

# From Evidence to Action

Gary S. Gronseth

Department of Neurology, University of Kansas Medical Center, Kansas City, Kansas 66160

---

**Summary:** Despite the availability of vast quantities of evidence from basic biomedical and clinical studies, a gap often exists between the optimal practice suggested by the evidence and actual practice. For many clinical situations, however, evidence is unavailable, of poor quality or contradictory. Out of necessity, clinicians have become accustomed to relying on non-evidence-based tools to make decisions. Out of habit, they rely on these tools even when high-quality evidence becomes available. Growing out of an increasing awareness of this problem, the evidence-based medicine (EBM) movement sought to empower clinicians to find the evidence most relevant to a specific clinical question. Various organizations have used EBM techniques to develop systematic reviews and practice guidelines to aid physicians in making evidence-based deci-

sions. A systematic review follows a process of asking a clinical question, finding the relevant evidence, critically appraising the evidence and formulating conclusions and recommendations. Results are mixed on whether educating physicians about evidence-based recommendations is sufficient to change physician behavior. Barriers to adopting evidence-based best practice remain, including physician skepticism, patient expectations, fear of legal action, and distorted reimbursement systems. Additionally, despite enormous research efforts there remains a lack of high-quality evidence to guide care for many clinical situations. **Key Words:** Diffusion of innovation, clinical practice guidelines, systematic reviews, quality of care.

---

## THE EVIDENCE–PRACTICE GAP

Physicians have at their disposal vast quantities of evidence from basic biomedical research and clinical studies. As discussed in other articles in this issue of *NeuroRx*®, much of this evidence is potentially useful for guiding clinicians' actions in patient care. Despite the availability of this evidence, however, a gap often exists between the optimal practice suggested by the evidence and actual practice.

One need not look hard to find examples of this evidence–practice gap. Many physicians still routinely use acute anticoagulation to treat noncardiogenic stroke, despite ample evidence that anticoagulation is not helpful, and indeed potentially harmful, in this setting.<sup>1</sup> Some physicians, including neurologists, fail to prescribe antiplatelet medications for the secondary prevention of stroke despite evidence of efficacy.<sup>2</sup> Patients with atrial fibrillation at high risk for stroke (e.g., those presenting with transient ischemic attacks or stroke) often do not

receive warfarin for secondary stroke prevention even when there are no contraindications.<sup>3</sup>

A more systematic look at the evidence–practice gap was provided by Schuster et al.<sup>4</sup> These authors demonstrated that only 70% of patients receive recommended acute care and 30% of patients receive contraindicated acute care.<sup>4</sup> This study and others provide ample evidence for the existence of an evidence–practice gap.

Many reasons for the existence of the gap have been postulated.<sup>5</sup> The most frequently cited is that it is impossible for the practicing clinician to keep up with the sheer volume of evidence provided by basic science and clinical research. Although there is no doubt the overwhelming quantity of evidence requiring assimilation plays a role, more fundamental barriers likely exist. After all, there are many venues including continuing medical education offerings and review articles in which large amounts of evidence are distilled and presented to practicing clinicians. The passive diffusion of evidence through these means is often not enough to change practice. Why?

Despite the large volume of basic science and clinical research studies, for many clinical situations evidence is unavailable, of poor quality, or contradictory. Out of necessity, clinicians have become accustomed to relying

---

Address correspondence and reprint requests to Gary Gronseth, M.D., Department of Neurology, University of Kansas Medical Center, 1033 Wescoe Pavilion, 3901 Rainbow Boulevard, Kansas City, KS 66160-7314. E-mail: ggronseth@kumc.edu.

on non-evidence-based tools to make decisions. Out of habit, they rely on these tools even when high-quality evidence becomes available.

To illustrate these points, consider a common clinical scenario. You are seeing a middle-aged, otherwise healthy patient in your office because they awoke this morning with complete right-sided facial weakness. After your clinical evaluation, you convince yourself the patient has Bell's palsy and consider starting oral steroids in hopes of improving the patient's chance of complete recovery of facial function. On your way to the medicine cabinet of the office to get a steroid dose pack you find a colleague and ask what dose of steroids she normally uses to treat Bell's palsy. To your surprise, she states she never uses steroids for Bell's palsy. Feeling somewhat defensive, you initiate a discussion to try to justify your position.

We have all been involved in such discussions with colleagues in which differences in practice habits are vetted. The kinds of arguments used to support one's position vary somewhat but tend to follow certain common themes. They highlight the non-evidence-based considerations used when making decisions.

One such argument might be: "Where I trained, we always used steroids to treat Bell's palsy." Of course, your colleague might retort she was trained differently. Falling back on how we were trained is not a very convincing argument. There is no doubt, however, that our training is a major reason why we practice the way we do. Our teachers, particularly the ones we respect and like, have a profound influence on our practice habits. Indeed, we tend to emulate them and adopt their practice habits. Unfortunately, the criteria we use in selecting the teachers we emulate often have more to do with the power of their personalities than the correctness of their actions.

You might attempt to convince your colleague of the correctness of your actions with a reasoned deductive argument based on established principles about the pathophysiology of Bell's palsy. For example: "Bell's palsy results from compression of the swollen, inflamed facial nerve within the temporal bone. Steroids reduce inflammation and swelling. Therefore, steroids should reduce the swelling and compression of the facial nerve and improve outcomes." The trouble with many deductive arguments in a field as complex as medicine is that it is often possible to develop a deductive argument to support the opposite action. For example: "Bell's palsy might be caused by a viral infection (perhaps herpes simplex). Steroids interfere with the immune response. Therefore, steroids should be avoided in patients with Bell's palsy because steroids will interfere with the immune response and may worsen outcomes." Deductive arguments are often unconvincing.

Many of us would appeal to our accumulated experience with Bell's palsy patients to make our case: "All of the Bell's palsy patients I have seen who got steroids did well. The few I've seen who didn't get steroids seemed to do poorly." Of course, your colleague may counter that her experience is different: "All of my Bell's palsy patients do well despite not getting steroids. Also, I can remember one diabetic patient I saw in the intensive care unit with severe ketoacidosis after receiving steroids for Bell's palsy." Arguments based on experience can be misleading. Our memories of cases tend to be biased to the more recent, those with extreme outcomes and those that support our preconceptions.<sup>6</sup> Such experience-based arguments are commonly dismissed as "anecdotal." Actually, however, experience-based arguments are a type of evidence-based argument, albeit a weak evidence-based argument.

Often we rely on the opinion of an expert to support our practice habits. After all, the expert should have more experience and superior knowledge of established principles. Here is a quote from a published expert you might use to convince your colleague that you are right in choosing steroids: "From a practical view point, treatment with steroids is appropriate for the management of Bell's palsy."<sup>7</sup> Your colleague, however, might find this quote to contradict the first expert: "Bell's palsy remains without a proven efficacious treatment."<sup>8</sup> It is often possible to find credible experts with conflicting opinions.

The non-evidence-based tools considered thus far—teachings, deduction, experience, and expert opinion—though flawed and unconvincing in the Bell's palsy scenario, seem to have at least some value in helping to inform our clinical decisions. Paradoxically, the most persuasive non-evidence-based arguments encountered in medicine have no value in informing our decisions.

Consider these arguments: "The use of steroids for Bell's palsy has become the standard of care in the community." Or, "The consequences of disfiguring facial weakness are so devastating that the use of steroids is mandatory." Or, "If I don't use steroids and the patient has a poor outcome, I'll be sued." These arguments are rhetorical. They are very persuasive but share one common feature; they all rely on logical fallacies.<sup>9</sup> The first statement makes an appeal to popularity; we should give steroids because everyone else gives steroids. The second statement begs the question: yes disfiguring facial weakness is bad but that does not mean steroids are effective at preventing disfiguring weakness. The third statement invokes an irrelevant outcome; whether we are sued or not is irrelevant to the question of whether steroids improve facial functional outcomes in patients with Bell's palsy. Although persuasive, such rhetorical arguments are irrelevant to the clinical question and are of no value in informing our decisions.

What about evidence-based arguments? Like many physicians, you might keep a set of files containing articles relevant to certain clinical entities. You could pull your Bell's palsy file. Within it you find an article by Shafshak et al.,<sup>10</sup> published in 1994, in which the authors prospectively studied a group of patients with Bell's palsy, some of whom received steroids and some of whom did not receive steroids. The patients that received steroids had better outcomes than the patients that did not. The authors concluded that steroids were beneficial. Should your colleague be convinced?

Your colleague has her own Bell's palsy file. In it she has an article from May et al.,<sup>8</sup> published in 1975, in which the authors randomly assigned patients with Bell's palsy to receive steroids or placebo.<sup>8</sup> There was no significant difference in facial functional outcomes between the two groups. The authors concluded steroids were ineffective. Like expert opinion, it would seem evidence can be contradictory.

At this point, you and your colleague realize that the two articles you have likely do not represent all of the relevant studies on the effectiveness of steroids in treating Bell's palsy. So you go to a nearby computer and perform a MEDLINE search. You find two reviews completed by the American Academy of Neurology (AAN)<sup>11</sup> and the Cochrane collaboration,<sup>12</sup> in which the literature regarding Bell's palsy was systematically identified and analyzed. Their conclusions: despite finding hundreds of articles with evidence pertinent to the question, both organizations have concluded that the benefit (or lack of benefit) of steroids in treating Bell's palsy has not been established. Hundreds of articles and no answer; why? Because all of the studies performed thus far have methodological flaws that preclude definitive conclusions.

In this setting, as in many clinical settings, both the non-evidence-based tools and the evidence-based tools do not provide a definitive answer. After dismissing rhetorical arguments and considering arguments based on training, experience, expert opinion, deduction, and evidence, we find we cannot convince our colleague (or, if we are honest, ourselves) that prescribing steroids for Bell's palsy is the correct action. None of the arguments, even the evidence-based one, seems definitive.

Of course, in the real world the various types of arguments listed above would not be systematically considered as we did here. Rather they would all be mixed together. Whether systematically considered or not, in the end a decision has to be made. In this case, as with most decisions in medicine, we would be forced to rely on our judgment.

### THE OTHER GAP: THE PRACTICE-EVIDENCE GAP

In the Bell's palsy scenario, the evidence-based argument is no more able to answer our question than the

non-evidence-based arguments. In this instance, there is a gap between what we need to know to practice medicine and the availability of evidence[em]a practice-evidence gap.

The absence of high-quality evidence is not unique to the study of Bell's palsy. Despite peer review, the literature is full of poorly designed, poorly executed and improperly interpreted studies. Indeed, for most controversial clinical questions in medicine, definitive trials to guide practice do not exist. Thus, under most circumstances clinicians must rely on non-evidence-based tools (teachings, expert opinion, experience, and deduction) to make their judgments on how best to care for patients.

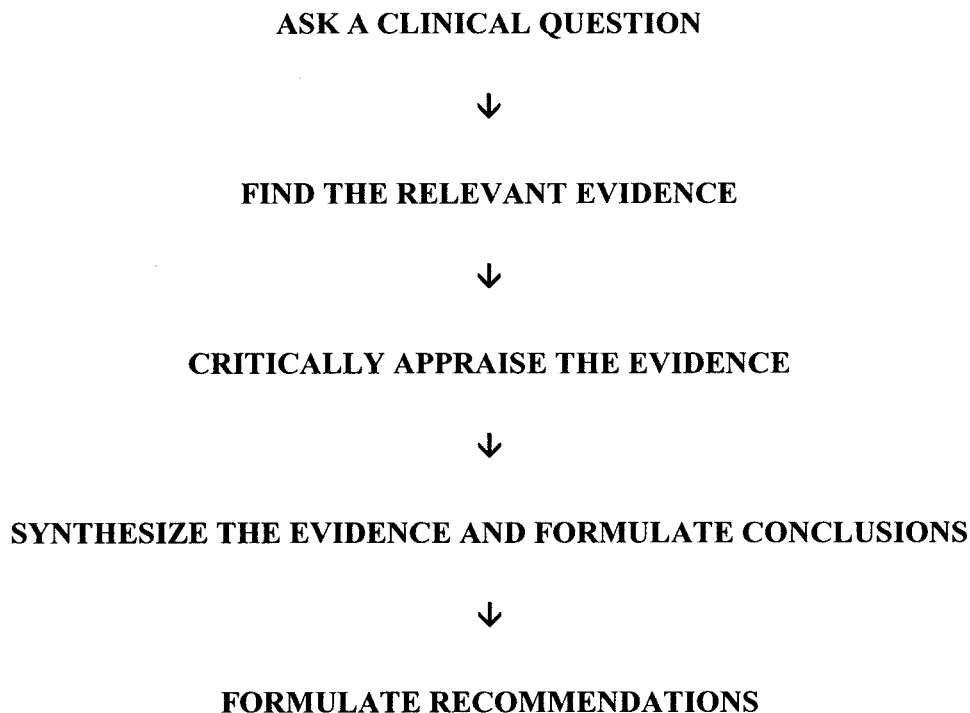
Under these circumstances, one should not be surprised that clinicians accustomed to practicing in an environment where the majority of published studies are flawed and contradictory, might be slow to adopt recommendations based on the occasional high-quality study. The evidence-practice gap exists in large part because vast quantities of poor-quality studies make it difficult for physicians to find and believe a definitive study.

### CLOSING THE EVIDENCE-PRACTICE GAP: IDENTIFYING THE BEST EVIDENCE

Occasionally, there is good evidence to inform clinical decisions. Finding it can be a challenge. Growing out of an increasing awareness of this problem, the evidence-based-medicine (EBM) movement began in the early 1990s. The movement was made possible by the explosion in informatics technologies. The wide accessibility of online bibliographic databases such as MEDLINE and the availability of bibliographic management software gave clinicians practical tools for managing the overwhelming quantity of evidence provided by basic science and clinical research. The EBM movement sought to empower clinicians to find the evidence most relevant to a specific clinical question, to identify the evidence most likely to be correct by applying well established concepts of clinical trial design (many of which have been discussed in previous sections), and to change their own practice behaviors to improve patient outcomes.

The initial thrust of the EBM movement was to help individual clinicians make better evidence-based decisions. Seeing the potential of this approach for closing the evidence-practice gap, various organizations including many specialty societies, the Cochrane collaboration and Agency for Healthcare Research and Quality (AHRQ) have also used EBM techniques to inform their audiences. The resultant products—systematic reviews and practice guidelines—are made available to physicians to aid them in making evidenced-based decisions for their patients.

Systematic reviews consist of a highly formalized review of the literature designed to answer a specific clin-



**FIG. 1.** Evidence-based medicine process of asking a clinical question to formulating a recommendation.

ical question. The systematic review is distinct from a typical literature review in that it is complete and transparent. By “complete” it is meant that it critically appraises all the relevant evidence, not just the evidence that might support a specific opinion. By “transparent” it is meant that the methods used are explicitly stated including the criteria used to identify and critically appraise the evidence. Furthermore, evidence and expert opinion are explicitly distinguished. The conclusions of systematic reviews serve as succinct summaries of the literature pertinent to specific aspects of care. Often, but not necessarily, the conclusions of a systematic review serve as the foundation for evidence-based practice recommendations (a practice guideline).

Practice guidelines are specific recommendations for patient management that assist physicians in clinical decision-making. A practice guideline is a series of specific, practice recommendations that answer an important clinical question (e.g., What pharmacological interventions should I consider for patients with new-onset Bell’s palsy to improve facial functional outcomes?).

Although there are some differences in the EBM approach taken by different organizations, they have much in common. This section will emphasize the process employed by the AAN. Differences between the process of the AAN and that of other organizations will be highlighted where appropriate. Asking and answering a specific clinical question forms the backbone of the EBM process. It follows the progression outlined in Figure 1.

#### **Forming an author panel**

Although EBM techniques were initially designed to allow individual clinicians to identify and critically appraise the literature, many organizations enlist the help of content experts to develop their systematic reviews and practice guidelines. The experts are often helpful at each step in the process, including formulating the relevant clinical questions and identifying evidence that might be missed through database search strategies. Often, a group with some methodological expertise (e.g., the quality standards subcommittee of the AAN) provides oversight of the author panel to ensure recommendations that represent the experts’ opinions are clearly distinguished from those that follow from the evidence.

Other organizations (such as those sponsored by AHRQ) employ methodological experts to develop the systematic review. Under these circumstances, oversight is provided by a group of content experts to ensure clinical relevance of the final product.

#### **Formulating specific clinical questions**

Beyond selecting a general topic area, formulating specific answerable clinical questions is a critical step in the EBM process. The clinical questions follow a common template with the three following 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.), and 3) outcome: the outcome(s) to be considered.

Questions are formulated so that the three elements are easily identified. For example, for patients with Bell's palsy (population), do oral steroids given within the first three days of onset (intervention) improve facial functional outcomes (outcome)?

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. Of course, 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.

Many organizations such as the Cochrane collaboration (perhaps because the techniques for assessing and synthesizing the evidence from therapeutic studies are more developed) have thus far limited their reviews to therapeutic questions; whereas others, such as the AAN, have considered all types of questions.

### **Finding the relevant evidence**

A comprehensive, unbiased search for relevant evidence is one of the key differences between a systematic review and a traditional review.<sup>13</sup> To ensure transparency and allow reproducibility, the systematic review explicitly enumerates the databases searched and the search terms employed.

The authors of a systematic review develop criteria for including or excluding articles during the literature search and article review processes. The criteria are developed before beginning the search. The article selection criteria employed usually include content criteria and methodological criteria. Content criteria focus on including articles relevant to the patient population, in-

tervention, and outcomes specified in the clinical question. Methodological criteria are usually used to exclude evidence with a high risk of bias (e.g., all studies without a concurrent control group).

### **Critical appraisal of the evidence**

All of the studies ultimately included in a systematic review have to be critically assessed relative to quality. Major study characteristics requiring formal assessment include the studies power (random error), risk of bias (systematic error), and generalizability (relevance to the clinical question).

Bias, or systematic error, is the tendency of the study to inaccurately measure the effect of the intervention 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, one can estimate the risk of bias of a study. The risk of bias is usually measured using an ordinal classification scheme in which studies are classified as having a relatively high or low risk of bias. The specifics of the schemes used by differing organizations vary but most share similar elements. The risk of bias of a study can only be judged relative to a specific clinical question. Because methodologies and sources of bias are so different, therapeutic, diagnostic, or prognostic accuracy and screening questions are necessarily judged by different standards. For AAN guidelines, the risks of bias in studies undergoing analysis for a therapeutic question are measured using a four-tiered classification scheme (Table 1).

The quality-of-evidence classification schemes employed by most organizations, including the AAN and the Cochrane collaboration, are designed to account for bias only. Random error (low study power) can be dealt with separately by using statistical techniques such as inspection of 95% confidence intervals when determining the results of the studies. Some organizations include considerations of the statistical power of a study when assigning a quality-of-evidence grade. Regardless of the technique used, both systematic and random errors are explicitly considered in a systematic review.

Usually, issues of generalizability (sometimes referred to as external validity) are dealt with during the article selection phase of the systematic review process. Articles that are not relevant to the clinical question are excluded. Issues of generalizability can resurface, however, if patients included in a study seem atypical of the patient population at large. This is particularly true if the results of the studies are inconsistent.

### **Synthesizing evidence: formulating conclusions**

Synthesizing the evidence after it has been identified and formally appraised can be challenging. The goal at

**TABLE 1.** American Academy of Neurology Evidence Classification Scheme for Therapeutic Articles with Linkage of the Class of Evidence to the Strength of Conclusions and Recommendations

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

\*Wording relevant to diagnostic, prognostic, and screening questions is indicated in parentheses.

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

<sup>‡</sup>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).

this step in the process is to develop a succinct statement that summarizes the evidence to answer the specific clinical question. Ideally, the summary statement indicates the effect of the intervention on the outcome and the quality of the evidence that it is based upon. The conclusion should be formatted in a way that clearly links it to the clinical question. For example, patients with new-onset Bell’s palsy oral steroids given within the first 3 days of palsy onset probably improve facial functional outcomes.

When all of the studies demonstrate the same result, have a low risk of bias, and are consistent with one another, developing the conclusion is straightforward. Often, however, this is not the case. Organizations differ in the approach they take to evidence synthesis under these circumstances.

The largest source of variation in evidence-synthesis methodologies is the reliance placed on consensus. Organizations relying on a consensus-based approach use a panel of content experts who have been exposed to (in

some cases actually perform) a systematic review of the literature. The expert panel, after considering the evidence, crafts the conclusions and recommendations. The consensus method used can be formal (e.g., a modified nominal group process<sup>14</sup>) or informal. Other organizations, including the Cochrane collaboration, use more formal techniques, including meta-analyses, and insist that only conclusions directly supported by the evidence can be made. Between these two extremes is a spectrum of methodologies combining some consensus and formal evidence synthesis.

Formal evidence synthesis methodologies generally consider four kinds of information when formulating the conclusion: the quality of the evidence (usually indicated by the class-of-evidence grade), random error (the power of the study as manifested, for example, by the width of the 95% confidence intervals), the effect size measured (how much improvement did the intervention make in the outcome), and the consistency of the results between studies.

There is a natural tension that exists between synthesis methodologies designed to reduce random error and those designed to minimize bias. The best approach to minimize random error is to include all studies in a meta-analysis.<sup>15</sup> In a meta-analysis the results of studies are combined statistically to obtain a single, precise estimate of the effect of the intervention on the outcome. However, this approach fails to account for bias. The precise estimate obtained may be inaccurate if poor-quality studies are included in the meta-analysis. We are aware of no specialty societies that currently employ meta-analysis without regard to bias in formulating conclusions. However, examples of such indiscriminate meta-analysis can be found throughout the literature and is one reason meta-analysis is sometimes distrusted.<sup>16</sup>

The best approach for minimizing bias in developing conclusions is to exclude all but the highest quality of evidence from the analysis. Excluding a large number of studies can, however, increase the random error of the estimated effect of the intervention on the outcome (i.e., decrease the precision of the estimate of the effect of the intervention on the outcome). Employing this technique under some circumstances might mean that no evidenced-based conclusions can be developed, as there may be no high-quality studies. In other circumstances the high-quality studies that exist may be insufficiently powered to answer the question. To minimize bias, the Cochrane collaboration considers only randomized controlled trials in the formulation of their conclusions. Observational studies, with their inherently higher risk of bias, are routinely excluded.

Intermediate approaches between the indiscriminate meta-analysis and exclusion of all less-than-high quality studies can be used. In some situations, the AAN uses a technique known as a cumulative meta-analysis. In this

technique, the highest quality studies are combined in a meta-analysis first. If the confidence intervals of the resulting pooled effect size are sufficiently narrow, the analysis stops and conclusions are crafted. If the pooled effect size is imprecise, the next highest quality studies are added and the meta-analysis repeated. The procedure is repeated until a sufficiently precise estimate of the effect of the intervention is obtained or until there are no further studies to analyze.

Most organizations performing systematic reviews formally indicate the level of certainty of their final conclusions. This level of certainty is generally linked to the quality of evidence upon which the conclusions are based. The linkage used by the AAN is illustrated in Table 1. For example, an AAN conclusion labeled "Level A" corresponds to a high level of certainty and indicates that the conclusion is based solely on studies with a low risk of bias. A conclusion labeled "Level C" corresponds to a low level of certainty based on studies with a moderately high risk of bias.

Problems arise in the formal synthesis of evidence when studies of similar quality show inconsistent results. "Inconsistent result" means that random error does not explain the difference in the results of the studies. Informally, inconsistency can be determined by inspecting the confidence intervals generated from the studies. If the intervals do not overlap, the studies are inconsistent. Formally, inconsistency is determined with a test of homogeneity obtained during a meta-analysis. Sometimes differences between the study populations, interventions, and outcome measures are sufficient to explain the inconsistencies. At times, however, 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.<sup>17</sup> In such circumstances, the only option may be to rely on the consensus opinion of a panel of experts who have reviewed the evidence to formulate conclusions.

### The formulation of 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, 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 B).

Additional factors to be considered when formulating the recommendations that are not considered in formulating conclusions include the magnitude of the effect of the intervention on the outcome, the availability of alternative interventions, and the relative value of various outcomes. The relative value of outcomes is a subjective determination of the desirability of the outcomes.

Usually, organizations developing practice guidelines do not use formal synthesis methods in developing recommendations. Rather, a panel of experts, which may include patients with the condition and other stakeholders, come to a consensus recommendation after reviewing the evidence. The consensus process employed is usually informal.

There are formal synthesis techniques that can be used to develop practice recommendations. These include decision analysis and cost-effectiveness analysis. Currently, no organization routinely uses these techniques when developing recommendations.

Under most circumstances, there is a direct link between the level of evidence used to formulate conclusions and the strength of the recommendation. Thus, in the AAN's scheme a Level A conclusion supports a Level A recommendation (Table 1). There are important exceptions to the rule of having a direct linkage between the level of evidence and the strength of recommendations. Some situations in which it may be necessary to break this linkage are, as follows:

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

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

### **CLOSING THE EVIDENCE-PRACTICE GAP: CHANGING BEHAVIOR**

When high-quality evidence supporting practice recommendations can be found, organizations face the challenge of changing the behavior of physicians to follow the guidelines. Thus far, education has been the main tool organizations have used to try to change physician behavior.

#### **Educating the skeptical**

Systematic reviews and practice guidelines are disseminated to physicians in multiple ways. Most guidelines developed by specialty societies are published in that specialty's journal. The AAN guidelines for example are published in the journal *Neurology*. Guidelines are also

commonly posted or linked to various websites. Additional "marketing" methods commonly employed include press releases, slide presentations, summary versions of the guideline, algorithms and other tools to help physicians incorporate guideline recommendations into their practices. Finally, guidelines meeting certain criteria for methodological rigor are published in compendiums such as the National Guidelines Clearinghouse.

Despite the aggressive dissemination and marketing of systematic reviews and practice guidelines, evidence is mixed on whether awareness of the recommendations alone change physician behavior and narrow the evidence-practice gap.<sup>18</sup>

If dissemination methods make clinicians aware of the best evidence and thus best practice, why won't some change? Is it that they do not believe the evidence identified by the EBM processing techniques or are there other barriers preventing them from implementing best practice?

As previously discussed, clinicians are accustomed to using non-evidence-based means such as experience and expert opinion to inform their judgment. Clinicians are also accustomed to practicing in circumstances where evidence has not provided the answers. Thus, when they are exposed to a systematic review where evidence is complexly processed, perhaps we should not be surprised that they are skeptical of the recommendations that result. Clinicians question the validity of EBM process and the resultant guidelines.<sup>19</sup>

Well, is EBM valid? To an extent asking this is akin to asking if neurology as a discipline is valid. The question is too broad. There is certainly ample evidence that specific principles and practices upon which the discipline of neurology is based are valid. Likewise, there is ample evidence that the fundamental principles upon which EBM is based are valid. The benefit of considering all the evidence seems self-evident. Concepts used to appraise the quality-of-evidence follow directly from principles of good study design that are already empirically validated. The statistical principles relating to the analysis of random error are also well established. Based on this, I conclude EBM is valid.

In my opinion, physicians question the validity of EBM as presently practiced because it too easily dismisses the validity of the non-evidence-based means clinicians use in making decisions. Surely, the commonly encountered fallacious arguments should be dismissed. But there is value in the rest: what we learned, experience, an expert's opinion and the well-formulated deductive argument. Until systematic reviews are able to assess and incorporate these non-evidence-based sources of knowledge as effectively as they process evidence, clinicians will remain skeptical. Such skepticism will only be overcome when clinicians become convinced that



EBM provides an additional tool to extend and refine their judgment.

### Other barriers

Even when clinicians become convinced of the validity of the evidence regarding best practice, other barriers to changing practice persist. Patient expectation is one such barrier. One reason clinicians over-prescribe antibiotics for probable viral infections, despite evidence that it is not helpful and likely harmful, is that patients demand it.

Another barrier commonly cited is the fear of being sued. Clinicians believe certain practices, even when not supported by the evidence, represent the “standard of care” and if they do not follow them, they risk being sued. Thus, despite evidence to the contrary, head imaging is routinely ordered for patients with uncomplicated headaches and normal neurological exams. As previously discussed, the concern of legal action is irrelevant to the question of best practice. This concern however, is very effective in persuading a physician not to change behavior.

Economic barriers also exist. Financial demands may lead a clinician to order a highly reimbursed but questionably beneficial diagnostic procedure. We are just beginning to understand these other barriers to best practice.

The development and dissemination of systematic reviews and practice guidelines is one of the first steps in the attempt to close the evidence–practice gap. Although provider, patient, and organizational education will remain the focus in the near future, more innovative techniques for encouraging providers to change their actions to those supported by best evidence need to be developed. Indeed, the informatics revolution that provided the tools that made the initial EBM movement practical may soon provide the tools necessary for the next step. The widespread availability of the electronic medical record may better inform clinicians of best practices at the point of care delivery and could facilitate quality improvement efforts by encouraging compliance with guideline recommendations. Some even envisage a reimbursement system that rewards best-evidence practice habits.

### CLOSING THE PRACTICE–EVIDENCE GAP: INFLUENCING THE RESEARCH AGENDA

Despite vast research efforts, for many clinical situations there is a practice–evidence gap: there is little high quality evidence to guide care. Often, this is not explicitly evident until a systematic review is performed that exposes the gaps in evidence.<sup>20</sup> Identifying these evidence gaps and communicating them to researchers and organizations funding research is potentially as valuable

as the attempts to inform clinicians of evidence-based best practice.

For the most part, organizations producing systematic reviews have dealt with these research concerns as an afterthought by adding a “Future Research Recommendations” section to their documents. However, just as the practice recommendations of systematic reviews need to be aggressively marketed to clinicians, research recommendations need to be aggressively marketed to researchers. The AAN, along with other organizations, is starting to formalize this process. Hopefully, researchers will not be as resistant to the research recommendations emanating from systematic reviews as clinicians have been to the practice recommendations.

### REFERENCES

1. Coull BM, Williams LS, Goldstein LB, Meschia JF, Heitzman D, Chaturvedi S et al. Anticoagulants and antiplatelet agents in acute ischemic stroke: report of the Joint Stroke Guideline Development Committee of the American Academy of Neurology and the American Stroke Association (a Division of the American Heart Association). *Neurology* 59:13–22, 2002.
2. Awtry EH, Loscalzo J. Aspirin. *Circulation* 101:1206–1218, 2000.
3. Hart R. Practice parameter: stroke prevention in patients with nonvalvular atrial fibrillation. *Neurology* 51:671–673, 1998.
4. Schuster MA, McGlynn EA, Brook RH. How good is the quality of health care in the United States? *Milbank Q* 76:517–563, 1998.
5. Davis DA. Physician education, evidence and the coming to age of CME. *J Gen Intern Med* 705–706, 1998.
6. Nisbett R, Ross L. Human inference. Englewood Cliffs, NJ: Prentice-Hall, 1980.
7. Stankiewicz JA. A review of the published data on steroids in idiopathic facial paralysis. *Otolaryngol Head Neck Surg* 97:481–486, 1987.
8. May M, Wette R, Hardin WB, Sullivan J. The use of steroids in Bell’s palsy: a prospective controlled study. *Laryngoscope* 86: 1111–1122, 1976.
9. Copi IM. Informal fallacies. In: Introduction to logic, pp 72–107. New York: Macmillan, 1972.
10. Shafshak TS, Essa AY, Bakey FA. The possible contributing factors for the success of steroid therapy in Bell’s palsy: a clinical and electrophysiological study. *J Laryngol Otol* 108: 940–943, 1994.
11. Grogan PM, Gronseth GS. Practice parameter: steroids, acyclovir, and surgery for Bell’s palsy (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 56:830–836, 2001.
12. Salinas RA, Alvarez G, Alvarez MI, Ferreira J. Corticosteroids for Bell’s palsy (idiopathic facial paralysis). *Cochrane Database Syst Rev* (1):CD001942, 2002.
13. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Tracking down the evidence to solve clinical problems. In: Clinical epidemiology, a basic science for clinical medicine, pp 335–358. Boston: Little, Brown and Company, 1991.
14. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 2:1–7, 1998.
15. Petitti DB. Statistical methods in meta-analysis. In: Meta-analysis, decision analysis and cost-effectiveness analysis (Petitti DB, ed), pp 90–114. New York: Oxford University, 1994.

16. Ramsey MJ, DerSimonian R, Holtel MR, Burgess LP. Corticosteroid treatment for idiopathic facial nerve paralysis: a meta-analysis. *Laryngoscope* 110:335–341, 2000.
17. Petitti DB. Advanced issues in meta-analysis. In: Meta-analysis, decision analysis and cost-effectiveness analysis (Petitti DB, ed), pp 70-90. New York: Oxford University, 1994.
18. Gifford DR, Mittman BS, Fink A, Lanto AB, Lee ML, Vickrey BG. Can a specialty society educate its members to think differently about clinical decisions? Results of a randomized trial. *J Gen Intern Med* 11:664–672, 1996.
19. Franklin GM, Zahn CA. AAN clinical practice guidelines: above the fray. *Neurol* 59:975–976, 2002.
20. Gronseth GS, Barohn RJ. Practice parameter: thymectomy for autoimmune myasthenia gravis (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* 55:7–15, 2000.