# SYLLABUS FOR BEDSIDE EBM COURSE 2017

**Gary Gronseth**
University of Kansas

**Michael Glantz**
Penn State Hershey

**Melissa Armstrong**
University of Florida

**Steve Messé**
University of Pennsylvania

**Tamara Pringsheim**
University of Calgary

## Course Description

Welcome to the 2017 "Bedside Evidence-based Medicine" course. The primary goals of this program is to de-mystify evidence-based medicine (EBM) concepts and equip participants with the knowledge and tools needed to translate evidence (a published article, a research presentation, clinical trial data) into action (a specific bedside intervention).

The program will minimize traditional didactic presentations, and will instead emphasize small group problem-solving using the peer instruction model. The program will consist of a series of short didactic presentations each followed by a small group exercise. Each participant will be assigned to a small group ($\leq$ 10 participants). Each small group will be led by a course faculty member with expertise in evidence-based medicine. Total class size will be limited to 50 participants.

Each faculty-mentored group will complete exercises illustrating specific steps in the EBM process including: formulating evidence-answerable questions, efficiently searching the literature to find high quality evidence, assessing a studies risk of bias, interpreting a studies results for significant and meaningful effects, synthesizing the results of several studies and using more than evidence to make patient-centered decisions.

Participants should come to the course with a laptop computer with the ability to connect to the internet wirelessly. Wireless internet access will be provided throughout the course. The laptop should also have a USB port. The computer should also have the Microsoft Excel and Power Point programs.

To illustrate EBM concepts participants will review the evidence pertinent to a current controversial topic. Although you will be exposed to EBM processes that are specific to one topic, the concepts that will be reviewed will be applicable to everyday clinical practice.

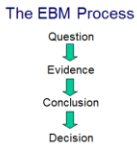
## Objectives

Upon Completion of this Program Participants will understand the principles of evidence-based medicine and critical review, and how these apply to the bedside practice of neurology; will be able to deconstruct published articles addressing therapeutic questions, assess their accuracy and rigor, and make appropriate evidence-based clinical decisions based on their assessment; will reduce their susceptibility to being misled by bias and the misuse of statistical procedures; and will feel comfortable developing, investigating, and publishing the findings of their own clinical research questions.

## Overview of the task

During the course, a specific clinical vignette describing a patient dilemma will be presented and discussed. Your job, will be to make a decision regarding a therapeutic intervention for the patient. The decision regarding the best course of action will be informed by the EBM process.

*Overview of the EBM process.* The EBM process follows a structured process. Asking a question is the first step in the EBM process. To answer a controversial clinical question, the EBM method would next require looking for strong evidence. So, what is evidence?




Evidence in an EBM context is information from any study of patients with the condition who are treated with the intervention of interest and are followed to determine their outcomes. To find such studies the EBM method uses comprehensive searches of online databases such as MEDLINE. The systematic literature search maximizes the chance that we will find all relevant studies.

When a study is found, we need to determine the strength of the evidence it provides. For this purpose EBM provides validated rules that determine the likelihood that an individual study accurately answers a question. Studies likely to be accurate provide strong evidence. Rating articles according to the strength of the evidence provided is especially necessary when different studies provide conflicting results. The study providing the strongest evidence should carry more weight.

After the finding and determining the strength of evidence of all the relevant studies, the next step in the EBM process is to synthesize the evidence to answer the question. The process of synthesizing the evidence must take into account, the risk of bias of the studies, the precision (or power) of the studies as well as what the studies demonstrated relative to the effect of the intervention.

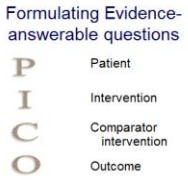
Finally, a decision must be made. Even when there is strong evidence to inform a question, judgments about how the evidence applies to your patient are crucial.

## Developing evidence-answerable questions

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

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

Clinical questions must have four components. These are often referred to by the mnemonic *PICO*:




1. Population: The type of person (patient) involved
2. Intervention: The exposure of interest that the person experiences (e.g. therapy, positive test result, presence of a risk factor.)
3. Co-intervention: An alternative type of exposure that the person could experience (e.g. no therapy, negative test result, absence of a risk factor)
4. 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 symptomatic myasthenia gravis, patients with Bell's palsy or patients with amyotrophic lateral sclerosis (ALS). The population of interest may also consist of patients at risk for a disease, for example patients with suspected multiple sclerosis or those 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). Likewise, 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 whether this intervention should be done. An example is, should patients with myasthenia gravis be treated with steroid-sparing agents?

Co-intervention
The co-intervention is the alternative to the intervention of interest. It is sometimes referred to as the comparator intervention. For therapeutic questions the co-intervention could be no treatment (or placebo) or an alternative treatment.

Of course, there are circumstances where there may be many potential treatment alternatives.

Outcomes
The outcomes to be assessed should be clinically relevant to the patient.

When specifying outcomes it is important to specify all of the outcomes that are relevant to the patient population and intervention. For example, the question might deal with the efficacy of a new anti-platelet agent in preventing subsequent ischemic strokes in patients with non-cardioembolic stroke. Important outcomes needing explicit consideration include the risk of subsequent ischemic stroke—both disabling and nondisabling—death, bleeding complications—both major and minor—and other potential adverse events.

In addition to defining the outcomes that are to be measured, the clinical question might 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.

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

*Therapeutic*. The easiest type of question to conceptualize is the therapeutic question. The clinician must decide whether to use a specific treatment. The relevant outcomes of interest are the effectiveness, safety, and tolerability of the treatment. The strongest study for determining the effectiveness of a therapeutic intervention is the masked, randomized, controlled trial (RCT). Because of the limited time available for this course, we will concentrate on therapeutic questions only.

*Diagnostic and Prognostic Accuracy.* There are many important questions in medicine that do not relate directly to the effectiveness of an intervention in improving outcomes. Rather than deciding to perform an intervention to treat a disease, the clinician may need to decide whether he or she should perform an intervention to determine the presence or prognosis of the disease. The relevant outcome for these questions is not the effectiveness of the intervention for improving patient outcomes. Rather, the outcome relates to improving the clinician's ability to *predict* the presence of the disease or the disease prognosis. The implication of these questions is that improving clinicians' ability to diagnose and prognosticate indirectly translates to improved patient outcomes.

Questions of diagnostic and prognostic accuracy follow a similar format. 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 rule out other causes of peripheral facial palsy.)

*Population Screening* There is another common type of clinical question worth considering. These questions have a diagnostic flavor but are more concerned with diagnostic yield than with diagnostic accuracy. This type of question is appropriate to the situation where a diagnostic intervention of established accuracy is employed. An example is, In patients with new-onset peripheral facial palsy should a physician routinely obtain a head MRI to identify sinister pathology within the temporal bone causing the facial palsy? There is no concern with the

diagnostic accuracy of head MRI in this situation. The diagnostic accuracy of MRI in revealing 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: the frequency with which the MRI reveals clinically relevant abnormalities in this patient population. The implication is that if the yield were high enough, clinicians would routinely order the test.

*Causation.* Occasionally, a guideline asks a question regarding the cause and effect relationship of an exposure and a condition. Unlike diagnostic and prognostic accuracy questions that look merely for an association between a risk factor and an outcome, causation questions seek to determine whether an exposure causes a condition. An example is, does chronic repetitive motion cause carpal tunnel syndrome? Another example is, does natalizumab cause progressive multifocal leukoencephalopathy? The implication is that avoidance of the exposure would reduce the risk of the condition. As in these examples, causation most often relates to questions of safety.

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

## The Evidence

### Finding the Evidence

A comprehensive literature search to find relevant evidence is the next step in the EBM process. The comprehensive search is performed to ensure, as much as possible, that all relevant evidence is considered. This helps to reduce the risk of bias being introduced into the process.

It is no exaggeration to state that the evidence based medicine movement started because of the availability of access to electronic databases of the medical literature. Because of the ease of access to databases like MEDLINE it became practical for a clinician to find high quality evidence. Being able to effectively search MEDLINE is an important skill to allow the effective practice of EBM

Many clinicians, equate electronically searching the medical literature with "Google." There is no doubt that the ability to "google" information is often useful. It is important to realize that the results of such searches are at best haphazard. It can be useful for finding quick background information on an unfamiliar topic just before rounds but it would be inappropriate to rely on a Google search to answer a focused clinical question. For well formulated evidence-based question one needs to use a data base such as MEDLINE.

MEDLINE is the most commonly searched database. There are other databases besides MEDLINE. EMBASE is one. It is particularly good for pharmaceutical studies.

MEDLINE is huge containing many separate citations (a citation is an entry corresponding to a published article). MEDLINE does not contain all published medical studies. For example, it only contains about 80% of randomized controlled trials relevant to a specific topic. Initially, MEDLINE only contains citations of articles published since 1966. However, articles have been added retrospectively so that MEDLINE now includes articles published at earlier dates. Abstracts are only available for articles published since 1974.

For the most part, MELINE is limited to peer-reviewed articles published in scholarly journals. It generally does not index books, conference paper or abstracts (though it does index some abstracts).

The National Library of Medicine maintains the MEDLINE database. There are several different interfaces to access the MEDLINE database. The most familiar are PUBMED and OVID. PUBMED and OVID do not maintain there own databases. Rather they provide interfaces to accessing MEDLINE and other databases. The NLM maintains its own interface to MEDLINE. It is important to understand the difference between an interface to access a database (the client) and the database itself (the server).

MEDLINE is the same database regardless of which client (interface) is used to access it. The interface allows you to search MEDLINE with the controlled vocabulary of the indexed medical subjects or via natural language (similar to Google).

The controlled vocabulary that indexes MEDLINE is known as MeSH (Medical Subject Heading). Having some familiarity with how MeSH works will increase the effectiveness of your searches. MeSH is a controlled list of medical subjects. It has a hierarchical structure with appropriate subject headings organized underneath appropriate broader headings.


This figure illustrates the hierarchical structure of MeSH headings contained with the Cerebrovascular Disorders heading. Cerebrovascular Disorders is MeSH subject under Brain Diseases. Those headings with "+" have additional, more specific, subjects within the structure. Each sublevel contains more and more specific subjects. You can easily explore the structure of MeSH if you google MeSH.

The advantage of MeSH over natural language is that it is standardized. Some interfaces, such as OVID, automatically map to the appropriate MeSH heading. PubMed will attempt to map search terms to MeSH terms and also uses "text word" searches.

MeSH is a dynamic thesaurus. Terms are continually added by indexers. MeSH indexers will use the most specific term available. An article on brain ischemia will be indexed on Brain ischemia and not cerebrovascular disorders. Other indexed terms beyond the articles topic are also included.

The hierarchical structure of the MeSH can be referred to as a tree structure. To ensure all sub terms contained within the tree structure are searched you need to "explode" the term. Thus, if you are interested in all "Epilepsy" as a search term and do not explode it, your search will exclude all of the articles that describe specific epilepsies. To be sure to include these, you need to "explode" the term epilepsy. By default, PubMed "explodes" search terms.

You can view a citations corresponding MeSH headings within PUBMED or OVID. Doing so can give you clues as to how the indexers have tagged the article and help you find other pertinent articles.

Subheadings should not be confused with the hierarchical structure of MeSH. They refer to specific subtopics such as diagnosis or therapy pertinent to the MeSH.

One can use free text or natural language searching for retrieving articles regarding subjects that do not have a MeSH heading. Such searches look for the appearance of the word or phrase within the title or abstract (similar to the way Google works). Text word and MeSH searches can be combined to ensure all relevant articles are retrieved (just in case the indexers missed it.)

Search terms are combining terms using Boolean operators to expand or narrow a search. The "and" operator is exclusionary. Both terms must be indexed in order for the article to be retrieved. "Or" is an inclusionary search strategy. An article with either term present will be included. For example, a search strategy that combines dystonia or myoclonus or tremors will include all articles with any of these hyperkinetic movement disorders. Usually "or" will be used for similar terms to include related articles. You usually should not use "and" and "or" on the same line in a search: it is too confusing.

The PICO format of your question is very useful to develop an effective search strategy. The search terms can be derived directly from the question. Usually the patient and intervention portions of the PICO question define the search.

The relevance of a search can be increased by using limits. Realize that there is a compromise between comprehensiveness and relevance, i.e. the sensitivity vs. the specificity of the search. If you are looking for a quick relevant answer, use limits. If you want to review all of the pertinent literature, do not use limits.

One powerful way of limiting your search is to use clinical filters (called clinical queries in PubMed). There is a lot of stuff on MEDLINE that is not relevant to clinical practice. Search filters help you to quickly discard that stuff. They are very effective at developing a search that will find clinically relevant articles. You can develop your own search filters and save them if you would like. Usually, you will use filters developed and validated by others.

It is best to use validated search filters. These have been empirically validated to provide high sensitivity (few relevant articles missed) high specificity (few irrelevant articles included) searches. Realize that the filters are most sophisticated for searches pertinent to therapeutic questions. The sensitivity of filters for diagnostic questions, for example, is not nearly as well studied. Combining a PICO focused search with a clinical filter quickly retrieves relevant articles.

During the EBM course you will develop a PubMed search strategy to identify studies that are relevant to the clinical question raised at the beginning of the course.

Rating the strength of a study

Now that you have found the relevant evidence, what do you do with it?

An important step in the EBM process 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 measure the bias of a study directly. (If it were, it would imply we already knew the answer to the clinical question.) However, using well-established principles of good study design, we can estimate a study's risk of bias.

The AAN has developed a tool to rate the risk of bias in studies using a four-tiered classification scheme. 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 moderately high risk of bias, and studies graded Class IV are judged to have a very high risk of bias. The classification rating is also known as the level of evidence.

The advantage of using a scheme like the AAN's is that the rules for determining the strength of evidence are pre-specified. This helps to reduce the common, conclusion-motivated process of finding something/anything "wrong" with a study whose findings are not consistent with our pre-determined beliefs. Or, believing the results of a study that are consistent with our beliefs no matter how flawed it is. The pre-specified level of evidence schemes let you know how much confidence you should have in the study's results—regardless of what those results are.

It is important to remember that the classification scheme the AAN employs accounts only for systematic error. Random error (low study power) is dealt with separately. A study's risk of bias can be judged only relative to a specific clinical question. Therapeutic, diagnostic or prognostic accuracy, screening, and causation questions are judged by different standards.

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 I or Class II, studies should use concurrent controls. Studies using non-concurrent controls are at best rated Class III.

*Treatment Allocation.* To reduce the risk of bias, authors of a therapeutic article must ensure that treated and untreated patient groups are similar in every way except for 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.

An important study characteristic that ensures patients are truly randomly allocated to different strategies is concealed allocation. Concealed allocation prevents investigators from manipulating treatment assignment. Examples of concealed allocation include use of consecutively numbered, sealed, opaque envelopes containing a pre-determined, random sequence for treatment assignment or use of an independent center that an investigator contacts to obtain the treatment assignment. By comparison, examples of unconcealed allocation include flipping

a coin (e.g., heads = treatment A, tails = treatment B) or assigning patients to treatment categories on the basis of the date (e.g., treatment A on odd-numbered days, treatment B on even-numbered days). These unconcealed allocation methods can be easily manipulated to control treatment allocation. For example the coin can be flipped again, or the patient can be told to come back the next day.

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

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

*Cross-overs.* 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 cross over from the treated group to the comparison group or from the comparison group to the treated group, confounding differences can be introduced. When this happens, it is important that the investigators analyze the results using intent-to-treat principles. Put simply, such principles entail analysis of the results according to whichever group (treatment or comparison) each patient was originally assigned.

*Masking.* For a study to be graded Class I or II, an investigator who is unaware of the patient's original treatment assignment must determine the outcome. This is termed masked or blinded outcome assessment.

For a study to be graded Class III, a study investigator who is not one of the treating providers must determine the outcome. Such independent outcome assessment, although not as effective in reducing bias as masking, nonetheless has been shown to be less bias prone than having the unmasked treating physician determine the outcome. A patient's own assessment of his or her outcome (e.g., a seizure diary or completion of a quality-of- life questionnaire) fulfills the criteria for independent assessment.

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 observer expectation bias (e.g., patient survival or a laboratory assay). Oftentimes determining whether an outcome is objective requires some judgment.

This figure, though somewhat simplified, describes the essence of the AAN's classification of therapeutic evidence scheme.

During the EBM course you will have the opportunity to rate several studies pertinent to the clinical question raided at the beginning of the course.

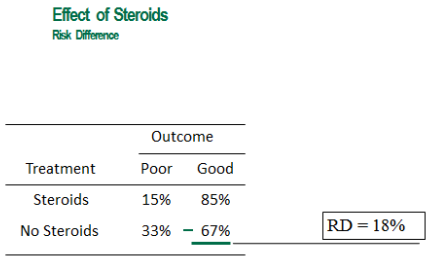
**AAN Classification of therapeutic evidence**

| Class | Controlled | Masked |
|---|---|---|
| I | randomized | Single or objective |
| II | matched | Single or objective |
| III | comparative | independent |
| IV | - | - |

## Conclusions: Interpreting effects and evidence synthesis

*Measures of effect.* The effectiveness of an intervention is usually best measured by using discrete, categorical variables rather than continuous variables. Categorical variables aid in the ur interpretation of the clinical importance of an effect. For example, the proportion of patients with Bell's palsy who have complete facial functional recovery is a more easily interpreted measure of patient outcome than the overall change in the median values of the House-Brackman facial function score.

Measuring patient outcomes using categorical variables involves counting patients. An example is, how many patients on drug X improved, and how many did not improve? Counting patients in this manner often enables construction of a contingency table. The accompanying figure is a simple, fictitious two-by-two contingency table showing the numbers of patients with Bell's Palsy improving on steroids versus placebo.




From this it is a relatively straightforward process to calculate numeric values that express the strength of association between the intervention and the outcome. Examples are the relative risk of a poor outcome in treated patients versus untreated patients (the proportion of treated patients with a poor outcome divided by the proportion of untreated patients with a poor outcome) or the poor-outcome risk difference (the proportion of treated patients with a poor outcome minus the proportion of untreated patients with a poor outcome). The risk difference (RD) is shown in the figure.

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

*Measures of Statistical Precision.* Regardless of the clinical question type or the outcome variable chosen, it is critical that some measure of random error (i.e., the statistical power of each study) be included in the estimate of the outcome. Random error results from chance. Some patients improve and some do not regardless of the intervention used. In any given study, more patients may have improved with treatment than with placebo just because of chance. Statistical measures of precision (or power) gauge the potential contribution of chance to a study's results. In general, the larger the number of patients included in a study, the smaller the contribution of chance to the results.

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

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

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

Relative to measures of statistical precision, 95% confidence intervals are preferred over *p* values.

*Interpreting a study*. Armed with the measure of association and its 95% confidence interval, we are in a position to interpret a study's results. Often the temptation here is to determine merely whether the study was positive (i.e., showed a statistically significant association between the intervention and outcome) or negative (did not show a statistically significantly association). In interpreting study results, however, four, not two, outcomes are possible. This derives from the fact that there are two kinds of differences you are looking for: whether the difference is statistically significant and whether the difference is clinically important. Henceforth we will use the term *significant* to mean *statistically significant* and the term *important* to mean *clinically important*. From these two types of differences, four possible outcomes can be seen:

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

For example an RCT of patients with symptomatic intracranial atherosclerosis demonstrates that 10% of patients who had a stent had strokes whereas 20% of patients who did not have a stent had strokes (risk difference 10%, 95% confidence intervals 5%–15%). This difference is statistically significant (the confidence intervals of the risk difference do not include 0) and clinically important (no one would argue that a finding of 10% fewer strokes is unimportant).

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

A separate RCT enrolling a large number of patients with symptomatic intracranial atherosclerosis demonstrates that 10.0% of patients who had a stent had strokes whereas 10.1% of patients who did not have a stent had strokes (risk difference 0.1%, 95% confidence intervals 0.05%–0.015%). This difference is statistically significant but arguably not clinically important (there are only 1 in 1000 fewer strokes in the patients with stenting).

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

A third RCT enrolling a large number of patients with symptomatic intracranial atherosclerosis demonstrates that 5% of patients who had a stent had strokes whereas 5% of patients who did not have a stent had strokes (risk difference 0%, 95% confidence intervals -0.015% –0.015%). This difference is not statistically significant. Additionally the 95% confidence intervals are sufficiently narrow to allow us to confidently exclude a clinically important effect of stenting.

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

Our last hypothetical RCT of patients with symptomatic intracranial atherosclerosis demonstrates that 5% of patients who had a stent had strokes whereas 5% of patients who did not have a stent had strokes (risk difference 0%, 95% confidence intervals -10% –10%). This difference is not statistically significant. However, the 95% confidence intervals are too wide to allow us to confidently exclude a clinically important effect of stenting. Because of the lack of statistical precision, the study is potentially consistent with an absolute increase or decrease in the risk of stroke of 10%. Most would agree that a 10% stroke reduction is clinically meaningful and important.

Let us consider these outcomes one at a time.

Scenario 1 represents the clearly positive study and scenario 3 the clearly negative study.

Scenario 2 usually results from a large study. The study has a very high degree of precision and can show even minor differences. The minor differences may not be important. The study should be interpreted as showing no meaningful difference.

Scenario 4 results from a small study. The study is so underpowered that it is unable to show significant differences even when there might be important differences. It would be inappropriate to interpret this study as negative.

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

*Synthesizing evidence.* At this step often multiple papers pertinent to a question have been analyzed. These collective data must be synthesized into a conclusion.

The AAN has adopted a modified version of the GRADE process for evidence synthesis. You will have the opportunity to use this process during the course. The process highlights the concepts that should be considered when evaluating a body of evidence pertinent to a particular clinical question.

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

First, anchor the confidence level to the risk of bias. A level of confidence in the evidence relative to each outcome is then assigned. Four levels of confidence are available. Although it may differ from the final designation of confidence, the confidence in the evidence is initially anchored to the class of evidence:

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

Consider factors that potentially downgrade the confidence in the evidence. Next, explicitly consider factors that could downgrade the confidence in the evidence to support recommendations. These factors include the number and consistency of the informative studies, the statistical precision (or power) of the studies, the directness (or generalizability) of the evidence, the biological plausibility of the effectiveness of the intervention and the presence of potential reporting bias. In general, the level of confidence in the evidence should be downgraded only by one level for these factors.

Consider factors that potentially upgrade your confidence in the evidence. Such factors include a large magnitude of effect (an effect can be so large that it likely overcomes any bias in a study), the presence of a dose response relationship (this lends plausibility to the biological effect of the intervention), the direction of bias (if the direction of bias moves in one direction but the measured effect moves in the other, we can be more confident that the observed effect is real. This is especially true of studies that show no effect, as the direction of bias for most studies is toward showing an effect). Importantly, these factors can result only in upgrading of the confidence in the evidence by one level.

<u>Make a decision (considering more than evidence)</u>

It is evident that numerous factors can influence the final decision. Explicitly keeping track of these varying factors and their relative importance can be difficult. To assist in this process the AAN uses a graphical tool called a "Clinical Contextual Profile." The rows convey each of the factors to consider in developing practice recommendations. A similar process can be used to inform a difficult clinical decision relevant to individual patients.

| Domain | Ratings | | | | Consensus |
|---|---|---|---|---|---|
| Confidence in Inference | Very low | Low | Moderate 10 | High | Yes |
| Benefit relative to Harm | Harm ≥ Benefit 2 | Benefit > Harm 7 | Benefit >> Harm 4 | Benefit >>> Harm 0 | Yes |
| Importance of outcomes | Not Important or Unknown 2 | Mildly Important 7 | Very Important 4 | Critically Important 0 | Yes |
| Variation in preferences | Large 2 | Moderate 7 | Modest 4 | Minimal 0 | Yes |
| Feasibility | Rarely 2 | Occasionally 11 | Usually 0 | Always 0 | Yes |
| Cost relative to net benefit | Very Large 2 | Large 11 | Moderate 0 | Small 0 | Yes |
| Strength of recommendation | R/U | C | B | A | |

If time permits, we will use this tool to develop practice recommendations pertinent to the clinical dilemma introducedten at the beginning of the course.

<u>Selected References</u>


1. Clinical Practice Guideline Process Manual**,** 2011 Ed. St. Paul, MN: The American Academy of Neurology (the syllabus contains excerpts from this manual)
2. American Academy of Neurology Evidence-based Medicine (EBM) Training Course (EBM toolkit) (the syllabus contains excerpts from the EBM toolkit). http://tools.aan.com/education/ebm/
3. PubMed tutorial. http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/cover.html
4. GRADE working group. http://www.gradeworkinggroup.org/
5. Cook DJ, Greengold NL, Ellrodt AG, Weingarten SR. The relation between systematic reviews and practice guidelines. Ann Intern Med 1997;127:210-216.
6. Franklin GM, Zahn CA. AAN clinical practice guidelines: above the fray. Neurol 2002;59:975-6.

7. 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. *Journal of General Internal Medicine. 11(11):664-72, 1996 Nov.*
8. Richardson WS. The well-built clinical question: a key to evidence-based decisions. ACP Journal Club 1995;Nov/Dec:A12-A13.