

Evaluating the Literature for the Best Evidence

Evidence-based medicine involves the “science” of medicine, in which physicians use the best available medical evidence and literature to make decisions about patient care, whereas the “art” of medicine requires physicians to make medical decisions without direct evidence. Each year, billions of dollars are spent in translating evidence-based research into practice. When multiple studies confirm the utility of a new test or treatment, medical organizations incorporate that evidence into clinical practice guidelines, and experts include it in review articles and book chapters. Increasingly, when there is an overwhelming amount of evidence to support a particular decision or intervention, physicians are being held accountable for making the right, or evidence-based, patient care decisions. Clinicians must be able to access the literature to answer specific clinical questions, which may or may not be answered by approved practice guidelines.

Accessing the Medical Literature

Clinicians can begin their literature search with a resource such as PubMed (www.ncbi.nlm.nih.gov/entrez), a free online service from the National Library of Medicine, which includes most important, published medical research findings found in MEDLINE (<http://medline.cos.com>) and published practice guidelines. PubMed allows users to search for articles using filters to limit irrelevant articles. The filters, Clinical Queries, and Systematic Reviews can be accessed through the Clinical Queries hyperlink on the left side-bar menu. Clinical Queries assist in the retrieval of articles on therapy, diagnosis, etiology, or prognosis, whereas the Systematic Reviews retrieve a broader selection of articles consisting of systematic reviews and meta-analyses, guidelines, and other types of reviews (**Figure 4**).

Other evidence-based, user-friendly sources exist for physicians who want clinically useful information or to keep their knowledge base current. UpToDate® is an online medical textbook that is updated frequently and integrates findings from research studies. The Physicians' Information and Education Resource (PIER), an online resource from the American College of Physicians, is a current information source that translates research findings into clear, concise recommendations that are graded according to standard, accepted criteria (<http://pier.acponline.org/index.html>). The Cochrane Collaboration (www.cochrane.org) is an international, nonprofit organization that performs frequent, evidence-based reviews of randomized trials, also known as meta-analyses. Meta-analyses use statistical methods to summarize individual studies and provide a single, combined measure on the effect of treatments. These reviews are also indexed in PubMed. Most clinical practice guidelines can also be found at the National Guideline Clearinghouse by using a standard Internet search engine or by going to <http://guidelines.gov>.

Key Points

- PubMed is a free online service from the National Library of Medicine that includes most of the important medical research publications found in MEDLINE in addition to published practice guidelines.
- Other evidence-based sources of useful clinical information include UpToDate, The Physicians' Information and Education Resource, The Cochrane Collaboration, and the National Guideline Clearinghouse.

Evaluating the Medical Literature

It is useful to understand the relative value of conclusions from different types of publications. The design of a study often limits the strength of the conclusions that can be drawn from it. The most useful publications systematically review the results of multiple randomized trials, such as those by the Cochrane Collaboration (<http://www.cochrane.org>). Randomized controlled trials are also useful because they control for extraneous variables and, through blinding, control for possible placebo effects. Applying evidence from a single randomized trial can be problematic because many of these early findings may not have been reproducible in subsequent studies. Cohort studies such as the Framingham study, which follow many individuals for several years and typically control for extraneous variables, are also potentially useful. An example of a limitation of this type of study, however, is the consistent finding in cohort studies that hormone replacement therapy was cardioprotective, a finding that was disproved after several large randomized controlled trials were conducted. Other study designs, such as case-control or case series, may be useful for generating hypotheses but have weaknesses inherent in their designs that hinder their ability to provide firm conclusions. However, randomized trials are often not possible to conduct, and consistent findings from other study designs may be deemed to be acceptable. Such is the case with the recommendation for Pap testing in cervical cancer screening. This screening approach, although widely recommended, has not been the subject of a randomized-controlled trial study design.

It is important to know whether the clinical information that is found during a literature search is evidence-based. Systematic review articles or meta-analyses that summarize the literature and describe the explicit article-selection criteria used are less likely to be biased in the selection of articles that support their conclusion than articles or meta-analyses that do not. These article-selection criteria should include a search of large bibliographic databases, such as MEDLINE and EMBASE (<http://www.embase.com>), and other measures to ensure that important studies have not been overlooked or improperly assessed.

Any source of information that provides specific clinical recommendations should also describe how the evidence was used to reach the stage of formal recommendation or guideline. Often, this process involves a system of letters and numbers representing each recommendation that reflects the amount and strength of the supporting evidence. For example, the United States Preventive Services Task Force (USPSTF see <http://www.ahrq.gov/>) uses a system that reflects the strength for or against a given recommendation (A, B, C, D, I) (**Table 6**) (1).

Key Points

- Clinical recommendations should be accompanied by a description of how the evidence was used to reach the stage of formal recommendation or guideline.
- Systematic review articles or meta-analyses that describe the explicit article-selection criteria used are less likely to be biased than articles or meta-analyses that do not.

Understanding the Newer Measures of Clinical Significance

Although practice guidelines provide actual recommendations, the medical literature provides the data that inform the clinical decision-making process. Data are presented with measures of clinical significance. Beyond sensitivity, specificity, predictive value, and odds ratios (**Table 7**), several newer epidemiologic measures are now used in research literature, meta-analyses, and review articles to describe the significance of diagnostic tests and the efficacy of treatments.

Likelihood Ratios

The utility of a diagnostic test has typically been represented by *sensitivity* and *specificity*. Sensitivity refers to the percentage of people with disease who are correctly diagnosed, whereas specificity refers to the percentage of people without disease who are correctly diagnosed. Based on sensitivity and specificity, the *likelihood ratio* (LR) has the advantage of summarizing in a single number the clinical utility of a test or physical examination finding. The LR is a function of a test, such as stress echocardiography, for confirming or excluding a diagnosis, such as coronary artery disease. The LR of a positive test or, a test that “rules in” a diagnosis, is calculated by using the formula

$\text{sensitivity}/(1 - \text{specificity})$

whereas the LR of a negative test, or a test that “rules out” a diagnosis, is calculated by using the formula

The pretest probability refers to the percentage of patients who have the target disorder as determined before the test is performed. The farther away the LR is from 1.0, the more clinically useful a test is in increasing or decreasing the pretest probability for confirming or excluding the target disorder. An LR between 1 and 3 for a positive test indicates a less-useful test than one with an LR between 5 and 7. For example, a positive result for a test with an LR of 8 adds approximately 40% to the pretest probability that a patient has a specific diagnosis (2). Positive LRs of 2, 5, and 10 increase the probability of disease by 15%, 30%, and 45%, respectively. Negative LRs of 0.5, 0.2, and 0.1 decrease the probability of disease by 15%, 30%, and 45%, respectively. When the pretest probability is very low or very high, only an extremely high LR would be needed to influence management decisions. Diagnostic tests are most useful when the pretest probability is within the intermediate range.

Number Needed to Treat

The *number needed to treat* (NNT), which represents the effect of treatments on patients, indicates how many patients must receive a treatment to produce one

additional improved outcome compared with the control treatment. The NNT is calculated from the absolute risk reduction (ARR), which is the difference in outcome observed between the placebo and active treatment. This is different from the relative risk reduction (RRR), which is a ratio of the difference in outcome between placebo and active treatment. The NNT is calculated by taking the inverse of the ARR, or the difference between event rates of patients who received treatment and those who did not. For example, if myocardial infarction occurred in 10% of patients in a clinical trial who received an experimental medication designed to prevent myocardial infarction, compared with 20% of patients who received placebo, the RRR would be 0.5 (0.1/0.2), the ARR would be 0.10 (0.2-0.1), and the NNT would be 10, the inverse of 0.20 minus 0.10. This means that 10 patients would need to be treated to prevent one myocardial infarction with the use of the new medication. The lower the NNT, the more effective the treatment.

Confidence Intervals

Most studies now provide estimates of event rates in terms of *confidence intervals* (CI). This number, which is a real calculation of events that occurred in a study, indicates that if a study were repeated 100 times, a result within the specified range of values would be expected 95% of the time (when using a 95% CI). In studies, this is typically represented by a number, followed by a parenthetical range of two other numbers, for example, "4.5 (95% CI, 3.8 to 5.4)." Therefore, if a study's result were 4.5, and the study were repeated 100 times, 95% of the time, the result would be between 3.8 and 5.4. Larger studies typically have narrower CIs.

Key Points

- Likelihood ratios (LRs) can be used to approximate the probability of disease after a test is performed.
- Positive LRs of 2, 5, and 10 increase the probability of disease by 15%, 30%, and 45%, respectively.
- The number needed to treat indicates how many patients require treatment to produce one additional improved outcome compared with the control treatment.
- Confidence intervals (CI) indicate that if a study were repeated 100 times, a result within the specified range of values would be expected 95% of the time.

References

1. U.S. Preventive Services Task Force Ratings: Strength of Recommendations and Quality of Evidence. Guide to Clinical Preventive Services, Third Edition: Periodic Updates, 2000-2003. Agency for Healthcare Research and Quality, Rockville, MD. Accessed at <http://www.ahrq.gov/clinic/3rduspstf/ratings.htm>.
2. McGee S. Simplifying likelihood ratios. J Gen Intern Med. 2002;17:646-9. [PMID: 12213147]

American College of Physicians. Medical Knowledge Self-Assessment Program 14. Philadelphia: American College of Physicians; 2006.

Table 6. U.S. Preventive Services Task Force Grades and Recommendations and Strength of Overall Evidence

Grades and Recommendations*		Strength of Overall Evidence*	
Grade	Recommendation	Grade	Definition
A	The USPSTF strongly recommends that clinicians provide [the service] to eligible patients. The USPSTF found good evidence that [the service] improves important health outcomes and concludes that benefits substantially outweigh harms.	Good	Evidence includes consistent results from well-designed, well-conducted studies in representative populations that directly assess effects on health outcomes.
B	The USPSTF recommends that clinicians provide [the service] to eligible patients. The USPSTF found at least fair evidence that [the service] improves important health outcomes and concludes that benefits outweigh harms.	Fair	Evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes.
C	The USPSTF makes no recommendation for or against routine provision of [the service]. The USPSTF found at least fair evidence that [the service] can improve health outcomes but concludes that the balance of benefits and harms is too close to justify a general recommendation.	Poor	Evidence is insufficient to assess the effects on health outcomes because of limited number or power of studies, important flaws in their design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes.
D	The USPSTF recommends against routinely providing [the service] to asymptomatic patients. The USPSTF found at least fair evidence that [the service] is ineffective or that harms outweigh benefits.		

From U.S. Preventive Services Task Force Ratings: Strength of Recommendations and Quality of Evidence. *Guide to Clinical Preventive Services, Third Edition: Periodic Updates, 2000-2003*. Agency for Healthcare Research and Quality, Rockville, MD. <http://www.ahrq.gov/clinic/3rduspstf/ratings.htm>

Table 7. Glossary of Statistical Terms

Statistical Term	Definition	Commonly Used Abbreviation/Presentation
<i>P</i> value	The measured probability of achieving results observed in a study, assuming the null hypothesis is exactly true. (The null hypothesis asserts that no true association or difference in the study outcome between comparison groups exists in the larger population from which the study population was obtained; in other words, the results occurred by chance). A <i>P</i> value < 0.05 is often considered significant; it means that less than a 5% probability exists that a finding [null hypothesis rejected] was due to chance alone.	<i>P</i> value
Relative risk	The probability of developing an outcome within a specified period if a risk factor is present, divided by the probability of developing the outcome in that same period if the risk factor is absent. The relative risk is applicable to randomized clinical trials and cohort studies; for case-control studies, the odds ratio can be used to approximate the relative risk if the outcome is infrequent. The relative risk should be accompanied by CIs.	RR
Odds ratio	The ratio of two odds; it may have different definitions depending on the study and, therefore, should be defined. For example, it may be the odds of having the disease if a particular risk factor is present to the odds of not having the disease if the risk factor is not present, or the odds of having a risk factor present if the person has the disease to the odds of the risk factor being absent if the person does not have the disease. The odds ratio typically is used for a case-control or cohort study. For a study of incident cases with an infrequent disease, the odds ratio approximates the relative risk. The odds ratio is usually expressed by a point estimate and 95% CI.	OR

[About Entrez](#)

[Text Version](#)

[Entrez PubMed](#)

[Overview](#)

[Help](#)

[FAQ](#)

[Tutorial](#)

[New/Noteworthy](#)

[E-Utilities](#)

[PubMed Services](#)

[Journals Database](#)

[MeSH Database](#)

[Single Citation Matcher](#)

[Batch Citation Matcher](#)

Clinical Queries

[Special Queries](#)

[LinkOut](#)

[My NCBI](#)

[Related Resources](#)

[Order Documents](#)

[NLM Mobile](#)

[NLM Gateway](#)

[TOXNET](#)

[Consumer Health](#)

[Clinical Alerts](#)

[ClinicalTrials.gov](#)

[PubMed Central](#)

[Privacy Policy](#)

This page provides the following specialized PubMed searches for clinicians:

- [Search by Clinical Study Category](#)
- [Find Systematic Reviews](#)
- [Medical Genetics Searches](#)

After running one of these searches, you may further refine your results using PubMed's [Limits](#) feature.

Results of searches on these pages are limited to specific clinical research areas. For comprehensive searches, use [PubMed](#) directly.

Search by Clinical Study Category

This search finds citations that correspond to a specific clinical study category. The search may be either broad and sensitive or narrow and specific. The search filters are based on the work of [Haynes RB et al.](#) See the [filter table](#) for details.

Search

Go

Category

Scope

- | | |
|--|--|
| <input type="radio"/> etiology | <input checked="" type="radio"/> narrow, specific search |
| <input type="radio"/> diagnosis | <input type="radio"/> broad, sensitive search |
| <input checked="" type="radio"/> therapy | |
| <input type="radio"/> prognosis | |
| <input type="radio"/> clinical prediction guides | |

Find Systematic Reviews

For your topic(s) of interest, this search finds citations for systematic reviews, meta-analyses, reviews of clinical trials, evidence-based medicine, consensus development conferences, and guidelines.

For more information, see [Help](#). See also [related sources](#) for systematic review searching.

Search

Go

Category

- All
- Diagnosis
- Differential Diagnosis
- Clinical Description
- Management
- Genetic Counseling
- Molecular Genetics
- Genetic Testing

[Write to the Help Desk](#)

[NCBI](#) | [NLM](#) | [NIH](#)

[Department of Health & Human Services](#)

[Privacy Statement](#) | [Freedom of Information Act](#) | [Disclaimer](#)