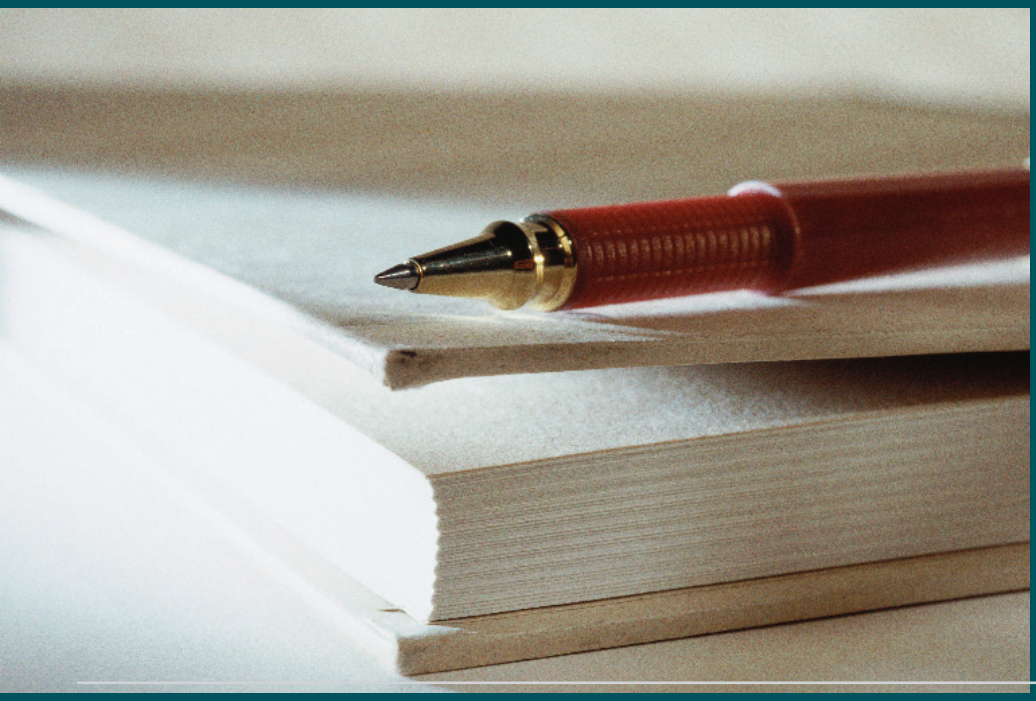
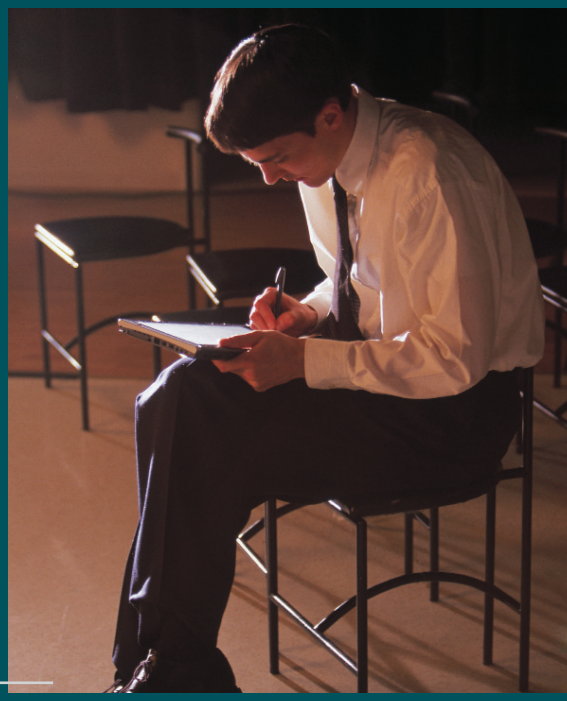
AMERICAN ACADEMY OF NEUROLOGY

CLINICAL PRACTICE GUIDELINE PROCESS MANUAL

2004 Edition





Prepared by

Wendy Edlund
Gary Gronseth, MD
Yuen So, MD, PhD, and
Gary Franklin, MD, MPH

For the

Quality Standards Subcommittee and the Therapeutics and Technology Assessment Subcommittee

CLINICAL PRACTICE GUIDELINE PROCESS MANUAL

2004 Edition

Prepared by Wendy Edlund; Gary Gronseth, MD; Yuen So, MD, PhD; and Gary Franklin, MD, MPH

For the

Quality Standards Subcommittee (QSS) and the Therapeutics and Technology Assessment Subcommittee (TTA)

For more information contact:
Wendy Edlund
Associate Director, Clinical Practice
American Academy of Neurology
1080 Montreal Avenue
St. Paul, MN 55116
(651) 695-2716
Fax: (651) 695-2791
wedlund@aan.com

The authors thank James Stevens, MD, FAAN, Michael Glantz, MD, FAAN and Richard Dubinsky, MD, MPH, FAAN and other members of the QSS and TTA for their careful review of drafts of this manuscript and Robert Miller, MD for his work in the preparation of a prior edition of this process manual.

Members of QSS

Jacqueline French, MD, FAAN (Co-Chair)
Gary Gronseth, MD (Co-Chair)
Charles E. Argoff, MD
Eric Ashman, MD
Christopher Bever, Jr., MD, MBA, FAAN
John England, MD, FAAN
Robert Holloway, MD, MPH, FAAN
Donald Iverson, MD, FAAN
Steven Messé, MD
Leslie A. Morrison, MD
Samuel Wiebe, MD
Dean M. Wingerchuck, MD, FAAN
Theresa Zesiewicz, MD, FAAN
Stephen Ashwal, MD, FAAN (Ex-Officio)
Deborah Hirtz, MD (Ex-Officio)
Gary Franklin, MD, MPH (Ex-Officio)
David Thurman, MD (Ex-Officio)

Members of TTA

Janis Miyasaki, MD, FAAN (Co-Chair)
Yuen So, MD, PhD (Co-Chair)
Richard Camicioli, MD
Vinay Chaudhry, MD, FAAN
Richard Dubinsky, MD, MPH, FAAN
Cynthia Harden, MD
Cheryl Jaigobin, MD
Irene L. Katzan, MD
Kenneth Mack, MD, PhD

Approved by the AAN Board of Directors January 2004

TABLE OF CONTENTS

INTRODUCTION

This manual has been created to provide instructions for developing an AAN clinical practice guideline. AAN practice guidelines consist of a formalized review of the literature that serves as the foundation for evidence-based practice recommendations. The literature review of a guideline is distinct from a typical literature review in that it is systematic and transparent. The recommendations of the guideline are also distinct in that they are fundamentally evidenced-based. Detailed instructions on how to complete a systematic review and formulate recommendations are outlined herein.

While reading this manual it is important to keep sight of the "big picture" of what is to be accomplished.

Essentially, authors are to identify a clinical question for which AAN members could benefit from evidence-based guidance. Then they are to answer the question by employing a methodology most likely to lead to the correct answer.

Asking and answering the question forms the backbone of both the guideline development process and the resulting manuscript. Both should clearly follow the progression of:

ASK A CLINICAL QUESTION

↓

FIND AND ANALYZE RELEVANT EVIDENCE

↓

STATE CONCLUSIONS

↓

MAKE RECOMMENDATIONS

First, authors identify a clinical question that needs to be answered. The question should address an area of quality concern, controversy, confusion, or variation in practice. The question must be answerable with sufficient scientific data. Answering the question must have the potential to improve clinical care and patient outcomes.

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

Third, the authors make conclusions that synthesize and summarize the evidence in answer to the clinical question.

Finally, the authors 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 graded based on the quality of supporting data.

OVERVIEW OF AAN CLINICAL PRACTICE GUIDELINES

Overall Objective: The AAN develops clinical practice guidelines to assist its members in clinical decision making—particularly in situations of controversy or variation in practice.

Background: Both the Quality Standards Subcommittee (QSS) and the Therapeutics and Technology Assessment Subcommittee (TTA) develop practice guidelines using the processes discussed in this manual. Both subcommittees form expert panels to critically assess all of the relevant literature on a given topic or technology. Evidence is rated based on quality of study design, and clinical practice recommendations are developed and stratified to reflect the quality and the content of the evidence. The QSS develops practice parameters, which are guidelines with a patient-centric focus. The TTA develops technology assessments and practice advisories, which are guidelines with an intervention-centric focus.

Key Audiences:
Primary: Neurologists
Secondary: Patients, payers, federal agencies, (e.g., CMS), other healthcare providers, clinical researchers

Definitions:

- **Practice Parameters** are strategies for patient management that assist physicians in clinical decision making. A practice parameter is a series of specific, evidence-based practice recommendations that answer an important clinical question (e.g., What pharmacological interventions reduce sialorrhea in patients with ALS?).

- **Practice Advisories** present evidence-based practice recommendations for emerging and/or newly approved therapies or technologies based on evidence from at least one Randomized Controlled Trial (RCT). The evidence may demonstrate only a limited clinical response, the statistical evidence may be weak, significant cost-benefit questions may exist, or there may be substantial (or potential) disagreement among practitioners or between payers and practitioners (e.g., Based on initial studies, is rt-PA safe and effective in the treatment of stroke?).

AAN Guideline Development Process

Select Guideline Topic
↓
Form Balanced Panel of Experts
↓
Develop Clinical Questions
↓
Comprehensively Review the Literature
↓
Rate Articles and Summarize Findings
↓
Write Guideline that Makes Explicit, Supported Practice Recommendations
↓
Distribute Draft for Extensive Peer Review by AAN, *Neurology*, and Others
↓
Obtain AAN Approval of Guideline

- **Technology Assessments** are statements that assess the safety, utility, and effectiveness of new, emerging, or established therapies and technologies in the field of neurology (e.g., What is the safety and utility of SPECT in evaluating neurologic patients?)

Common Uses of AAN Guidelines

- Improve health outcomes for patients
- Stay abreast of the latest in clinical research
- Provide medico-legal protection
- Advocate for fair reimbursement
- Determine whether one's practice follows current, best evidence
- Reduce practice variation
- Affirm the role of neurologists in the diagnosis and treatment of neurological disorders
- Influence public or hospital policy
- Promote efficient use of resources
- Identify research priorities based on gaps in current literature

1. TOPIC DEVELOPMENT

Topic Development takes one to six months, and produces a Project Development Plan—the project blueprint—for accepted topics.

1.1 Topic Nominations

Any AAN member, Committee, Section, or an outside organization (e.g., an organization responsible for generating health policy) may request the development of a guideline. All topic suggestions must be submitted in writing in the form of a justification statement (see Table 1).

Topics are evaluated quarterly by the QSS/TTA Topic Review Panel based upon neurologists' need for guidance, the availability of evidence to provide guidance, and the potential to improve patient care and outcomes.

The Topic Review Panel assigns accepted topics to either QSS or TTA based on whether the topic is best addressed from the perspective of the patient (QSS) or a technology or therapy (TTA).

QSS and TTA initiate projects based upon the criteria listed in Table 1 and the availability of resources.

1.2 Forming the Author Panel

QSS or TTA assign a committee member to serve as the project facilitator. The facilitator identifies a lead author. Together the facilitator and lead author assemble an author panel, being careful to seek a variety of perspectives and to avoid bias.

The author panel usually consists of five to ten individuals. Under most circumstances the panel should include members with expertise relevant to the topic, including panel members that are nationally recognized experts on the topic being addressed (i.e., have authored clinical publications in high impact journals).

It may also be useful to appoint a general neurologist and/or an expert in guideline methodology to the panel.

The facilitator and lead author should also seek input from other medical specialties, as appropriate. This can be accomplished through formal collaboration with another organization (to be coordinated by AAN staff) or by appointing non-neurologists to the panel without formal collaboration with corresponding medical specialty organizations.

Table 1
Completing the Justification Statement

The justification statement should be approximately one double-spaced page in length.

It should outline the problem to be addressed and include examples of specific clinical questions that could be answered in the guideline.

Because availability of evidence is crucial to the project's success, it is essential to include sample citations that represent the best available evidence on the topic.

The justification statement should examine the following factors, on which QSS or TTA will base its decision for approval:

- Relevance to neurology
- Prevalence of condition
- Health impact of condition for the individual
- Socioeconomic impact
- Extent of practice variation
- Quality of available evidence
- External constraints on practice (e.g., access issues, reimbursement issues, paucity of data for setting policy, health policy gaps, resource constraints)
- Urgency for evaluation of new therapy or technology
- Potential for significant benefit, risk, or abuse

It may be difficult to form an expert panel devoid of potential conflict, thus it is **essential to balance the panel** between those with and without conflicts (financial, research, academic, etc.). The AAN forbids commercial participation in guideline projects.

The author panel roster is communicated to QSS or TTA at its next meeting. Any subsequent changes to the author panel must be communicated to QSS or TTA.

Information regarding AAN guideline projects, including contact information of lead authors, is shared with the Cochrane Collaboration; their staff may contact authors to inform them of pertinent Cochrane reviews. AAN Sections may also be alerted to the development of the guideline and provided an author panel roster.

Conflict of Interest
Panel members must sign a conflict of interest statement (Appendix 3). All real or potential conflicts for the past five years must be noted; conflicts will be disclosed in the guideline.

Authorship
All participating panel members, including the facilitator, are listed as authors. The lead author and facilitator determine the order of authorship and arbitrate any questions regarding who qualifies for authorship.

Roles and Responsibilities
Facilitator: Member of QSS or TTA assigned to help guide the project. Provides advice on process issues—particularly the classification of evidence and translation of evidence to practice recommendations. Reports to QSS or TTA on project progress.

Lead Author: Project chair. Sets timeline, assigns tasks to panel members, and coordinates activities (e.g., literature review and drafting the guideline).

Author Panel Member: Active participant in the project. Usually reviews articles, classifies evidence, and writes portions of the document.

AAN Staff: Provides administrative support and advice, facilitates meetings and group communications, coordinates resource allocation (e.g., medical librarian), and liaisons to the journal and approval bodies.

1.3 Completing the Project Development Plan

A Project Development Plan (PDP) worksheet is provided in Appendix 1. The PDP provides a framework for each panel to define the project and receive feedback at an early stage in the process. The following information is presented in the completed PDP:

- Potential clinical questions
- Terms and databases to be used in the literature search
- Inclusion and exclusion criteria for article selection
- Project timeline

Developing Potential Clinical Questions

Statement of the Clinical Question
Authors should select questions that can be answered based on published, peer-reviewed evidence. It may be helpful to perform a preliminary literature search to determine the availability of evidence to answer the questions being considered.

Clinical questions should have three basic components:

1. **Population:** The type of person (patient) involved
2. **Intervention:** The type of exposure that the person

experiences (therapy, test, risk factor, prognostic factor, etc.)
3. **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 ALS. The population of interest may also consist of patients at risk for a disease, for example patients with suspected multiple sclerosis or people at risk for stroke.

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). Similarly, it may be necessary to explicitly indicate that the population of interest includes or excludes children.

Intervention
The intervention defines the treatment or diagnostic procedure being considered. The question almost always asks if this intervention should be done. For example, should patients with new onset Bell's palsy be treated with steroids?

An example from the perspective of a diagnostic question would be, should patients with new onset Bell's palsy routinely receive brain imaging?

More than one intervention can be explicitly or implicitly listed in the question. For example, in patients with ALS which interventions improve sialorrhea? This more general question implies that authors will look at all potential interventions for treating sialorrhea.

It may be important to be highly specific in defining the intervention. For example, authors might specify a specific dose of steroids for the treatment of Bell's palsy of interest. Similarly, authors 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.

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 because they rarely predict clinically important outcomes accurately. 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. Guidelines on treatments should measure adverse effects, as well as beneficial effects.

In addition to defining the outcomes that are to be measured, the clinical question should state *when* the outcomes should be measured. The interval must 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 three elements are easily identified. For example:

Population: For patients with Bell's palsy

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

Outcome: improve long-term facial functional outcomes.

Examples of focused, answerable

Table 2
Types of Clinical Questions: Therapeutic, Diagnostic and Prognostic Accuracy, and Screening

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

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 best study for determining the effectiveness of a therapeutic intervention is the masked, randomized, controlled trial.

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 if she should perform an intervention to determine the presence of the disease, or to determine the 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 prognosis of the disease. 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? There is clearly an intervention of interest in this question: facial nerve conduction studies. There is also an outcome: an improved ability to predict the patient's long-term facial functioning. The answer to this question would go a long way in helping clinicians to decide if they should offer facial nerve conduction studies to their patients with Bell's palsy.

The best study for measuring the accuracy of facial nerve conduction studies for determining prognosis in Bell's palsy would not be a randomized controlled trial. Rather, it 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 those of prognostic accuracy. 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 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 if 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 form patients with other causes of peripheral facial weakness, we would recommend that clinicians routinely test taste in this clinical setting.

There is another common type of important clinical question worth considering. These questions have a diagnostic flavor but they are more concerned with diagnostic yield than diagnostic accuracy. This type of question is appropriate to the situation where a diagnostic intervention of established accuracy is employed. For example: In patients with new onset peripheral facial palsy should we routinely obtain head MRIs to identify sinister pathology within the temporal bone causing the facial palsy? There is no issue regarding the diagnostic accuracy of head MRI in this situation. The diagnostic accuracy of MRI in demonstrating temporal bone pathology is established. The clinical question here 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: how often does the MRI reveal 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 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.

Determining early the type of question to be asked for a guideline is critical for guiding the process. The kind of evidence to be sought to answer the question and how that evidence will be graded relative to quality follow directly from the type of question.

clinical questions are presented in the PDP in Appendix 1; further description of the types of clinical questions is presented in Table 2.

Scope of the Question

The scope of the question can be relatively broad or narrow. Overall, the AAN seeks focused, answerable clinical questions for guidelines. A focused question makes the project more manageable and leads to recommendations that are more pertinent to clinical care.

Guidelines are not textbooks on how to diagnose and manage particular diseases. Rather, they are summaries of the published literature pertinent to specific aspects of care.

The clinical question should address features of the patients and interventions that are believed to significantly affect outcome. Taking too narrow of a focus may limit the amount of data in the review and thereby increase the risk for false-positive and false-negative results.

Revising the Clinical Question
Although the clinical question, and thus the criteria for what is to be addressed in the guideline, must be set before data collection begins, it may be necessary to revise the question based on the availability of data.

Care should be taken to avoid making changes to the clinical question that would be likely to introduce bias.

For example, the question should not be changed on the basis of the results of individual trials. It may, however, be reasonable to change the criteria if alternative, acceptable ways of defining the study population or intervention are discovered.

Selecting the Search Terms and Databases
The second and third sections of the PDP are devoted to developing the search strategy. First, it is essential to identify the search terms and databases that will result in capturing the articles that can best answer the clinical questions.

Preliminary Literature Search
Authors are encouraged to perform a preliminary literature search in order to become familiar with the breadth of literature available on the topic. This will 1) assist with the identification of search terms and search strategies, and 2) identify a set of articles against which to check the accuracy and completeness of future searches. The authors may also contact the AAN's medical librarian, who can identify and suggest appropriate terms and databases, as well as ensure a broad and inclusive search.

Search Terms
It is incumbent on the author panel to 1) define terms, 2) identify synonyms, acronyms, and special jargon, and 3) ensure that all elements of the search question are identified and the relationships between the concepts are described. Authors should be sure to include appropriate synonyms from other nationalities and disciplines.

Medical Subject Headings (MeSH) terms, a controlled vocabulary, should be used specifically for searching MEDLINE. Several MeSH terms for common concepts in evidence-based medicine are identified in Appendix 4. Authors should pair relevant terms from that list with MeSH vocabulary representing the particular disease entity, patient population, transaction, and/or desired outcomes being investigated. These terms may be augmented by terms representing quality of life or psychological aspects.

In some cases the MeSH term should be "exploded" in order to retrieve more specific related terms, e.g., clinical trial (exploded) would also retrieve clinical trial, phase I; clinical trial, phase II; etc. MeSH also has subheadings that describe frequently discussed aspects of a subject. In addition, MEDLINE includes useful "publication types" (e.g., controlled trial, review, etc.) which can be included in the search. MeSH vocabulary can also be supplemented by text words for further searching of MEDLINE or other databases. Please be aware that MEDLINE indexers are usually not experts in the topic or

research methodology and may make indexing mistakes.

Databases
The PDP must stipulate which medical databases will be searched.

A MEDLINE search will likely uncover only 30 to 80% of published RCTs on a topic. Therefore, it is recommended that authors search MEDLINE, EMBASE, and Science Citation Index or Current Contents. (See Appendix 5.)

In consultation with a professional medical librarian, the author panel should determine whether it is appropriate to search additional databases, based on the topic being investigated. 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. A brief description of the major databases is provided in Appendix 5.

Selecting the Inclusion and Exclusion Criteria

The author panel should develop criteria for including or excluding articles during the literature search and article review processes. This is an essential step in the process, and should be adapted to the specific clinical issue being addressed.

The criteria must be developed prior to beginning the search. 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.

The author panel should develop an explicit list of inclusion and exclusion criteria by evaluating each of the following issues and any other issues that are pertinent to the specific topic being addressed. The QSS or TTA facilitator can provide valuable assistance in completing this portion of the PDP.

Languages
Authors are urged to include all languages in the search, unless there is a specific reason for excluding non-English articles. English abstracts are available for many non-English articles. It is usually possible to obtain a translation of a non-English paper through a university or the Internet.

Type of Subjects
Usually, the search is limited to papers concerned with human subjects. However, for some topics, it may be appropriate to include experimental articles from the laboratory.

Authors should also state whether studies pertaining to related diseases should be sought (e.g., sialorrhea in cerebral palsy for a guideline on the management of sialorrhea in ALS). Depending upon the condition, issues surrounding diagnostic criteria may require clarification, as well.

Intervention
The type of intervention should be explicit.

Outcome Measures
Outcome measures that will be examined should be included. Authors 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 randomized controlled trials (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. Authors should only use methodological selection criteria if it will result in obtaining articles that are clearly superior.

Peer Review and Publication
Articles are eligible for inclusion in AAN guidelines if they have been published in a peer-reviewed journal; supplements and book chapters should not be used. Exceptions to this rule can occur if 1) authors evaluate the data presented in the non-peer-reviewed source using the AAN classification of evidence system, and 2) the data is not available in a peer-reviewed source. Authors will be asked to justify the inclusion of the non-peer-reviewed references.

In-press articles may be included if they will be published prior to the guideline and authors are able to review the data.

Relevance
The study must be relevant to the clinical question.

Setting the Project Timeline

A worksheet is provided on the PDP to help formulate the project timeline. AAN staff use the dates provided to develop an official project timeline that takes into account upcoming committee meeting dates and the availability of resources.

1.4 Submission of Project Development Plan

The completed PDP is submitted to staff and the facilitator.

The facilitator carefully reviews the plan and suggest revisions, as appropriate.

Staff formalizes the timeline and allocates resources to the project, including opening an account with the AAN's medical librarian.

2. DATA REVIEW AND ANALYSIS

Upon approval of the PDP, the author panel completes the literature review and data analysis steps of the project. These steps take one to six months to complete and produce first a master list of articles to be included in the guideline, then an evidence table outlining the key characteristics, quality, and results of each selected study.

2.1 Performing the Literature Search

Authors should execute the search strategy outlined in the PDP, as follows.

Consulting a Research Librarian
One panel member should complete the literature search in consultation with an AAN-appointed professional research librarian. To ensure that the guideline is based upon the best evidence, the librarian should run comprehensive searches on several major databases, interpret all aspects of the clinical question, interactively query the databases to define and refine the search and then apply quality filters to the results.

The literature search results should be obtained as a list of titles and abstracts. Authors who use EndNote or other reference management software (see Table 3) are encouraged to receive and track the literature search results electronically. The notes fields in these programs are helpful to track the review and classification of evidence processes.

Table 3
A Note on Bibliographic Management

It is possible to manage the references manually. However, authors are urged to utilize reference management software, such as EndNote or Reference Manager.

Such software places citations and abstracts into a maneuverable database so that authors can easily access, track and reference selected articles.

Possible uses include:

- Importing items or other documents into the database
- Searching the database
- Copying and inputting citations into the document
- Reformatting the citation
- Placing the fields in the order and with the punctuation desired
- Identifying and eliminating duplicates
- Cutting and pasting to create a bibliography
- Making personal annotations to citations
- Identifying key words, scanning the database to search for key words
- Applying the key words to new articles that are brought in
- Grouping articles according to levels of evidence or other criteria
- Tracking which articles authors have in printed format.

Documenting the Literature Search

It is essential that the search be carefully documented and reported in the guideline. The documentation should include the following information:

- Date searches were conducted
- Question that was posed
- Definition of terms
- Databases searched
- Dates included in search
- History of what was searched (terms and combinations of terms)
- Explicit description of the inclusion and exclusion criteria

Authors should also document the evaluation and decision-making process for including or excluding articles, the success of the search, and any revisions or modifications to the search. This is important for 1) ensuring the methods presented in the paper are reproducible, and 2) for post-publication determinations of whether the guideline should be updated.

Evaluating the Accuracy of the Literature Search; Identifying Additional Articles

Upon receipt of the search results, the lead author should critically evaluate the quality and accuracy of the search.

Authors should:

- Ensure the articles are on target and no essential concepts related to the question were missed
- Ensure that all of the articles identified in the preliminary search are included in the results
- Have panel members identify additional relevant articles (published or in press)
- Identify additional articles from reference lists, particularly the reference lists of review articles and guidelines
- Determine whether it is necessary to broaden or narrow the search
- Ensure that new or changed aspects of the question are accounted for in follow-up searches

2.2 Selecting Articles for Inclusion

A two-step process (Table 4) is used to exclude articles that do not meet the inclusion criteria. All identified abstracts are reviewed for relevance to the clinical question and adherence to the inclusion criteria. Then the same process is applied to the selected articles.

AAN staff is available to help distribute the abstracts and articles and track panel member responses.

Reviewing the Abstracts

Every abstract should be reviewed by at least two panel members. The lead author may select two panel members to review all abstracts, or distribute the abstracts evenly among all panel members. It is essential that the inclusion and exclusion criteria be distributed to the panel members with the abstracts.

Panel members review the abstracts and determine which are pertinent to the clinical question and meet the inclusion criteria. It is best to be inclusive at this stage in the process—if an article may be relevant, it should be obtained. If it is unclear whether the article meets the inclusion criteria, it should be obtained.

Panel members submit a list of articles to be obtained to the panel chair. The panel chair then develops a master list of articles to be obtained. It is not necessary to settle disagreements between the two reviewers for each abstract; it is best to obtain any article considered to meet the inclusion criteria by either reviewer.

The panel chair should be careful to document the number of abstracts reviewed and the number of abstracts excluded.

Obtaining and Reviewing the Articles

AAN staff obtains and distributes the selected articles. Each article should be read independently by two panel members. The panel chair may choose to distribute the articles randomly or according to topic.

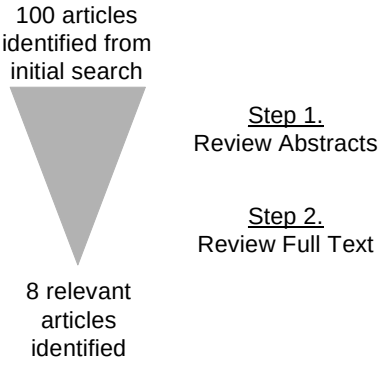
Panel members should review each article for pertinence to the clinical question and adherence to the inclusion criteria set forth in the PDP. This is a screening review of the article; panel members are not extracting data at this point. It is best to be exclusive at this stage in the process. If it 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.

Panel members submit a list of articles to be included in the guideline to the panel chair. The panel chair compiles a master list of articles to be included and resolves any disagreements regarding inclusion of individual articles.

The panel chair distributes the master list of articles to the full author panel; panel members should refer to the criteria listed in section 2.1, "Evaluate the Accuracy of the Literature Search; Identify Additional Articles," to ensure that all relevant articles have been identified.

AAN staff sends a copy of all selected articles to all panel members.

Table 4
Two-step literature review process



2.3 Data Extraction and Classification of the Evidence

The extraction of data and classification of evidence are crucial tasks. Many of the concepts to be discussed in this section may be unfamiliar to panel members without a methodological background. Panel members should seek the assistance of the facilitator in completing these steps, as necessary.

The study characteristics—or elements—to be extracted from each article vary depending on the clinical question. Some generalizations can be made, however. In general, the characteristics extracted will fall into one of the following categories:

- Citation information
- Items relevant to the study generalizability
- Elements relevant to the quality of evidence contained within the study
- Elements relevant to the study outcomes

Citation Information

Citation information should be extracted from each article. This should include:

- Name of the first author
- Year of publication
- Journal
- Country in which study performed

Elements Relevant to Generalizability

Authors should also extract from the studies elements that assist in the judgment of the relevance of the study to the clinical question and the generalizability of the results. They can be directly related to the three elements of the clinical question.

Elements relating to the patient population should include:

- Source of patients (e.g., neuromuscular referral center)
- Inclusion criterion
- Age of the patients (e.g., mean and standard deviation)
- Gender of the included population (e.g., proportion female)

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

- Dose of medication used
- Timing of the intervention
- Nature of the 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:

- Scale used to determine the outcome (e.g., 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 should also be extracted. The items extracted will vary according to the type of question.

For therapeutic questions, critical elements include:

- Use of a comparison (control) group
- Method of treatment allocation (randomized versus 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:

- Study design (case-control versus cohort survey)

- Spectrum of patients included (narrow spectrum versus wide spectrum)
- Proportion of patients where 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:

- Study design (prospective vs. retrospective)
- 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)

Patient Relevant Outcomes

Finally, patient relevant outcomes need to be extracted. These consist of a quantitative measure of the outcome of interest. Regardless of the type of question, clinically relevant outcomes are usually 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.

For a therapeutic question, quantitative outcomes in the treated population are usually measured relative to an untreated population. Common measures of effectiveness include the relative rate of an outcome (e.g., the proportion of patients with good facial outcomes in patients with Bell's palsy receiving steroids divided by the proportion of good outcomes in those not receiving steroids), or the rate difference (e.g., the proportion of patients with good facial outcomes in patients with Bell's palsy receiving steroids minus the proportion of good outcomes in those not receiving steroids.) See Appendix 7 for examples on calculating these effect measures.

For articles of diagnostic or predictive accuracy, relative risks, positive and negative predictive values, likelihood ratios, and sensitivity and specificities are the outcome variables of interest. See Appendix 7 for examples on calculating these accuracy measures. For screening procedures, the quantitative measure of effect will be the proportion of patients with a clinically significant abnormality identified. (See Appendix 7). Regardless of the type of clinical question or the outcome variable chosen, it is critical that some measure of random error (i.e., the statistical power of each study) be included in the estimate of the outcome. Including 95% confidence intervals of the outcome measure of interest is usually the best way of measuring the amount of random error within each study.

Sometimes authors of the studies being considered might not have calculated the pertinent outcome measures or their confidence intervals. In such circumstances, panel members need to calculate them. In doing so, guideline authors are encouraged to seek help from the facilitator or the methodological experts on the subcommittees. Additionally, the companion CD of this handbook contains an Excel spreadsheet that will make these calculations for you (to be available soon).

Developing a Data Extraction Form
The extraction of the study characteristics described above can be facilitated by development of a data extraction form. The panel chair develops a data extraction form to apply to each article identified for inclusion. Sample data extraction forms are provided in Appendices 7 and 8.
It may be helpful for the facilitator or a member of the QSS or TTA to hold a conference call with all panel members to provide instruction prior to the start of data extraction.

Generally, the form should include the following:

- Name of first author
- Citation information: date of publication, journal
- Country of completion of work
- Study type (RCT, CCT etc.)
- Conclusions
- Methods of statistical evaluation
- Patient characteristics (age, gender, inclusion, exclusion)
- Therapeutic intervention (specific drug used, sensitivity analysis, dose/regimen)
- Fidelity and monitoring of treatment (adherence/compliance, loss to follow up and dropouts)
- Outcomes (patient related, adverse effects)

Authors should extract data from each article that was selected for inclusion using the data extraction form.

Panel members should submit the completed data extraction sheets to the panel chair. Disagreement regarding the extracted elements, the classification of evidence, or assessment of effect size should be resolved by consensus among panel members.

Classifying the Evidence

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

For AAN guidelines, the risk of bias in included studies is measured using a four-tiered classification scheme (appendix 9). In this scheme, studies graded class I are judged to have a low risk of bias; studies graded class II are judged to have a moderate risk of bias; studies graded class III are judged to have a moderate to high risk of bias; studies graded class IV are judged to have a very high risk of bias. The classification grade is also known as the level of evidence.

Panel members assign a classification for each study based on that study's extracted quality-of-evidence characteristics.

The classification scheme employed by the AAN accounts for systematic error only. Random error (low study power) is dealt with separately using confidence intervals.

The risk of bias of a study can only be judged relative to a specific clinical question. Therapeutic, diagnostic or prognostic accuracy, and screening questions are judged by different standards.

Appendix 9 describes the study characteristics needed to attain the various risk-of-bias grades in paragraph form. The next four sections explain in more detail each study characteristic (or element) that contributes to a study's final classification for each of the four

types of studies (therapeutic, diagnostic, prognostic, screening).

Classifying Evidence for Therapeutic Questions

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

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 graded class IV.

To be graded class II, studies should use concurrent controls. Studies using non-concurrent controls, such as those using patients as their own controls (e.g., a before-after design) or those using external controls, are graded class III.

Treatment Allocation

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

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

Additionally, to be graded class I, panel members should assure themselves that the randomization scheme effectively balanced the treatment and comparison group for confounding baseline differences.

Finally, panel members should be convinced that the allocation process was sufficiently concealed so that investigators could not manipulate treatment assignment.

Occasionally, panel members will encounter an article where instead of assigning patients to treatment or comparison groups randomly, investigators attempt to match each treated patient with an untreated, comparison patient with similar baseline characteristics. Such matched studies are graded class II.

Completeness of Follow-Up

Patients enrolled in studies are sometimes lost to follow-up. Losses to follow-up occur for nonrandom reasons. Such losses may introduce confounding differences between the treated and untreated groups. Thus, to be graded class I, more than 80% of patients within the study should have complete follow up.

For various reasons, sometimes patients initially assigned to the treatment group do not receive treatment and patients assigned to the comparison group receive treatment. If patients crossover 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, this means the investigators analyze the results according to whichever group (treated or comparison) the patient was originally assigned.

Masking

For a study to be graded class II or I, an investigator who is unaware of the patient's original treatment assignment must determine the outcome. To be graded class III, an investigator who is not part of the treatment team (i.e., independent) must determine the outcome. The requirement for masked

or independent assessment can be waived if the outcome measure is objective. An objective outcome is one that is unlikely to be affected by expectation bias (e.g., survival or a laboratory assay).

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 without the disease or outcome of interest. Quantitative measures of accuracy cannot be calculated from studies without a comparison group. These studies are judged to have a high risk of bias and are graded 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 would have a positive test, others a negative test. The cohort would then have the actual presence or absence of the disease 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.

Studies of diagnostic accuracy are often done backwards. Rather than starting with a group of patients suspected of having the disease, investigators often start by selecting a group of patients who clearly have the disease (cases) and a group of patients who clearly do not have the disease (control). The test is then performed on both cases and controls and measures of diagnostic accuracy are calculated. Although this case-control study is easier to execute, its retrospective design introduces several potential biases. Thus, at best, such studies can only be graded class II.

Patient Spectrum

One of the dangers of the case-control design is that sometimes only patients who clearly have the disease or clearly do not have the disease might be included. Including such unambiguous cases can exaggerate the diagnostic accuracy of the test. To avoid this, it is important for a study employing a case-control design to include a wide spectrum of patients. A wide spectrum would include patients with mild forms of the disease and patients with clinical conditions that could be easily confused with the disease. Studies employing a case-control design with a wide spectrum of patients can be graded class II, those with a narrow spectrum, class III.

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 true presence of the disease or outcome. This reference standard should be independent of the diagnostic test or prognostic predictor in question. The reference standard could consist of pathological, laboratory, or radiological 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 grade studies without a valid reference standard class IV.

Completeness
Ideally, all patients enrolled into the cohort should have the diagnostic test result (presence of the prognostic variable) and the true presence or absence of the disease (outcome) measured. A study should be downgraded to class II if less than 80% of subjects have these variables measured.

Masking
For a study to be graded class II or I, 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 (outcome). In the case of the case-control design, to obtain a class II grade, an investigator who is unaware of the presence or absence of the disease (outcome) should perform the diagnostic test (measure the prognostic predictor) of interest.

To obtain a grade of class III, the investigators performing the diagnostic test (or measuring the prognostic predictor) should be different than the investigator who determines the true presence or absence of disease (or the outcome).

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

Classifying Evidence for 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 sub-optimal, incomplete measurement. Thus, data collection should be prospective to classify a study class I or II.

Setting
Studies are often performed by highly specialized centers. Such centers, because they tend to see more difficult and unusual cases, are often non-representative of the patient population considered in the clinical question. In general, because of the potential non-representativeness of patients, panel members should grade studies from referral centers class III. (Occasionally, the screening question's population of interest is primarily patients referred to specialty centers. For example, some rare or difficult-to-treat conditions may only be managed at referral centers. Under these circumstances, such studies can be graded class II.)

Patients recruited from non-referral centers such as primary care clinics or general neurology clinics are more representative. These studies can be graded class II. Population-based studies tend to be the most representative and can be graded class I.

Sampling
The ideal methods of selecting patients for a study designed to answer a screening question are to 1) take all patients or 2) take a statistical sample of patients. This ensures that the patients are representative. Thus, a consecutive sample, a random sample, or a systematic sample of patients (e.g., every other patient) warrant a class I or II grade. Because patients may potentially be non-representative, a study using a selective sample of patients can only be graded class III. For example, a study looking at the yield of MRI in patients with headache that

included patients who happened to have head MRIs ordered would be class III. This sample is selective. A study performing MRIs on all consecutive patients presenting with headache is not selective and would earn a class II or I grade.

Completeness
For reasons similar to that discussed under sampling, it is important that all patients included in the cohort undergo the test of interest. If less than 80% of subjects receive the intervention of interest, the study can be graded no better than class III.

Masking
To be graded class I or II for a screening question, the interpretation of the intervention of interest (usually a diagnostic test) should be done without knowledge of the patient's clinical presentation. To attain a class II grade, someone other than the treating physician should do the interpretation of the diagnostic test.

The requirement for independent or masked assessment can be waived if the interpretation of the diagnostic test is unlikely to be changed by expectation bias (i.e., is objective).

2.4 Development of the Evidence Tables

The author panel should develop evidence tables with the data extracted from each study using the data extraction forms. The rows of the table correspond to each included study. The headings of the table correspond to the extracted study characteristics. It is essential to include the class of evidence determined for each study. Example evidence tables can be found in Appendix 10.

Potential table headings are provided below:

- Author, year
- Level of evidence (Class I, II, III, or IV)
- Main purpose of study
- Study population: N, gender, mean age, diagnosis
- Intervention
- Outcome measures
- Results
- Number needed to treat and number needed to harm

It is recommended that the tables be created in Microsoft Excel for easy manipulation.

The evidence tables should be submitted to QSS or TTA with each draft of the guideline.

2.5 Formulating Conclusions

The goal at this step in the process is to develop a succinct statement that summarizes the evidence in answer to the specific clinical question. Ideally, this summary statement should indicate the magnitude of the effect and the level of evidence that it is based upon. The conclusion should be formatted in a way that clearly links it to the clinical question.

For example, in answer to the clinical question:

> For patients with new onset Bell's palsy,
>
> Do oral steroids given within the first three days of onset
>
> Improve long-term facial outcomes?

The conclusion may read:

> For patients with new onset Bell's palsy,

> Oral steroids given within the first three days of onset of palsy
>
> Are probably safe and effective to increase the chance of complete facial functional recovery (rate difference 12%) (two class I and two class II studies).

In this example, the level of evidence upon which the conclusion is based is indicated in two ways: 1) the term "*probably* safe and effective" indicates that the effectiveness of steroids is less than certain, and 2) the number and class of evidence upon which the conclusion is based are clearly indicated in parentheses.

The level of certainty directly relates to the *lowest* class of evidence used to develop the conclusion. Thus, if the conclusion were based on…

- Class I studies only, it would read:

  > Are *established as* safe and effective…

- Class II studies only or class I and II studies, it would read:

  > Are *probably* safe and effective…

- Class III studies only, or a combination of class III, II, and I studies, it would read

  > Are *possibly* safe and effective…

- Class IV studies, there is insufficient evidence to support a conclusion of effectiveness (or lack of effectiveness), thus it would read:

  > For patients with new onset Bell's palsy, there is insufficient evidence to determine if oral steroids…are effective in improving facial functional outcomes.

Four kinds of information need to be considered when formulating the conclusion:

- The class of evidence
- The effect (was the study positive or negative)
- Random error (the power of the study as manifested by the width of the confidence intervals)
- The consistency between studies

When all of the studies demonstrate the same positive result, are of the same class, and are consistent with one another, developing the conclusion is straightforward.

Often, however, this is not the case.

Conflicting Evidence

Consider a hypothetical example where the search strategy identified one class I study, one class II study, and one class III study looking at the effectiveness of steroids in Bell's palsy. The class I study is positive and the class II and III studies are negative. What should the author panel do? One approach would be to treat each study like a vote. Since the majority of studies (2/3) show no benefit, they 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 error or random error.

The authors should consider systematic error first. In this example, the different risks of bias in 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 discarded, and, if possible, the conclusion formulated should be based solely on the class I study.

The conclusion would be worded:

> Oral steroids are *probably* safe and effective to…

(The "probably effective" conclusion is supported when there is a *single* class I study used to formulate the recommendation. If we changed this example slightly and included *two or more* positive class I studies, the conclusion would read "established as effective.")

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 benefit from steroids and two studies did not.

Systematic error does not obviously explain the difference, since all studies are class I. Therefore, the authors must consider the random error of the studies by looking at the confidence intervals. If the confidence intervals of all of the 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. The solution in this circumstance is to reduce the random error. Meta-analysis is a technique that reduces random error (but not systematic error). In this circumstance, the combined estimate of the effect of steroids should be used to develop the conclusions.

The level of the conclusion (established, probably, 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 using the terminology "established as effective."

Methodological experts on the committee can help authors perform a meta-analysis, if necessary. Alternatively, authors can use the spreadsheet on the included CD-ROM (to be available soon).

Another situation to consider is if all three of the class I studies in the hypothetical 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 "established as ineffective." 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 power 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.

Consider a third hypothetical example. Here the search strategy identifies five articles looking at the effectiveness of steroids in Bell's palsy. Two studies are class I, two class II, and one class III. The studies are inconsistent in that the class III and class II studies are positive and the class I studies are negative.

The authors should first consider systematic error and consider the studies with the lowest risk of bias—the class I studies. They should next consider random error within these studies. 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 narrows the confidence intervals, but the combined confidence interval is still too wide to exclude a benefit.

Next authors should consider the class II studies. They should perform a meta-analysis that includes both class I and class II studies. The meta-analysis shows a statistically significant benefit of steroids. They can now formulate a conclusion.

The example conclusion used at the beginning of this section would be appropriate to this evidence. Because class II evidence was used in the conclusion formulation, "probably effective" is used to indicate the level of certainty.

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 concluded that there is insufficient evidence to make conclusions.

Combining studies in a meta-analysis is often a useful way to reduce random error. However, differences in study design, patient populations, or outcome measures sometimes make combining studies in a meta-analysis inappropriate. Methodological experts of the subcommittees can guide panel members in this situation.

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 screening questions. The conclusions are worded slightly differently in that the term "useful" is substituted for "effective." Thus, a conclusion regarding the prognostic accuracy of facial compound motor action potential in identifying patients at increased risk of poor facial function might read:

> For patients with new onset Bell's palsy
>
> The measurement of facial compound motor action potentials is probably useful
>
> To identify patients at increased risk for poor facial functional recovery (sensitivity 85%, specificity 75%) (three class II studies).

2.6 Formulating Recommendations

The formulation of recommendations flows from the conclusions. Similar to conclusions, recommendations are best formatted in a way that clearly shows how they answer the clinical question.

For example, a recommendation resulting from the Bell's palsy conclusion presented in section 2.5 would read:

> For patients with new onset Bell's palsy
>
> Clinicians *should consider* giving oral steroids within the first 3 days of palsy onset
>
> To improve facial functional outcomes (Level A).

Additional factors to be considered when formulating the recommendations (that are not considered in formulating

conclusions) include the *magnitude* of the effect (benefit or harm of therapy, accuracy of tests, yield of studies) and the relative *value* of various outcomes.

Under most circumstances, there is a direct link between the level of evidence used to formulate conclusions and the strength of the recommendation. This linkage is illustrated in Appendix 9.

Thus, an "established as" (two class I) conclusion supports a "should be done" (level A) recommendation; a "probably effective" (two class II) conclusion supports a "should be considered" (level B) recommendation; a "possibly effective" (two class III) conclusion supports a "may be considered" recommendation.

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 would read, "should *not* be done." There are important exceptions to the rule of having a direct linkage between the level of evidence and the strength of recommendations. Some situations where it may be necessary to break this linkage are listed below:

- A statistically significant but marginally important benefit of the intervention is observed
- The intervention is exorbitantly costly
- Superior and established alternative interventions are available
- There are competing outcomes (both beneficial and harmful) that cannot be reconciled

Under such circumstances it may be appropriate to downgrade the level of the recommendation.

There are sophisticated techniques designed to reconcile conflicting outcomes. These include decision analysis and cost effectiveness analysis. Generally, such analyses are beyond the scope of a guideline.

3. WRITING, REVISION, AND APPROVAL

It is essential that authors set and adhere to a timeline for the remainder of the project. The timing of these steps are closely linked to committee meeting schedules. As the relevant committees meet three or four times per year, each missed deadline delays the project by three to four months. Missing several deadlines may require repeating earlier steps to update the literature review or to account for future revisions to this process. Table 5 outlines approximate timelines for these steps.

3.1 Drafting the Guideline

The author panel should translate the evidence tables into a guideline manuscript following the Guideline Format provided in Appendix 11.

Authors should adhere to the clinical question → evidence → conclusions → recommendations flow discussed in the Introduction.

Getting Ready to Write

Before authors begin to write the guideline, they should review all of section 3, as well as the "Instructions for Authors" and "Suggestions to Authors" at *www.neurology.org*. The manuscript will be evaluated by both AAN and the journal. It is essential to know the expectations of each.

Writing

Usually, the panel chair assigns specific topics to each panel member; panel members develop the first draft of their assigned section. The panel chair then integrates all of the sections into a cohesive document.

Guideline Format

The author panel should follow the structure provided in the Guideline Format outlined in Appendix 11.

Drafts should be double-spaced with text in 12-point font size. Pages should be numbered. Drafts should be no more than 16 pages. Each draft should be labeled with the date and step in the process, as noted in Appendix 11.

Adherence to Billings' rules is quite helpful: 1) have something to say, 2) say it; 3) stop as soon as you have said it.[1] It is a Herculean effort to reduce an 80-page treatise to the 12-16 pages suitable for publication in *Neurology*. It is better to be brief from the start.

Essential Elements

Once authors have completed the draft guideline following the instructions in Appendix 11, they should verify that they have done the following:

- Utilized the headings provided in the Guideline Format and followed the instructions for each section
- Described the literature review process so that it is replicable
- Reported the number of studies identified and reviewed
- Presented and referenced each article in the text and in an evidence table
- Referenced each major point with both the article on which it is based and the level of evidence (e.g., class I)
- Answered the clinical questions in the conclusions
- Ensured that the conclusions flow from the evidence and that the recommendations flow from the conclusions
- Included a quality of evidence label (e.g., level A) on each recommendation
- Included Recommendations for Future Research, as detailed in section 3.2
- Dated the draft
- Included appropriate disclosure statements

Table 5
AAN Review and Approval Process

Manuscript Drafted
↓
QSS/TTA Review
(quarterly meetings)
↓
Revision and Resubmission
(If first draft not approved to move forward, the draft is revised and resubmitted to QSS/TTA on a quarterly basis until it is approved)
↓
External Review
(two months)
↓
QSS/TTA Approval
(quarterly meetings)
↓
***Neurology* Peer Review**
(two months)
↓
Practice Committee Approval
(meets three times a year)
↓
Board of Directors Approval
(meets three times per year)

[1] Billings JS. An address on our medical literature. BMJ 1881; Aug 13: 262-268)

3.2 Developing Recommendations for Future Research

The future research section of the guideline is an important vehicle for identifying areas that were found deficient based on the thorough, systematic literature analysis.

The panel should hold a conference call or face-to-face meeting to critically analyze the gaps and flaws uncovered in the comprehensive review of the literature and identify and prioritize future research directions based on the potential for impacting care.

The future research section should include:

- A brief explanation of why the standardized literature review and guideline development process places the guideline author panel in an ideal situation to assess the need for future research within that topic
- An explicit summary of study design issues that were found to be "pitfalls" in the existing literature. For example, the need for multi-center studies, the need for adequate sample sizes, the need for randomized studies, the need for more comprehensive or reliable outcomes measures, and so forth
- A rank ordering of future research recommendations, prioritized by a set of criteria that could include, but are not necessarily limited to:
    - The potential the research has to positively impact patient outcomes
    - Impact on the burden of disease:
        - Prevalence of target disease
        - Percentage of patients with target disease affected by results of study
        - Significance of therapeutic impact that could be detected by the trial
        - Potential impact of trial on quality of life
        - Economic impact
        - Availability of alternative evaluations or treatments:
        - Whether evaluation/treatment is new or unique
        - Whether evaluation/treatment is already in use but has not been evaluated for effectiveness
    - Likelihood of success:
        - Can a study be designed which is practical and feasible?
        - Are there ethical constraints to doing a study?
    - Availability of adequate scientific justification for undertaking a study at this time:
        - Is the evaluation/treatment scientifically reasonable?
        - Are appropriate outcome measures available?
        - Are further pilot studies or data needed?

The recommendations should be reassessed as the project reaches completion.

3.3 Committee Review

The draft is submitted to QSS or TTA for review. QSS or TTA carefully reviews the manuscript and often requests modifications. The most common requests for revision pertain to deviations from the established Guideline Format, incorrect translation of the evidence to conclusions or the conclusions to recommendations, the presence of ambiguities, and the length of the manuscript.

3.4 External Review

Obtaining Reviews
Once the draft guideline receives QSS or TTA approval, AAN staff sends it out for review to the following groups:

- Appropriate physician organizations
- Members of the AAN Member Reviewer Network
- Appropriate AAN Sections or Committees
- Domestic and international subject matter experts
- AAN's Ethics, Law and Humanities Committee or legal counsel, when appropriate

Staff collects the responses and forwards them to the facilitator and lead author.

Responding to the Reviews
The author panel should revise the document, as appropriate, and develop a Revision Table that lists each comment, the reviewer, and how the comment was addressed in the document (see example in Appendix 12). The Revision Table must be submitted to QSS or TTA with the final guideline draft. The Revision Table will accompany the document when it is sent to *Neurology*, the Practice Committee, and the AAN Board of Directors.

Authors are encouraged to utilize revision format (underline and strike out) for this and subsequent drafts for which the changes are minor.

The revised manuscript and Revision Table should be submitted to staff.

3.5 Committee Approval

Staff submits the revised document with the Revision Table to QSS or TTA for an official vote.

QSS or TTA approval may be contingent upon additional revision.

3.6 Journal Review

Once QSS or TTA has approved the document, it is sent to *Neurology* for peer review.

The Editor obtains peer review and sends the comments to the lead author and staff. Authors are encouraged to consider all of the revisions suggested by the journal peer reviewers. Authors should contact the facilitator if the reviewers' comments are in conflict with AAN guideline requirements.

The lead author should submit the revised draft to AAN staff (not directly to the journal) with a cover letter denoting the panel's responses to all of the journal reviewers' comments. Staff resubmits the guideline to the journal.

The journal may request a second or third round of reviews prior to accepting the manuscript for publication.

3.7 Practice Committee and Board of Directors Approval

Once the article has been accepted for publication in *Neurology*, the guideline is submitted to the Practice Committee, then the Board of Directors, for approval.

Requests for revision during the approval process are reviewed by the QSS or TTA chairs. Substantive revisions may require reapproval by QSS or TTA.
Once the Board of Directors has approved the guideline, the statement becomes the official policy of the AAN.

3.8 Seeking Endorsements

It may be appropriate to seek endorsement of the guideline from other

organizations. Authors should inform staff of organizations that should be contacted.

4. GUIDELINE DISSEMINATION

The guideline is:

- Published in *Neurology*
- Posted on the AAN Website
- Sent to all AAN members in an annual mailing
- Announced in *AANnews*
- Submitted to guideline compendiums such as the National Guidelines Clearinghouse

The Practice Improvement Subcommittee may undertake additional guideline dissemination and implementation projects. These may include a press release, slide presentation, patient version of the guideline, physician summary of the guideline, algorithms, and other tools to help members incorporate guideline recommendations into their practices.

5. RESPONDING TO CORRESPONDENCE

If Letters to the Editor are received, authors and facilitators should work together to craft a response. The response should be reviewed by the subcommittee chair prior to submission to the journal.

6. GUIDELINE UPDATING PROCESS

All guidelines are evaluated on an annual and triennial basis to ensure their continuing validity. QSS and TTA make formal decisions whether to reaffirm, update, or retire guidelines as needed based on the annual screening review and triennial detailed review including an updating literature search. Decisions are communicated to the AAN membership through the Website and other means.

If it is determined that an update is warranted, QSS or TTA forms a new author panel, which may include members of the initial author panel. The project then follows the same process as outlined in this manual.

APPENDICES

Appendix 1

Project Development Plan Worksheet

1. Clinical Question Development:

a) Problem/Issue to be addressed:

b) To what patient population does this apply?

c) What is the intervention (therapy, test, risk factor)?

d) What are the outcomes of interest?

e) State one or more answerable clinical questions that include the population, intervention and outcomes of interest:

Examples:

- *What is (are) the best medication(s) for controlling seizures while minimizing side effects and providing a good quality of life for a patient who requires treatment for epilepsy?*
- *Does anticonvulsant prophylaxis decrease the risk of developing late seizures in patients with head injury?*
- *In patients with Bell's palsy, do steroids improve facial function outcomes?*

Clinical Questions:

2. Criteria for Literature Search:

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 and Index words for the intervention

c) Databases to be searched (e.g. MEDLINE, EMBASE, Current Contents)

d) Years to be included in the search

3. Inclusion and Exclusion Criteria:

a) Include all languages (*AAN will pay for non-English translations*)
☐ YES or ☐ NO, English only

Please provide an explanation if you selected English only:

b) Selected Study Population
Human studies ☐ YES or ☐ NO
Animal studies ☐ YES or ☐ NO

c) Disease in question or closely related to diseases to be included

d) Interventions to be:
   i) Included:
   ii) Excluded:

e) Outcomes to be:
   i) Included:
   ii) Excluded:

f) Types of studies to be included:
   i) **RCT** - ☐ YES or ☐ NO
   ii) **Cohort** - ☐ YES or ☐ NO
   iii) **Case Control** - ☐ YES or ☐ NO
   iv) **Case Series (**n must be greater than _____**)**
      (1) ☐ YES or ☐ NO
   v) **Review Papers** - ☐ YES or ☐ NO
   vi) **Meta-analyses** ☐ YES or ☐ NO

g) Standard exclusion criteria:
   i) Not relevant to the clinical question
   ii) Unrelated disease
   iii) Outside of study population
   iv) Articles not peer reviewed

h) Additional exclusion criteria:

4. Project Timeline: (Enter dates based on your availability and the guidelines provided)

a) Complete panel formation by _____ (usually takes two to four weeks)
b) Literature search _____ (select a timeframe of one to two weeks, during which you will have time to complete the search with the librarian and review and distribute the abstracts; AAN staff will have the librarian contact you to begin this step)
c) Panel review of the literatures _____ (two-step process of reviewing abstracts then selecting articles – typically six to eight weeks)
d) Data extraction and development of evidence tables _____ (takes three to eight weeks depending on the total number of articles to analyze and tabulate)
e) Drafting the guideline _____ (takes four to eight weeks)
f) Goal for submitting the first draft to QSS or TTA
    i) TTA
        (1) ☐ February 2007
        (2) ☐ April 2007
        (3) ☐ June 2007
        (4) ☐ September 2007
    ii) QSS
        (1) ☐ December 2006
        (2) ☐ April 2007
        (3) ☐ June 2007
        (4) ☐ September 2007

Appendix 2

Suggested Supplementary Materials

Regarding Evidence-Based Medicine and Reviews:

Cochrane Handbook (available at www.update-software.com/ccweb/cochrane/hbook.htm)

Counsell, Carl. Formulating Questions and locating primary studies for inclusion in systematic reviews (Academia and Clinic: Systematic Review Series). Ann Intern Med, 1997;127:380-387.

Evidence-Based Medicine (Sackett et al, 1997)

Evidence-Based Principles and Practice (McKibbon, 1999)

National Guideline Clearinghouse at www.guidelines.gov

The CATbank at http://cebm.jr2.ox.ac.uk/docs/catbank.html

Regarding Using EndNote to Search Remote Databases:

www.biomed.lib.umn.edu/endref.html

Regarding Using EndNote to Create a Bibliography:

www.biomed.lib.umn.edu/end.html

AAN Guideline Author Conflict of Interest Policy

The American Academy of Neurology is committed to producing independent, critical and truthful clinical practice guidelines (CPG) of drugs and technologies that impact patients who have or are at risk for neurological diseases. To the extent possible, the AAN believes that those who produce the CPGs and those who have a financial stake in the success or failure of the products appraised in the CPGs should be kept separate and distinct. However, it may be difficult to form an expert panel devoid of potential conflicts, therefore the AAN carefully balances known biases of panel members, reviews conflict of interest disclosure, limits participation, and provides strong oversight of the panels.

Balancing the Panel

When it is not feasible to form an expert panel devoid of potential conflict, it is essential to balance the panel between those with and without conflicts (financial, research, academic, etc.). At least half of the panel members should not have a significant conflict of interest. For guidelines of broad scope, panel members should not all be from the same institution or study group. If there is a recognized, credible controversy on the chosen guideline topic, both sides should be represented on the panel.

Attainment of Conflict of Interest Disclosures

Panel members must complete and sign a conflict of interest statement (attached) annually. All potential conflicts for the past year for the author, spouse and minor children must be disclosed.

Oversight by the QSS and TTA

The author panel roster and conflict of interest forms—and all subsequent changes to either—are communicated to QSS or TTA. QSS and TTA reserve the right to make changes to the author panel to ensure balance and avoid bias.

Conflicts that Limit Participation

QSS or TTA may choose to not appoint an individual as a lead author or the lead of a section of a guideline if the individual has any of the following relationships to the issues or products being assessed: having stock or ownership; being compensated for expert testimony; being a pioneer or inventor; holding a patent or intellectual property; or any substantial direct or indirect compensation that can be viewed as a conflict.

The AAN forbids commercial participation in guideline projects. Being a current employee of a pharmaceutical company or a device manufacturer prevents participation.

Disclosure of Potential Conflicts of Interest

Substantial conflicts of interest will be disclosed in the published guideline as determined by QSS and TTA and as required by journal standards.

In addition, the following paragraph will be included in all future guidelines:

> The American Academy of Neurology is committed to producing independent, critical and truthful clinical practice guidelines (CPG). Significant efforts are made to minimize the potential for conflicts of interest to influence the recommendations of this CPG. To the extent possible, the AAN keeps separate those who have a financial stake in the success or failure of the products appraised in the CPGs and the developers of the guidelines. Conflict of interest forms were obtained from all authors and reviewed by an oversight committee prior to project initiation. AAN limits the participation of authors with substantial conflicts of interest. The AAN forbids commercial participation in, or funding of, guideline projects. Drafts of the guideline have been reviewed by at least three AAN committees, a network of neurologists, *Neurology* peer reviewers and representatives from related fields. The AAN Guideline Author Conflict of Interest Policy can be viewed at **www.aan.com**.

Statement of Potential Competing Interests (Financial, Equity, Intellectual Property, Research, Advocacy) with AAN Practice Parameters or Technology Assessments

Name of AAN Member: Date of Statement:

Name of Practice Parameter or Technology Assessment:

In accordance with the action by the American Academy of Neurology Board of Directors, authors and expert panelists for each QSS and TTA project are required to disclose all possible conflict of interest with respect to the topic being studied. Use additional pages if necessary to complete this form.

1. **PAST OR PRESENT FINANCIAL RELATIONSHIPS**: Please list below all pharmaceutical, medical device, biotechnology, or medical consulting companies in which you or your immediate family member(s) have or have had financial, equity, or intellectual property interests, currently and in the 1 year prior to the date of this document.

Name of Company	Type of Relationship (Please check (√) if yours or write "FM" if family member, defined as spouse and minor children)		
	Financial*	Equity**	Intellectual Property
For interests ≤ $10,000			
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
For interests > $10,000 to $25,000			
1.			
2.			
3.			
For interests > $25,000			
1.			
2.			
3.			

* Fees for consulting, speaker's bureaus, advisory boards, or other committees. Include fees paid to you directly or indirectly to you through a University account that is under your control (eg, discretionary account).
** Do NOT include mutual funds.

2. **FUTURE STOCK OPTIONS/PATENT RIGHTS:** Please list <u>all</u> stock options and/or patent rights that you or your family member(s) have in a pharmaceutical, medical device, or biotechnology company.

Name of Company	Type of INTEREST (Please check (√) if yours or write "FM" if family member)	
	Future Stock Options	Patent Rights
1.		
2.		

3. List any grant or contract that either provides salary support paid to you through your institution or supports your research without salary support, currently and in the 1 year prior to the date of this document. Only include research that could reasonably be considered related to this parameter or assessment.

Name of Sponsor*	Brief Description of Research
1.	
2.	
3.	
4.	
5.	
6.	
7.	

* List Government (eg, NIH, FDA, AHRQ), or Foundation source or name of private company (eg, pharmaceutical, medical device or biotechnology company).

4. List all diagnostic procedures you currently perform in your clinical practice and estimate the percentage of your clinical effort devoted to this procedure in the past 1 year. Only include procedures that could reasonably be considered related to this parameter or assessment.

Procedure	Estimated Percent of Your Clinical Effort
1.	
2.	
3.	
4.	

5. In the past 1 year, did you serve as an officer, director, partner, manager or employee of **<u>any</u>** pharmaceutical, medical device, or biotechnology company?
☐ No ☐ Yes

   If yes, specify company and role.

6. In the past 1 year, have you received payment for expert testimony in a legal proceeding on a topic that could reasonably be considered related to this parameter or assessment?
☐ No ☐ Yes

   If yes, specify content area of your testimony.

7. In the past 1 year, have you received payment for an advocacy role in government or non-profit organization on a topic that could reasonably be considered related to this parameter or assessment?
   ☐ No ☐ Yes

   If yes, specify advocacy role.

The statements I have made are true, complete, and correct. I give my permission to disclose this information to appropriate AAN Leaders and any one who specifically requests it in writing.

By typing your name in the space provided, you are submitting the electronic equivalent of a legal signature. You are also asserting that you completed the application. To verify the contents of the application, the signatory must enter his/her name in the space provided. Acceptable "signatures" should be preceded and followed by the forward slash (/) symbol. Acceptable "signature" should be as follows: /John Doe/.

Electronic Signature Date

Return this form by e-mail to Thomas Getchius at tgetchius@aan.com or by fax to 651-361-4905.

Appendix 4

Evidence-Based Medicine-Related Terms for Searching MEDLINE

	MeSH Terms	MeSH subheadings	Textwords	MEDLINE publication types
Etiology	epidemiologic studies (exp) case-control studies cohort studies risk risk assessment risk factors odds ratio	chemically induced complications congenital embryology epidemiology etiology genetics immunology microbiology parasitology secondary transmission	cohort risk causa$ predispos$	
Diagnosis	sensitivity and specificity double blind method single blind method	Used with disease terms or anatomical terms: diagnosis radiography radionuclide imaging ultrasonography Used with diagnostic techniques or methodologies: diagnostic use	diagnosis diagnos$ sensitivity specificity predictive	
Therapy	clinical trials (exp) research design (exp) comparative study placebos double blind method	Used with disease terms: therapy diet therapy drug therapy nursing prevention and control radiotherapy rehabilitation surgery transplantation Used with drugs and other therapeutic agents or procedures: therapeutic use administration and dosage adverse effects contraindications poisoning toxicity	therap$ treat$ manag$ placebo$ random$	clinical trial randomized controlled trial multicenter study
Prognosis	prognosis cohort studies (exp) disease progression mortality (exp) morbidity (exp) time factors survivors	complications mortality	natural history prognos$ course cohort surviv$ outcome$	Practice guidelines clinical guidelines consensus development reports
Overview/ Meta-analysis	meta-analysis		metaanaly$ meta-analy$ overview	meta-analysis

$ indicates that the root term may be altered to include such terms as diagnostics, diagnosing, etc.

Appendix 5

Major Literature Databases

MEDLINE®	
Type: Bibliographic citations with author abstracts. **Materials Covered**: International coverage of over 3,800 journals. **Dates of Coverage**: 1966 to present, updated monthly. **Producer/Publisher:** U.S. National Library of Medicine.	MEDLINE covers the fields of medicine, public health, nursing, dentistry, veterinary medicine, and the preclinical sciences. MEDLINE encompasses information from three print indexes, Index Medicus, Index to Dental Literature, and International Nursing Index as well as other sources of coverage in the areas of allied health, biological and physical sciences, humanities and information science as they relate to medicine and health care.
EMBASE®	
Type: Bibliographic citations with abstracts. **Materials Covered:** International coverage of over 3,500 journals. **Dates of Coverage:** 1980 to present, updated weekly or monthly depending on access. **Producer/Publisher:** Elsevier Science	The Excerpta Medica database is a major biomedical and pharmaceutical database indexing over 3,500 international journals in the following fields: drug research pharmacology; pharmaceutics; toxicology; clinical and experimental human medicine; health policy and management; public health; occupational health; environmental health; drug dependence and abuse; psychiatry; forensic medicine; biomedical engineering/instrumentation. EMBASE is one of the most widely used biomedical and pharmaceutical databases because of its currency and in-depth indexing. It is particularly strong in coverage of drug-related literature, European journals, and conference proceedings. Frequent updates allow access to the latest medical and pharmacological trends. The database currently contains over 6 million records, with more than 375,000 citations and abstracts added yearly.
Science Citation Index Expanded	
Type: Bibliographic citations, plus some author abstracts. Each citation also includes a list of references cited in the source article. The Citation Index enables the reader to take a known paper and find other papers that cite it. The Source Index enables the reader to discover what a particular author has published during the period covered. **Materials Covered:** Articles, reviews, letters, etc. from over 5,300 major journals across 164 scientific disciplines. **Dates of Coverage**: Varies, depending on access system. Updated weekly. **Producer/Publisher:** The Institute for Scientific Information	The sciences, including agriculture, astronomy, biochemistry, biology, biotechnology, chemistry, computer science, materials science, mathematics, medicine, neuroscience, oncology, pediatrics, pharmacology, physics, plant sciences, psychiatry, surgery, veterinary science, and zoology.

Current Contents	
Type: Journal table of contents and bibliographic citation with author abstracts and author addresses. **Materials Covered:** Clinical Medicine – Provides access to more than 900 of the world's leading journals in clinical medicine, including disciplines such as anatomy, anesthesiology, clinical psychiatry and psychology, internal medicine, nuclear medicine, oncology, pediatrics, and surgery. Includes complete bibliographic information for each article, review, letter, note, and editorial listed. Life Sciences -- Indexes more than 1,200 of the world's leading journals in the life sciences, including disciplines such as biochemistry, biophysics, endocrinology, genetics, immunology, microbiology, molecular biology, neuroscience, pharmacology, physiology, and toxicology. Provides complete bibliographic information for each article, review, letter, note, and editorial listed. **Dates of Coverage:** 1994 to present, updated weekly. **Producer/Publisher**: Institute for Scientific Information	Current Contents is a multidisciplinary current awareness service for scholarly journals. This online product provides access to all seven Current Contents printed editions. Of particular interest are Clinical Medicine and Life Sciences.
BIOETHICSLINE®	
Type: Bibliographic citations with abstracts available on selected citations. **Materials Covered:** English language; journal articles, monographs, chapters in monographs, newspaper articles, court decisions, bills, laws, audiovisual materials, and unpublished documents. **Dates of Coverage:** 1973 to present, updated quarterly. **Producer/Publisher:** Bioethics Information Retrieval Project of the Kennedy Institute of Ethics at Georgetown University for the U.S. National Library of Medicine.	BIOETHICSLINE covers the ethical, legal and public policy issues surrounding health care and biomedical research. Topics include euthanasia and other end-of-life issues, organ donation and transplantation, allocation of health care resources, patient rights, professional ethics, new reproductive technologies, genetic intervention, abortion, behavior control and other mental health issues, AIDS, human experimentation, and animal experimentation. Citations are derived from the literature of law, religion, the social sciences, philosophy, and the popular media as well as the health sciences.
CINAHL®	
Type: Bibliographic citations with author abstracts and cited references. Full text is available from selected state nursing journals, nursing standards of practice and nurse practice acts. **Materials Covered:** More than 900 journals, including virtually all English-language nursing journals, selected foreign-language journal titles, publications of the American Nurses Association and the National League for Nursing, books, book chapters, educational software, audiovisuals, pamphlets, dissertations, selected conference proceedings and research instruments are covered. **Dates of Coverage:** 1982 to present, updated monthly. **Producer/Publisher:** Cinahl Information Systems.	CINAHL, Cumulative Index to Nursing and Allied Health, has a multidisciplinary scope covering nursing, 17 allied health disciplines, biomedicine, consumer health, health sciences librarianship and selected standards of professional practice. The allied health disciplines include cardiopulmonary technology, emergency services, health education, medical/laboratory technology, medical assistant, medical records, occupational therapy, physical therapy, radiologic technology, respiratory therapy, surgical technology and physicians assistants.

International Pharmaceutical Abstracts	
Type: Bibliographic citations with specially written abstracts on journal articles and full text of the meeting abstracts of the American Society of Health- Systems Pharmacists (ASHP). **Materials Covered:** Articles from 850 primary journals from throughout the world and all U.S. state pharmacy journals. **Dates of Coverage:** 1970 to present, updated monthly. **Producer/Publisher**: American Society of Health-Systems Pharmacists.	International Pharmaceutical Abstracts (IPA) provides information on all phases of the development and use of drugs and on professional pharmaceutical practice. In early 1985 coverage was expanded to include state pharmacy journals that deal with state regulations, salaries, guidelines, manpower studies, laws, and more. The scope of the database ranges from the clinical, practical, and theoretical to the economic and scientific aspects of the literature. Comprehensive information is included for drug therapy, toxicity, and pharmacy practice as well as legislation, regulation, technology, utilization, biopharmaceutics, information processing, education, economics, and ethics as related to pharmaceutical science and practice. A unique feature of abstracts reporting clinical studies is the inclusion of the study design, number of patients, dosage, dosage forms and dosage schedule.
Health Services Technology Assessment Texts (HSTAT)	
Type: Full text of documents. **Materials Covered:** Quick-reference guides for clinicians, consumer brochures, and evidence reports sponsored by the Agency for Healthcare Research and Quality (AHRQ); AHRQ technology assessment reports; National Institutes of Health (NIH) consensus development conference and technology Assessment reports; NIH Warren G. Magnuson Clinical Center research protocols; HIV/AIDS Treatment Information Service (ATIS) resource documents; Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment (SAMHSA/CSAT) treatment improvement protocols; and the Public Health Service (PHS) Preventive Services Task Force Guide to Clinical Preventive Services. It also provides a link to the Centers for Disease Control and Prevention (CDC) Prevention Guidelines Database. **Dates of Coverage:** 1994 to present **Producer/Publisher:** National Library of Medicine's (NLM) Information Technology Branch of the Lister Hill Center. It is part of the expanded Health Services Research Information Program coordinated by NLM's National Information Center on Health Services Research and Health Care Technology (NICHSR). NICHSR works closely with AHCPR to improve the organization and dissemination of the results of health services research, including practice guidelines and technology assessments.	HSTAT is a free, electronic resource that provides access to documents, including clinical practice guidelines useful in health care decision making. HSTAT is accessed at http://text.nlm.nih.gov/ .

PsycINFO	
Type: Bibliographic citations and abstracts. **Materials Covered:** Articles from more than 1,300 international journals in psychology and related fields. **Dates of coverage**: 1967 to the present, updated monthly. **Producer/Publisher**: American Psychological Association.	All areas of psychology, including experimental and developmental, communications, social processes and issues, personality, physical and psychological disorders, professional issues, applied psychology, educational psychology, behavioral literature in such related fields as law, business and medicine.
BIOSIS Previews	
Type: Bibliographic citations, many with abstracts. **Materials Covered:** Journal articles, books, research reports, conference proceedings. **Dates of Coverage**: 1980 to present, updated monthly. **Producer/Publisher:** Biosis, Inc.	Biological and medical sciences, including biochemistry, biophysics, biotechnology, botany, environment, microbiology, and zoology.

Appendix 6

Costs Associated with Guideline Development

Several steps of this process require financial resources to complete. Authors are not expected to incur any out-of-pocket expenses. However, authors must authorize all expenditures through AAN staff. The following table should provide a guide for determining how to handle expenses.

Expense	Cost	Who pays?	How to initiate
Author panel conference calls	Approximately $200 per call	AAN will coordinate and pay for conference calls.	Contact AAN staff at (651) 695-2805
Author panel meetings at AAN Annual Meeting or other meetings	Varies	AAN will pay room rental for the work group to meet. AAN may provide beverages and snacks dependent on budget constraints.	Contact facilitator or AAN staff several months prior to the meeting.
Other author panel meetings	Approximately $1,000 per person	AAN does not have budget resources to support author panel meetings other than at the AAN Annual Meeting or with special approval.	Contact AAN staff to request a special budget allotment. This action may require AAN Board of Directors approval.
Literature searches	MEDLINE approximately $150 per search; EMBASE approximately $500 per search	Authors are encouraged to take advantage of free services available to them. AAN pays for authorized literature searches.	For AAN assistance, contact staff at (651) 695-2805. Staff initiates contact with librarian service. Authors should then contact the librarian service directly to execute the search.
Obtain articles	Approximately $6 per article; approximately $200-$300 per focused topic	Authors are encouraged to take advantage of free services available to them. AAN pays for retrieval of articles.	Submit list of articles to be retrieved to AAN staff. guidelines@aan.com or 651-695-2805
Attend QSS or TTA meeting to present paper	Approximately $1,000 per person	AAN often invites authors to attend a single QSS or TTA meeting to present a draft guideline.	Upon invitation.

Appendix 7

Common Formulas for Calculating Effect Sizes

Therapeutic questions:

	Good	Poor
Treated	**A**	**C**
Untreated	**B**	**D**

Relative Rate = [A / (A + C)] / [B / (B + D)].

Rate Difference = [A / (A + C)] - [B / (B + D)].

Diagnostic (Prognostic) Accuracy Questions:

	Disease (outcome) present	Disease (outcome) absent
Test (predictor) Positive	**A**	**C**
Test (Predictor) Negative	**B**	**D**

Relative Risk = [A / (A + C)] / [B / (B + D)].

Sensitivity = A / (A + B)

Specificity = D / (C + D)

Positive Predictive value = A / (A + C)

Negative Predictive value = D / (B + D)

Screening questions:

	Condition present	Condition absent
Tested	**A**	**C**

Yield = A / (A + C).

Appendix 8a

Sample Data Extraction Form
(for established diagnostic tests)

Panel Member_____
Paper relevant to project? Y N

Author: _____
Year: _____ Journal: _____
Title: _____

Article Funding Source: _____

Type of Article (circle one)

Review article
Meta-analysis
RCT
Cohort
Class IV
Case Control
Observational Case Series (n=____)

Classification of Evidence (circle one)

Class I
Class II
Class III

Study Characteristics:

Subjects			***Controls***	
_____		Number of subjects and controls	_____	
Yes	No	Normals?	Yes	No
Yes	No	Patients with competing diagnoses?	Yes	No
Yes	No	Patients with other neurologic diagnoses?	Yes	No

Blinding

Blinded to diagnosis?	Yes	No
Blinded to outcome?	Yes	No

Gold standard comparison? _____

Prospective, retrospective, other, or indeterminate? (circle one)
If other, explain _____

Can a 2X2 table be constructed from data? If yes, complete table and calculate:
Sensitivity_____
Specificity_____
Positive predictive value_____
Negative predictive value_____
Statistical significance_____
Magnitude_____
Do authors give likelihood ratios?
Are ROC curves available?

Appendix 8b

Sample Data Extraction Form
(for therapeutics)

Panel Member_____
Paper relevant to project? Y N

Author: _____
Year: _____ Journal: _____
Title: _____
_____

Article Funding Source: _____

Type of Article (circle one)

Review article
Meta-analysis
RCT (prospective)
Cohort Study (nonrandomized)
Case Series
Observational Case Series (n=____)
Case Control Series (retrospective)

Classification of Evidence (circle one)

Class I
Class II
Class III
Class IV

1. Purpose of the study:
   _____
2. Sample Size: _____
3. Loss to follow-up: _____
4. Type(s) of patients studied:_____
5. Were standardized diagnostic criteria applied? YES / NO
   a. If YES, what standardized criteria were used? _____
6. Type(s) of controls_____
7. Intervention_____
8. Outcome _____
   a. Positive (describe) _____
   _____
   b. Negative (describe, including significant AE's) _____
   _____
   c. If a review, model, or meta-analysis, what is the main utility for the guideline?
9. Comments (special reasons to include, noteworthy findings, etc.) _____

_____

_____

Appendix 9

Definitions for Classification of Evidence

Suggested wording	Translation of evidence to recommendations	Rating of Therapeutic Article
(Note: Wording relevant to diagnostic, prognostic and screening questions are indicated in parenthesis.) **Conclusion:** A = Established as effective, ineffective or harmful (or established as useful/predictive or not useful/predictive) for the given condition in the specified population **Recommendation:** Should be done or, should not be done	Level A rating requires at least two consistent Class I studies*	**Class I:** Prospective, randomized, controlled clinical trial with masked outcome assessment, in a representative population. The following are required: a) primary outcome(s) clearly defined b) exclusion/inclusion criteria clearly defined c) adequate accounting for drop-outs and cross-overs with numbers sufficiently low to have minimal potential for bias d) relevant baseline characteristics are presented and substantially equivalent among treatment groups or there is appropriate statistical adjustment for differences.
Conclusion: B = Probably effective, ineffective or harmful (or probably useful/predictive or not useful/predictive) for the given condition in the specified population **Recommendation:** Should be considered or, should not be considered	Level B rating requires at least one Class I study or two consistent Class II studies	**Class II:** Prospective matched group cohort study in a representative population with masked outcome assessment that meets a-d above OR a RCT in a representative population that lacks one criteria a-d.
Conclusion: C = Possibly effective, ineffective or harmful (or possibly useful/predictive or not useful/predictive) for the given condition in the specified population **Recommendation:** May be considered or, may not be considered	Level C rating requires at least one Class II study or two consistent Class III studies	**Class III:** All other controlled trials (including well-defined natural history controls or patients serving as own controls) in a representative population, where outcome is independently assessed, or independently derived by objective outcome measurement.**
Conclusion: U = Data inadequate or conflicting. Given current knowledge, treatment (test, predictor) is unproven **Recommendation:** None	Studies not meeting criteria for Class I – Class III	**Class IV:** Evidence from uncontrolled studies, case series, case reports, or expert opinion.

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

**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).

Rating of Diagnostic Article	Rating of Prognostic Article	Rating of Screening Article
Class I: Evidence provided by a prospective study in a broad spectrum of persons with the suspected condition, using a reference (gold) standard for case definition, where test is applied in a blinded evaluation, and enabling the assessment of appropriate tests of diagnostic accuracy. All patients undergoing the diagnostic test have the presence or absence of the disease determined.	**Class I:** Evidence provided by a prospective study of a broad spectrum of persons who may be at risk for developing the outcome (e.g. target disease, work status). The study measures the predictive ability using an independent gold standard for case definition. The predictor is measured in an evaluation that is masked to clinical presentation and, the outcome is measured in an evaluation that is masked to the presence of the predictor. All patients have the predictor and outcome variables measured.	**Class I.** A statistical, population-based sample of patients studied at a uniform point in time (usually early) during the course of the condition. All patients undergo the intervention of interest. The outcome, if not objective, is determined in an evaluation that is masked to the patients' clinical presentations.
Class II: Evidence provided by a prospective study of a narrow spectrum of persons with the suspected condition, or a well designed retrospective study of a broad spectrum of persons with an established condition (by "gold standard") compared to a broad spectrum of controls, where test is applied in a blinded evaluation, and enabling the assessment of appropriate tests of diagnostic accuracy.	**Class II:** Evidence provided by a prospective study of a narrow spectrum of persons at risk for having the condition, or by a retrospective study of a broad spectrum of persons with the condition compared to a broad spectrum of controls. The study measures the prognostic accuracy of the risk factor using an acceptable independent gold standard for case definition. The risk factor is measured in an evaluation that is masked to the outcome.	**Class II.** A statistical, non-referral-clinic-based sample of patients studied at a uniform point in time (usually early) during the course of the condition. Most patients undergo the intervention of interest. The outcome, if not objective, is determined in an evaluation that is masked to the patients' clinical presentations.
Class III: Evidence provided by a retrospective study where either persons with the established condition or controls are of a narrow spectrum, and where the reference standard, if not objective, is applied by someone other than the person that performed the test	**Class III:** Evidence provided by a retrospective study where either the persons with the condition or the controls are of a narrow spectrum. The study measures the predictive ability using an acceptable independent gold standard for case definition. The outcome, if not objective, is determined by someone other than the person who measured the predictor.	**Class III.** A sample of patients studied during the course of the condition. Some patients undergo the intervention of interest. The outcome, if not objective, is determined in an evaluation by someone other than the treating physician.
Class IV: Any design where test is not applied in an independent evaluation OR evidence provided by expert opinion alone or in descriptive case series (without controls).	**Class IV:** Any design where the predictor is not applied in an independent evaluation OR evidence provided by expert opinion or case series without controls.	**Class IV.** Expert opinion, case reports or any study not meeting criteria for class I to III.

Retrospective: a case control study. Prospective: a cohort survey. Objective: a measurement unlikely to be affected by expectation bias.

Appendix 10

Sample Evidence Table

Design characteristics and outcomes in controlled studies of patients with Bell's Palsy treated with steroids

Author Year	Class	Blind	Cohort Size	Completion Rate %	Steroid Dose Duration Rx	Follow-up months	Severity %	Duration days	NH %	RR Good Recovery (CI)	RR Complete Recovery (CI)
May 1976[7]	I	Yes	51	100	Prednisone 410 mg 10 days	6	47	2	81	0.99 (0.76-1.30)	0.92 (0.60-1.4)
Taverner 1954[8]	I	Yes	26	100	Hydrocortisone 1 gm 8 days	NS	23	9	67	1.07 (0.64-1.80)	_
Brown 1982[9]	I	Yes	82	100	Unnamed 400 mg 10 days	12	0	3	73	1.20 (0.97-1.50)	1.20 (0.97-1.49)
Wolf 1978[10]	I	No	239	100	Prednisone 760 mg 17 days	12	31	5	98	1.02 (0.99-1.06)	1.09 (0.98-1.22)
Austin 1993[11]	I	Yes	76	71	Prednisone 405 mg 10 days	6	22	5	83	1.21 (1.05-1.39)	1.71 (1.00-2.95)
Shafshak 1994[12]	II	Yes	160	100	Prednisolone 420 mg 10 days	12	91	6	69	1.24 (1.03-1.49)	1.76 (1.08-2.87)
Adour 1972[6]	II	No	304	85	Prednisone 216 mg 12 days	1	NS	14	64	1.39 1.20-1.62	1.58 (1.25-2.00)
Prescott 1988[13]	II	No	879	66	Prednisolone 520 mg 8 days	9	51	7+	92	1.04 (0.99-1.09)	1.04 (0.99-1.09)

Completion rate: percentage of subjects followed to study completion. Severity: Percentage of patients with complete palsy. Duration: Maximum duration of palsy before starting steroids. NH: Natural history, percentage of non-steroid treated patients attaining a good outcome. RR: relative rate of steroid treated patients attaining outcome compared to non-steroid treated patients. CI: 95% confidence intervals. NS: Not stated.

Design characteristics and outcomes in controlled studies of patients with Bell's palsy treated with Acyclovir

Author Year	Class	Blind	Cohort Size	Completion Rate %	Dose Duration Rx	Follow-up months	Severity %	Duration days	NH %	RR Good Recovery (CI)	RR Complete Recovery (CI)
Adour 1996[15]	I	Yes	99	83	400 mg x 5 qd 10 days	12	20	3	76	1.22 (1.02-1.45)	1.21 (0.98-1.49)
De Diego 1998[16]	I	No	101	89	800 mg tid 10 days	3	1	4	94	0.83 (0.71-0.98)	_
Ramos 1992[17]	I	No	30	100	1000 mg qd 5 days	NS	63	NS	100	1.00*	_

Completion rate: Percentage of subjects followed to study completion. Severity: Percentage of patients with complete palsy. Duration: Maximum duration of palsy before starting steroids. NH: Natural history, percentage of non-acyclovir treated patients attaining a good outcome. RR: relative rate of acyclovir treated patients attaining outcome compared to non-acyclovir treated patients. CI: 95% confidence intervals. NS: Not stated. *All patients with good recovery.

Design characteristics and outcomes in controlled studies of patients with Bell's palsy treated with Facial Nerve Decompression

Author Year	Class	Blind	Cohort Size	Completion Rate %	Surgical Approach	Follow-up months	Severity %	Duration days	NH %	RR Good Recovery (CI)	RR Complete Recovery (CI)
Brown 1982[9]	II	No	92	100	Vertical, Stylomastoid, Midcranial fossa	12	100	14	47	1.21 (0.97-1.5)	1.30 (0.89-1.90)
Gantz 1999[18]	II	No	70	100	Mid cranial fossa & meatal foramen	7	100	14	42	2.19	2.96
May 1981[19]	II	No	60	100	Transmastoid, Vertical	6	92	14	6	1.14 (0.79-1.65)	6.4 (0.92-45)
May 1985[20]	II	No	38	100	Transmastoid, Extralabyrinthine, Subtemporal	6	100	14	23	0.87 (0.24-3.07)	_
Fisch 1981[21]	II	No	27	100	Midcranial fossa & meatal foramen	12-36	100	21	15	3.30 (0.82-12.90)	_

Completion rate: Percentage of subjects followed to study completion. Severity: Percentage of patients with complete palsy. Duration: Maximum duration of palsy before starting steroids. NH: Natural history, percentage of non-surgical patients attaining a good outcome. RR: relative rate of surgically treated patients attaining outcome to non-surgically treated patients. CI: 95% confidence intervals. NS: Not stated.

Appendix 11

Guideline Format

Cover Page:

Practice Parameter or Assessment: Title (An Evidence-Based Review)
Report of the TTA or QSS
of the American Academy of Neurology
List authors' names

Date and Stage of Draft (e.g., July 5, 2003 Second Draft for QSS Review)
Contact Information for Lead Author

Manuscript:

ABSTRACT
Up to 240 words should summarize the guideline as follows:

Objective: Summary of clinical focus
Methods: Description of process
Results: Status, quality and content of evidence
Recommendations: Summarize recommendations

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 and abuse
  - Extent of practice variation
  - Urgency
  - Controversy regarding validity or applicability
- Clinical Question Statement

DESCRIPTION OF THE ANALYTICAL PROCESS
This section should present the exact, replicable process the authors used to develop the guideline, including:

- How the panel was selected, including disclosure of information, funding and outside input (e.g. reviewers).
- Description of literature review
  - How the literature search was conducted (search terms, databases searched, other search strategies, languages included, dates covered). Describe bibliographic or other search techniques 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, etc.).
    - Inclusion and exclusion criteria and process for "weeding out" articles
    - State 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 eventually included in the guideline.
    - State 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 using a data extraction form
    - Classification of evidence definitions

    - Development of evidence tables
- Internal and external review of the document (may be footnoted)

ANALYSIS OF EVIDENCE
This section is the scientific body of the paper and should include a detailed narrative description of the evidence and the statistical analysis applied to it, as appropriate to the topic. If more than one clinical question is addressed, it is appropriate to deal with the questions separately.

There are two types of guidelines, those on diagnostic tests and those on therapies. For each, the following should be presented:

For diagnostic tests:

- Results
- Levels of evidence
- Statistical analysis (meta-analysis, sensitivity and specificity, positive and negative predictive values, ORs, relative rates, and numbers needed to treat/harm.
- Relevance (selection criteria, complications, contraindications, test 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, ORs, 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
This section summarizes the evidence in answer to the clinical question. The conclusions should be directly linked to the evidence (e.g., four class II studies show…) If there are a number of conclusions, they should be bulleted.

RECOMMENDATIONS
This section translates the conclusions into action statements. Each recommendation must be clearly linked to the evidence and include a quality of evidence label (e.g., A). Recommendations should not be broader or narrower than the clinical question. If there are a number of recommendations, they should be bulleted.

RECOMMENDATIONS FOR FUTURE RESEARCH
This section presents the identified gaps in the literature.

TOOLS
Tables, algorithms, or figures should be presented if they help communicate—but not alter—the evidence-based recommendations.

DISCLAIMER
The following disclaimer must appear on all guidelines:

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

ACKNOWLEDGMENTS

REFERENCES

Appendix 12

Sample Revision Table

#	Reviewer	Criticism	Action
1	R.F. Nelson (AAN Ethics Committee)	1. Clarify the diagnostic criteria 2. PEJ vs PEG. 3. "Breaking the News" is a flippant term 4. Editorial changes suggested	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	J. Belsch	1. Many aspects of symptomatic care are not covered 2. Some evidence from only 1 or 2 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.	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	M. Swash	1. Delete the option on laryngectomy for recurrent aspiration. 2. The word "entrapment" with respect to tracheostomy/ventilator without proper planning is unclear. 3. Extensive editing.	1. No change; evidence supports its consideration in patients with both aphonia and recurrent aspiration. 2. The word "entrapment" is dropped and the phrase clarified. 3. Selectively accepted.

Appendix 13

Policy for Dealing with Grey Literature and with Data from Published Reviews in the Creation of Clinical Practice Parameters

1. Background

A systematic appraisal of all the relevant evidence is crucial to assembling evidence based practice parameters, which must be free from publication bias. The latter is defined as the tendency to publish articles containing positive findings, especially "new" results, in contrast to reports that do not yield "significant" results, or results that do not accord with previously published findings (1). Frequently excluded studies are those with limited distribution, those not included in bibliographical retrieval systems such as PubMed, small studies, those with negative results, and those published in non-English languages. Strong empirical evidence demonstrates that excluding these articles results in a substantial risk of bias and threatens the validity of evidence based summaries, such as systematic reviews, meta-analyses, and practice parameters (2-8). In particular, it can lead to overestimation of the effects of interventions (9).

Another important practical issue facing authors of clinical practice parameters pertains to dealing with the evidence already assembled in existing reviews on the topic at hand. For example, should existing meta-analyses be given an evidence level, and should their summary results be taken at face value in drafting recommendations of practice parameters?

Based on analysis of the published evidence and of the methodology of relevant QSS practice parameters, the following recommendations are proposed.

2. Publication Bias

A) Every effort should be made to conduct as exhaustive and comprehensive a literature search as possible.
B) The assembly of the literature search strategy should involve the assistance of a library expert and focus on sensitivity (inclusiveness) rather than specificity (exclusiveness).
C) The practice parameter should include text or a table describing the literature search strategy, sources consulted for published and grey literature, numbers of studies retrieved and number included for full review. This material may be published in the text of the manuscript or on-line.
D) QSS will develop a system to assist authors in applying statistical methods to assess the presence and magnitude of publication bias in their data set.

2.1 Published Literature

A) The literature search should include at least the following databases, based on relevance to the topic at hand: National Library of Medicine (eg., PubMed), CINAHL, PsychLit, EMBASE, and the systematic review and the registry of randomized trials components of the Cochrane Database.
B) Authors should review citations of pertinent studies, reference sources and reviews for additional studies
C) Every effort should be made to include non-English language studies. Support for translation of articles may be obtained from QSS. Alternatively, two author panel members fluent in the language of the article may independently appraise it and agree on its eligibility, level of evidence and data abstraction.

2.2 Grey Literature

This can be defined as evidence that is unpublished, has limited distribution, or is not included in bibliographical retrieval systems (9).

A) Every effort should be made to assess this evidence in the assembly of evidence summaries for practice parameters. An exhaustive literature search as described in 2.1.1 will go a long way towards identifying some of this literature. Additional sources may include sponsors of industry-funded studies, known lead investigators, the FDA registry, and the National Institute for Clinical Excellence (NICE).
B) Evidence does not have to be published in peer-reviewed journals in order to be considered for inclusion. The rigorous process of critically appraising and rating of the evidence which is inherent in the creation of all clinical practice parameters will be applied equally to peer-reviewed and non-peer-reviewed evidence.
C) At times, additional, unpublished information pertaining to specific clinical trials may be uncovered by searching Clinical Trial Registries. Most often such registries include only generalities on design and aims of the trials. However, it is important to consult trial registries because they may shed light on ongoing, concluded, or published data from specific studies. This is important not only to avoid publication bias, but also to determine if and when new evidence may be available to mandate updating of a practice parameter.
D) The **Current Controlled Trials (CCT)** registry at http://www.controlled-trials.com/ (last consulted July 08, 2006) is a useful starting point. This UK based index contains 4,548 registered randomized trials. It is regularly updated by some European, NIH and Canadian research federal funding agencies. For example, it includes trials listed under the NIH's ClinicalTrials.gov (http://clinicaltrials.gov). The CCT assigns a unique number to all RCTs – the ISRCTN (International Standard Randomised Controlled Trial Number). This simplifies identification and provides a means of unambiguously tracking a trial throughout its life cycle. Records of trials to which ISRCTNs have been assigned are available at http://www.controlled-trials.com/isrctn/browse/A/1/10/results.asp.
E) QSS will work on designing a standardized search strategy for grey literature.

3. Using Data from Existing Systematic Reviews

Reviews can be categorized as traditional reviews, systematic reviews and meta-analyses. Traditional reviews include publications such as chapters, editorials, and expert reviews. Systematic reviews 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 a systematic review plus statistical pooling of the results into a single summary measure, such as an odds ratio, relative risk, or risk difference. In addition, systematic reviews or meta-analyses may be embedded in such studies as economic evaluations and decision analyses.

Systematic reviews and meta-analyses are of particular importance to our practice parameter process. These studies contain much of the work required for a practice parameter (eg., literature search, study selection, critical appraisal, and summary of results). Therefore, it is tempting to use their results at face value. However, there are important disadvantages to this approach. Specifically there are often small but important differences in the specific question addressed, the literature search, the definitions of clinical conditions and interventions, the thresholds for assessing outcomes, and the dates of the literature review. Furthermore, evidence ratings systems are usually different and studies may not be described in sufficient detail to apply QSS ratings.

We compared the methodology and conclusions of an existing meta-analysis on prognosis of coma following cardiopulmonary arrest (10) and a QSS practice parameter on exactly the same topic (11). This showed that even with such similarity of topics, there were important differences in the literature search strategy, languages of articles included, eligibility criteria, definition of outcomes, types of predictors included in the analysis, date of the literature search, number of eligible articles, rating of the evidence, and most importantly, on the overall results and recommendations. Accordingly we propose the following recommendations.

A) All systematic reviews on the topic at hand should be acknowledged in the section on results of the literature search.
B) The references cited in systematic reviews should be independently assessed for eligibility, and they should be critically appraised and graded.
C) Results and summary results (meta-analyses) of systematic reviews will not be used in drafting recommendations.
D) Results of individual studies as described within published systematic reviews will not be used at face value in drafting recommendations.
E) If an existing systematic review arrives at different results than those obtained by QSS-TTA supported guidelines, this should be acknowledged and explained in the text of the practice parameter.

Reference List

(1) A dictionary of epidemiology., 3 ed. New York: Oxford University Press, 1995.

(2) Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337(8746):867-872.

(3) Moher D, Fortin P, Jadad AR, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;347(8998):363-366.

(4) Ioannidis JP, Cappelleri JC, Sacks HS, Lau J. The relationship between study design, results, and reporting of randomized clinical trials of HIV infection. *Control Clin Trials* 1997 Oct;18(5):431-444.

(5) Callaham ML, Wears RL, Weber EJ, Barton C, Young G. Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. *JAMA* 1998;280(3):254-257.

(6) Link AM. US and non-US submissions: an analysis of reviewer bias. *JAMA* 1998;280(3):246-247.

(7) Misakian AL, Bero LA. Publication bias and research on passive smoking: comparison of published and unpublished studies. *JAMA* 1998;280(3):250-253.

(8) Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *Br Med J* 2000 Jun 10;320(7249):1574-1577.

(9) McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000 Oct 7;356(9237):1228-1231.

(10) Zandbergen EG, de Haan RJ, Stoutenbeek CP, Koelman JH, Hijdra A. Systematic review of early prediction of poor outcome in anoxic-ischaemic coma. *Lancet* 1998 Dec 5;352(9143):1808-1812.

(11) Wijdicks EF, Hijdra A, Young GB, Bassetti C, Wiebe S. Practice Parameter: Prediction of outcome in comatose survivors after cardiopulmonary resuscitation (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 67[2], 203-210. 2006.