from the perspective of individual patients. These different perspectives derive from very different motivations. For example, although the primary purpose of a CEA would be to determine the cost effectiveness of an intervention from a societal perspective, a guideline would not likely recommend against the use of the intervention because it is cost ineffective. Rather, the information regarding cost relative to benefit provided by the CEA would be used by guideline developers to add nuance to their recommendations. For example, because of the excessive cost of a well-studied, highly effective treatment, guideline developers may decide to add a recommendation that clinicians counsel patients regarding the availability of less-studied but more affordable alternative treatments.

When guideline developers choose to look to CEAs to inform their recommendations, it is important that all relevant CEAs be identified and reviewed to lower the risk of introducing bias. Thus, in this circumstance, a systematic review of CEAs will be completed. In addition, in a manner analogous to the assessment of primary studies, CEAs will be rated for their risk of bias through use of the CEA risk of bias classification scheme enumerated in [Appendix 2](#). Finally, under extraordinary circumstances when the guideline panel determines that a formal CEA is critical to their deliberations and none is found during the literature review, the panel may request that the AAN sponsor the completion of a CEA. Completion of a CEA would require input from external experts.

Synthesis of All Factors and Determination of a Recommendation Level

It is evident that numerous factors can influence the wording and rating of a practice recommendation. Keeping track of these varying factors and their relative importance can be difficult. The opinion of the practice advisory/guideline development panel with regard to the importance of these factors is elicited by a questionnaire during the recommendation development process ([Appendix 5](#)). In addition, the AAN uses tables (the rationale profiles) to help determine the strength of recommendations and the inferences supporting them ([Appendix 5](#)). The rows indicate each of the factors to consider in developing recommendations. The columns are labeled to aid the developer in making a judgment with regard to the magnitude or importance of that factor relative to development of the recommendation. The output of the tools is indicated by the recommendation level, shown in the bottom row of the second table.

The first table is designed to determine the overall confidence in the inference (bottom row of the first table: Very Low, Low, Moderate, High). The confidence in the inference derives from the lowest score that results from the questionnaire factors indicated in the rows above (rationale logical, evidence statements accurate, axioms true, related evidence strong and applicable, internal inferences logical). The recommendation level is anchored to the confidence in the inference derived from this table.

The second table reviews factors potentially modifying the recommendation level. Any of these factors can be used to downgrade the recommendation level. For example, if patient values relative to potential benefits or risks of the outcome are judged to be highly variable, it is reasonable to downgrade our confidence that adherence to the recommendation will improve outcomes (because what is desirable varies from patient to patient). Likewise, one would not be confident that attempted adherence to a recommendation to use an intervention that has limited availability would improve outcomes—it might be appropriate to downgrade a recommendation or an intervention with limited availability. Moreover, even though one may be highly confident in the evidence relative to a specific intervention, if the relative value of the benefit vs. risk is low, it may be appropriate to downgrade the recommendation level.

Only one factor—the relative value of benefit vs. risk—can be used to upgrade the recommendation level from that determined by the confidence in the inference. If the relative value of the benefit vs. the risk is judged by the development panel to be large or moderate, a lower recommendation level can be upgraded one level. On the basis of the relative value of benefits vs risks, the recommendation level can be upgraded only by one level and can attain a Level **A** rating only if the confidence in the inference is high.

Counseling Recommendations

It is worthwhile to note that there is not necessarily a direct linkage between the evidence strength and the recommendation level. Rather, the criterion for the EVID statements is that they accurately describe the evidence that was systematically reviewed. This accommodates circumstances where weak evidence supports a strong recommendation, which most commonly occurs in cases of counseling recommendations. With counseling recommendations, clinicians are being advised to inform patients about the current limitation of the evidence. In this circumstance, even weak evidence can support strong counseling recommendations.

Knowing When Not to Make a Recommendation

When there is insufficient evidence to support or refute the effectiveness (or usefulness) of an intervention, no recommendation can be made. In such circumstances, the following statement is made to highlight the lack of evidence:

*- Because of insufficient evidence, no recommendation can be made (Level **U**).*

If the available evidence is insufficient to justify any practice recommendations, the developers can publish an SR rather than a practice advisory/guideline. Highlighting the gaps in evidence in such circumstances becomes particularly important. In the absence of recommendations, the developer relabels the document type as a *practice advisory* or *practice guideline* rather than a *focused systematic review* or *comprehensive systematic review*.

Even when there is high-quality evidence, a recommendation need not necessarily follow. For example, there may be major concerns of generalizability or clinical applicability within the evidence base that would call into question the usefulness of any associated recommendations. In these circumstances, a formal recommendation is not required. A placeholder within the document where the recommendation would normally appear still needs to be present. This placeholder section would briefly explain why a recommendation was not made. In most circumstances, the limitations of the evidence resulting in the absence of a recommendation would be explicated in the published practice advisory/guideline.

Making Suggestions for Future Research

Often after formally reviewing the evidence, the practice advisory/guideline developers are in a unique position to suggest future research to fill in the evidence gaps. The future research section of the practice advisory/guideline is important for identifying areas that were found deficient on the basis of the thorough, systematic literature analysis.

Logistics of the AAN Guideline and Case Definition Development

This section describes the logistics of AAN guideline development. It encompasses such topics as how to propose a guideline topic, how to conduct a literature search and review, and how to format and write an AAN guideline for publication. This section also details the characteristics of the guidelines and the processes involved in guideline development.

The AAN processes for developing guidelines are overseen by the GDDI. The GDDI reports

to the AAN Practice Committee, and GDDI members are appointed to two-year terms by the AAN president. Each GDDI member serves in a particular role on a guideline development panel as determined by his or her particular expertise. In addition to their expertise in at least one representative neurologic subspecialty, GDDI members each have expertise in SR methodology or guideline methodology, or both.

Expert development panels are formed for each project under development in order to critically assess all relevant literature on a given topic or technology. Evidence is rated on the basis of quality of study design (SR), and practice recommendations are developed and stratified to reflect the quality and content of the evidence. Practice advisories/guidelines developed by the GDDI are written with a patient-centric focus or an intervention-centric focus.

Understanding Common Uses of AAN Guidelines and Case Definitions

AAN guidelines (and, in a limited sense, case definitions) have the following uses:

- Improving health outcomes for patients
- Assisting clinicians in staying abreast of the latest in clinical research
- Advocating fair reimbursement
- Determining whether one's practice follows current, best evidence
- Reducing practice variation
- Affirming the role of neurologists in the diagnosis and treatment of neurologic disorders
- Influencing public, payment, or hospital policy
- Promoting efficient use of resources
- Identifying research priorities on the basis of gaps in current literature
- Informing quality measures

The following are the key audiences of an AAN guideline or case definition:

- Primary:** Neurologists
- Secondary:** Patients, patient advocacy organizations, policymakers (e.g., payers, federal agencies), clinical researchers, other health care providers

Distinguishing Between the Types of AAN Evidence-based Documents

The AAN develops several types of evidence-based documents to assist its members in clinical decision making, particularly in situations of controversy or variation in practice. *Appendix 6* depicts the steps of the development process for each of the five AAN evidence-based document types. A table showing the foundational elements of each document type is shown in *Appendix 7*.

The AAN is committed to producing SRs and practice guidelines that are compliant with the 2011 Institute of Medicine's standards (IOM standards).^{1,2} However, SRs and practice guidelines produced in full compliance with IOM standards take considerable resources and time to develop. Recognizing there may be clinical questions that warrant timely evidence-based review and guidance, the AAN approved development of focused SRs

and practice advisories that are partially compliant with the IOM standards. The developers, in consultation with the GDDI and methodologists, may recommend opting out of any or all of the standards outlined below. Any decision made to opt out of these milestones will be documented in the final manuscript for publication.

- Provide a public comment period for the protocol and refine each question on the basis of feedback (practice guideline standard 7)
- Engage a librarian/information specialist to perform the literature search (SR standard 3.1.1)
- Assign an independent librarian or other information specialist to review the search results (SR standard 3.1.3)
- Conduct a "hand search" of the selected journal and conference abstracts (SR standard 3.2.4)
- Conduct a web search (SR standard 3.2.5)
- Search for studies in languages other than English (SR standard 3.2.6)
- Search the "grey" literature databases, clinical trial registries, and other sources of unpublished information about studies (SR standard 3.2.1)
- Invite study researchers and sponsors to clarify information in their studies and to provide unpublished data (SR standard 3.2.3)
- Train screeners with written documentation and test and retest screeners to improve accuracy and consistency (SR standard 3.3.4)
- Provide a public comment period for the report and publicly report on the disposition of comments (SR standard 5.2.2)

Identifying the Five Document Types

Comprehensive SRs (Evidence Reports)

Comprehensive SRs are evidence-based documents that draw conclusions on the basis of a comprehensive analysis of all the available evidence. They include at least three clinical questions. These documents do not include practice recommendations. They often serve as the foundation for practice advisories and practice guidelines; in some cases, however, SRs are published as standalone AAN evidence-based documents. SRs provide neurologists with information about the state of the evidence and often serve as an impetus for researchers to design studies to address the current knowledge gaps. The SRs are developed in full compliance with the 2011 IOM standards.

Focused SRs

The AAN develops focused SRs in circumstances where only one or two

clinical questions are posited. In contrast, comprehensive SRs address three or more clinical questions. Moreover, focused SRs do not include practice recommendations. Focused SRs are developed in partial compliance with the IOM standards.

Practice Guidelines

These document types make actionable practice recommendations based on SRs developed with a methodologic rigor equivalent to or greater than the AAN's SR process. As with SRs, practice guidelines are documents that assess the safety, utility, and effectiveness of new, emerging, or established therapies and technologies in the field of neurology. Contrary to SRs, practice guidelines also address strategies for patient management that assist physicians and patients in clinical decision making, focusing on a series of specific, evidence-based practice recommendations that answer one or more important clinical questions. Practice

guidelines are developed in full compliance with the IOM standards.

Practice Advisories

Based on SRs, practice advisories also make actionable recommendations. The AAN develops these documents to provide guidance in less time than is involved with the full development process. These documents are narrowly focused and typically limited to one or two clinical questions. Practice advisories are developed in partial compliance with the IOM standards.

Case Definitions

Case definitions are documents developed for conditions for which there is no validated reference standard. In these circumstances, evidence cannot adequately define the condition; therefore, these documents are developed using a formal, validated expert consensus approach (e.g., modified Delphi).

Nominating the Topic

Guideline topics may be submitted by any AAN member, AAN committee, or AAN section, or any external organization (e.g., an organization responsible for generating health policy). To submit a topic, individuals may use the Topic Nomination Form available at AAN.com/guidelines/home/development.

Periodically, the AAN Institute Board of Directors will select a broad topic for the development of a set of practice guidelines. Broad topics targeted for practice guideline development in the past include muscular dystrophies, PD, MS, epilepsy, dementia, and headache.

The AAN GDDI evaluates and votes on nominated topics quarterly using a ranking tool known as the Topic Nomination Priority Score (TNPS). The GDDI Chair designates one GDDI member who has content expertise in the area of the nominated topic, and who has no relevant conflicts of interest, to rank the proposed topic (or the Chair may select a topic expert with no relevant conflicts from outside GDDI, if necessary). To do so, the GDDI member or designated expert uses the TNPS tool to rank the position of the topic in the development priority hierarchy (i.e., the degree of impact the topic has for the AAN membership and people with neurologic

conditions). The following criteria are taken into account:

- Relevance to neurologists
- Disease prevalence
- Degree of practice variation or controversy
- Project feasibility (amount of evidence, AAN and GDDI resources, whether collaboration with one or more external societies is required)
- Impact on patient care and outcomes

Not all topics are accepted for development. In addition, the GDDI limits the existing topic list to 25 topics awaiting development at any one time.

Appendix 8 presents the TNPS tool.

Collaborating with Other Societies

After a topic is approved, the GDDI may decide that the project would benefit from the perspective of other, related medical specialty societies. Obtaining this perspective is accomplished in the following ways:

Full collaboration: The development panel reflects equal representation from the collaborating societies. The societies involved sign a formal letter of agreement outlining terms of copyright ownership,

simultaneous publication, division of costs before project initiation, and the method to be used.

Invited participation: AAN staff works with the societies to obtain a nominee.

This individual serves as the official representative from the organization, providing updates to the organization's board of directors. The organization has the opportunity to review and comment on the guideline during public comment periods and to endorse the guideline before publication.

DID YOU KNOW?

The AAN often receives requests for neurologic representation on external societies' SR/practice guideline development panels. In such cases, AAN guideline staff works with the GDDI leadership, guideline methodologists, and the AAN President to formally appoint a representative with the authority to provide neurologic perspective on the AAN's behalf. Such an appointment does not imply endorsement; rather, it affords the AAN an opportunity to learn from other societies' processes for developing SRs and practice guidelines.

Forming the Development Panel

The GDDI assigns committee members to serve on a project facilitation team. A facilitation team member is designated the lead facilitator. The facilitation team helps with reviewing and rating the articles, rating the evidence, and serving as the liaison to the GDDI. In rare cases, the lead developer of the SR or case definition will be the person who submitted the topic. The facilitation team, with the help of AAN staff and the GDDI, assembles a development panel, being careful to seek a variety of perspectives, avoid bias, and avoid financial and intellectual conflicts. The AAN's policy for avoiding conflicts is described in the *Implementing the AAN Conflict of Interest Policy for Guidelines and Case Definitions* section. For SRs and practice guidelines fully compliant with the IOM standards, the development panel should always include a patient advocate or a representative from a patient/consumer organization. At least one current or former patient also should be

included, unless exceptional circumstances make this impracticable (e.g., the neurologic condition of interest causes severe impairment that would make participation impracticable). At a minimum, patient involvement should include participation in clinical question formulation and review of the draft practice guideline.

The panel should be capable of defining the clinical question(s) and performing the technical aspects of the SR development. The panel also should be multidisciplinary in composition, with at least one individual from the following three groups: experts in SR methodology, including risk of bias assessment, study design, and data analysis; librarians or information specialists trained in searching bibliographic medical and scientific databases for published articles, including studies; and clinical content experts to validate the questions and the search results. Clinical content experts will not

review and rate the evidence unless they are past or present members of the GDDI. Other relevant users and stakeholders should be included as feasible. A single member of the development panel can have multiple areas of expertise.^{1,2}

The size of the panel (which includes the facilitation team) will depend on the number and complexity of the question(s) being addressed. For SRs and practice guidelines, the development panel usually numbers between five and 10 individuals. The number of individuals with particular expertise needs to be carefully balanced so that one group of experts is not overly influential.

Often, it is useful to have nationally recognized experts who are familiar with the literature pertaining to the topic being addressed (i.e., have authored clinical publications in high-impact journals). Participants with these credentials increase the credibility of publication.

Implementing the AAN Conflict of Interest Policy for Guidelines and Case Definitions

The AAN is committed to producing independent, critical, and trustworthy guidelines and case definitions. The AAN fulfills this commitment by convening experts that conduct in-depth review and develop conclusions and recommendations based on the best available evidence in a manner that minimizes the influence of industry and other relevant entities. The AAN makes best efforts not to include individuals with conflicts of interest in the development of AAN guidelines but recognizes that this is not always practicable and may preclude necessary thought leaders from participating.⁸ Therefore, management and

disclosure of guideline developer and reviewer relationships are conducted in compliance with the AAN Conflicts of Interest Policy, Principles Governing Academy Relationships with External Sources of Support, and the Council for Medical Specialty Societies Code for Interactions with Companies.⁹ The following procedures implement the relevant policies and outline the process followed through each phase of guideline development and review.

Disclosing Relationships and Determining Relevance

Prospective panel members for AAN guideline projects must disclose all financial and certain nonfinancial relationships with industry (including for-profit entities that develop, produce, market, or distribute drugs, devices, services, or therapies used to diagnose, treat, monitor, manage, or alleviate health conditions), as well as relevant relationships with other entities (including payers, government entities, and not-for-profit organizations) and intellectual biases by completing the AAN's Relationship

Disclosure Form ([Appendix 9](#)) before commencing work on or reviewing an AAN guideline. The form describes the categories or types of relationships to be reported. Members of GDDI, other applicable AAN committees, and members of the Board of Directors who review AAN guidelines are required to make the same disclosures. The term *relationship disclosure* is preferred to *conflict of interest disclosure*, as not all relationships necessarily imply conflict or bias.

All relationships with industry must be disclosed regardless of the perceived relevance to the guideline topic. However, to assist the reviewers, prospective panel members are asked to highlight relationships that they deem to be “relevant” to the guideline’s topic (see the following description of relevance). Regarding relationships with non-industry entities or intellectual biases, only those relationships or potential biases that are relevant to the guideline topic must be disclosed. Intellectual biases may include “academic activities that create the potential for an attachment to a specific point of view that could unduly affect an individual’s judgment about a specific recommendation” (Guyatt et al., 2010, p. 739), examples of which are receipt of a grant or participation in research or article(s) directly related to that recommendation.^{2,8,10} In addition, a strong intellectual conflict would be judged to exist if a potential panel member had a strong preexisting opinion that would not be changed by strong evidence.

The guideline facilitation team, composed of an AAN EBM methodologist, AAN staff, and GDDI leadership, reviews each form before the prospective panel member is officially invited to begin work on the guideline project. The facilitation team reviews the relationship disclosures for any relevant relationships that may constitute a conflict of interest. Relevant relationships may include any of the following:

- A relationship or interest that relates to the same or similar topic, intellectual property or asset, or issue addressed in the guideline
- A relationship of the person or an immediate family member having a reasonable possibility of financial, professional, or other personal gain or loss as a result of the guideline content
- A relationship with an “affected” company within industry, meaning there is a reasonable likelihood of

direct regulatory or commercial impact (positive or negative) on the company as a result of care delivered in accordance with guideline recommendations. Affected companies will generally be identified before commencement of the guideline project by the facilitation team, who will be assisted by the prospective panel members highlighting relationships they deem to be relevant, per the above disclosure process

The guideline facilitation team will consider the relevance of the relationship and the degree of influence when determining whether a conflict of interest exists. Depending on the severity of the conflict, mitigation or management steps may include not inviting the prospective panelist to participate or restricting the panelist’s involvement in the development process (as described in the following paragraphs).

The relevance and severity of an intellectual bias can be difficult to objectively identify and measure. The AAN’s guideline development methodology, including the modified GRADE process for developing conclusions and the modified Delphi process for developing recommendations, has been designed to effectively mitigate intellectual bias as much as possible.

Identifying Relationships Considered Conflicting That Preclude Panel Involvement

Although some relationships may be appropriately managed with the mitigation techniques described in this section, others constitute conflicts of interest incapable of being managed and inconsistent with the AAN’s goal of producing an independent guideline. Relationships that render an individual ineligible to serve on a guideline panel include any of the following:

- Serving on a speakers bureau on behalf of an affected company in industry (this is a compensated role as a presenter for which any of the following circumstances are met: the company has a contractual right to dictate or control the content, the company created the slides/presentation for the speaker, or the presenter is expected to act as the company’s agent or spokesperson for the primary purpose of disseminating company or product information)

- Being employed, or having been employed during the year before panel appointment, by a company in industry
- Holding significant ownership interest (shares greater than \$50,000 in value or an equity interest in a privately held company greater than 5 percent) in an affected company¹¹

In addition, GDDI may choose not to appoint an individual as a lead author or as lead of a section of a guideline if the individual has any of the following relationships to the issues or products being assessed: having any stock or stock ownership, being compensated for expert testimony, being a pioneer or having any substantial direct or indirect compensation or other relationship that GDDI deems as creating a conflict.

Understanding Panel Composition and Responsibilities

The AAN requires that a majority (51 percent) of the members of a development panel be free of conflicts of interest relevant to the subject matter of the practice advisory/guideline. When developing an SR, the AAN requires that the entire panel be free of conflicts of interest relevant to the subject matter of the guideline.

The AAN requires the panel chair (or at least one chair if there are co-chairs) be free of financial conflicts of interest relevant to the subject matter of the guideline, and to remain free of such conflicts for at least one year after the guideline is published.

Panel members must update their Relationship Disclosure Form (see [Appendix 9](#)) at least annually but also promptly at any time a relationship changes. All relationships that existed during the development of the guideline will be disclosed as described in the following paragraphs. Panel members in the majority are encouraged to remain free of conflicts of interest from the time of their appointment through publication to preserve the balance of the panel. However, if an individual’s relationships change during that period such that he or she is no longer eligible to serve in the panel majority, the GDDI leadership will shift the individual to the panel minority. If that is not feasible given the panel composition, the individual must resign from the panel. The nonconflicted chair (or co-chair) of the panel

must remain free of conflicts from the time of his or her appointment through one year after the guideline is published, or relinquish his or her role as chair or co-chair.

The AAN prohibits guideline developers from speaking about the guideline they authored or serving as an expert witness about the guideline on behalf of a company in industry, if that company could be positively or negatively affected by care provided in adherence with the guideline, for a period of one year after the AAN's publication of the guideline.

For guidelines of broad scope, panel members should not all be affiliated with the same institution or study group. If there is a recognized, credible controversy regarding the chosen guideline topic, both perspectives should be represented on the panel.

The GDDI reserves the right to make changes to the author panel composition at any time to ensure balance and avoid bias.

PITFALL

Special care should be taken when inviting nationally recognized experts in the field to serve as guideline developers. These developers may have predetermined ideas about the state of the evidence and the structure of the conclusions and recommendations, and may not agree with the AAN's strict study-rating criteria. In addition, they may have authored many of the studies of interest and be unable to provide an unbiased perspective.

Managing Conflicts: Restrictions for Conflicted Developers and Reviewers

To promote the independence and integrity of the guideline development, review, and approval process, certain restrictions are placed on conflicted panelists and reviewers.

When developing an SR, the AAN requires the entire panel be free of conflicts of interest relevant to the subject matter of the guideline in these critical stages of development.

When developing guideline recommendations, GDDI leadership may require that panel members free of conflicts of interest lead the formulation of recommendations in collaboration with the clinical experts on the panel who may be conflicted, such that the responsibility for the final presentation of evidence summaries and rating the quality of evidence rests with the nonconflicted panel members.^{2,8} Because the panel majority is free of conflicts of interest, the entire panel may vote on the guideline recommendations.

AAN guidelines will be reviewed and approved only by committee and board members who do not have a conflict of interest, as determined by the Reviewing Authority in accordance with the AAN's Conflicts of Interest Policy at tools.aan.com/apps/disclosures/index.cfm?event=committee:intro (with consideration of the elements measured when determining conflicts of interest for guidelines, as described here).

Disclosing Conflicts at Publication

The AAN's Conflicts of Interest Policy and this Implementing the AAN Conflict of Interest Policy for Guidelines and Case Definitions section of the manual will be cited in the published guideline, along with the relevant relationship disclosures of the authors. In addition, to promote further transparency, a link to the full list of disclosed relationships will be included.

DID YOU KNOW?

Relationships deemed relevant to the guideline topic are disclosed within the publication. The *Neurology* journal publishes a link to the complete list of author relationships in conjunction with the publication.

Identifying Violations of Conflict of Interest Policy for Guidelines

An AAN guideline developer's or reviewer's failure to accurately, honestly, and fully complete the Relationship Disclosure Form or adhere to the responsibilities described in this Implementing the AAN Conflict of Interest Policy for Guidelines and Case Definitions section of the manual may face sanctions by the AAN, including any or all of the following:

- Exclusion from developing future AAN guidelines or case definitions
- Exclusion or removal from participation on AAN boards, committees, subcommittees, work groups, task forces, guideline or quality measurement panels, or other AAN positions
- Disciplinary action under the AAN's Disciplinary Action Policy at [AAN.com/membership/professionalism-and-disciplinary-program/](https://aan.com/membership/professionalism-and-disciplinary-program/)

DID YOU KNOW?

Developers of AAN guidelines and case definitions are prohibited from speaking about the guideline they authored or serving as an expert witness about the guideline on behalf of a company in industry, if that company could be positively or negatively affected by care provided in adherence with the guideline, for a period of one year after the AAN's publication of the guideline.

Undertaking Authorship

All participating panel members, including the facilitator, are listed as authors. The lead developer and facilitator determine the order of authorship and arbitrate any questions regarding who qualifies for authorship. The journal has strict guidelines regarding who should and should not be considered an author. At the time of journal submission, all development panel members will be required to complete a form affirming their contribution to the project as involving either study design/conceptualization, data/statistical analyses, or writing/revising of the manuscript. Developers whose work does not fit within any of these categories may not be recognized as authors but may be acknowledged as contributors in the acknowledgments section of the publication.

DID YOU KNOW?

All AAN guideline and case definition developers perform the work of guideline or case definition authorship on behalf of the AAN. Therefore, the AAN is the sole owner of the rights to the guideline or case definition. Developers are required to transfer copyright to the AAN before work begins.

Understanding Roles and Responsibilities

Lead Facilitator

A GDDI member is assigned to guide the project and advises on process matters—

particularly the classification of evidence and translation of evidence to practice recommendations. This person reports to the GDDI quarterly on project progress and may serve as the lead developer of the guideline or case definition. The AAN requires the panel chair (or at least one chair if there are co-chairs) be free of conflicts of interest relevant to the subject matter of the guideline, and to remain free of such conflicts of interest for at least one year after the guideline is published.

Facilitators

Other GDDI members are assigned to the facilitation team to assist with the project.

Lead Developer (If Different from the Facilitator)

The lead developer works with the facilitator to set the timeline, assign tasks to panel members, and coordinate activities (e.g., literature review and drafting of the guideline or case definition).

Development Panel Member

Each development panel member is an active participant in the project who usually reviews articles, classifies evidence, and writes portions of the guideline or case definition.

Patient Representative

The patient representative is considered a member of the development panel. Patient representatives participate in formation of the clinical questions; review of the

draft protocol, comprehensive SR, and practice guideline; and consensus voting on recommendations (optional). Patient representatives have the option to participate in additional aspects of development, as approved by the lead author, including literature review and manuscript drafting.

Note: *Developers of focused SRs, practice advisories, and case definitions have the option of including patient representatives on the development panel.*

EBM Methodologist

This person provides methodologic and statistical guidance throughout the project, including assisting in forming clinical questions, developing data extraction forms, training developers on the AAN classification of evidence schemes, and adjudicating discrepancies in the rating of articles.

AAN Staff

AAN staff members provide administrative support and advice, facilitate meetings and group communications, provide manuscript management and copyediting (including for styles and standards), coordinate resource allocation (e.g., medical librarian), and coordinate the journal approval and publication process. When a new development panel is formed, AAN staff distributes roles and responsibilities documents to the panel members (see [Appendix 10](#)).

Completing the Project Protocol

A project protocol is an objective document that details how the development panel will implement the guideline development process. The following information is included in the draft protocol (see [Appendix 11](#) for a template):

- Development panel, designations/credentials, affiliations, and disclosures
- Justification for development
- Analytic framework used to help elicit the essential clinical questions

- Clinical questions structured in PICO format (described in the second section of this manual)
- Terms and databases to be used in the literature search
- Inclusion and exclusion criteria for article selection
- Proposed project timeline

The development panel submits to the GDDI a protocol draft of the SR/practice guideline or case definition, and AAN staff makes it

publicly available on the AAN website for comment. Note that the GDDI may decide to post protocols for focused SRs and practice advisories on a per-case basis. After development panel members receive input from multiple reviews, they modify and finalize the protocol. They then develop a table of comments for the peer review and a list of corresponding changes (or the reasoning for changes not made).

Developing Clinical Questions

The AAN seeks to develop focused, answerable clinical questions for guidelines and case definitions. Focused questions make the project more manageable and lead to conclusions and recommendations that are more pertinent to clinical care. Developers should select questions that can be answered on the basis of published, peer-reviewed evidence but also should realize that AAN staff will make every effort to identify additional, relevant data in the grey literature.

Each clinical question should address characteristics of the patients and interventions that are believed to affect outcome significantly. Taking too narrow a focus may unnecessarily limit the amount of evidence available for review. Conversely, development panels that take too broad a focus or ask too many questions risk becoming overwhelmed with too much evidence, thus encumbering the process.

Remember, guidelines and case definitions are not textbooks or comprehensive summaries about how to diagnose and manage particular diseases; rather, they are analyses of the published literature pertinent to specific aspects of care.

TIP

It may be helpful to perform a preliminary literature search to determine the availability of evidence to answer the questions being considered and to become familiar with the breadth of literature available on the topic.

Developing the Search Strategy: Selecting the Search Terms and Databases

Search Terms

The developers should preliminarily identify the search terms that will ensure articles are obtained that can best answer the clinical questions. Developers should be sure to include appropriate synonyms from other nationalities, ethnicities, and disciplines, as well as variant spellings of terms (for terms that previously had or have variant spellings).

TIP

Developers should not be concerned with identifying all search terms at this stage. During the literature search process, the contracted medical librarian will suggest refinements and seek clarification of terms to ensure that the most comprehensive search is performed. It is essential for the content experts on the panel to identify (a priori) a few key, relevant articles to ensure that these are identified by the search.

Databases

The developers should identify the databases to be searched. A MEDLINE search will likely uncover only 30 percent to 80 percent of published RCTs on a topic.¹² Therefore, it is recommended that developers search multiple databases, including MEDLINE, EMBASE, the Cochrane Database of Systematic Reviews, and Science Citation Index or Current Contents. In consultation with a professional medical librarian, the development panel should determine on the basis of the topic being investigated whether it is appropriate to search other, additional databases. Some databases to consider are Bioethicsline, Cumulative Index to Nursing and Allied Health Literature (CINAHL), International Pharmaceutical Abstracts (IPA), Health Services Technology Assessment Texts (HSTAT), Psychological Abstracts, and BIOSIS. No literature search for a guideline or case definition should be limited to only one database; a minimum of two databases searched is required in the AAN process. *Appendix 7* outlines the elements of each type of evidence-based document, including the required number of databases to be searched.

Selecting Inclusion and Exclusion Criteria

The development panel should develop criteria for including or excluding articles during the literature search and article review processes.

The criteria must be developed before search initiation. However, they may be revised as necessary (e.g., if too few or too many studies are identified) as literature search results are

obtained, provided that care is taken to avoid making changes that would introduce bias. If modifications are made, they should be reflected in the protocol as necessary.

The development panel should develop an explicit list of inclusion and exclusion criteria by evaluating each of the issues described next and any other issues that are pertinent to the specific topic being addressed. The GDDI facilitation team can provide valuable assistance in completing this portion of the protocol.

Languages

Developers are encouraged to include all languages in the search, unless there is a specific reason for excluding non-English-language articles. Abstracts are available in English for many non-English-language articles. It is usually possible to obtain a translation of a non-English-language article through a university, the Internet, or AAN staff.

Relevance

The type of study participants, interventions, and outcomes specified in the search strategy must be relevant to the clinical question.

Type of Study Participants

Usually, the search is limited to articles concerned with human study participants. However, for some topics it may be appropriate to include experimental articles from the laboratory. Nonetheless, it is important to remember that animal research does not inform conclusions or recommendations.

Intervention

The type of intervention should be made explicit in the search strategy.

Outcome Measures

Outcome measures that will be examined should be included. Developers should consider whether the timing of follow-up for the outcome should be specified.

Types of Studies

The types of studies to be included in the search should be stipulated. If there is a large literature base, it may be appropriate to limit the search to RCTs (Class I) and controlled clinical trials (Class II). If the literature base is small, case-control studies—and possibly retrospective case

series—may be included. Developers should use methodologic selection criteria only if doing so will result in obtaining articles that are clearly superior in quality.

Setting the Project Timeline

AAN staff members use dates provided in the protocol to develop an official project

timeline that takes into account upcoming committee meeting dates and the availability of resources.

Performing the Literature Search

After the GDDI approves the protocol, the development panel should conduct the literature search.

Consulting a Research Librarian

A medical librarian contracted by the AAN will develop and perform a comprehensive literature search based on the information given in the protocol regarding search terms and databases. The librarian will interactively query the database to define and refine the search as necessary. The guideline or case definition lead developer and facilitator perform a quick review of the results on the basis of a preliminary search strategy to ensure that key articles thought to be pertinent to the search are identified. When the search strategy is finalized through this iterative process, the strategy is sent to an independent research librarian for peer review.

Documenting the Literature Search

The literature search results are kept on file at the AAN. The following data are captured:

- Date(s) searches were conducted
- Search terms/strategy used
- Database(s) searched
- Date ranges included in search
- Explicit description of the inclusion and exclusion criteria

Documenting this information ensures the methods presented in the manuscript are transparent and reproducible, which is essential for producing a minimally biased document. The entire search strategy for each question is published in the *Neurology* journal as an appendix accompanying each guideline and case definition.

Ensuring the Completeness of the Literature Search: Identifying Additional Articles

Upon receipt of the search results, the lead developer and facilitator should critically evaluate the completeness of the search.

Developers should do the following:

- Ensure no essential concepts related to the question were missed
- Ask panel members to identify additional relevant articles (published or in press)
- Identify additional articles from reference lists, particularly the reference lists of review articles and meta-analyses

Using Data from Existing Traditional Reviews, SRs, and Meta-analyses

Review articles can be categorized as traditional reviews, SRs, and meta-analyses. Traditional reviews include publications such as book chapters, editorials, and expert reviews. SRs follow a rigorous methodology to address focused questions, apply explicit eligibility criteria, conduct exhaustive literature searches, and critically appraise the evidence. Meta-analyses consist of an SR plus statistical pooling of the results into a single summary measure, such as an odds ratio, relative risk, or risk difference. In addition, SRs or meta-analyses may be embedded in such studies as economic evaluations, decision analyses, practice advisories/guidelines, and case definitions.

SRs and meta-analyses are of particular importance in the development processes of AAN practice advisories, practice guidelines, and case definitions. SRs and meta-analyses contain many of the elements required for these three types of guidance documents (e.g., literature search, study selection, critical appraisal, and summary of results). Therefore,

it is tempting to accept the study results at face value. However, there are several important disadvantages to this approach. Often, small but important differences can be identified in the specific question(s) addressed, the literature search, the definitions of clinical conditions and interventions, the thresholds for assessing outcomes, and the dates of the literature review. Furthermore, the evidence-rating systems of other organizations usually differ from the AAN's rating systems, and studies may not be described in sufficient detail to be rated according to AAN classification of evidence criteria.

Because of these disadvantages, usually traditional reviews, SRs, and meta-analyses discovered during the literature search process will be used as follows:

- SRs on the topic can be acknowledged in the clinical context section of the manuscript. This is encouraged when well-known SRs have conclusions that contradict the conclusions and recommendations of AAN guidelines and case definitions.
- The references cited in SRs should be independently assessed for eligibility and then critically appraised and rated. (The reference lists of the selected SRs are compared with the results received from the literature review. Discrepancies are identified by AAN staff and sent to the lead developer for input.)
- Results and summary results (meta-analyses) of SRs should not be used in drafting recommendations.
- Results of individual studies as described within published SRs should not be used at face value when recommendations are drafted.
- Differences in the results obtained by existing SRs and those of AAN practice advisories/guidelines and case definitions should be acknowledged and explained in the text of the document.

Although not usually the case, at times SRs previously published elsewhere may address the same specific questions of a planned

AAN practice advisory/guideline or case definition, and the methodologic quality of the review may be substantially equivalent to that followed for an AAN SRs. Development panels of AAN practice advisory/guidelines and case definitions should assess the previously published SRs formally for methodologic quality using an assessment tool such as the Revised Assessment of Multiple Systematic Reviews.¹³ In these circumstances, after receiving GDDI approval, the development panel can use the published SR as the basis of an AAN practice advisory/guideline or case definition.

Minimizing Reporting Bias: Searching for Non-peer-reviewed Literature

Often it is tempting to exclude non-peer-reviewed sources of evidence such as supplements, book chapters, and studies that are unpublished or are not included in bibliographic retrieval systems (so-called grey literature). Substantial empirical evidence demonstrates that excluding such evidence sources introduces bias. One major reason for this is that negative studies (i.e., studies not showing an effect of an intervention) are less likely to be published in peer-reviewed journals. These non-peer-reviewed sources provide important evidence that is not available in the peer-reviewed sources. Thus, every effort should be made

to assess this evidence to determine whether critical studies are being missed.

Reporting bias, including publication bias, presents a fundamental obstacle to the scientific integrity of guidelines and case definitions. To minimize the effect of reporting bias, the AAN endorses a literature search process that includes not only easy-to-access bibliographic databases but also other information sources that contain grey literature, particularly trial data and other unpublished reports. The search should be comprehensive and include both published and unpublished research. Additionally, panel members of guidelines and case definitions are encouraged to contact authors of primary studies under review, to clarify unclear reports or to obtain unpublished data that are relevant.

Selecting Articles

A two-step process is used to exclude articles that do not meet the inclusion criteria. All abstracts identified through the literature search are reviewed for relevance to the clinical questions and adherence to the inclusion criteria. The same process is applied to the selected articles.

Obtaining and Reviewing Titles and Abstracts

AAN staff distributes the abstracts and tracks panel member responses. Every abstract should be reviewed by at least two panel members. The lead developer may choose to have two or more selected panel members review all abstracts or to have the abstracts distributed evenly among all panel members. When the number of developers reviewing the abstracts has been determined, AAN staff will use its document review database to assign the abstracts systematically to the developers, to ensure that each abstract is reviewed by two individuals working independently of each other.

Panel members review the abstracts and determine which are pertinent to the clinical questions and meet the inclusion criteria. It is best to be inclusive at this stage of the process. If it is unclear whether an article is relevant or meets the inclusion criteria, it should be obtained for full-text review. If either reviewer indicates that an abstract is relevant, the associated article will be

included for the full-text review. AAN staff will document the number of abstracts reviewed, the number excluded, and the reason(s) for exclusion.

Tracking the Article Selection Process

To ensure transparency of the development process for guidelines and case definitions, AAN staff will track the disposition of every article identified by the search strategy. The tracking should explicitly identify the reason for the exclusion of studies. AAN staff will maintain a record of excluded study citations and the reason(s) for exclusion. After article selection, a flow diagram depicting the disposition of articles will be constructed.

Obtaining and Reviewing Articles

After all abstracts have been reviewed, AAN staff works with the developers to obtain and distribute the selected articles. Each article should be read by two panel members working independently of each other. The panel chair may choose to distribute the articles at random, by topic area, or by another method.

Panel members should review each article for pertinence to the clinical questions and adherence to the inclusion criteria set forth in the protocol. Typically, this is a screening

review of the article, and data are not yet extracted; however, on occasion developers will extract data during article screening. It is best to be exclusive at this stage in the process. If it is unclear whether an article meets the inclusion criteria, it is appropriate to seek clarification through discussion with other panel members or by contacting the author of the study. However, if the panel members choose to do the latter, they must contact all authors of studies for which they have similar questions, to avoid introducing bias.

If the panel members cannot agree on inclusion of a study, the study should be sent to an independent reviewer for adjudication. The adjudicator can be any of the following individuals deemed nonconflicted: the lead developer, any facilitator, a GDDI member, or the AAN EBM methodologist.

Panel members send AAN staff a list of articles to be included in the guideline or case definition. AAN staff works with the lead developer to compile a library of articles to be included and resolves any disagreements regarding inclusion of individual articles.

Extracting Study Characteristics

The study characteristics—or elements—to be extracted from each article vary depending on the clinical question of interest. In general, the characteristics extracted will correspond to one of the following categories:

- Citation information (depending on the software used, AAN staff will prepopulate this information)
- Items relevant to the study generalizability
- Elements relevant to the quality of evidence presented in the study
- Elements relevant to the study outcomes

TIP

Data extraction and evidence classification are crucial tasks. Many of the concepts discussed in this section are often unfamiliar to panel members who lack a methodologic background. Panel members should seek the assistance of the facilitator in completing these steps, as necessary.

Developing a Data Extraction Form

The extraction of the study characteristics just described can be facilitated by development of a data extraction form. The AAN EBM methodologist, in conjunction with the lead developer and the GDDI lead facilitator, develops a data extraction form to apply to each clinical question. Sample data extraction forms are provided in [Appendix 12](#). It may be helpful for the facilitator or a GDDI member to convene a conference call with all panel members to provide instruction before the start of data extraction.

Data from each article should be extracted by at least two panel members working independently of each other. Panel members complete these forms electronically, which are automatically submitted to AAN staff. Disagreement regarding the extracted elements, classification of evidence, or assessment of effect size should be resolved by consensus among panel members. If consensus cannot be obtained, the GDDI lead facilitator and a methodologist can arbitrate.

Constructing the Evidence Tables

Evidence tables are developed from the data extraction forms. Each table row corresponds to an included study. Each table column corresponds to extracted characteristics of an included study. It is essential to include the class of evidence determined for each study. A sample evidence table can be found in [Appendix 3](#).

TIP

Tables are created in an electronic spreadsheet for easy manipulation and review of the data. Complete evidence tables are required for each manuscript draft submission to the GDDI.

Constructing Evidence Synthesis Tables

The developers document the modified GRADE process⁴ using an evidence synthesis table (see [Appendix 13](#) for an example). This table serves as a useful outline for the “Analysis of Evidence” section of the document.

Drafting the Document

The development panel should translate the content of the evidence tables into manuscript form following the format provided in [Appendix 14](#). Developers should use the following structural flow as described in the Preface:

Clinical question → Evidence → Conclusions → Recommendations*

*A recommendation section will be created only for a practice advisory/guideline.

Getting Ready to Write

Before developers begin writing the document, they should review [Appendix 14](#) in its entirety, as well as the “Information for Authors” at neurology.org. In addition to undergoing review by AAN committees and the AAN Institute Board of Directors, the manuscript will be evaluated by both the AAN guideline staff and *Neurology*

journal staff. It is essential to understand the expectations of each. The journal editorial policy limits the length of a guideline/case definition print publication to a maximum of 3,500 words of text, 250 words of abstract, and 40 references (subsequent references are published as electronic references, or e-references, on the journal website); however, developers of AAN guidelines and case definitions should focus primarily on adhering to the development requirements for these documents regardless of word count while being mindful of the need for succinctness in summarizing the evidence. With final approval of the developers, AAN staff can help write the shorter version of the guideline/case definition that will be published in both the online and print editions of the journal.

Usually, the lead developer assigns specific topics to each development panel member; panel members develop the first draft of their assigned

sections. The panel chair then integrates all of the sections into a cohesive document.

DID YOU KNOW?

Many AAN Guideline and case definition manuscripts exceed the *Neurology* journal length limit for print. In these cases, an accompanying summary version must be drafted. The summary manuscript will present salient information from the full-length version and will be published both online and in print as the main article. The full-length manuscript version will be published online only, as a data supplement to the summary article. In the full-length manuscript, references are presented as “e-references.”

Formatting the Manuscript

The development panel should follow the structure provided in the manuscript format outlined here. AAN staff members with writing/editing expertise are available to assist in organizing the document, including populating standard text, numbering and formatting the references, and writing the abstract.

Drafts should be double-spaced and paginated, with text presented in Times New Roman 12-point font and line numbers included. Each draft should be labeled with the date and step in the process, as noted in [Appendix 14](#).

TIP

It is critical to be as transparent as possible in describing the process followed or results obtained in the development of the guideline or case definition. A long version of the final document will be published online, and a shorter version will be published in both online and print editions of the journal.

Essential Elements

Cover and Front Matter Pages

The cover and front matter pages of the manuscript should include: the title; author names, degrees (MD, PhD, etc.), and institutional affiliations; correspondence address; abstract word count; title character count (with spaces); manuscript word count; endorsements (if any) from external organizations; date of draft; a statement disclosing study funding source(s); the relevant relationship disclosures of the authors with a hyperlink to the full list of disclosed relationships (AAN staff can assist in writing conflict of interest disclosures, as the forms are kept on file); and abbreviations of terms used in the manuscript.

Abstract

The abstract, although the first part of the manuscript presented, should be written last. The abstract is a brief, 250-word summary of the paper, highlighting the important points and findings. It is extremely difficult to write a 250-word summary of a manuscript not yet written. AAN staff is available to assist in drafting the abstract.

TIP

No information should be presented in the abstract that is not found in the manuscript, and all important points from the manuscript should be mentioned in the abstract. The abstract often is the only part of the article that physicians read, some of whom are reading it to determine whether to read the entire article.

The abstract should contain four sections: objective, methods, results (conclusions), and recommendations. These are described as follows:

Objective: A brief statement (preferably limited to a phrase or sentence in length) of the purpose of the guideline or case definition (e.g., to perform an evidence-based review of the safety and efficacy of botulinum neurotoxin in the treatment of adult and childhood spasticity). The objective usually derives from the clinical question(s).

Methods: A one- or two-sentence statement regarding the literature search strategy (including databases and years searched, if space permits) and the method of evidence classification. If space allows, the AAN development process version should be indicated.

Results: Information from the conclusions sections of the manuscript.

Recommendations: A summary of the recommendations in the paper and their levels. Not all recommendations need to be presented. If there are many recommendations, it may be best to present only those with the strongest levels of evidence. Note that recommendations are not incorporated into the SR document.

Introduction

The introduction should be brief (preferably not more than one or two pages). It should include background on the topic (including prevalence, where applicable) and a brief description of gaps and controversies (i.e., a justification for this publication), and should end with a statement of the clinical questions to be examined in the rest of the manuscript. Explicitly state any assumed principles of care.

Description of the Analytic Process (Methods)

The description of the analytic process describes the exact process that the panel used to create the document. It is important that the description be detailed enough to be transparent and replicable. This section should describe panel formation (usually a brief sentence stating that the AAN convened an expert panel made up of neurologists and any other types of specialists, if applicable), the literature review dates and databases searched, the secondary search strategy (usually examining the references of review articles), inclusion/exclusion criteria used, how articles were reviewed, the classification of evidence schemes used, the process by which developers resolved disagreements in classification, and any modifications to the schemes employed that were specific to this question. (Note that the complete search strategy will be presented in an online appendix.) This section should also describe the outcomes of interest, the measure of effect preferred, and the measure of statistical precision used, and should identify what was considered a clinically important effect. Moreover, it is important to include a brief description of the process for developing the conclusions and modified Delphi approach to crafting recommendations.

Analysis of Evidence

This section is best organized by clinical question. Each clinical question is listed as a subheading under which the relevant evidence for that question is presented. Each subsection (clinical question) should provide the number of citations retrieved for that question at each stage of review: first abstracts, then full-text articles, and then the articles selected for incorporation into the paper. Describe the evidence, briefly justify the evidence rating, and provide appropriate quantitative measures of effect size including measures of statistical precision (e.g., 95 percent confidence intervals) where feasible and meta-analyses results, if appropriate.

Conclusions (Evidence Synthesis)

Conclusions are a key component of the evidence synthesis. Conclusions are succinct statements that synthesize and summarize the evidence to answer the clinical question(s). At this step, multiple articles pertinent to a question have been analyzed and summarized in an evidence table. Ideally, this summary statement will indicate the magnitude of the

effect and the class of evidence on which it is based. The conclusion should be formatted in a way that clearly links it to the clinical question.

A conclusions section should follow each clinical question subsection in the analysis of evidence section. GDDI and the supervising committees will make the final determinations regarding the appropriateness of the conclusions as written.

For each conclusion, mention the number of supporting studies and the class of evidence and statistical precision of those studies. If the confidence in the evidence was changed during the modified GRADE process,⁴ indicate the reason for the change. An example is, drug A is probably useful to reduce the symptoms of disease X (two adequately precise Class III studies, confidence upgraded for magnitude of effect).

Putting the Evidence into a Clinical Context

This is an optional section following the conclusions designed to highlight important clinical issues not necessarily discussed in the formal evidence review. Topics which may be discussed in these sections include special populations and comorbidities/multimorbidities deserving special attention, noteworthy limitations not discussed in the evidence review, remaining controversies, and current common practice patterns.

Care must be exercised regarding the wording of this section in order to avoid the inclusion of any commentary that could be construed as recommendations based not on the evidence but rather on prevailing practice or opinion. To prevent this potential undermining of the careful, rigorous process used to develop AAN Guidelines and case definitions, the following process should be followed:

- First, consider whether the point to be made would be most appropriately addressed in the introduction rather than in a separate clinical context section.
- Leave the evidence-based recommendations unchanged.
- Include a description of the clinical context issue in paragraph form. Include critical issues only. No new recommendations can be made in the clinical context section.

If clinical context includes discussion of commonly used therapies or procedures excluded from the guideline/case definition

because of lack of evidence, such therapies or procedures should be identified not as “standard of care” but rather simply as “common practice” and must include a relevant reference citation.

Practice Recommendations (Included Only in Practice Advisories/ Guidelines)

Recommendations are presented as a separate section after all of the evidence for all questions has been presented. Each recommendation is preceded by a rationale.

Rationale: A description of the logic chain supporting the recommendation, including the factors that influenced each recommendation/subset of recommendations, should be summarized in a section preceding the respective recommendation(s). The primary purpose is to explain the rationale for the formulation of the specific recommendation.

The information in the rationale may include alternatives for which there was limited evidence, risk-benefit profiles, limits to the generalizability of the evidence, magnitude of benefit, harms, cost, and outcomes.

Recommendations: The recommendations should flow from the rationale and should use suggested AAN recommendation language (see *Appendix 4* for suggested language). For each recommendation, a strength-of-recommendation label (e.g., Level **A**) must be included. Recommendations should be written to support patient-centered outcomes* and should include a statement of harm, if appropriate.**

*Avoid wording recommendations as such: “Therapy X should be prescribed by clinicians.” Instead, restructure the wording as such: “Clinicians should prescribe therapy X in order to reduce spasticity in children and adolescents with cerebral palsy.” The latter example presents a patient-centered outcome, which will aid clinicians in applying the practice advisories/guidelines in their practice.

**A recommendation should include a statement of harm especially when there are important or severe side effects, defined as those that may be life-threatening, are common and affect safety or quality of life, or are covered by a US Food and Drug Administration black box warning.

Suggestions for Future Research

The completion of the SR and analysis of the literature position the development panel favorably to recommend areas of future research. The future research section should present a summary of study design concerns that were found to be limitations in the existing literature, such as the need for multicenter studies, adequate sample sizes,

randomized studies, and more comprehensive or reliable outcomes measures. This section should also address the need for more studies on therapies for which evidence was deemed inadequate or conflicting.

Disclaimer

This is a stock language statement provided by AAN staff (see *Appendix 14*).

Acknowledgments

The acknowledgments section is optional and is reserved for those who assisted in manuscript development but who do not qualify as authors under the *Neurology* journal authorship policy. People who are frequently acknowledged are research assistants, editors, AAN staff, or reviewers who made significant comments.

DID YOU KNOW?

The *Neurology* journal requirements for word count in AAN guidelines and case definitions include only the body of the paper, from the introduction through the suggestions for future research. Word count does not include the text of appendices, stock AAN language, references, or tables or figures.

Tables/Figures

In general, tables and figures may be published in print or online as necessary, although the *Neurology* journal limits the number of each that may appear in the print article but does not limit the number that may be published online as data supplements.

Two types of tables, evidence tables and evidence synthesis tables, are only published online as data supplements. Although both of these table types provide information about the evidence, they serve different functions. Evidence tables present all the study data needed to understand the assigned evidence ratings. Evidence synthesis tables are provided to show the detailed results of the modified GRADE process that yields the final conclusion levels. In this latter table type, the developers present key data such as study classification, factors assessed during the voting process (e.g., precision), and the initial and final degrees of confidence in the evidence.

Both of these table types are presented as part of the manuscript at each stage of review.

Appendices

The appendices include the following:

- GDDI mission statement and member roster (as constituted at the time of GDDI approval of the manuscript)

- Complete search strategy
- Classification of evidence scheme
- Rules for determining the confidence in the evidence
- Evidence synthesis tables
- Steps and rules for formulating the recommendations
- Rationale profiles presenting the recommendations and their supporting rationale

The methodologist provides the evidence synthesis tables and recommendation tools; AAN staff provides the remaining content. Appendices are published online only.

References

Because of journal space requirements, the number of print references is limited to 40. The remaining references will be published online as e-references, the process for which AAN staff will coordinate.

Reviewing and Approving Guidelines and Case Definitions

Stages of Review

AAN staff and the GDDI will review the guideline/case definition at several stages during the development process. These stages are outlined below:

Stage	Reviewer
General topic	GDDI
Development panel* composition	GDDI leadership and AAN staff
Protocol	GDDI, AAN staff, public
Evidence table draft	GDDI, AAN staff
Guideline/case definition draft	GDDI, AAN staff, public
Guideline/case definition draft post-public comment†	GDDI, AAN staff, <i>Neurology</i> peer reviewers
GDDI-approved guideline/case definition	AAN staff, Practice Committee, <i>Neurology</i> peer reviewers, AAN Institute Board of Directors

* The development panel includes the facilitator.

† Public comment is optional for focused SRs, practice advisories, and case definitions. When developers opt not to submit the document for public comment, the document will undergo an external review phase to obtain input from targeted stakeholders.

These levels of review are described in more detail next.

TIP

- Use “people-first” language. For example, say, “patients with dementia” rather than “demented patients.”
- The word *data* is plural (as in “data are,” not “data is”).
- When referring to the class of a study, make sure to use Roman numerals (Class I, II, III) to avoid potential confusion from use of multiple numeric values, as in “2 Class 2 studies.”
- Always capitalize the word *class* when referring to classification of a specific study (e.g., “Class I”) and *level* when referring to the level of a specific recommendation (e.g., “Level A”).

AAN Staff-level Review

All initial draft documents—including the protocol and those for the evidence tables, SR, practice advisory/guideline, and case definition—are first reviewed by AAN staff and the AAN EBM methodologist. These reviews ensure that drafts submitted to the GDDI meet AAN requirements for methodologic quality and formatting. Often, this step involves AAN staff queries to the developer.

The most common revision requests (both for AAN staff-level review and for GDDI review) pertain to the following:

- Poorly constructed clinical questions
- Incorrect classification of the evidence
- Missing articles

- Significant deviations from the established format
- Incorrect translation of the evidence to conclusions
- Incorrect translation of the conclusions to recommendations
- Manuscript too long

Author Response to Reviews

At each phase of review, the development panel should revise the document as appropriate and populate a revision table that lists each reviewer, the reviewer comment, and how the comment was addressed in the document (see example in [Appendix 15](#)). Note that at the discretion of the lead developer and GDDI facilitator, comments that are identical or very similar may be included as a single comment to which the developers respond. The revision table must be submitted to the GDDI with each manuscript draft. The table will also accompany the document when it is sent to the Practice Committee, the *Neurology* journal, and the AAN Institute Board of Directors.

Developers are encouraged to use electronic word-processing formatting tools (underline and strikethrough, changes tracker) for this draft and for subsequent drafts for which the changes are minor.

The revised manuscript and revision table are then submitted to AAN staff.

Initial GDDI review

After review and approval from AAN staff and the AAN EBM methodologist, draft documents are submitted to the GDDI for review at one of its quarterly in-person meetings. The GDDI reviews the documents

carefully and often requests revisions. AAN staff and the facilitator compile the list of requested revisions in a revision table (see [Appendix 15](#)), and developers are asked to respond to all comments and revise the documents accordingly before the next GDDI meeting. The typical timeframe for manuscript revision is six to eight weeks.

Public Comment

When the GDDI grants initial approval of the draft protocol, SR, or practice guideline (or, when applicable, the focused SR/practice advisory), AAN staff posts it for review and comment on [AAN.com](#) for 30 days. The documents are shared publicly because the AAN realizes that SR and practice guideline development groups are limited to a small number of individuals for expediency and efficiency in the development process. The AAN will not limit review to a predefined external reviewing group; rather, any individual will be able to access the document on [AAN.com](#) for review and comment.

The AAN realizes that, although the document may be publicly available at [AAN.com](#) for 30 days, organizations and interested stakeholders may not always be aware of its availability for comment. Thus, the AAN will make best efforts to identify individuals and groups within the following segments who may have an interest in the document, notify them of the document's availability and encourage them to comment:

- AAN membership
- Members of AAN sections
- Members of AAN committees, subcommittees, task forces, and work groups
- Domestic and international subject matter experts
- Other physician organizations
- Patient advocacy organizations

AAN staff collects the responses and forwards them to the facilitator and lead developer. The responses are presented in a revision table (see [Appendix 15](#)), and developers are required to respond to all reviewer comments. The developers decide whether to make changes to the manuscript on the basis of reviewer comments; however, the developers must adequately defend this decision in the revision table.

GDDI Re-review (Post-public Comment)

AAN staff sends the GDDI the revised documents and revision table (reflecting input from public comment, as applicable) for review and a vote at the next GDDI meeting. GDDI approval may be contingent on additional requested revisions.

Journal Review

The *Neurology* journal solicits reviewers from its network to review and comment on the manuscript. Comments are sent directly to the lead developer and AAN staff. The lead developer drafts a revision letter presenting all comments from *Neurology* peer reviewers. Developers are encouraged to consider all revisions suggested by the journal peer reviewers. Developers are to notify the facilitator if the reviewers' requested changes conflict with AAN requirements for guidelines or case definitions, particularly if reviewers request substantial revisions to the wording of conclusions or recommendations. The lead developer then submits the revised draft to AAN staff (not directly to the journal) with the completed revision letter denoting the panel's responses to all of the journal reviewers' comments. The revised draft must show all changes made to the manuscript, using an electronic editing tool (e.g., changes tracker, strikethrough font, or highlighted font). AAN staff then submits the manuscript to GDDI.

Endorsement

It may be appropriate to seek guideline endorsement from relevant external organizations. AAN staff obtains from the developer the names of organizations to approach for endorsement. The desired organizations typically are identified at the public comment stage, and outreach takes place at the point of first submission to the *Neurology* journal.

GDDI Re-review and Approval

AAN staff sends the revised documents, the revision table (reflecting input from public comment), and the *Neurology* peer review comments for a review and an official vote at the next quarterly GDDI meeting.

Practice Committee Review and Approval

When the GDDI gives final approval of

the manuscript, AAN staff submits the manuscript to the Practice Committee for approval. The Practice Committee may have additional revision requests, and if these revisions are substantial, the changes are reviewed by the GDDI Chair. Substantial revisions—particularly those that change the conclusions and recommendations—may require GDDI reapproval.

In some instances, the guideline developers, GDDI Subcommittee members, and Practice Committee members may disagree substantially with requested changes received from *Neurology* peer review that cannot be resolved with manuscript revisions. In cases of disagreement, the AAN EBM methodologist, GDDI Chair, and *Neurology* Editor-in-Chief may convene a meeting to discuss whether the disagreement warrants publication of a report on the pertinent area(s) of controversy. If these individuals determine such a report is needed, the development panel will generate a discussion section for inclusion in the final publication to highlight the point of disagreement. The *Neurology* journal may choose to write a separate editorial or companion document for simultaneous publication that articulates how the areas of controversy related to the guideline or case definition affect the field.

Journal Re-review

After the Practice Committee approves, AAN staff edits the manuscript and resubmits it to the journal for a second review and subsequent approval.

The journal may request additional rounds of reviews before accepting the manuscript for publication.

AAN Institute Board of Directors Approval

When the manuscript has been accepted for publication in the *Neurology* journal, AAN staff submits it to the AAN Institute Board of Directors for approval.

Requests for revision by the Board during the approval process are reviewed by the GDDI Chair. Substantive revisions may require reapproval by the GDDI, Practice Committee, and *Neurology* journal.

Taking Next Steps (Beyond Publication)

Undertaking Dissemination

At a minimum, the following steps are taken to promote an AAN guideline or case definition:

- Published in *Neurology* journal
- Posted on the AAN website
- Announced by email to all AAN members or a subset of members (e.g., AAN Neuromuscular Section)
- Announced in *AANnews*[®] and *AANe-news*[®]
- Submitted to guideline compendia such as the National Guidelines Clearinghouse

The GDDI, AAN guideline staff, or AAN communications staff may undertake additional dissemination and implementation efforts. These may include strategic outreach to clinicians, patients, and the public. AAN communications staff may launch a media publicity campaign, including tactics such as issuing a press release. The GDDI and AAN guideline staff may develop tools for clinical audiences, including a slide presentation, summary of the guideline or case definition for clinicians, and algorithms, to help members incorporate in practice the conclusions or recommendations. Tools for patients also may be developed, such as a summary of the guideline or case definition for patients and their families or caregivers.

Responding to Correspondence

Because AAN staff members coordinate the journal submission and publication process, they receive any related letters to the editor. For any letters received, developers and facilitators should work together to draft a response letter. The response letter is reviewed internally by AAN staff before its submission to the journal. For correspondence that addresses the development process, the GDDI leadership will also review the response.

TIP

Developers should not be discouraged if they receive a negative letter to the editor about their publication. The AAN views such correspondence as opportunities to educate *Neurology* journal readership on EBM principles.

Updating Guidelines and Case Definitions

Published guidelines and case definitions can become out of date.¹⁴ Therefore, the AAN approved the system described next for evaluating guidelines and case definitions to ensure that those that are out of date are identified and updated in a timely manner.

Triennial Review: Updating the Literature Search and Assessing Methodologic Soundness

Guidelines and case definitions are assessed every three years to determine whether new literature has been published that would warrant an update. The following steps are taken:

1. The facilitator and all other developers are notified of triennial review and asked if they are aware of any relevant new evidence.
2. A GDDI member, assisted by AAN staff, performs a literature search update and reviews the obtained studies for applicability to the clinical questions. The search should specifically seek to identify new evidence that would change the guideline conclusions or recommendations.
3. The GDDI member and a methodologist review the guideline for methodologic soundness.
4. The GDDI reviews the results of steps 1 through 3 and determines which of the three following actions should be taken:
 - **Reaffirm:** If methodology is still sound, and either (1) there is no new evidence or (2) there is new evidence but it would not change conclusions or recommendations
 - **Update:** If new evidence would change conclusions or recommendations and an update

is warranted

- **Retire:** If new evidence would change conclusions or recommendations but an update is not warranted

If an update is deemed warranted, GDDI forms a new development panel, which may include members of the initial development panel. The project then follows the same process as outlined in this manual, with two modifications. First, in most circumstances, the new search should cover just the evidence since the time of the last literature search. Second, development panels are strongly urged to rerate all the articles reviewed for the previous guideline/case definition version, using any changes to the classification of evidence schemes.

On occasion, the GDDI will decide not to revise a document in need of updating. In these circumstances, the document will be retired. The GDDI has the authority to retire a document without the prior authorization of the AAN Institute Board of Directors.

Decisions regarding the update status will be communicated to the AAN membership through the AAN website. All documents triennially reviewed by the GDDI that do not require an update are reaffirmed. Documents that require updating will be designated as such on AAN.com, including the status and date of the update action.

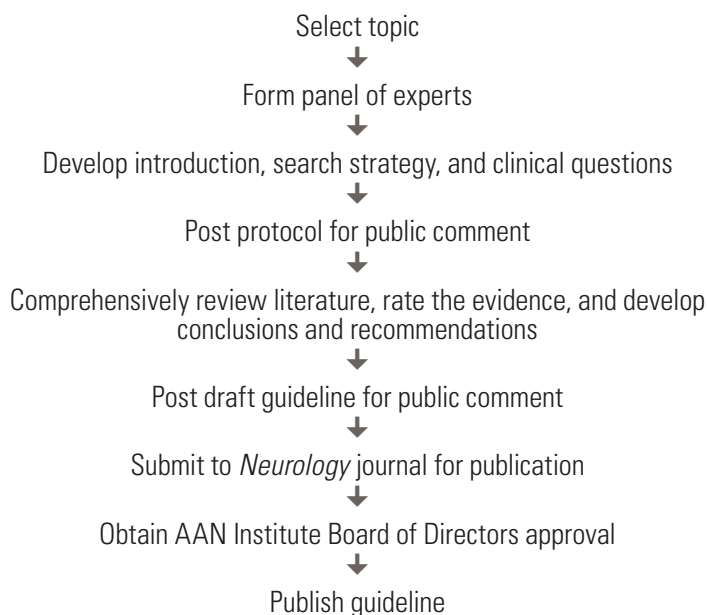
Any guidelines or case definitions that have not been either updated or reaffirmed by five years after the previous publication or reaffirmation will be retired automatically.

Figure 7 summarizes the AAN guideline/case definition update process.

DID YOU KNOW?

When updating an AAN guideline or case definition, the GDDI may choose not to invite one or more members of the previous development panel. The reasons for this may include presence of author conflict of interest; revised, more limited scope for the update, which may require fewer developers; or the decision to have the GDDI develop the update on its own. The GDDI will notify former developers that an update to their previously authored document is under way and will give the developers an opportunity to provide input on the scope of the update.

Figure 7. Steps in AAN Guideline Development



Appendices

Appendix 1: The American Academy of Neurology Institute

The American Academy of Neurology Institute, of which the American Academy of Neurology

is the sole voting member, is a 501(c)(3) tax-exempt nonprofit organization with its primary

purpose centered on education (including practice guidelines), science, and research.

Appendix 2: Classification of Evidence Schemes

Criteria for Rating Therapeutic Studies

Class I Criteria

- Randomized controlled clinical trial (RCT) in a representative population
- Triple-masked studies (i.e. the patient, treating provider, and outcome assessors are unaware of treatment assignment)
 - Relevant baseline characteristics of treatment groups (or treatment order groups for crossover trials) are presented and substantially equivalent between treatment groups, or there is appropriate statistical adjustment for differences
- Additional Class I criteria:
 - a. Concealed allocation
 - b. No more than two primary outcomes specified
 - c. Exclusion and inclusion criteria clearly defined
 - d. Adequate accounting of dropouts (with at least 80 percent of participants completing the study) and crossovers
 - e. For noninferiority or equivalence trials claiming to prove efficacy for one or both drugs, the following are also required*:
 - i. The authors explicitly state the clinically meaningful difference to be excluded by defining the threshold for equivalence or noninferiority
 - ii. The standard treatment used in the study is substantially similar to that used in previous studies establishing efficacy of the standard treatment (e.g., for a drug, the mode of administration, dose, and dosage adjustments are similar to those previously shown to be effective)
 - iii. The inclusion and exclusion criteria for participant selection and the outcomes of participants on the

standard treatment are comparable with those of previous studies establishing efficacy of the standard treatment

- iv. The interpretation of the study results is based on a per-protocol analysis that accounts for dropouts or crossovers
- v. For crossover trials, both period and carryover effects are examined and statistical adjustments performed, if appropriate

Class II Criteria

- RCT that lacks one or two Class I criteria a–e (see above)
- Cohort studies employing methods that successfully match treatment groups on relevant baseline characteristics (e.g., propensity score matching) meeting Class I criteria b–e (see above)
- Randomized crossover trial missing one of the following two criteria:
 - a. Period and carryover effects described
 - b. Baseline characteristics of treatment order groups presented
- All relevant baseline characteristics are presented and substantially equivalent across treatment groups (or treatment order groups for crossover trials), or there is appropriate statistical adjustment for differences
- Masked or objective** outcome assessment

Class III Criteria

- Controlled studies (including studies with external controls such as well-defined natural history controls)
- Crossover trial missing both of the following two criteria:
 - a. Period and carryover effects
 - b. Presentation of baseline characteristics

- A description of major confounding differences between treatment groups that could affect outcome**
- Outcome assessment performed by someone who is not a member of the treatment team

Class IV Criteria

- Studies not meeting Class I, II, or III criteria

*Numbers i–iii in Class Ie are required for Class II in equivalence trials. If any one of the three is missing, the class is automatically downgraded to Class III

**Objective outcome measurement: an outcome measure that is unlikely to be affected by an observer's (patient, treating physician, investigator) expectation or bias (e.g., blood tests, administrative outcome data)

Criteria for Rating Cost-effectiveness Studies

Class I Criteria

- Conflicts of interest of authors described
- Less than 50% of authors have relevant conflicts
- Patient relevant outcomes used (not just surrogate measures)
- At least 10 of the following criteria fulfilled:
 - a. The source of funding was described
 - b. The methods employed were clearly stated as predefined
 - c. The primary outcome was clearly defined
 - d. Model assumptions were clearly stated
 - e. If the cost analysis examined the superiority of one treatment over another, the basis papers included a head-to-head comparison of the treatments of interest
 - f. If a comparator was described, it was based on a representative modern treatment for the disease of interest

- g. The authors reported the method used to derive utility weights
- h. The variable estimates were based on trials meeting AAN Class I or II criteria
- i. The time horizon was sufficiently long for all relevant and important outcomes
- j. The health outcomes and/or measures were valid and reliable
- k. If there was data uncertainty, a sensitivity analysis was performed
- l. If there was uncertainty in the parameters, then in addition to a sensitivity analysis, the authors included parameter uncertainty (i.e., a stochastic model) in the model

Class II Criteria

- Conflicts of interest of authors described
- Less than 50% of authors have relevant conflicts
- Patient relevant outcomes used (not just surrogate measures)
- At least eight of the criteria from Class Ia–II fulfilled

Class III Criteria

- Conflicts of interest of authors described
- More than 50% of authors have relevant conflicts
- Patient relevant outcomes used (not just surrogate measures)
- At least six of the criteria from Class Ia–II fulfilled

Class IV Criteria

- Studies not meeting Class I, II, or III criteria

Criteria for Rating Causation Studies

Class I Criteria

- Cohort survey with prospective data collection
- All relevant confounding characteristics are presented and substantially equivalent between comparison groups, or there is appropriate statistical adjustment for differences
- Outcome measurement is objective or determined without knowledge of risk factor status
- Additional Class I criteria:
 - a. No more than two primary outcomes

specified

- b. Exclusion and inclusion criteria clearly defined
- c. Accounting of dropouts indicated (with at least 80 percent of participants completing the study)

Class II Criteria

- Cohort study with retrospective data collection or case-control study. Study meets Class I criteria a–c (see above)
- All relevant confounding characteristics are presented and substantially equivalent across comparison groups, or there is appropriate statistical adjustment for differences
- Masked or objective outcome assessment

Class III Criteria

- Cohort or case-control study designs
- A description of major confounding differences between risk groups that could affect outcome**
- Outcome assessment masked, objective,** or performed by someone other than the investigator who measured the risk factor

Class IV Criteria

- Studies not meeting Class I, II, or III criteria

**Objective outcome measurement: an outcome measure that is unlikely to be affected by an observer's (patient, treating physician, investigator) expectation or bias (e.g., blood tests, administrative outcome data)

Criteria for Rating Prognostic Accuracy Studies

Class I Criteria

- Cohort survey with prospective data collection
- Inclusion of a broad spectrum of persons at risk for developing the outcome
- Outcome measurement is objective or determined without knowledge of risk factor status
- Additional Class I criteria:
 - a. Exclusion/inclusion criteria clearly defined
 - b. Both the risk factor and the outcome measured in at least 80 percent of participants

Class II Criteria

- Cohort study with retrospective data collection or case-control study. Study meets Class I criteria a and b (see above)
- Inclusion of a broad spectrum of persons with and persons without both the risk factor and the outcome
- Presence of the risk factor and outcome are determined objectively, or the outcome is determined without knowledge of the risk factor and the presence of the risk factor is determined without knowledge of the outcome

Class III Criteria

- Cohort or case-control study
- Narrow spectrum of persons with or without the disease
- Presence of the risk factor and outcome are determined objectively, or the outcome is determined without knowledge of the risk factor and the presence of the risk factor is determined without knowledge of the outcome

Class IV Criteria

- Studies not meeting Class I, II, or III criteria

Criteria for Rating Diagnostic Accuracy Studies

Class I Criteria

- Cohort survey with prospective data collection
- Inclusion of a broad spectrum of persons suspected of having the disease
- Disease status determination is objective or made without knowledge of diagnostic test result
- Additional Class I criteria:
 - a. Exclusion/inclusion criteria clearly defined
 - b. Both the diagnostic test and disease status measured in at least 80 percent of participants

Class II Criteria

- Cohort study with retrospective data collection or case-control study. Study meets criteria a and b (see Class I)
- Includes a broad spectrum of persons with and persons without the disease

- The diagnostic test result and disease status are determined objectively or without knowledge of one another

Class III Criteria

- Cohort or case-control study
- Narrow spectrum of persons with or without the disease
- The diagnostic test result and disease status are determined objectively, or the outcome is determined without knowledge of the risk factor and the presence of the risk factor is determined without knowledge of the outcome

Class IV Criteria

- Studies not meeting Class I, II, or III criteria

Criteria for Rating Population Screening Studies

Class I Criteria

- Study of a cohort of patients at risk for the outcome from a defined geographic area (i.e., population based)

- Outcome is objective
- Additional Class I criteria:
 - a. Exclusion and inclusion criteria clearly defined
 - b. At least 80 percent of patients undergo the screening of interest

Class II Criteria

- Non-population-based nonclinical cohort (e.g., mailing list, volunteer panel) or a general medical/neurology clinic/center without a specialized interest in the outcome. Study meets Class I criteria a and b (see above)
- Outcome is objective

Class III Criteria

- Referral cohort from a center with a potential specialized interest in the outcome

Class IV Criteria

- Studies not meeting Class I, II, or III criteria

Appendix 3: Sample Evidence Table

Design Characteristics and Outcomes in Class I and Class II Controlled Studies of Patients with Bell's Palsy Treated with Antiviral Agents and Steroids Relative to Patients Treated with Steroids Alone

Author, y	Cohort size	Age, y	Steroid dose duration Rx	Severity, % ^a	Duration, d ^b	Follow-up, mo	Completion rate, % ^c	Blind	Class	NH % ^d	RD good recovery (CI)	RD complete recovery (CI)
Engström 2008 ⁸	422	Median 39 (IQR 23–54)	Prednisolone ^e 60 mg daily X 5, taper	Med HB 4 IQR 3-5	3	12	99	Yes	I	56	—	15% (8%–21%)
Sullivan 2007 ¹¹	551	Mean 44 (16.4 SD)	Prednisolone 25 mg BID	Mean HB 3.6 ± 1.3	3	9	90	Yes	I	82	—	12.8% (7.2%–18.6%)
Lagalla 2002 ¹⁵	58	Range 15–84	Prednisone 1 g IV X 3 d then 0.5 g IV X 3 d	24	3	12	100	Yes	II	75	7% (-14% to 27%)	—
May 1976 ¹⁶	51	53% >30	Prednisone 410 mg 10 d	47	2	6	100	Yes	II ^f	81	-0.75% (-18% to 22.5%)	—
Taverner 1954 ¹⁷	26	Mean 40 (range 12–76)	Hydrocortisone 1 g 8 d	23	9	NS	100	Yes	II ^f	67	5.25% (-27% to 55%)	—

Abbreviations: CI = confidence interval; HB = House Brackman score; IQR = interquartile range; NH = natural history; NS = not stated; RD = risk difference (positive values results favoring steroids).

^aPercentage of patients with complete palsy.

^bMaximum duration of palsy before steroids started.

^cPercentage of subjects followed to study completion.

^dPercentage of patients not treated with steroids who attained a good outcome.

^ePrednisolone and prednisone are dose-equivalent steroids.

^fDowngraded by one Class from the rating in the original practice parameter because of no description of allocation concealment

Note: Table 1 published in: Gronseth GS, Paduga R. Evidence-based guideline update: Steroids and antivirals for Bell palsy. Report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology* 2012;70:2209–2013.¹⁵ Used with permission.

Appendix 4: Tools for Building Conclusions and Recommendations

The following tools are provided to assist in the development of conclusions and recommendations. Keep in mind that all mandatory elements must be included (or obviously implied) in each conclusion and recommendation statement. The exact wording and order of the elements can vary from those suggested for grammatical and stylistic considerations. See the examples below for the wording of conclusions when there is insufficient evidence. Note that these examples are hypothetical.

Conclusions

Elements of Conclusions

Mandatory Elements	Suggested Verbiage
Patient population	<i>For patients with condition X, it is</i>
Strength of evidence (pick one) - Strong - Moderately strong - Weak - Insufficient	<ul style="list-style-type: none"> - <i>highly likely (highly probable) that</i> - <i>likely (probable) that</i> - <i>possible that</i> - <i>insufficient evidence to support or refute that</i>
Intervention (co-intervention)	<i>Intervention A (compared with intervention B)</i>
Effect	<p>(For therapy/causation)</p> <ul style="list-style-type: none"> - <i>is (not) effective in reducing the risk of</i> - <i>(does not) increase(s) the risk of</i> <p>(For prognosis/diagnosis/screening)</p> <ul style="list-style-type: none"> - <i>is (not) useful (predictive) in identifying</i> - <i>patients at increased risk for</i> - <i>patients with</i> - <i>a (treatable important) cause of</i>
Outcome	Outcome Y (if possible, include a magnitude of effect)
Evidence summary	(Indicate number of studies and their classifications)

Examples

Therapeutic

For patients with Bell’s palsy [condition], it is highly likely that prednisolone [intervention A] (compared with placebo [intervention B]) is effective for reducing the risk [effect] for incomplete facial functional recovery [outcome]—risk reduction 12 percent (two Class I studies) [evidence summary].

For patients with Bell’s palsy [condition], it is highly likely that antivirals [intervention A] (compared with placebo [intervention B]) are not effective for reducing the risk [effect] for incomplete facial functional recovery [outcome] (two Class I studies) [evidence summary].

Causation

For persons at risk of developing multiple sclerosis (MS) [condition], it is possible that low serum vitamin D levels [intervention] increase the risk [effect] for the development of MS [outcome]—odds ratio 1.23 (two Class III studies) [evidence summary].

For young children [condition], it is likely that immunizations [intervention] do not increase the risk [effect] for autism [outcome] (multiple Class II studies) [evidence summary].

Diagnostic Accuracy

For patients with rapidly progressing dementia [condition], it is likely that CSF 14-3-3 assays [intervention] are useful for identifying [effect] patients with prion disease [outcome]—sensitivity 80 percent, specificity 85 percent (multiple Class II studies) [evidence summary].

For patients with symptoms and signs suggesting carpal tunnel syndrome [condition], it is highly likely that the flick sign [intervention] is not useful for identifying [effect] patients with carpal tunnel syndrome [outcome]—sensitivity 80 percent, specificity 20 percent (multiple Class I studies) [evidence summary].

Prognostic Accuracy

For patients with cryptogenic ischemic stroke [condition], it is possible that the presence of patent foramen ovale (PFO) [intervention] is useful for identifying [effect] patients at increased risk for recurrent ischemic stroke

[outcome] (two Class III studies) [evidence summary].

For patients with ischemic stroke [condition], it is highly likely that elevated serum homocysteine levels [intervention] identify [effect] patients at increased risk for recurrent stroke [outcome]—relative risk 1.6 (two Class I studies) [evidence summary].

Population Screening

For children with global developmental delay (GDD) [condition], it is possible that routine MRI of the head [intervention] is useful for identifying [effect] a cause of the GDD [outcome]—yield 4.5 percent (multiple Class III studies) [evidence summary].

For patients meeting International Headache Society (IHS) criteria for migraine and a normal neurologic examination [condition], it is likely that routine head imaging (MRI or CT) [intervention] is not useful for identifying [effect] important abnormalities [outcome]—yield 0.5 percent (single Class I study) [evidence summary].

Insufficient Evidence

For patients with Alzheimer disease [condition], there is insufficient evidence to support or refute the effectiveness of coenzyme Q [intervention] for slowing [effect] cognitive decline [outcome] (Class IV studies only) [evidence summary].

For patients with post–cardiac arrest brain anoxia [condition], there is insufficient evidence to support or refute the usefulness of visual evoked potentials [intervention] for identifying [effect] patients at low risk of recovery [outcome] (inadequately powered Class II study) [evidence summary].

Recommendations

Elements of Recommendations

Mandatory Elements	Suggested Verbiage		
When (in what circumstances and in what patient population)	(For/In) patients with condition X		
Who (the person performing the action of the recommendation statement)	Clinicians		
Level of obligation (A, B, C)	A: <ul style="list-style-type: none"> - Must (not) prescribe, offer (Rx) - Must (not) test, counsel, monitor (Scrn, Dx, Px) - Must avoid (causation) 	B: <ul style="list-style-type: none"> - Should (not) offer, prescribe - Should (not) test, counsel, monitor - Should avoid 	C: <ul style="list-style-type: none"> - May offer, prescribe - May test, counsel, monitor, educate* - May avoid - May choose not to offer, prescribe - May choose not to test, counsel, monitor
What (do what): Intervention (co-intervention): Intervention A (as compared with intervention B)	Describe specific intervention/test		
To precipitate what: (outcome)	Outcome Y		
Level of evidence: (Level Z)	(Level Z)		

* In the special case of negative Level **C** recommendations, we add the word choose because the term may not connote a higher level of obligation than is intended.

Note: Levels **U** and **R** are excluded here because they do not denote a level of clinical obligation.

Examples

In addition to the recommendation statement, note that recommendations also require a rationale section that precedes the recommendation statement. This section describes the logic supporting the recommendation and any pertinent considerations such as generalizability and cost.

Therapeutic

In patients with new-onset Bell's palsy [when], clinicians [who] must [level of obligation] prescribe prednisolone [what/intervention] to reduce the risk of incomplete facial functional recovery [to precipitate what outcome] (Level **A**) [recommendation strength].

In patients with Bell's palsy, clinicians must not offer antivirals (compared with placebo) to reduce the risk of incomplete facial functional recovery (Level **A**).

Clinicians should offer patients with mild to moderate Alzheimer disease cholinesterase inhibitors to slow the rate of cognitive decline (Level **B**). For patients with mild to moderate Alzheimer disease, clinicians should offer cholinesterase inhibitors to modestly slow the rate of cognitive decline (Level **B**).

Clinicians may choose not to offer mycophenolate to patients with generalized myasthenia gravis who are taking steroids, to allow more rapid tapering of steroids (Level **C**).

Causation

Persons at risk for developing MS may avoid low serum vitamin D levels to decrease their risk of developing MS—odds ratio 1.23 (Level **C**).

Parents and clinicians should not avoid immunizations in young children to decrease the risk of autism (Level **B**).

Diagnostic Accuracy

Clinicians must inform families and patients with rapidly progressing dementia that the presence of CSF 14-3-3 protein increases the likelihood of prion disease (Level **A**).

Clinicians should inform families and patients with rapidly progressing dementia that the presence of CSF 14-3-3 protein increases the likelihood of prion disease (Level **B**).

Clinicians may choose to inform families and patients with rapidly progressing dementia that the presence of CSF 14-3-3 protein increases the likelihood of prion disease (Level **C**).

Clinicians may choose not to inform patients with symptoms and signs suggesting carpal tunnel syndrome with a flick sign that they are more likely to have carpal tunnel syndrome (Level **C**).

Prognostic Accuracy

Patients with cryptogenic ischemic stroke should be counseled that the presence of PFO is not useful for identifying patients at increased risk of recurrent ischemic stroke (Level **B**).

Patients with cryptogenic ischemic stroke may be counseled that the presence of PFO is not useful for identifying patients at increased risk of recurrent ischemic stroke (Level **C**).

Clinicians should inform patients with cryptogenic stroke that the presence of a PFO does not increase their risk of subsequent stroke (Level **B**). Clinicians must inform patients with cryptogenic stroke that the presence of a PFO does not increase their risk of subsequent stroke (Level **A**). Clinicians must not inform patients with cryptogenic stroke that the presence of a PFO increases their risk of subsequent stroke (Level **A**).

Population Screening

For children with GDD, clinicians may order routine MRI of the head to identify a cause of the GDD (Level **C**). Clinicians may offer MRI of the head to children with GDD to identify the cause of the GDD (Level **C**).

For patients who meet IHS criteria and have a normal neurologic examination, clinicians should not offer head imaging to identify important abnormalities (Level **B**).

For patients who meet IHS criteria and have a normal neurologic examination, clinicians should not routinely perform head imaging to identify important abnormalities (Level **B**).

Insufficient Evidence

For patients with Alzheimer disease, there is insufficient evidence to make recommendations regarding the use of coenzyme Q for slowing cognitive decline (Level **U**).

For patients with post–cardiac arrest brain anoxia, there is insufficient evidence to make recommendations regarding the usefulness of visual evoked potentials for identifying patients at low risk of recovery (Level **U**).

Appendix 5: Tools for Determining the Strength of the Recommendation

Modified Delphi Process Questionnaire

- Assuming all premises in the rationale are true, does the recommendation logically follow from the premises?
 1 Yes
 2 No
- Do the evidence-based premises (EVID) accurately reflect the conclusions of the systematic review relative to the direction, magnitude and confidence in the effects?
 1 Yes
 2 No
 3 There are no evidence-based premises
- Do you agree that all axiomatic premises (PRIN) supporting the recommendation are true?
 1 Yes
 2 No
 3 There are no axiomatic premises
- Do you agree that the premises based upon evidence from related conditions (RELA) are strong and applicable to this condition?
 1 Yes
 2 No
 3 There are no premises based upon evidence from related conditions
- Do the inferred premises (INFER) logically follow from the other premises?
 1 Yes
 2 No
 3 There are no inferred premises
- What is your judgment as to the balance between health-related benefits and health-related harms (risks/burdens) attained by compliance with the recommendation? Consider both the number of people who will be affected as well as the magnitude of the benefits and harms. Ignore cost and resource use in your assessment.
 1 Benefits greatly outweigh harms
 2 Benefits moderately outweigh harms
 3 Benefits slightly outweigh harms
 4 Benefits and harms are balanced or harms outweigh benefits
- How important are the outcomes that will be affected by the recommendation? If multiple outcomes are affected, rate the outcome with the highest importance?
 1 Critically important
 2 Very Important
 3 Mildly Important
 4 Not important or importance unknown
- How much variation in patient preferences relative to complying with the recommendations do you expect (e.g., based on personal values, would many patients prefer not to comply with the recommendation)?
 1 Minimal variation in preferences
 2 Modest variation in preferences
 3 Moderate variation in preferences
 4 Large variation in preferences
- Are the proposed interventions discussed in the recommendation feasible (e.g., is the intervention available)?
 1 Always feasible
 2 Usually feasible
 3 Occasionally feasible
 4 Rarely feasible
- What is your judgment of the incremental cost (or resource use) relative to the net benefits of complying with the recommendation?
 1 Cost is small relative to the net benefits
 2 Cost is moderate relative to the net benefits
 3 Cost is large relative to the net benefits
 4 Cost is very large relative to the net benefits

Rationale Profile Tables

Tool for Determining the Strength of the Inference Supporting a Recommendation

Strength of Inference

Domain	Agreement				Consensus
Rationale logical	< 50%	50% to < 80%	80% to < 100%	100%	Yes
Evidence statements accurate	< 50%	50% to < 80%	80% to < 100%	100%	Yes
Axioms true	< 50%	50% to < 80%	80% to < 100%	100%	Yes
Related evidence strong and applicable	< 50%	50% to < 80%	80% to < 100%	100%	Yes
Internal inferences logically follow	< 50%	50% to < 80%	80% to < 100%	100%	Yes
Confidence in inference	Very low	Low	Moderate	High	

Note: Of the elements listed, the one that ranks lowest determines the strength of the evidence.

Tool for Determining the Strength of the Recommendation

Strength of Recommendation

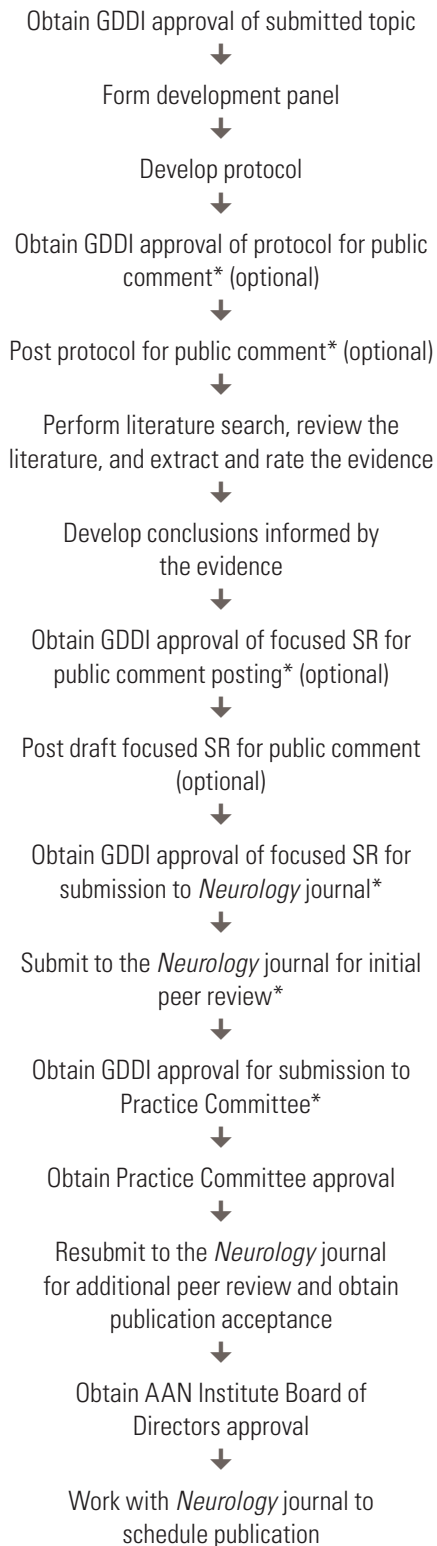
Domain	Ratings				Consensus
Confidence in inference (and evidence)	Very low	Low	Moderate	High	Yes
Benefit relative to harm	Harm ≥ benefit 0	Benefit > harm 2	Benefit >> harm 0	Benefit >>> harm 5	No
Importance of outcomes	Not important or unknown 1	Mildly important 2	Very important 2	Critically important 2	No
Variation in preferences	Large 1	Moderate 0	Modest 2	Minimal 4	Yes
Feasibility	Rarely 1	Occasionally 1	Usually 3	Always 2	No
Cost relative to net benefit	Very large 1	Large 1	Moderate 3	Small 2	No
Strength of recommendation	R/U	C	B	A	

Notes: The strength of the recommendation is anchored to the strength of the inference. The recommendation strength can be downgraded for any modifier; it can be upgraded only by one level for a moderate to large benefit relative to harm.

Domains include the premises and factors on which the recommendations are based. Consensus is considered to have been reached if 80% or more of the guideline author panel agree on the strength of a given domain. For nonpremise domains, intensity of shading corresponds to the number of panel members who were in agreement (shading of greater intensity indicates a larger number of panel members who reached agreement).

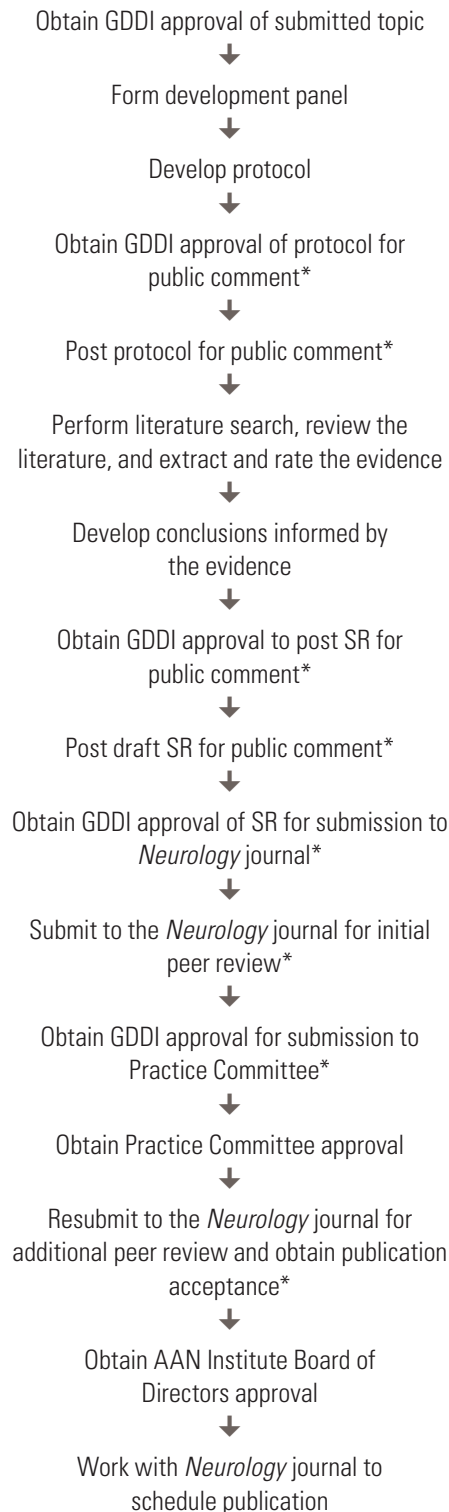
Appendix 6: Steps in the AAN Development Process for Evidence-based Documents

Focused SR Development Process



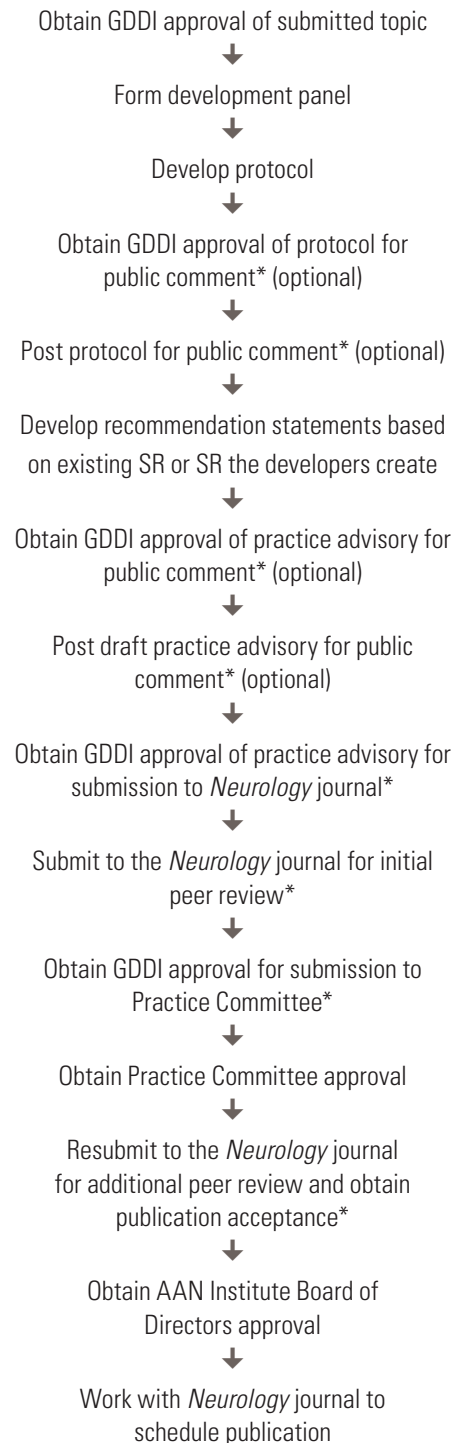
*Minor/major revisions to the draft manuscript are involved at this stage

Comprehensive SR Development Process



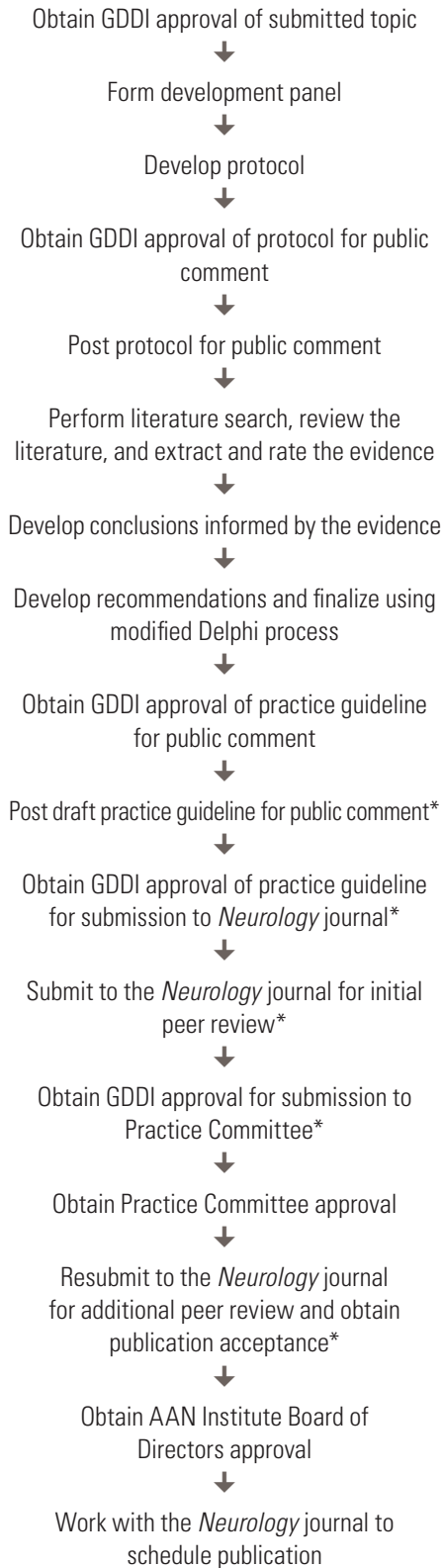
*Minor/major revisions to the draft manuscript are involved at this stage

Practice Advisory Development Process



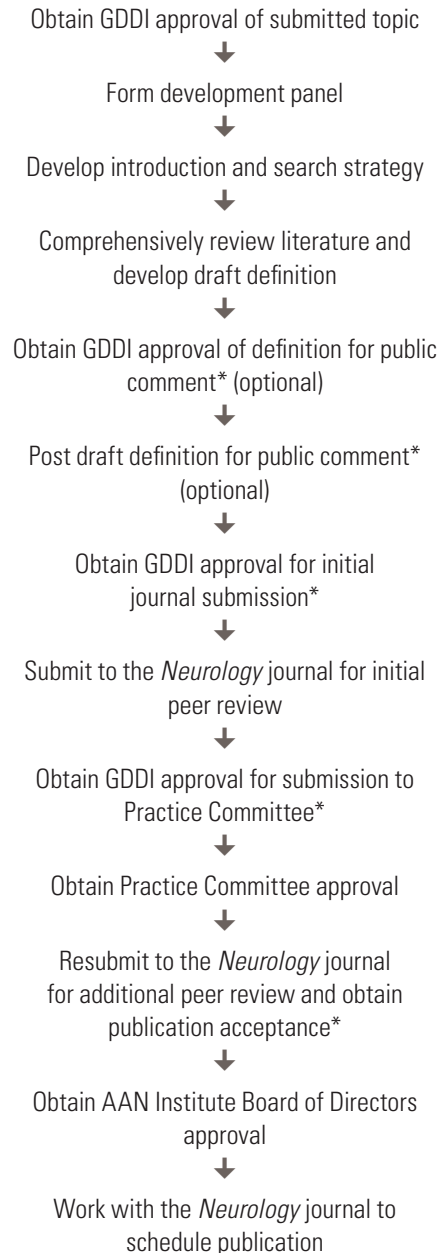
*Minor/major revisions to the draft manuscript are involved at this stage

Practice Guideline Development Process



*Minor/major revisions to the draft manuscript are involved at this stage

Case Definition Development Process



*Minor/major revisions to the draft manuscript are involved at this stage

Appendix 7: Elements of AAN Evidence-based Documents

	Case Definition	Systematic Review		Guideline	
		Focused Systematic Review	Comprehensive Systematic Review	Practice Advisory	Practice Guideline
Number of Clinical Questions Typically Addressed*	—	< 2	> 2	< 2	> 2
Number of Databases Searched	Minimum of 1	Minimum of 1	At least 2 required	Minimum of 1	At least 2 required
Search of Grey Literature	Optional	Optional	Yes	Optional	Yes
Practice Recommendations Included	No	No	No	Yes	Yes
Public Comment Period Included†	Optional	Optional	Yes	Optional	Yes
Patient/Patient Advocate on Panel Included	Optional	Optional	Yes	Optional	Yes

*At times, more than two questions may be asked in a focused SR or practice advisory.

†When developers opt not to submit the document for public comment, the document will undergo an external review phase to obtain input from targeted stakeholders.

Appendix 8: Topic Nomination Priority Scoring Tool

American Academy of Neurology Topic Nomination Priority Score Instructions

The Topic Nomination Priority Score (TNPS) should be completed by an AAN GDDI member, or designated expert, who has content expertise in the area of the nominated topic and does not have any relevant conflicts of interest. The individual nominating the topic will remain anonymous.

The GDDI member, or designated expert, assigned to complete the TNPS should review the information submitted in the Guideline Topic Nomination Form. The member may supplement this with his or her own literature search if needed. After reviewing the relevant information, the member will grade each question below on a scale of 1 to 5 according to the provided instructions.

1. Title/topic:

2. How relevant to neurologists is the proposed guideline question(s)?

(1 = minimally relevant, 5 = extremely relevant)

1 2 3 4 5

3. What is the prevalence of this disease or condition?

(1 = extremely rare, 5 = extremely common)

1 2 3 4 5

4. What is the amount of practice variation or controversy?

When answering this, please consider both scientific issues and cost. Are nearly all neurologists handling this issue the same way, or are many neurologists handling it differently? What is the cost of the screening test or therapy (or other relevant intervention for the proposed question)? What are the cost implications of the guideline? Are there articles on the cost-effectiveness of the proposed guideline?

(1 = minimal practice variation or controversy, minimal cost implications; 5 = significant practice variation or controversy, significant cost implications)

1 2 3 4 5

5. How feasible is the proposed project?

Please consider the amount of evidence and published data to answer the proposed question, whether a preexisting systematic review might be able to inform

the guideline, the number of questions proposed, and whether the AAN could collaborate with an associated society to complete the guideline.

(1 = not feasible, 5 = easy to complete in a timely manner)

1 2 3 4 5

6. How might this guideline improve patient care and outcomes?

When weighing this, please consider whether this guideline might inform the creation of a related AAN quality measure or whether it lends itself to use of certain dissemination and implementation tools.

(1 = small impact, 5 = large impact)

1 2 3 4 5

7. Please provide any additional comments regarding your prioritization of this topic.

The total score will be calculated in the spreadsheet fed by the online form. The lowest possible score is 5, and the highest possible score is 25.

American Academy of Neurology Evidence-based Guideline Topic Nomination Form

Please complete each of the required fields. The information you provide will be used to grade and prioritize the topic's importance. Please include specific references where requested.

1. Title/topic:
2. Clinical question(s) – Please provide in PICO format. For assistance, please see page 5 in AAN's Clinical Practice Guidelines Process Manual (AAN.com/Guidelines/Home/Development)
3. Background on the topic and disease:
 - a. Why is this topic important and relevant?
 - b. What is the prevalence of this disease?
 - c. Describe the amount of practice variation or controversy.
 - d. Provide cost information relevant to the proposed guideline: What is the cost of the screening test or therapy (or other relevant intervention for the proposed question)? What are the cost implications of the guideline? Are there papers regarding the cost-effectiveness of the proposed guideline? Please provide references regarding cost.
4. How much evidence and published data are there to answer the proposed question? Please provide references to key systematic reviews or seminal high-quality research papers. Please comment on whether there are existing guidelines on this topic.
5. How might this guideline improve patient care and outcomes?
6. What implementation and development tools might be relevant to the proposed guideline (e.g., clinician summary, patient summary, algorithm, app)? Could this guideline be used to develop an AAN quality measure? Please describe.
7. Please suggest proposed authors for this guideline.

Appendix 9: Relationship Disclosure Form

Introduction:

The Relationships and Conflicts of Interest Policy ("Policy") of the American Academy of Neurology ("AAN") and the American Academy of Neurology Institute ("AANI") (collectively, "Academy") requires completion of this relationship disclosure form ("Form") by individuals serving the Academy in various capacities, as defined in the Policy ("Person").

Completed Forms will be reviewed in accordance with the process described in the Policy and, as applicable, the specific review processes described for certain Academy activities, including the section entitled, "Implementing the AAN Conflict of Interest Policy for Guidelines and Case Definitions," in the 2017 Edition of the Clinical Practice Guideline Process Manual.

Information provided on completed Forms may be compared to information publicly available through the Open Payments program.

Information provided will be disclosed on applicable Academy publications, websites, or products, or at CME activities, as required by the Academy, the Accreditation Council for Continuing Medical Education, or the Council for Medical Specialty Societies' Code for Interactions with Companies.

A Person's failure to complete and sign the Relationship Disclosure Form by the deadline given by the Academy, or failure to accurately, honestly and fully complete the Relationship Disclosure Form may face sanctions by the Academy ("Sanctions"). Sanctions may include:

1. Exclusion from submitting abstracts or presenting at Academy CME activities;
2. Exclusion from publishing in Academy publications;
3. Exclusion or removal from participation on Academy boards, committees, subcommittees, work groups, task forces, clinical practice guideline or quality measurement panels, or other Academy positions;
4. Disciplinary action under the AAN's Disciplinary Action Policy; or
5. Sanctions as otherwise determined by the Academy.

Instructions:

Individuals completing the Form must disclose their financial relationships and those of Immediate Family Members (as defined below), currently and from the past two years regardless of whether these relationships are related to the topic of the Academy activity, educational presentation or course, submitted manuscript, clinical practice guideline (and related products), or performance measure. For authors, if the study period of the manuscript exceeded two years, financial relationships occurring outside the two-year window relevant to the topic must also be disclosed.

Completed Forms will be kept on file and must be updated annually if the Person continues to serve the Academy in the applicable role. Additionally, the Relationship Disclosure Form must be promptly updated and re-submitted at any time if any answers provided are no longer correct, current or complete.

Definitions:

Commercial Interest: any entity developing, producing, marketing, re-selling, or distributing health care goods or services, including drugs, devices, services or therapies, consumed by, or used on, patients to diagnose, treat, monitor, manage, and alleviate health conditions. Commercial Interest does not include non-profit entities, entities outside of the healthcare sector, or entities through which physicians provide clinical services directly to patients, unless the provider of clinical service is owned, or controlled by, a Commercial Interest.

Compensation: Anything of monetary value including a salary, honorarium, stipend, gift, or payment of travel-related expenses.

Expert Witness: A person who has provided expert medical testimony. According to the AAN’s Qualifications and Guidelines for the Physician Expert Witness, “Medical expert testimony encompasses the following:

1. Medical evaluation of a party to a legal proceeding, including personal interview and examination and/or review of medical records or other pertinent data (including laboratory tests and imaging studies)
2. Formulating an expert opinion based on such evaluation; and

3. Communicating such an opinion to attorneys, courts, licensing boards, peer review bodies or other lawful agencies, whether in the form of testimony in court, deposition, answers to interrogatories, or affidavit.”

Immediate Family Member: A Person’s spouse or partner and anyone who the Person has a significant relationship with and a reasonable belief that the individual would benefit financially from the Person’s manuscript, guideline, measure, educational work or other Academy work, or role with the Academy, because of their relationship to the Person.

1. PERSONAL COMPENSATION

Currently or within the past two years, I or my Immediate Family Member, receive or received personal compensation for the following:

a. Serving on a scientific advisory board or data safety monitoring board.

No Yes *If yes, list specific disclosures below*

Commercial Interest	Role	Year(s)

b. Gifts (other than travel or compensation for consulting or for educational efforts)

No Yes *If yes, list specific disclosures below*

Commercial Interest	Brief description of gift	Year(s)

c. Travel Expenses

No Yes *If yes, list specific disclosures below*

Commercial Interest	Purpose of travel	Year(s)

d. Serving as a journal editor, an associate editor, or editorial advisory board member. This may include a journal published by your national medical/scientific organization. Please include regardless of whether you receive compensation

No Yes *If yes, list specific disclosures below*

Full journal name	Role	Year(s)

e. Patents held or pending that may accrue revenue, whether revenue has been received to date

No Yes *If yes, list specific disclosures below*

Brief description of invention/technology	Patent (or application) number(s)	Status (filed/issued)	Year(s)

f. Royalties from publishing		
<input type="radio"/> No <input type="radio"/> Yes <i>If yes, list specific disclosures below</i>		
Full title of work	Full name of publisher	Year(s) of publication (or receipt of royalties)

g. Honoraria for speaking engagements or educational activities		
<input type="radio"/> No <input type="radio"/> Yes <i>If yes, list specific disclosures below</i>		
Commercial Interest	Role/activity	Year(s)

h. Commercial interest appointments and consultancies		
<input type="radio"/> No <input type="radio"/> Yes <i>If yes, list specific disclosures below</i>		
Commercial Interest	Role/activity	Year(s)

i. Speakers' bureau	
<input type="radio"/> No <input type="radio"/> Yes <i>If yes, list specific disclosures below</i>	
Commercial Interest	Year(s)

j. Employment		
<input type="radio"/> No <input type="radio"/> Yes <i>If yes, list specific disclosures below</i>		
Commercial Interest	Position	Year(s)

k. Other activities not covered in designations above		
<input type="radio"/> No <input type="radio"/> Yes <i>If yes, list specific disclosures below</i>		
Commercial Interest, other relevant for-profit entity, or relevant non-profit entity	Brief description of activity	Year(s)

2. RESEARCH SUPPORT

Currently or within the past two years, I or my Immediate Family Member receive or received research support from the following:

a. Commercial interests

No Yes *If yes, list specific disclosures below*

Commercial Interest	Role	Year(s)

b. Government entities

No Yes *If yes, list specific disclosures below*

Government entity	Grant number(s)	Role	Year(s)

c. Academic entities (for authors, academic entities other than those attributed in the author affiliations of the manuscript)

No Yes *If yes, list specific disclosures below*

Academic entity	Role	Year(s)

d. Foundations or societies

No Yes *If yes, list specific disclosures below*

Full name of Foundation or Society	Role	Year(s)

3. STOCK, STOCK OPTIONS AND ROYALTIES

Currently or within the past two years, I or my Immediate Family Member receive or received the following:

a. Stock or stock options or receive/received expense compensation for serving on a board of directors

No Yes *If yes, list specific disclosures below*

Commercial Interest	Relationship (e.g., 'hold stock')	Year(s)

b. License fee payments

No Yes *If yes, list specific disclosures below*

Invention/technology	Source of payment	Year(s)

c. Royalty payments or have contractual rights for receipt of future royalty payments from technology or inventions

No Yes *If yes, list specific disclosures below*

Invention/technology	Source of payment	Year(s)

d. Stock or stock options in Commercial Interest sponsored research with which you or your Immediate Family Member was involved as an investigator. (Exclude investments in mutual funds held by you or Immediate Family Members.)

No Yes *If yes, list specific disclosures below*

Commercial Interest	Year(s)

a. Stock or stock options in Commercial Interest. (Exclude investments in mutual funds held by you or Immediate Family Members.)

No Yes *If yes, list specific disclosures below*

Commercial Interest	Year(s)

4. EXPERT TESTIMONY

Currently or within the past two years, I or my Immediate Family Member receive or received:

Financial or material research support or compensation from giving expert testimony, acting as a witness or consultant, or preparing an affidavit for any legal proceeding involving a Commercial Interest (do not include proceedings for individual patients)

No Yes *If yes, list specific disclosures below*

Commercial Interest	Activity (e.g., gave expert testimony)	Year(s)

5. NON-FINANCIAL DISCLOSURE

I have chosen to declare one or more non-financial interests (e.g., special interest groups you represent or others that may be affected by your service for the Academy, if your paper is published, or that could be perceived as biasing the study, clinical practice guideline, quality measure, or your presentation, as applicable)

No Yes *If yes, list specific disclosures below*

Interest	Activity	Year(s)

4. FINANCIAL GAIN

(ONLY authors of Academy publications, clinical practice guidelines, or quality measures required to answer.)

Some published work has potential for financial gain for the study investigators or the sponsor. The following question seeks to provide transparency regarding any financial benefits to investigators or sponsors. Do you perform clinical procedures or imaging studies in your practice or unit that overlap with the content of this study, clinical practice guideline, or quality measure, and would your sponsor or this part of your practice or unit benefit if the conclusions were widely followed?

No Yes *If yes, list specific disclosures below*

Procedure	Practice unit (e.g., 35%)	Year(s)

I have completed this Relationship Disclosure Form fully and to the best of my ability. I understand that the information may be disclosed on applicable Academy publications, websites, products, or at applicable education programs, as required by the Academy.

By my electronic signature (name preceded and followed by the forward slash symbol [/]; e.g., /John Doe/) below, I verify the completeness and accuracy of the contents of this form.

Signature: _____

Date: _____

Appendix 10: Development Process: Roles and Responsibilities

Lead Developer Responsibilities

- Coordinates the project from start to finish with direction from the GDDI facilitator
- Communicates deadlines and reminders to development panel
- Consults with methodologist on project development as needed
- Drafts protocol and revises after committee and public reviews, with input from the development panel as requested
- Drafts manuscript and revises after committee and public reviews, with input from the development panel as requested
- Completes author revision table reflecting changes made on the basis of committee and public review feedback
- Signs off on completeness and accuracy of all manuscript content, including the following:
 - PICO format of clinical questions
 - Search strategy (inclusion/exclusion criteria)
 - Balance of panel with respect to intellectual and financial conflict of interest
 - Classification of studies
 - Language used in conclusions
 - Recommendation levels
 - Evidence synthesis tables
 - Data tables, figures, and other nonstandard appendices
 - References
- Ensures studies that inform conclusions are discussed in the text. The class and citation of each of these studies should be listed in the discussion text. The conclusion text should include the class of each applicable study
- Reviews and approves (when applicable) all edits to the manuscript
- Completes, manages, and finalizes references, including full citations and numbering, at all stages; formatting will be done by AAN staff
 - **Note:** Staff will not populate the reference list and will renumber only if all references have been provided

- Must report current information to staff with regard to professional degrees (e.g., MD, PhD) and affiliations (institutions)
- Must remain nonconflicted with regard to SR/guideline topic throughout development process and notify AAN staff if conflict of interest changes such that a new lead developer would need to be appointed; must maintain current conflict of interest information with the AAN

Codeveloper Responsibilities

- Contributes as needed to project development, including participation in drafting of project protocol, review of abstracts and full-text articles*, data extraction*, classification of studies*, informal consensus process with evidence synthesis tables, recommendation development, modified Delphi voting for final recommendations, SR manuscript* and practice advisory/guideline development (drafting and revising), and development and confirmation of accuracy of tables, figures, and appendices
- Must report current information to staff with regard to professional degrees (e.g., MD, PhD) and affiliations (institutions)
- Must maintain current conflict of interest information with the AAN

*Only developers determined to be free of conflict of interest (as indicated in AAN conflict of interest policy) may participate in these development stages

GDDI Facilitator Responsibilities

- Provides process guidance to lead developer and development panel
- Participates in review of abstracts and full-text articles*, data extraction*, classification of studies*, informal consensus process with evidence synthesis tables, recommendation development, modified Delphi voting for final recommendations, SR manuscript* and practice advisory/guideline manuscript development (drafting and revising), and development and confirmation of accuracy

- of tables, figures, and appendices
- Adjudicates discrepancies in full-text review and data extraction
- Participates in manuscript development (drafting and revising)
- Consults with methodologist on project development as needed
- Reviews and approves (when applicable) all edits to the manuscript
- Presents draft manuscripts at GDDI meetings for review and communicates required edits to lead developer and codevelopers
- Maintains current conflict of interest information with the AAN

*Only developers determined to be free of conflict of interest (as indicated in AAN conflict of interest policy) may participate in these development stages.

AAN Methodologist Responsibilities

- Provides guideline/case definition methodology training, as needed, at milestone stages of project development
- Consults with lead developer and GDDI facilitator regarding balance of development panel with respect to intellectual and financial conflict of interest
- Signs off on the following:
 - PICO format of clinical questions
 - Search strategy (inclusion/exclusion criteria)
 - Conflict of interest on panel
- Reviews study classifications and conclusions and recommendations for accuracy and consistency with the process
- Adjudicates discrepancies in full-text review and data extraction
- Creates data extraction forms and trains development panel in exercise to extract data
- Populates the level of evidence for data extractions
- Performs statistical assessments as requested by developers or GDDI

- Coordinates and facilitates modified Delphi voting on recommendations and analyzes results
- Participates in modified Delphi voting on recommendations (if needed or requested)
- Reviews and approves draft manuscripts when they are submitted for GDDI meetings
- Provides feedback to manuscript developers on methodologic issues
- Keeps abreast of methodology of manuscripts and communicates any methodologic concerns to staff and development panel
- Maintains current conflict of interest information with the AAN

AAN Staff Responsibilities: Development

- Gathers and maintains conflict of interest statements, copyright assignments, and curriculum vitae
- Guides lead developer and development panel through literature search and review
 - Organizes abstract review and communicates results to lead developer, facilitator, and methodologist
 - Organizes full-text review, including searching for complete articles obtainable at no cost. Communicates results of article review to lead developer, facilitator, and methodologist
 - Works with methodologist and lead developer to arrange data extraction trainings
 - Works with methodologist to create data extraction form
 - Organizes data extraction assignments and corresponds with reviewers on reconciliations
 - Organizes completed data table of evidence from data extraction for lead developer, facilitator, and methodologist
- Coordinates manuscript reviews with the GDDI and communicates comments to the developers after reviews
- Coordinates external review and public comment, and communicates comments to the developers after reviews

- Populates data in the manuscript on literature search information, including number of abstracts reviewed and final articles selected and reviewed
- Coordinates protocol/manuscript review with AAN legal team and communicates comments to the developers after reviews
- Coordinates conference calls and meetings throughout the development project as needed

AAN Staff Responsibilities: Manuscript Editing

- Copyedits the manuscript (edits for style/usage)
 - Clarity, concision, word choice, sentence structure, spelling
 - Word count
 - Callouts to appendices, tables, figures
 - Citation numbering (numeric order)
 - Header levels
 - Other publication style considerations
- Assembles/formats front matter
 - Title page content (formatting)
 - Developers ensure title is final, affiliations current and accurate (as per lead developer)
 - Disclosures (wording, spelling)
 - Developers ensure accuracy, completeness, currency
 - Disclosures in manuscript: only those relevant to the manuscript; disclosures online through *Neurology* portal: comprehensive disclosures which will appear online only

Note: *Relevant disclosures for the manuscript are finalized on journal submission (first or second submission). Complete disclosures for website are obtained at the journal's provisional acceptance of the manuscript*
- Assembles/formats back matter and online-only material
 - Appendices, tables, figures (formatting only)
 - Checked at a high level for title, column and row labels, punctuation, abbreviations

- Standard copy (e.g., disclaimer, disclosure, conflict of interest statement)
- Developer contributions
- Standard categories from journal website. Editor assigns, lead developer/GDDI facilitator confirms
- Formats reference list
 - Numeric order (40 references maximum allowed for print—the remainder are e-references for online only)
 - All published articles/studies must be cited
 - Citations should be numbered in succession as they appear in the body of the manuscript
 - Citations in other matter (tables, appendices, figures): if appeared in manuscript, have same number; if did not appear in manuscript, are numbered starting with where left off numerically in manuscript
 - One-to-one correspondence numerically with citations in text
 - Style compliance
 - Renummer when needed (does not confirm accuracy of citation content or placement—accuracy of final citations and reference list are responsibility of the lead developer)
- Fact-checks (certain items) for accuracy/consistency
 - Correspondence of conclusions to recommendations in abstract and text; correspondence of recommendations in text and appendices
 - Identical wording of clinical questions when repeated, unless such wording is intentionally truncated for brevity (in abstract, introductory text, analysis of evidence)
 - Double-checking consistency of repeated content across two or more manuscripts*

*For projects with summary and full-length versions or projects with two or more manuscripts. This should also be done by the lead developer before submission to the editor.

The following presents at a glance the roles of participants in guideline and case definition development.

Roles of participants in guideline and case definition development process					
	Lead Developer	Codevelopers	GDDI Facilitator	Methodologist	AAN Staff
Coordinates project	P		I		I
Communicates deadlines to panel, including sending reminders	P	I	I	I	I
Drafts protocol and revises after public comments	P	I	I	I	I
Drafts manuscript before committee and public reviews	P	I	I	I	
Revises manuscript after committee and public reviews	P	I	I	I	I
Verifies completeness and accuracy of all guideline content, including classification of studies, language used in conclusions and recommendations, recommendation levels, and references	P	I	I	I*	
Maintains current developer designations and affiliations in manuscripts	P	P	I		I
Maintains current conflict of interest information on a yearly basis	P	P	P	P	I
Reviews and approves all edits to manuscript	P	I	P	I	
Completes, manages, and finalizes references (including full citations and numbering) at all stages <i>Note: Any published articles/studies must be cited. Developers should populate reference list. Staff will renumber and format the provided references as needed</i>	P	I	I		I
Reviews abstracts	P	P	I	I	I
Reviews full text of selected articles	P	P	I	I	I
Performs data extraction	P	P	I	I	I
Develops recommendations	P	P	P	P	I
Develops and ensures accuracy of tables, figures, and appendices	P	I	I		
Approves and presents draft manuscripts at GDDI meetings	I	I	P	I	I
Performs statistical assessments	I	I	I	P	
Gathers and maintains records of conflict of interest statements and curriculum vitae					P
Arranges phone calls/meetings					P
Copyedits manuscript					P
Assembles front and back matter					P

P = primary responsibility I = involved *Level of evidence

Appendix 11: Project Protocol Template

Protocol for Proposed [indicate intended final document type: focused systematic review, comprehensive systematic review, practice advisory, practice guideline] Project: Title

Proposal of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology [if codeveloped by another organization, indicate organization name here]

Authors

Jane Doe, MD¹; John Doe, MD²

1. Department of Neurology, Hospital, Minneapolis, MN
2. Department of Neurology, Hospital, St. Paul, MN

Correspondence to

American Academy of Neurology:

guidelines@aan.com

Approved by the Guideline Development, Dissemination, and Implementation Subcommittee on [indicate date approved]. [If codeveloped by another organization, indicate organization's approving body here.] All comments submitted during the 30-day public comment period in which this protocol is posted will be reviewed and addressed by the author panel members. Although all comments will be considered, author panel members will not specifically respond to individual comments online.

Study Funding

This [indicate document type: focused systematic review, comprehensive systematic review, practice advisory, practice guideline] protocol was developed with financial support from the American Academy of Neurology. Authors who serve as AAN subcommittee members or methodologists ([indicate member and methodologist initials here]), or who are AAN staff members ([indicate any staff member initials here]), were reimbursed by the AAN for expenses related to travel to subcommittee meetings where drafts of manuscripts were reviewed.

Disclosures

Description of AAN Document Types

This protocol is the planning document for one of four AAN document types: focused systematic review, comprehensive systematic review, practice advisory (based

on a systematic review), or practice guideline (based on a comprehensive systematic review). The term *guideline* is the general term that refers to all AAN evidence-based documents, with the exception of case definitions. Because it is for planning purposes only, this protocol document is not a substitute for the complete guideline.

Guideline Project Protocol

Guideline Project Development Plan

This proposed project will be developed in accordance with the processes described in the 2017 edition of the AAN clinical practice guideline development process manual. The developers of this guideline project intend to develop a [indicate intended final document type: focused systematic review OR comprehensive systematic review OR practice advisory based on a systematic review OR practice guideline based on a systematic review]. This protocol will be posted for public comment. Patient representatives will not be included on the panel; however, literature regarding patient preferences will be reviewed to frame the questions.

Guideline Project Timeline

The following is the tentative timeline for development of this [indicate intended final document type: focused systematic review OR comprehensive systematic review OR practice advisory based on a systematic review OR practice guideline based on a systematic review]:

- *Panel formation:*
- *Drafting of protocol:*
- *Approval of protocol by the AAN Guideline Development, Dissemination, and Implementation Subcommittee (GDDI):*
- *Protocol posted for public comment:*
- *Literature search:*

- *Panel review of abstracts:*

- *Review of full articles, data extraction, and development of evidence tables:*

- *Systematic review draft submitted to AAN GDDI and AAN Legal Department:*

- *Systematic review posted for public comment:*

- *Develop recommendations on the basis of systematic review and other pillars (principles, strong related evidence from other conditions, inferences):*

- *Submit draft guideline to AAN GDDI for review and approval for public comment:*

- *Post guideline for public comment:*

- *Submit to AAN GDDI for review and approval of final document (Neurology will do preliminary review concurrently; when approved by GDDI, the AAN Practice Committee also will review):*

- *Submission to Neurology:*

Composition of the Author Panel

In [indicate month and year project initiated], the AAN GDDI recruited a multidisciplinary panel consisting of [indicate number] AAN members to develop this guideline project protocol. The panel included content experts ([indicate content experts' initials]), a methodology expert ([indicate methodology expert's initials]), and AAN GDDI members ([indicate GDDI members' initials]). The clinicians were required to submit online conflict of interest forms and copies of their curriculum vitae (CV). The panel leadership, consisting of the lead developer ([indicate the lead developers' initials]), the AAN methodologist ([indicate the methodologist's initials]), and AAN staff persons ([indicate staff persons' initials]), reviewed the conflict of interest forms and CVs for financial and intellectual conflicts of interest. These documents were screened specifically to exclude both those individuals with a

clear financial conflict and those whose intellectual bias would diminish the credibility of the review in the eyes of the intended users. In accordance with AAN policy, the lead developer ([indicate lead developer's initials]) has no conflict of interest. One of the guideline developers ([indicate developer's initials]) was deemed to have conflicts of interest, but the conflict of interest was judged as not significant enough to preclude involvement of this developer as an author. The developer deemed to have conflicts of interest ([indicate developer's initials]) will not be permitted to review or rate the evidence. This individual will be consulted in an advisory capacity to help with the validation of the key questions, the scope of the literature search, and the identification

of seminal articles to validate the literature search. The panel member with conflicts of interest will be allowed to participate in the recommendation development process. The lead developer recommended the final panel composition to the AAN GDDI leadership, who reviewed the list of members and the panel leaders' conflict of interest forms, and provided final approval. This panel will be solely responsible for the final decisions about the design, analysis, and reporting of the proposed systematic review [if applicable, add the proposed subsequent document type: practice advisory or practice guideline]. The document will then be submitted for approval to the AAN GDDI.

Introduction to Proposed [indicate intended final document type] Project Topic

Rationale for This [indicate intended final document type]

The purpose of this [indicate intended final document type] is to systematically assess all high-quality randomized controlled trials that evaluate the efficacy of [intervention of interest] for [indications of interest] and the risks associated with the use of [intervention of interest]. The systematic review will then be used to develop recommendations regarding the use of [intervention of interest] in these conditions.

Clinical Questions

The systematic review for this [indicate intended final document type] addresses the following questions:

[Present text of clinical question(s) in list format]

Table [optional]

Question (type)*	Population	Intervention	Co-intervention	Outcome	Study design
1					
2					

*Question type refers to one of the following: screening, diagnostic, therapeutic, prognostic, causation.

**Rationale for the Clinical Questions
Consideration of Patient Preferences**

Considerations for Special Populations and Multiple Morbidities

Special populations and important comorbidities should be considered at the project development and PICO question stage in the following ways:

- Special populations relevant to the guideline must be identified (at a minimum, key population factors that must be considered are age, sex, race)
- Topic-specific comorbidities must be highlighted a priori
- For identified key populations, experts with related expertise should be included on the panel

Plan to Address Special Populations and Multiple Morbidities in the [indicate intended final document type]

Literature searches and review must incorporate identification of the special populations and relevant comorbidities. If relevant literature is identified for special populations or comorbidities, developers follow the usual guideline process for conclusions and recommendations, including specific conclusions and recommendations relating to the special populations and patients with comorbidities. If no literature is identified, one of two things will happen:

- The panel will develop recommendations for the special populations using related evidence
- The panel will highlight the special populations and important comorbidities in the clinical context section, particularly if special-population-specific recommendations cannot be made

Study Screening and Selection Criteria: Inclusion and Exclusion Criteria for Article Selection

- Types of participants
- Types of intervention
- Comparison group
- Types of outcome measures

Literature Search Strategy

- Inclusion and exclusion criteria
- Terms and databases to be used in the literature search
- Keywords
 - a. Key text words and index words for the condition or closely related conditions, if appropriate (linked by the word OR):
 - b. Key text words for identification of special populations and relevant comorbidities (linked by the word OR):
 - c. Key text words and index words for the intervention (linked by the word OR):
- Preliminary literature search [optional]

Disclaimer

Clinical practice guidelines, practice advisories, systematic reviews, and other guidance published by the American Academy of Neurology and its affiliates are assessments of current scientific and clinical information provided as an educational service. The information (1) should not be considered inclusive of all proper treatments or methods of care, or as a statement of the standard of care; (2) is not continually updated and may not reflect the most recent evidence (new evidence may emerge between the time information

is developed and when it is published or read); (3) addresses only the question(s) specifically identified; (4) does not mandate any particular course of medical care; and (5) is not intended to substitute for the independent professional judgment of the treating provider, as the information does not account for individual variation among patients. In all cases, the selected course of action should be considered by the treating provider in the context of treating the individual patient. Use of the information is voluntary. The AAN provides this information on an “as is” basis, and makes no warranty, expressed or implied, regarding the information. The AAN specifically disclaims any warranties of merchantability or fitness for a particular use or purpose. The AAN assumes no responsibility for any injury or damage to persons or property arising out of or related to any use of this information or for any errors or omissions.

Conflict of Interest

The AAN's Conflict of Interest Policy is available at tools.aan.com/apps/disclosures/index.cfm?event=committee:intro. All AAN guideline authors must meet the stipulations outlined in the policy in order to participate on a guideline development panel. This policy is further described in the 2017 AAN Clinical Practice Guideline Development Manual, available at [AAN.com/Guidelines/Home/Development](https://aan.com/Guidelines/Home/Development).

References

Appendix 12: Sample Data Extraction Forms

Atrial Fibrillation Rx Data Extraction Form-DRAFT

Patient Population:

For patients with nonvalvular atrial fibrillation (including these special populations: patients with intracranial hemorrhages [spontaneous transformation, posttraumatic, hypertensive, vascular malformation]); patients with intracranial or intraspinal (vascular) malformations; patient's status post-coronary artery bypass graft surgery (CABG); other special populations (those who are elderly, residents of nursing homes, those with end-stage renal disease, those with dementia)

Intervention:

What therapies (including antithrombotics: warfarin, aspirin, dabigatran, apixaban, rivarixaban, combination therapy); rate or rhythm control of atrial fibrillation (with medical therapy or ablation)

Comparative Intervention:

As compared with no therapy or another therapy

Outcomes:

Reduce the risk of ischemic stroke with the least risk of hemorrhage (including intracerebral hemorrhage)

Summary

Panel Member: _____

Article ID#: _____

Inclusion Criteria

- Human studies only
- Enrolled patients with atrial fibrillation
- Patients receiving different therapies to prevent ischemic stroke

Exclusion Criteria

- Case report, editorial, meta-analysis, or review (please specify)
- < 50 patients

Comparison Group

- The study compares outcomes between groups using different management strategies (e.g., ablation plus anticoagulation to anticoagulation alone).
- To be considered an RCT, patients should have been randomized to different management strategies.

Relevance

Study is relevant to question?

Yes No

If no, STOP. Explain

Design

- RCT
 Nonrandomized trial that includes a comparison group

If the study does not include a comparison group, STOP. (Study does not meet inclusion criteria.)

For the therapeutic rating:

If an RCT, maximum Class I.

If not a randomized trial, maximum Class II.

➔ **Maximum Therapeutic Class** I II

Sample Size

Total patients enrolled: _____

If total less than 50 patients, STOP. (Study does not meet inclusion criteria.)

Outcome Assessment

1. Was any outcome assessment blinded to management strategy?

Yes No Not stated

2. Was any outcome objective?

Yes No Not stated

3. Was any outcome assessed independently?

- Yes No Not stated

Comments regarding outcome assessment:

If 1 or 2 = YES, maximum is Class I

If only 3 = YES, maximum is Class III

If all = NO/NOT STATED, STOP: Class IV

➔ Maximum Therapeutic Class

- I II III IV

Outcomes

- Ischemic stroke
- Bleeding

Objective

The determination of the outcome is unlikely to be affected by observer expectations. Consider the following outcomes objective: death, disabling stroke, major hemorrhage.

Independently

The investigator determining the outcome was different than that identified by the treating physician.

Other Therapeutic Study Characteristics

1. Was treatment allocation concealed (check "no" if not an RCT)?

- Yes No Not stated

2. Primary outcome measure(s) was specified?

- Yes No Not stated

Record primary outcome(s) measure

Record secondary outcomes measure

3. Explicit inclusion and exclusion criteria were used

- Yes No Not stated

Summarize relevant criteria

4. Patients in different treatment arms were similar at baseline, or appropriate statistical adjustments were made for baseline differences

- Yes No Not stated

5. Less than 20 percent of patients were lost to follow-up

- Yes No Not stated

Percentage lost to follow-up: _____

If all = "yes," maximum is Class I.

If only three or four = "yes," maximum is Class II.

If < three = "yes," maximum is Class III.

➔ Maximum Therapeutic Class

- I II III

Concealed Allocation

Investigators could not manipulate treatment assignment. Examples of concealed allocation include consecutively numbered sealed, opaque envelopes containing a predetermined, random sequence for treatment assignment or an independent center that an investigator contacts to obtain the treatment assignment.

Final Rating: Select lowest maximum therapeutic class from above

- ➔ I II III IV

If Class IV, STOP

Demographics (for entire study population if possible. Otherwise list values for all groups)

Age:

Central tendency: Mean Median

Value: _____

Dispersion:

- SD SE Range
 Interquartile range

Value: _____

Sex percent female: _____

Special Atrial Fibrillation Populations Included (check all that apply; describe)

- Patients with intracranial hemorrhages (spontaneous transformation, posttraumatic, hypertensive, vascular malformation)

Elderly _____

Nursing home residents _____

Patients with end-stage renal disease _____

Patients with dementia _____

- Other _____
- Other _____
- Other _____
- Other _____

Type(s) of Management Strategies

(check all that apply; describe)

- Aspirin _____
- Clopidogrel _____
- Clopidogrel plus aspirin _____
- Warfarin _____
- Dabigatran _____
- Apixaban _____
- Rivarixaban _____
- Triflusal and acenocoumarol _____
- Medication(s) for rate or rhythm control _____
- Ablation for rate or rhythm control _____
- Other _____
- Other _____
- Other _____
- Other _____

Describe Management Strategy Comparison Groups, Including the Number in Each Group (there should be at least two)

Number	Description of Group
Group 1 _____	_____
Group 2 _____	_____
Group 3 _____	_____
Group 4 _____	_____

Outcomes Described

Thromboembolic Events (check all described)

- Ischemic stroke _____
- Transient ischemic attack _____
- All ischemic stroke _____
- Fatal ischemic stroke _____
- Disabling ischemic stroke _____
- Nondisabling ischemic stroke _____
- Other _____
- Other _____
- Other _____

Comments: _____

Bleeding Events (check all described)

- Minor bleeding _____
- Major bleeding _____
- Intracranial bleeding _____
- Death secondary to hemorrhage _____
- Gastrointestinal hemorrhage _____
- Other bleeding events _____
- Other bleeding events _____

Other Outcomes (check all described)

- All-cause death _____
- Other _____
- Other _____
- Other _____
- Other _____

Results (briefly summarize the study results)

Comments (provide any special reasons to include, noteworthy findings, reason for classifying, etc.)

Appendix 13: Sample Evidence Synthesis Table

Evidence synthesis: Dabigatran 150 mg twice daily vs. dose-adjusted warfarin

Outcome	Study	Class	Effect size (95% CI)	Starting confidence in evidence	Precision	Consistency	Directness	Plausibility	Reporting bias	Magnitude of effect	Dose response	Direction of bias	Final level of confidence in evidence
All stroke and SE	Connolly 2009	I	RR 0.66 (0.53 to 0.82)										
Summary stroke and SE		1 Class I	RR 0.66 (0.53 to 0.82)	Moderate	NC	NC	NC	NC	NA	NC	NA	NC	Moderate
Major hemorrhage	Connolly 2009	I	RR 0.93 (0.81 to 1.07)										
Summary major hemorrhage		1 Class I	RR 0.93 (0.81 to 1.07)	Moderate	NC	NC	NC	NC	NA	NC	NA	NC	Moderate
Intracranial hemorrhage	Connolly 2009	I	RR 0.40 (0.27 to 0.60)										
Summary of intracranial hemorrhage		1 Class I	RR 0.40 (0.27 to 0.60)	Moderate	NC	NC	NC	NC	NA	NC	NA	NC	Moderate
GI hemorrhage	Connolly 2009	I	RR 1.50 (1.19 to 1.89)										
Summary GI hemorrhage		1 Class I	RR 1.50 (1.19 to 1.89)	Moderate	NC	NC	NC	NC	NA	NC	NA	NC	Moderate

Abbreviations: CI = confidence interval; GI = gastrointestinal; NA = not applicable; NC = no change; RR = relative rate; SE = systemic embolism.

Values < 1 favor dabigatran

Table adapted from appendix e-6, online data supplement, published in: Culebras A, Messe SR, Chaturvedi S, Kase CS, Gronseth G. Summary of evidence-based guideline update: Prevention of stroke in nonvalvular atrial fibrillation. Report of the Guideline Development Subcommittee of the American Academy of Neurology. Neurology 2014;82:716–724.¹⁶ Used with permission.

Appendix 14: Manuscript Format

Note: It is important to follow Neurology journal style format if submitting to that journal.

Cover Page

- [Include header with the date and stage of development for reference prepublication; this should be removed for publication]
- Document type [if an update to previous publication, indicate here]: Title
- Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology [if codeveloped with another organization, indicate organization name here]
- List developers' names, degrees, and professional affiliations
- Indicate corresponding developer information:
 - Address correspondence and reprint requests to American Academy of Neurology: guidelines@aan.com
- [For manuscripts to be published in print, indicate word count for abstract, word count for body of manuscript (excludes back matter), and title character count (including spaces)]
- [Indicate any endorsing organizations, if applicable, and the endorsement dates]

Page 2/Remaining Front Matter

- Author contributions list
- Study funding statement
- Conflict of interest disclosures
- Abbreviations list

Abstract

Maximum of 250 words; should summarize the guideline as follows:

Objective: Summary of clinical focus

Methods: Description of process

Results: Status, quality, and content of evidence

Recommendations: Summary of recommendations

Manuscript Body

Introduction

The introduction should concisely cover the following:

- Statement of purpose (including identification of audiences)
- Background and justification. An overview of the problem or topic area under study and the underlying justification for

pursuing the question. May include any or all of the following:

- Membership needs; the degree of interest and usefulness to Academy members, if known (e.g., by survey)
 - The potential for significant benefit or risk to patients
 - Extent of practice variation
 - Urgency
 - Controversy regarding validity or applicability
- Clinical question statement

Description of the Analytic Process

This section should present the exact, replicable process the developers used to develop the document, including:

- How the panel was selected, including disclosure of information, funding, and outside input (e.g., reviewers)
- An indication of the process manual edition followed. (If a combination of processes from more than one manual was followed, those processes explicitly identified, with the supporting manual editions clearly stated and cited)
- Description of literature review
- How the literature search was conducted (search terms, databases searched, other search strategies, languages included, dates covered). Bibliographic or other search techniques described in sufficient detail so that the process can be replicated
- How articles were selected for inclusion (e.g., all articles reviewed, only prospective studies selected)
 - Inclusion and exclusion criteria and process for “weeding out” articles
 - An indication of the number of articles identified in the search, the number excluded during the abstract review, the number excluded during the article review, and the number included in the final document
 - Statement of how abstracts and articles were reviewed (e.g., how many panel members reviewed each, how disagreements were resolved)
- Analysis of the data
 - Elements of evidence extracted from pertinent articles (developers use a data extraction form for this)
 - Classification of evidence definitions
 - Brief description of modified GRADE

process followed for deriving conclusions⁴

- Brief description of the modified Delphi process followed for constructing recommendations
- Mention of the evidence summary tables, with reference to an appendix presenting the evidence synthesis tables

Analysis of Evidence

This section is the scientific body of the manuscript and should include a detailed narrative description of the evidence and the statistical analysis applied to it (e.g. meta-analysis), as appropriate to the topic. If more than one clinical question is addressed, it is appropriate to deal with the questions one at a time, providing data analyzed, levels of evidence, and conclusions for each question.

For diagnostic tests/procedures:

- Results
- Levels of evidence
- Statistical analysis (meta-analysis, sensitivity and specificity, positive and negative predictive values, odds ratios, relative rates, and numbers needed to treat/harm)
- Relevance (selection criteria, complications, contraindications, test/procedure specifics)
- Clinical significance
- Availability of a reference standard (gold standard) for comparison

For therapies:

- Results
- Levels of evidence
- Statistical analysis (meta-analysis, sensitivity and specificity, positive and negative predictive values, odds ratios, relative rates, and numbers needed to treat/harm)
- Relevance (patient selection criteria, complications, contraindications, intervention details, protocols, difficulty with implementation, duration/frequency of treatment)
- Clinical significance

Conclusions (Evidence Synthesis)

The conclusions typically are presented as part of the Analysis of Evidence, after discussion of the evidence associated with each clinical question. The conclusions summarize the evidence in the context of

additional key factors to be considered in response to the clinical question. The panel members should apply the modified GRADE process described on [page 20](#) of this manual when constructing conclusion statements.

Putting the Evidence into Clinical Context

Optional for inclusion, this section helps the development panel put the information just presented into a clinical context. This section may also mention the recommendations (presented in a subsequent section) to some degree. The document should end with the recommendations, so the clinical context is presented here. This section should include comments on each of the following:

- Clinical context
- Special populations and comorbidities/multimorbidities
- Mention here if there are important modifiers (e.g., cost, availability) that are present but which the panel decided not to use to downgrade the recommendations
- Limitations
- Controversies (optional; whether controversy rises to level that it should be discussed will be decided by the guideline panel chairs)

Response to Public Comment

Internally, author panels are required to complete a table where authors respond to each comment made during the public comment process. This is saved with internal AAN documents pertaining to the guideline. For the guideline itself, authors must draft a brief paragraph (with the assistance of AAN staff for dates and details) explaining how they responded to public comment. An example of such a paragraph is the following: “A draft version of this guideline was posted for public comment on the AAN Web site from [indicate date range]. The guideline panel reviewed all public comments received and created a response table. In response to public comment, the panel revised the final recommendation to [indicate revised recommendation language]. The development panel also reexamined [indicate topic reexamined] and found [indicate result of reexamination].” The development panel may also choose to describe changes to the protocol resulting from public comment in the methods/approach section of the guideline.

Practice Recommendations

This section translates the conclusions into recommendations, or action statements. The recommendations are statements that result from a modified Delphi process performed by the panel members. The panel responds to key questions that capture major elements, including the available evidence, strong evidence from closely related conditions, inferences from one or more of the other statements, and established principles of care. The recommendations are rated to indicate their levels of strength (Level **A**, **B**, **C**, **U**, and **R**). Recommendations should not be broader or narrower than the clinical question. Each recommendation or set of recommendations is preceded by a rationale section that delineates the justification for the recommendations and their corresponding levels. If modifiers affected the recommendations, a discussion of those modifiers must also be included. Thus, for each recommendation, there will be two to three elements included:

- Rationale
- Discussion of modifiers (if applicable)
- Recommendation

Suggestions for Future Research

This section presents the identified gaps in the literature.

Tables/Figures

Tables, algorithms, or figures should be presented if they help communicate—but not alter—the evidence-based recommendations. In most cases, evidence tables are placed online.

Disclaimer

The following disclaimer must appear in all published documents:

Practice guidelines, practice advisories, comprehensive systematic reviews, focused systematic reviews and other guidance published by the American Academy of Neurology and its affiliates are assessments

of current scientific and clinical information provided as an educational service. The information: 1) should not be considered inclusive of all proper treatments, methods of care, or as a statement of the standard of care; 2) is not continually updated and may not reflect the most recent evidence (new evidence may emerge between the time information is developed and when it is published or read); 3) addresses only the question(s) specifically identified; 4) does not mandate any particular course of medical care; and 5) is not intended to substitute for the independent professional judgment of the treating provider, as the information does not account for individual variation among patients. In all cases, the selected course of action should be considered by the treating provider in the context of treating the individual patient. Use of the information is voluntary. AAN provides this information on an “as is” basis, and makes no warranty, expressed or implied, regarding the information. AAN specifically disclaims any warranties of merchantability or fitness for a particular use or purpose. AAN assumes no responsibility for any injury or damage to persons or property arising out of or related to any use of this information or for any errors or omissions.

Conflict of Interest Statement

The AAN’s Conflict of Interest Policy is available at tools.aan.com/apps/disclosures/index.cfm?event=committee:intro. All AAN guideline authors must meet the stipulations outlined in the policy in order to participate on a guideline development panel. This policy is further described in the 2017 AAN Clinical Practice Guideline Development Manual, available at [AAN.com/Guidelines/Home/Development](https://aan.com/Guidelines/Home/Development).

Acknowledgments

Appendices

Appendices will include GDDI mission statement and member roster, the complete search strategy employed, the

schemes for classification of evidence, the rules for determining the confidence in the evidence, the evidence tables, the evidence synthesis tables, the steps and rules for formulating recommendations, and the rationale of factors for developing practice recommendations. This section can be populated by AAN staff and the methodologist assigned to the project. All appendices are included in the complete manuscript, available as an online data supplement to the main article.

References

For the summary manuscript, the first 40 references are included in the main article online and in print; any subsequent references are labeled as e-references and published online only as a data supplement to the main article. For the complete manuscript, which is published online only as a data supplement to the main article, all references are presented as e-references at the back of the document.

Appendix 15: Sample Revision Table

#	Reviewer	Criticism	Action
1	A.B. Smith	<ol style="list-style-type: none"> 1. Clarify the diagnostic criteria 2. PEJ vs. PEG 3. "Breaking the News" is a flippant term 4. Editorial changes suggested 	<ol style="list-style-type: none"> 1. A sentence has been inserted about diagnostic criteria citing the World Federation of Neurology criteria 2. There is little evidence on PEJ and expert consensus was not achieved—no action 3. No change; the term was derived from the literature and from consensus of the task force 4. Selectively incorporated
2	X.Y. Jones	<ol style="list-style-type: none"> 1. Many aspects of symptomatic care are not covered 2. Some evidence from only one or two studies provides the basis for some recommendations, e.g., sialorrhea 3. We omitted data from Belsch and Shipman in a book chapter 4. The recommendation about invasive ventilation should be separated and expanded to include fully informing about burdens and benefits 	<ol style="list-style-type: none"> 1. No change; to be covered in future practice parameters 2. No change; this is the status of the evidence 3. No change; reference not added since no measures of quality of life or survival were made 4. So changed
3	Anonymous	<ol style="list-style-type: none"> 1. Delete the option on laryngectomy for recurrent aspiration 2. The word <i>entrapment</i> with respect to tracheostomy/ventilator without proper planning is unclear 3. Extensive editing 	<ol style="list-style-type: none"> 1. No change; evidence supports its consideration in patients with both aphonia and recurrent aspiration 2. The word <i>entrapment</i> is dropped and the phrase clarified 3. Selectively accepted

References

1. Institute of Medicine of the National Academies. Clinical Practice Guidelines We Can Trust: Standards for Developing Trustworthy Clinical Practice Guidelines (CPGs). Available at: iom.edu/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx. Released March 23, 2011. Accessed August 11, 2011.
2. Institute of Medicine of the National Academies. Finding What Works in Health Care: Standards for Systematic Reviews. Available at: iom.edu/Reports/2011/Finding-What-Works-in-Health-Care-Standards-for-Systematic-Reviews.aspx. Released March 23, 2011. Accessed August 11, 2011.
3. Kung J, Chiappelli F, Cajulis OO, et al. From Systematic Reviews to Clinical Recommendations for Evidence-based Health Care: Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. *Open Dent J* 2010;4:84–91.
4. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE Guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol* 2011;64:380–382.
5. Wijdicks EFM, Varelas PN, Gronseth GS, Greer DM. Evidence-based Guideline Update: Determining Brain Death in Adults. Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 2010;74:1911–1918.
6. Essaihi A, Michel G, Shiffman RN. Comprehensive Categorization of Guideline Recommendations: Creating an Action Palette for Implementers. Musen M, ed. *Proc Amer Med Informatics Assoc* 2003;Washington, DC:220–224.
7. Codish S, Shiffman RN. A Model of Ambiguity and Vagueness in Clinical Practice Guideline Recommendations. *AMIA Annu Symp Proc* 2005;146–150.
8. Schünemann HJ, Al-Ansary LA, Forland F, et al.; for the Board of Trustees of the Guidelines International Network. Guidelines International Network: Principles for Disclosure of Interests and Management of Conflicts in Guidelines. *Ann Intern Med* 2015;163:548–553.
9. Council of Medical Specialty Societies Code for Interactions with Companies. Available at: <https://cmss.org/wp-content/uploads/2016/02/CMSS-Code-for-Interactions-with-Companies-Approved-Revised-Version-4.13.15-with-Annotations.pdf>. Accessed September 15, 2016.
10. Guyatt G, Akl EA, Hirsh J, et al. The Vexing Problem of Guidelines and Conflict of Interest: A Potential Solution. *Ann Intern Med* 2010;152:738–741; p. 739.
11. Conflict of Interest Implementation for CancerLinQ. American Society of Clinical Oncology, Policy for Relationships With Companies 2013 JCO2013.49.5002. cancerlinq.org/sites/cancerlinq.org/files/Conflict-of-Interest.pdf. Accessed August 1, 2016.
12. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from cochrane-handbook.org. Accessed August 11, 2011.
13. Kung J, Chiappelli F, Cajulis OO, et al. From Systematic Reviews to Clinical Recommendations for Evidence-based Health Care: Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. *Open Dent J* 2010;4:84–91.
14. Shekelle PG. Updating Practice Guidelines. *JAMA* 2014;311:2072–2073.
15. Gronseth GS, Paduga R. Evidence-based guideline update: Steroids and antivirals for Bell palsy. Report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology* 2012;70:2209–2013.
16. Culebras A, Messé SR, Chaturvedi S, Kase CS, Gronseth G. Summary of evidence-based guideline update: Prevention of stroke in nonvalvular atrial fibrillation. Report of the Guideline Development Subcommittee of the American Academy of Neurology. *Neurology* 2014;82:716–724.



201 Chicago Avenue
Minneapolis, MN 55415

AAN.com

(800) 879-1960 or (612) 928-6000 (International)