What Are Evidence Based Guidelines?

David G. Nichols, M.D.
Johns Hopkins School of Medicine
Baltimore, Maryland, USA

You - the director

As the director of a pediatric anesthesia group, you wish to insure that your group is following the best practice based on evidence. You know clinical evidence can be assessed at various levels ranging from randomized controlled trials (RCT) up to the practice guidelines (Table 1). You choose to focus on practice guidelines, which you hope will reflect clinical trials, meta-analyses, and systematic reviews.

Table 1. Definitions of levels of evidence in evidence-based medicine

Practice guideline	-**"***a user-friendly statement that brings together the best external evidence and other knowledge necessary for decision-making about a specific health problem***".**[1]
Systematic review	-a summary of all evidence on a specific question after completing a thorough search and critical appraisal of the literature using a predefined, explicit methodology. The systematic review is often based on quantitative meta-analysis but may include case-controlled studies, cohort studies, case reports, or expert opinion.
Meta-analysis	- a statistical process of aggregating the data from a set of independently conducted studies that address a specific question. Ideally, all studies that meet a strict set of criteria (unbiased with comparable methods, populations, etc.) are included thereby generating much larger sample sizes than can be achieved by any single study.
Randomized controlled trial	-a type of quantitative and controlled study in which subjects are allocated at random to receive one of several clinical interventions.

The search for practice guidelines

You begin the search for appropriate practice guidelines on "pediatric" AND "anesthesia" and discover that there is no ideal site or database. Table 2 shows some major databases for practice guidelines. All have trade-offs between sensitivity and specificity. The Tripdatabase search yields 248 results, however, most have little to do with pediatric anesthesia (poor specificity). The National Guideline Clearinghouse is more specific, but also yields some results with little relevance to pediatric anesthesia. The Cochrane Library is a superb resource for systematic reviews but does not contain guidelines. PubMed is useful results but is sensitive to search methodology.

Table 2. Search results for the terms "pediatric" AND "anesthesia" in various databases

	Systematic Reviews	Guidelines
Tripdatabase… (www.tripdatabase.com)	356	248
National Guideline Clearinghouse (www.guideline.gov)	0	75
Cochrane Library (www.cochrane.org/reviews)	5	0
PubMed (Limit: Practice Guideline)	0	0
PubMed (Link: "Clinical Query")	8	1
PubMed (Search: "anesthesia"; Limit: Practice Guideline, Infant)	0	6

After reviewing the search results from the various databases, you conclude that the *National Guideline Clearinghouse (NCG) contains the most useful results*. Another major benefit of NCG is the ability to add guidelines into a personal collection ("my collection") and then compare them for quality using standardized criteria. You decide to focus on 2 guidelines: (1) *Procedure Guideline for Pediatric Sedation in Nuclear Medicine,* revised by Society of Nuclear Medicine

(SNM) in 2003[2]; and (2) *Preoperative Evaluation* revised by the Institute for Clinical Systems Improvement (ICSI) in 2006[3].

Interpretation of the Guidelines
You are tempted to simply take the major recommendations from each guideline and e-mail them to the members of your group. However, closer inspection of the guidelines suggests that you may actually disagree with them. Therefore, you are forced to dig a little deeper to determine the origin and quality of the recommendations.

Elements of a practice guideline
Practice guideline contains five major elements: purpose, identity, methods of development, knowledge components, and post development plan.

Purpose of the guideline
The purpose of the guideline is to answer a *clinical question*, which the guideline should state clearly. You must decide whether the stated objective and clinical question are relevant to your practice. The more specific the purpose and question are, the easier it is for the clinician to determine the guideline's applicability to clinical practice.

The statement of purpose should discuss the *background and rationale* behind the clinical question and why a guideline in this area is important. Since mortality from pediatric anesthesia is extremely low, other outcome measures such as morbidity (SNM guideline), adverse effects (SNM guideline), patient/parent satisfaction (ICSI guideline), quality-of-life, or economic cost (ICSI guideline) may justify the importance of the guideline. The guideline should list and weight explicitly all potential outcomes of the question under consideration.

A clear statement on the *target population* for the guideline as well as any excluded populations or clinical scenarios is extremely important. You discover that the SNM guideline concentrates exclusively on children, whereas the ICSI guideline focuses on children and adults with the bulk of the evidence devoted to adults.

Identity of the guideline
The identity segment of the guideline includes basic descriptors such as guideline title, release date, citation, and structured abstract. However, the most important identity descriptors are the names, titles, affiliations, and conflicts of interests of the panel members. A diverse panel membership (research methodologists, various clinical specialists, and public representatives) may reduce bias and increase accuracy and utility of the guideline. The ICSI guideline relied on a panel with multiple clinical specialties and research methodologists. The SNM panel did not list specialty affiliations. Neither panel included public representatives.

The greatest threat to the objectivity of a guideline comes from financial *conflicts of interest* - either by individual panel members or the sponsor (e.g., specialty society). A recent survey of panel members involving prescription guideline development found 35% had a conflict of interest.[4] Some even owned stock in companies whose products were the subject of the guideline. At a minimum, the guideline must disclose all potential conflicts of interest by individual members and the sponsoring entity. A variety of conflict management strategies may allow experts with financial relationships to still contribute to the guideline development process. The SNM guideline did not contain a statement on conflicts of interest. The ICSI guideline contains a full statement on conflicts of interest. ICSI is funded in part by Blue Cross and Blue Shield and other health plans. This discovery makes you wonder whether the guideline will adopt the perspective of the insurance industry.

The guideline development process
You now seek to determine whether the underlying evidence and the recommendations themselves are valid. This requires an understanding of how the panel used evidence to reach its conclusions - the guideline development process. It involves 4 steps: (1) identification of a specific practice question; (2) determination of the potential outcomes; (3) a search, summary, and grading of the evidence relevant to the clinician and patient; consensus on a guideline statement that best answers the practice question.

Evidence selection methodology
The *search* for the best evidence usually begins with electronic databases (e.g., MEDLINE, EMBASE, CINHAHL, and the Cochrane library). Either the "clinical query" or "special query" links in PubMed direct the user to a search engine for systematic reviews. Because database searches may miss some relevant studies, guideline panels will often add a "hand-search", where panel members cull further studies from the reference lists of reviewed articles or from interviews of experts in the field.

The *critical appraisal* of the selected literature follows the search. The Scottish Intercollegiate Guidelines Network (SIGN, www.sign.ac.uk) maintains a series of helpful checklists to determine the internal validity and quality of the evidence in systematic reviews (Table 3), meta-analyses, RCT's, and observational studies.

Table 3. Methodology checklist for systematic reviews and meta-analyses. (Adapted from the Scottish Intercollegiate Guidelines Network. www.sign.ac.uk)

Section 1: internal validity	
The review addresses an appropriate and clearly focused question.	Well covered. Adequately addressed. Poorly addressed. Not addressed. Not reported. Not applicable.
A description of the methodology is included.	
The literature search is sufficiently rigorous to identify all the relevant studies.	
Study quality is assessed and taken into account.	
There are enough similarities among the studies selected to make combining them reasonable.	
Section 2: overall assessment of the review	
How well was the review done to minimize bias?	Code as ++, +, or -
If coded as + or - , what is the likely direction in which bias might affect the review results?	
Section 3: description of the review	
What types of studies are included in the review?	RCT/CCT/Cohort/Case-Control/Other
How does this review help to answer your key question?	Summarize and comment.

CCT – controlled clinical trial

Grading of the evidence is the next step after search and appraisal. The effort to apply a formal taxonomy of "levels of evidence" began with the Canadian Task Force on Periodic Health Examination (CTFPHE).[5] Many modifications of this original numerical system have appeared including the one employed by ICSI. The plethora of grading systems has led to an international effort by the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) Working Group to simplify and unify the grading systems. Table 4 shows the GRADE system and roughly equivalent designation for strength of evidence in the CTFPHE and SIGN systems.

Table 4. GRADE system for estimating the strength of the evidence (modified from www.gradeworkinggroup.org). The *approximate* corresponding grades from Canadian Task Force on Periodic Health Examination (CTFPHE) and Scottish Intercollegiate Guideline Network (SIGN) are provided for comparison.

GRADE level	CTFPHE level	SIGN level	GRADE Study design	GRADE Conclusion
High	**I**	**1++, 1+, 2++**	**-RCT(s)** (valid) **-Observational studies (>2)** with very strong evidence of the association (RR >5 or <0.2) and no major threats to validity	Further research is *very unlikely* to change confidence in the estimate of the effect
Mod.	**II**	**1-, 2++, 2+**	**-RCT(s)** with: serious limitation, important inconsistency, some uncertainty of the directness, imprecise or sparse data, or high probability of reporting bias. **-Observational studies** (≥2) with: -strong evidence of the association (relative risk >2 or <0.5) -evidence of a dose response gradient -all plausible cofounders would have reduced the effect	Further research is *likely* to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	**III**	**2-**	**-Observational study** **-RCT** with very serious limitation to study quality or major uncertainty about directness	Further research is *very likely* to have an important impact on our confidence in the estimate of the set and is likely to change the estimate
Very low		**3,4**	**Case reports, expert opinion**	Any estimate of the effect is very uncertain

RR – relative risk

Consensus on the actual guideline

Even with strong evidence, it is often difficult for a panel to reach a consensus on a guideline statement. The major challenges include the different *perspectives* (patient, provider, insurer, society, etc.), the different relative weights placed on *benefits versus harms*, and the actual *consensus-building process* used by the panel. For instance, even though you may agree with the ICSI guideline to omit all preoperative laboratory testing in otherwise healthy patients, you wonder whether that guideline represents the perspective of insurers (who fund ICSI) or of patients. Neither SNM nor ICSI state explicitly whose perspective the guideline represents.

Guidelines addressing interventions that carry both benefits and harms should state how the panel members weighed the benefits versus harms. For instance, a guideline recommending postoperative analgesia with morphine carries the benefit of analgesia and patient comfort at the risk of vomiting and excessive sedation. It is helpful to know what incidence of vomiting and excess sedation the panel considered acceptable in exchange for the benefit of postoperative analgesia.

The number needed to treat (NNT) and the number needed to harm (NNH) provides an objective way to determine if the benefits of the treatment are worth the harms and costs mentioned in the guideline. NNT is the reciprocal of the absolute risk reduction provided by a therapy. If morphine administration reduces the incidence of postoperative pain from a control

event rate of 70% to an experimental event rate of 20%, the absolute risk reduction is 50% (70-20). You would only need to treat 2 (1/50%) additional patients before seeing the benefit of morphine. If the control event rate for postoperative vomiting is 10% which increases to 20% with morphine administration, the absolute risk increase is 10% (20-10). Hence 10 (NNH =1/10%) additional patients would have to get morphine before a patient developed vomiting. These trade-offs should be transparent and consider multiple different perspectives (provider and patient/parent).

A formal consensus-building process helps avoid bias and domination of the panel by a few vocal panel members. The *Delphi* method involves polling the panel members with a series of questionnaires. A facilitator tallies the results and provides feedback before the next round of polling. Through this iterative process a "correct" answer (consensus) is reached.

Grading the guideline recommendation

The final grade of the guideline depends foremost on the strength of the evidence, but also on the magnitude and consistency of benefits versus harms and on the relative value placed on different outcomes. As with grading the evidence, there is no uniform agreed-upon system to grade guidelines. Table 5 compares the GRADE and SIGN systems to grade guidelines.

Table 5. Approximate comparison of the GRADE and SIGN ratings to grade guidelines.

GRADE rating	SIGN rating	SIGN commentary*
Strong	A	At least one meta-analysis, systematic review, or RCT rated as 1++ and directly applicable to the target population; *or* a body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating an overall consistency of results
	B	A body of evidence including studies rated as 2++, directly applicable to the target population, and demonstrating overall consistency of results; or extrapolated evidence from studies rated as 1++ or 1+
Weak	C	A body of evidence including studies rated as 2+, directly applicable to the target population and demonstrating overall consistency of results; or extrapolated evidence of studies rated as 2++
	D	Evidence level 3 or 4

*See table 4 for explanations of SIGN evidence levels (1++…4).

The real question is whether the clinical importance of the recommendation based on the magnitude and consistency of benefits over harms is sufficient to persuade you to change your practice. After this careful consideration you decide that neither the SNM nor the ICSI guidelines are sufficiently compelling to change your practice.

References

1. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence Based Medicine. Churchill Livingstone. New York 1997.

2. Mandell GA, Majd M, Shalaby-Rana EI, Gordon I. Procedure guideline for pediatric sedation in nuclear medicine, 3.0. Reston (VA): Society of Nuclear Medicine; 2003 Jan 25.

3. Institute for Clinical Systems Improvement (ICSI). Preoperative evaluation. Bloomington (MN): Institute for Clinical Systems Improvement (ICSI); 2006 Jul.

4. Taylor R, Giles J. Cash interests taint drug advice. Nature 2005; 437: 1070-1

5. Canadian Task Force on the Periodic Health Examination: The periodic health examination. Can Med Assoc J 1979, 121:1193-254