CATALYST Consortium / TAHSEEN Project

EBM: Scenario-based FP/RH Training Curriculum for Physicians

Trainee's Guide

March 2004

Project funded by the United States Agency for International Development

The CATALYST Consortium is a global reproductive health activity initiated in September 2000 by the Office of Population and Reproductive Health, Bureau for Global Health, U.S. Agency for International Development (USAID). The Consortium is a partnership of five organizations: the Academy for Educational Development (AED), Centre for Development and Population Activities (CEDPA), Meridian Group International, Inc., Pathfinder International and PROFAMILIA/Colombia.

This publication was made possible through support provided by the Office of Population and Health, United States Agency for International Development, under the terms of contract No. HRN-A-00-00-00003-00. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the United States Agency for International Development.

Evidence Based Medicine: A Program for Clinicians

INTRODUCTION:

Medical practice is changing; emphasis on the use of the best evidence derived from medical literature in guiding medical practice has become a global demand.

The foundations of this change lie in developments in clinical research over the last 30 years. In 1960, the randomized clinical trial was an oddity. It is now accepted that no drug can enter clinical practice without a demonstration of its efficacy in clinical trials. Moreover, the same randomized trial method is being applied to surgical therapies and diagnostic tests. Meta-analyses are gaining increasing acceptance as a method of summarizing the results of a number of randomized trials, and ultimately may have as profound an effect on setting treatment policy as have randomized trials themselves. Crucial methodological advances have also been made in other areas, such as the assessment of diagnostic tests and prognosis.

A new philosophy of medical practice and teaching has followed these methodological advances, this is apparent on the following observed changes:

- The profusion of articles instructing clinicians on how to access, evaluate, and interpret the medical literature.
- The proposals to apply the principles of clinical epidemiology to day-to-day clinical practice.
- Major medical journals adopting a more informative structured abstract format which incorporates issues of methods and design into the portion of an article the reader sees first.

- Journals focusing on publications of high relevance and of methodological rigor.
- Appearance of textbooks which provide a rigorous review of available evidence, including a methods section describing both the methodological criteria used to systematically evaluate the validity of the clinical evidence and the quantitative techniques used for summarizing the evidence.
- Appearance of practice guidelines based on rigorous methodological review of the available evidence.

In Egypt, the demand for the incorporation of evidence based medicine into medical practice has been demonstrated in the conferences and conventions organized by different bodies; Universities, Syndicates, Scientific groups, MOHP hospitals. Even the media, has begun to show interest in this new method of medical thinking.

RATIONALE:

The influence of evidence-based medicine on clinical practice and medical education is increasing. Clinicians are required to keep updated with current best evidence to optimize patient care outcomes. The clinician needs to learn to access, select, appraise and appropriately utilize up-to-date medical information to be able to maintain the standards of his practice. The clinician should be skilled in navigating useful internet sites, and in the use of appropriate search engines for locating current best evidence.

GOAL:

The goal of this program is to educate clinicians in the practice of evidence-based medicine. This program is designed to provide clinicians with the skills and knowledge required to bring evidence based medicine into their daily practice, by stressing upon the examination of evidence from clinical research, and de-emphasizing intuition, unsystematic clinical experience, and patho-physiologic rationale as sufficient grounds for clinical decision-making.

PROGRAM DESCRIPTION:

This is a 36 hours program which will allow clinicians to attain the knowledge and develop the skills and attitudes required as the basis for the integration of Evidence Based Medicine into their clinical practice.

LEARNING OBJECTIVES:

Evidence-based medicine requires new skills of the physician, including efficient literature-searching, and the application of formal rules of evidence in evaluating the clinical literature.

By completing this program the participating clinician is expected to achieve the following objectives:

1. Define the terms: EBM, POEMs, DOEs, and Information Mastery.
2. Describe the importance of an evidence-based approach to clinical practice.
3. Apply Information Mastery to the management of the medical literature.
4. Define the importance of clinical questions.
5. Create well-formed clinical questions.
6. List the characteristics of different study types.
7. List the type of study required to answer each clinical situation.

8. Define the Priority of different questions according to certain criteria.
9. Develop a search strategy.
10. Define the terminology and methods of searching the MEDLINE database.
11. Perform a web search.
12. Identify the web sites where relevant information may be sought.
13. Develop skill in searching for relevant literature using the internet.
14. List the different methods of reasoning used by physicians during the process of diagnosis.
15. Identify the steps of evaluating literature dealing with diagnostic techniques.
16. Calculate sensitivity and specificity, positive and negative predictive values and likelihood ratios for different diagnostic tests.
17. Use the sensitivity and specificity, positive and negative predictive values and likelihood ratios in clinical decisions.
18. Distinguish between tests that rule-in and rule-out disease.
19. Appraise literature dealing with diagnostic tests using the worksheet.
20. Identify pitfalls in common approaches to clinical decision-making.
21. Determine if a study about therapy is relevant and valid.
22. Analyze study results on an “intention to treat" basis.
23. Distinguish between “relative risk reduction” and “absolute risk reduction”.
24. Calculate the “number needed to treat” (NNT) in a study.
25. Explain the steps of performing a meta-analysis: identification, selection, abstraction, and analysis.
26. Distinguish between random and fixed effects models when reading a meta-analysis
27. Evaluate the relevance, validity and usefulness of a meta-analysis.

28. Critically appraise a simple meta-analysis.

WHO SHOULD ATTEND:

Clinicians with basic computer skills who are capable of word processing and navigating the web.

CLASS SIZE:

The class size should be designed according to the computer and internet facilities available for the program, together with the number of available instructors. One computer with internet access is required for every two participants, while one instructor for every ten participants will usually suffice.

TRAINING APPROACH:

- Participants will be trained by professional educators.
- The training will consist of self-study, classroom instruction and hands-on computer lab training.
- Competency in performance will be achieved through computer-lab training guided by worksheets.
- The worksheets will be available for participants to be used during self-study.
- Participants will receive a **Trained EBM Clinician** certificate upon completion of the training.

COURSE CONTENTS:

Module I: INTRODUCTION TO EBM		
Introduction & Pretest	*30 m*	
What is EBM?	*20 m*	
Limitations of the current practice	*30 m*	
The argument for EBM	*30 m*	
Information Mastery	*80 m*	
POEMs and DOEs	*30 m*	
Quiz	*20 m*	
Total minutes allocated to Module I:		**240 m**
Module II: Asking Clinical Questions		
Introduction & Questions asked by Clinicians at the Point of Care.	*20 m*	
Clinical Questions and Lifelong Learning.	*10 m*	
Learning to Ask a Focused Clinical Question.	*30 m*	
Steps to improve the ability to ask and answer questions	*30 m*	
Total minutes allocated to Module II:		**90 m**
Module III: Searching for the Best Evidence		
Characteristics of different types of studies.	*90 m*	
Type of study to answer type of question.	*30 m*	
Prioritizing your questions.	*30 m*	
Search strategy.	*30 m*	
Searching Medline.	*60 m*	
Using the World Wide Web.	*60 m*	
Collection of important sites	*60 m*	
Total minutes allocated to Module III:		**360 m**
Module IV: Appraising Literature : 1.Diagnosis		
Introduction & Types of reasoning	*15 m*	

Evaluating the relevance of diagnosis article.	*15 m*	
Evaluating the validity of a diagnosis article	*15 m*	
Calculating Sensitivity, specificity positive and negative predictive values and likelihood ratios	*30 m*	
Using Sensitivity, specificity positive and negative predictive values and likelihood ratios	*30 m*	
Distinguish between results that rule in and rule out diseases.	*30 m*	
Appraise literature on diagnosis using worksheet	*105 m*	
Total hours allocated to Module VI:		**240 m**
Module V: Appraising Literature : 2. Therapy		
Introduction	*30 m*	
Critically reading a paper on therapy.		
Evaluating Relevance	*15m*	
Evaluating Validity	*15 m*	
Analyze study results	*30 m*	
Relative and absolute risk reduction	*30 m*	
Worksheet for evaluating an article about Therapy & Calculating NNT	*30 m*	
Total minutes allocated to Module V:		**150 m**
Module VI: Appraising Literature : 3. Prognosis		
Formulate question relative to prognosis	*20 m*	
Understanding a cohort study	*30 m*	
Understanding case control study	*30 m*	
Evaluating the relevance of studies.	*20 m*	
Evaluating the validity of studies	*20 m*	
Appraise article dealing with prognosis	*120 m*	
Total minutes allocated to module VI:		**240 m**

Module IV: Appraising Literature : 4. Integrative studies			
Introduction.	*20 m*		
The need and the development of meta-analyses.	*40 m*		
Steps of meta-analyses.	*30 m*		
Distinguish between random and fixed effects models	*60 m*		
Evaluating the relevance validity and usefulness of a meta-analysis.	*120 m*		
Critically appraise a simple meta-analysis.	*180 m*		
Total minutes allocated to Module IV:		**450 m**	
Post test and Conclusion		**30 m**	
Total minutes allocated to the Course		**1800 m**	**30 h**

Educational Components of the Curriculum

	Knowledge	Attitudes	Skills
Question	• The anatomy of a question • The Map • Basic clinical skills (H&P)	• Curiosity • Comfort with Uncertainty • Value Active Listening	• Formulate a Question
Searching	• MEDLINE • Understanding of search strategies (filters, etc.)	• Computer Phobia • Deal with aversion to these technologies	• Computer Literacy • Informatics • Tie question to specific information sources
Critical Appraisal	• Practical clinical epidemiology (User's Guides) • Survival Stats • Fatal Flaws (the hopeless to perfect scale)	• Innumeracy • Readiness to challenge authority (Challenge them to be critical, do not accept it as it must be so)	• Which article will answer your question • Apply these skills to real time settings
Application	• Getting the Individual Patient (NNT) • Going from pre-test to post-test probabilities (likelihood ratios) • Strength of inference	• The recognition that value judgements are implicit in every clinical decision • Judgements are being made all the time by physicians based on the MD's value system • Adding to the notion of patient's values when you go from evidence to practice • Readiness and willingness to change	• Solicit patient preferences • Assess co-morbidities and social support of patient • Assess where the patient's value system is on the paternalism to technical continuum

Course Schedule:

Day Hour	09.00-10.30	10.30-11.00	11.00- 13.00	13.00-13.30	13.30-15.00
Day 1	Introduction/ Pretest	Break	Module I	Break	Module I
Day 2	Module II	Break	Module II	Break	Module II
Day 3	Module III	Break	Module III	Break	Module III
Day 4	Module IV	Break	Module IV	Break	Module IV
Day 5	Module V	Break	Module V	Break	Module VI
Day 6	Module VI	Break	Module VI	Break	Closing and Post test

Levels of Evidence

Grade of recommendation	Level of Evidence	Therapy: Whether a treatment is efficacious/ effective/harmful	Therapy: Whether a drug is superior to another drug in its same class	Prognosis	Diagnosis	Differential diagnosis/symptom prevalence study	Economic and decision analysis
A	1a	SR (with homogeneity*) of RCTs	SR (with homogeneity**) of head-to-head RCTs	SR (with homogeneity*) of inception cohort studies; CDR† validated in different populations	SR (with homogeneity*) of Level 1 diagnostic studies; CDR† with 1b studies from different clinical centres	SR (with homogeneity*) of prospective cohort studies	SR (with homogeneity*) of Level 1 economic studies
	1b	Individual RCT (with narrow Confidence Interval‡)	Within a head-to-head RCT with clinically important outcomes	Individual inception cohort study with ≥ 80% follow-up; CDR† validated in a single population	Validating** cohort study with good††† reference standards; or CDR† tested within one clinical centre	Prospective cohort study with good follow-up****	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
	1c	All or none§		All or none case-series	Absolute SpPins and SnNouts††	All or none case-series	Absolute better-value or worse-value analyses‡‡
B	2a	SR (with homogeneity*) of cohort studies	Within a head-to-head RCT with validated surrogate outcomes ‡‡‡	SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs	SR (with homogeneity*) of Level >2 diagnostic studies	SR (with homogeneity*) of 2b and better studies	SR (with homogeneity*) of Level >2 economic studies
	2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Across RCTs of different drugs v. placebo in similar or different patients with clinically important or validated surrogate outcomes	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR† or validated on split-sample§§§ only	Exploratory** cohort study with good†††reference standards; CDR† after derivation, or validated only on split-sample§§§ or databases	Retrospective cohort study, or poor follow-up	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses
	2c	"Outcomes" Research; Ecological studies		"Outcomes" Research		Ecological studies	Audit or outcomes research
	3a	SR (with homogeneity*) of case-control studies	Across subgroup analyses from RCTs of different drugs v. placebo in similar or different patients, with clinically important or validated surrogate outcome		SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies	SR (with homogeneity*) of 3b and better studies
	3b	Individual Case-Control Study	Across RCTs of different drugs v. placebo in similar or different patients but with unvalidated surrogate outcomes		Non-consecutive study; or without consistently applied reference standards	Non-consecutive cohort study, or very limited population	Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations.
C	4	Case-series (and poor quality cohort and case-control studies§§)	Between non-randomised studies (observational studies and administrative database research) with clinically important outcomes	Case-series (and poor quality prognostic studies ***)	Case-control study, poor or non-independent reference standard	Case-series or superseded reference standards	Analysis with no sensitivity analysis

Levels of Evidence

D	5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"; or non-randomised studies with unvalidated surrogate outcomes	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

1. These levels were generated in a series of iterations among members of the NHS R&D Centre for Evidence-Based Medicine (Bob Phillips, Chris Ball, Dave Sackett, Brian Haynes, Sharon Straus and Finlay McAlister).
2. Users can add a minus-sign "-" to denote the level of that fails to provide a conclusive answer because of:
   - EITHER a single result with a wide Confidence Interval (such that, for example, an ARR in an RCT is not statistically significant but whose confidence intervals fail to exclude clinically important benefit or harm)
   - OR a Systematic Review with troublesome (and statistically significant) heterogeneity.
3. Grades of recommendation are shown as linked directly to a level of evidence. However levels speak only of the validity of a study not its clinical applicability. Other factors need to be taken into account (such as cost, easy of implementation, importance of the disease) before determining a grade. Grades that are currently in the guides link closely to the validity of the evidence - these will change over time to reflect better concerns that we highlight in the text of the guide or related CATs.

Notes

*	By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. As noted above, studies displaying worrisome heterogeneity should be tagged with a "-" at the end of their designated level.
†	Clinical Decision Rule. (These are algorithms or scoring systems which lead to a prognostic estimation or a diagnostic category)
‡	See comment #2 for advice on how to understand, rate and use trials or other studies with wide confidence intervals.
§	Met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but none now die on it.
§§	By poor quality cohort study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and non-exposed individuals and/or failed to identify or appropriately control known confounders and/or failed to carry out a sufficiently long and complete follow-up of patients. By poor quality case-control study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both cases and controls and/or failed to identify or appropriately control known confounders.
§§§	Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples.
††	An "Absolute SpPin" is a diagnotic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An "Absolute SnNout" is a diagnostic finding whose Sensitivity is so high that a Negative result rules-out the diagnosis.
‡‡	Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost. Worse-value treatments are as good and more expensive, or worse and equally or more expensive.
†††	Good reference standards are independent of the test, and applied blindly or objectively to applied to all patients. Poor reference standards are haphazardly applied, but still independent of the test. Use of a non-independent reference standard (where the 'test' is included in the 'reference', or where the 'testing' affects the 'reference') implies a level 4 study.
**	Validating studies test the quality of a specific diagnostic test, based on prior evidence. An exploratory study collects information and trawls the data (e.g. using a regression analysis) to find which factors are 'significant'.
***	By poor quality prognostic cohort study we mean one in which sampling was biased in favour of patients who already had the target outcome, or the measurement of outcomes was accomplished in <80% of study patients, or outcomes were determined in an unblinded, non-objective way, or there was no correction for confounding factors.
****	Good follow-up in a differential diagnosis study is >80%, with adequate time for alternative diagnoses to emerge (eg 1-6 months acute, 1 - 5 years chronic)
‡‡‡	Surrogate outcomes are considered validated only when the relationship between the surrogate outcome and the clinically important outcomes has been established in long-term RCTs.

Levels of Evidence

References

1. Canadian Task Force on the Periodic Health Examination: The periodic health examination. CMAJ 1979;121:1193-1254.
2. Sackett DL. Rules of evidence and clinical recommendations on use of antithrombotic agents. Chest 1986 Feb; 89 (2 suppl.):2S-3S.
3. Cook DJ, Guyatt GH, Laupacis A, Sackett DL, Goldberg RJ. Clinical recommendations using levels of evidence for antithrombotic agents. Chest 1995 Oct; 108(4 Suppl):227S-230S.
4. Yusuf S, Cairns JA, Camm AJ, Fallen EL, Gersh BJ. Evidence-Based Cardiology. London: BMJ Publishing Group, 1998.

Module I

INTRODUCTION TO EVIDENCE BASED MEDICINE

CONTENTS:

- **Objectives of the module**
- **What is EBM?**
- **Limitations of the current practice**
- **The argument for EBM**
- **Information Mastery**
- **POEMs and DOEs**
- **Quiz**

Objectives for the Introduction to Evidence-Based Medicine

At the end of this module, the participant will be able to:

- Define the terms: EBM, POEMs, DOEs, and Information Mastery.
- Describe the importance of an evidence-based approach to clinical practice.
- Apply Information Mastery to the management of the medical literature.

What is Evidence-Based Medicine?

Evidence-based medicine (EBM) is an important change in the way physicians practice, teach, and do research. It was initially proposed by Dr. David Sackett and colleagues at McMaster University in Ontario, Canada.

Dr. Sackett defines EBM as:

"...the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."

Evidence alone is not sufficient for the compassionate, effective care of patients. The practice of EBM requires that it be integrated with clinical expertise, which brings the following important elements:

- Expertise in performing the history and physical examination.
- Knowledge of the patient, the family, and the community which creates a context for therapeutic decision-making.
- A relationship with the patient informed by his or her beliefs and values.
- Practical knowledge of the availability of resources in the community.

The balance, and even tension, between evidence and clinical expertise is summarized by Sackett:

"Without clinical expertise, practice risks becoming tyrannized by external evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient. Without current best external evidence, practice risks becoming rapidly out of date, to the detriment of patients."

Taking an evidence-based approach to the care of patients is an intellectually exciting style of practice, which leads down a path of exploration and lifelong learning. It gives a framework for knowing when to adopt new tests

and therapies, when to discard old ones, and how to look for answers to the clinical questions that arise daily during in the care of patients.

Limitations of Current Clinical Practice

Do we need to change our current practices? Our patients like us; our outcomes are generally good; and we keep passing our exams. Besides, change is hard work. Aren't we busy enough learning about how to survive in our community?

Unfortunately, there is good evidence that the quality of care we give our patients could be better. Such evidence comes from:

- Clinical examples, in which lack of good evidence has led to harm for patients.
- Common patterns of thinking that introduce bias.
- The wide variation in current clinical practice among physicians
- The difficulty of managing medical information, when results conflict and thousands of articles are published every month.
- Our knowledge declines over time as we get further from medical school.

Practice without the best evidence:

Much of what we do is NOT supported by reliable external evidence, be it from randomized trials, systematic evaluation of a diagnostic test, or careful follow-up of large numbers of patients. Consider the following examples:

Back to sleep

Physicians in western countries traditionally recommended that babies sleep on their stomachs. It was thought that by sleeping on their backs infants were at risk for regurgitation and aspiration, leading to sudden infant death syndrome (SIDS). In the 1980's, some physicians asked the question, "Is there any evidence to support the practice of sleeping babies

on their stomachs?" As it turned out, case-control and ecologic studies found a dramatic decrease in SIDS deaths among children sleeping on their backs, leading to the national "Back to Sleep" program. Had someone asked the question 20 years earlier, tens of thousands of lives might have been saved.

Eye-patches for corneal abrasion

The standard practice of ophthalmologists has always been to patch the eyes of patients who present with a simple, uncomplicated corneal abrasion, as well as provide a mydriatic agent and antibiotic eyedrops. However, no one ever asked if patching was beneficial - it was just "common sense". There have been at least five randomized controlled trials of patch vs no patch, and each has come up with the same answer: patches offer no benefit, and may even slow healing and increase patient discomfort.

Anti-arrhythmic agents

In the late 1980's, *encainide* and *flecainide* were marketed as anti-arrhythmic agents based on their ability to suppress ventricular arrhythmias. However, a large randomized controlled trial (CAST) showed that mortality was considerably higher among treated patients than among controls, This especially demonstrates the problem of relying on "intermediate" or "disease-oriented" outcomes such as arrhythmia suppression, rather than looking at more important measures such as mortality, morbidity, and quality of life.

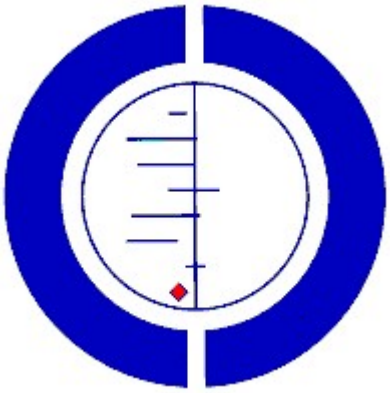
Steroids in prematurity

In 1973, a small study demonstrated that steroids given to women expected to deliver prematurely reduced the likelihood of death in their infants. Six further studies in the next 10 years had mixed results, primarily because they were all quite small. Had a meta-analysis been done in 1983, it would have shown that the overall results of all the trials combined supported a beneficial effect of steroids. However, it took

another decade and seven more studies before these results were accepted and began to change practice. Had a systematic review of the literature been performed in 1983, it might have changed practice much sooner and saved thousands of lives.

The results of the first seven studies of steroids in prematurity, and the summary results (the diamond at the bottom of the diagram) are memorialized in the logo of the ***Cochrane Collaboration****, shown at right. The bars represent the confidence interval of the odds ratio of infant death, and the vertical line is an odds ratio of 1.0. Thus, lines completely to the left of the vertical bar represent a statistically significant benefit of steroids in preventing death.*

The Cochrane Collaboration supports the evidence-based practice of medicine by performing systematic reviews of the literature to answer important clinical questions.



The failure of common sense

There are many more examples of "obvious" interventions which fail to help our patients live longer or better lives. How can common sense fail us so badly? As human beings, we are "wired" to respond to cues from the environment in certain, predictable ways. In particular, it has been very successful from an evolutionary perspective to look for causality in our environment:

“I shook the tree, and a coconut fell on my head. Hmm....maybe shaking the tree caused the coconut to fall on my head! Better not shake trees without first watching for falling coconuts...”

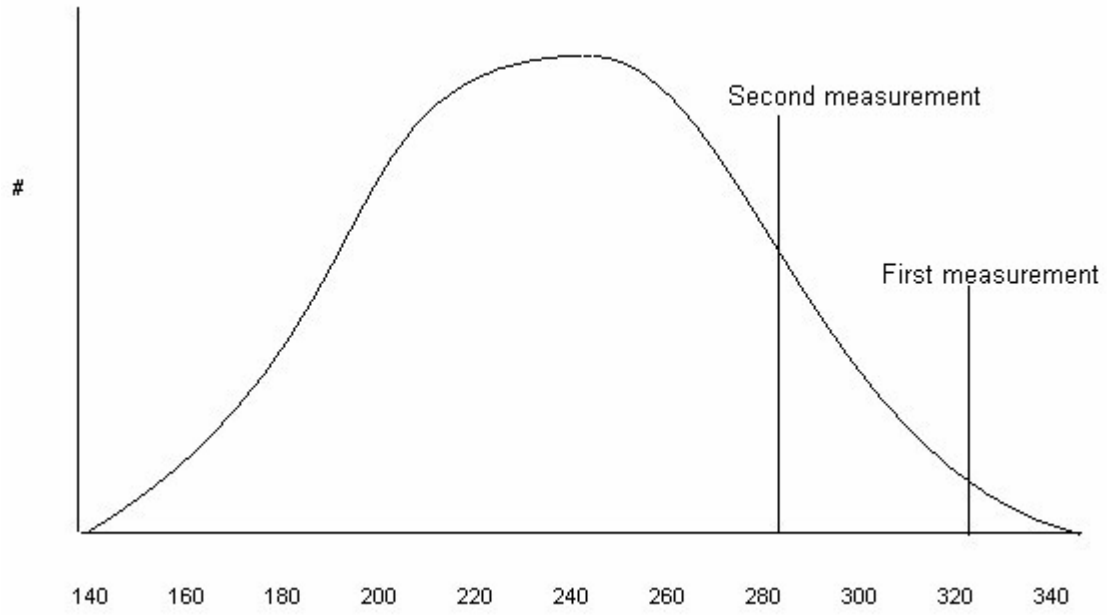
Consider a similar example from the medical field:

“I gave my patient who has had bronchitis for 4 or 5 days an antibiotic, and 3 days later she felt better. Hmm...maybe giving the patient the antibiotic caused her to feel better! I better give all of my patients with bronchitis an antibiotic...”

- This ***"causality heuristic"*** is driven by our belief in ourselves as physicians, our belief in the value of pharmaceutical interventions (which have been so

successful for other conditions), and our patients' desire to see their decision to visit the physician validated.

- There are many other heuristics or ***"rules of thumb"*** which guide, and sometimes mislead, clinical decision-making. For example, another common flaw in decision-making is the **"availability heuristic"**. Certainly, we've all had the experience of diagnosing an unusual condition, then looking for it a little harder than usual in the next few patients. For example, if you diagnose *antiphospholipid antibody syndrome* in a patient with repeated fetal losses, then expect to find the antibodies in each and every similar condition.
- Physicians can also be led astray by "***ignoring the prior probability of disease"***. For example, the likelihood that chest pain is caused by coronary artery disease is extremely small in an otherwise healthy 20 year old, but may exceed 50% in an older patient with multiple risk factors. The youth may require only a careful history and reassurance; while for the older patient catheterization may be an appropriate initial study. While having a single diagnostic strategy for all patients with chest pain would be convenient, it would not be good practice.
- **"Regression to the mean"** is often unrecognized and can lead to inappropriate diagnoses and interventions. Consider a patient with a slightly elevated liver function test on the first measurement:



Clearly, it is much more likely, just given random variation, that the second measurement will be lower than higher. There is much more of the curve to the left than to the right of the initial measurement. Assumptions of an improving medical situation built only on the test result can be misleading in such situations, and the variance in the results should be understood as part of the picture, not as the whole picture.

Variation in current practice

As new tests and therapies are developed, how do physicians decide which to adopt? Without a clear, consistent framework, these decisions are typically driven by the ***practice patterns of local "opinion leaders"***, ***advertising***, ***pharmaceutical representatives***, ***specialists who may see a different spectrum of patients***, and other potentially biased sources. The result is a huge variation in practice patterns among regions, hospitals, and even departments in the same hospital.

Managing medical information

Each month, thousands of medical journals publish tens of thousands of articles. Even if you only consider the 90 or so clinically oriented journals of most interest to primary care physicians, they publish over 15,000 articles per year. If a

physician is to read 40 articles every day of the year, he would still fall way behind!

Clearly, physicians can't read everything. In addition, much of the literature is in apparent conflict with a sort of "ping-pong" game played as issues are debated in the pages of journals. Results are often presented which are premature, use inappropriate outcomes, or represent communications to other researchers, not clinicians. Most importantly, many published studies suffer from serious flaws which invalidate their results.

Unless specially trained in critical appraisal, physicians may be misled by these invalid results. A systematic, rational strategy is needed to deal with this information overload.

Knowledge declines over time

Although medical knowledge declines as physicians get further from medical school, their ability to practice the "art" of medicine improves. Physicians get to know their patients, sharpen their diagnostic skills, and are exposed to an ever increasing number of patients and problems. Wouldn't it be great if our medical knowledge also improved as we moved through our careers?

The argument FOR evidence-based medicine

Taking an evidence-based approach to practice, teaching, and research can help you address some of the limitations of current medical practice. It can help you:

- Stay up to date with the current literature.
- Communicate effectively with consultants.
- Make the best use of other sources of information, such as pharmaceutical representatives and colleagues.
- Make the best use of information from the history, physical examination, and diagnostic testing.
- Avoid common pitfalls of clinical decision-making

Using the best evidence

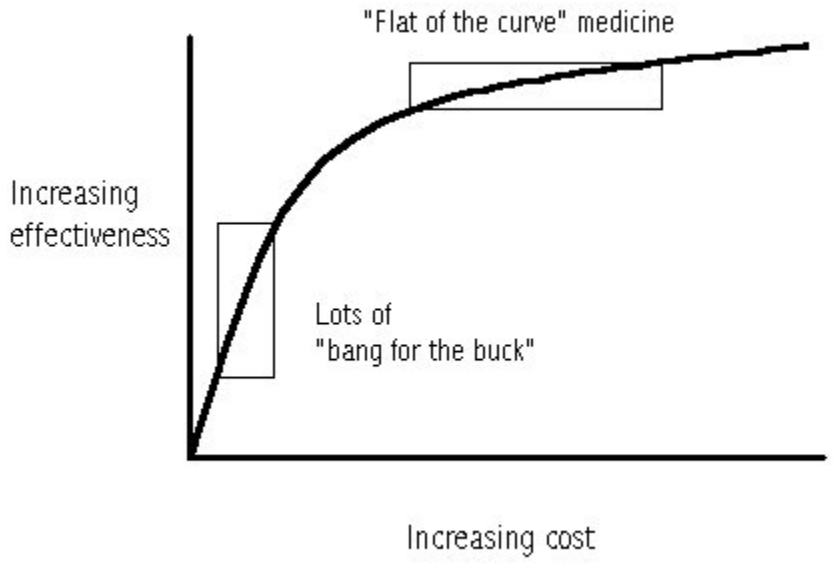
Central to an evidence-based approach is ***the ability to ask the right question***. Consider the example of patching eyes for a corneal abrasion. Identifying the best way to manage corneal abrasions begins when you ask the question *"Is there any evidence that patching eyes improves outcomes for my patients with corneal abrasion."* Learning how to ask these questions is a skill, and also a habit for lifelong learning.

Next, you (or a colleague) have to ***search ALL of the existing literature***, to see if anyone has attempted to answer the question. An evidence-based approach requires that you systematically and exhaustively search the medical literature to identify all of the relevant evidence. Finding one or two recent and conveniently available articles that happen to support your thesis is not evidence-based medicine! Searching the literature effectively and efficiently is another skill for practicing EBM.

Once you have found some articles that attempt to answer your clinical question, you have to ***appraise their validity***. Validity can be internal or external. A study with good internal validity has used the right study design and appropriate methods to answer the question. External validity can be thought of as "*generalizability*" - the results can be generalized to your patients, practice, or community.

Finally, you have to decide whether the results can be ***applied to your practice***. We all operate within constraints imposed by economics, health systems, patient preferences, culture, and availability of resources.

An evidence-based approach is sometimes confused with cost-effectiveness. Although an evidence-based approach should increase effectiveness, it may also result in higher costs. The curve below plots increasing effectiveness against increasing cost:



Once a more effective intervention is identified, the decision to implement may depend on where it lies on the "health-policy" curve shown above. If near the lower left corner, where you get a large improvement in benefit for a small cost, it may make sense. If near the "flat of the curve", you (or the MHOP, or

government) may decide that the incremental benefit is not worth the additional cost. For example, a recent randomized trial demonstrated a small improvement in symptoms among patients with herpes simplex labialis who used a fairly expensive antiviral cream. While some western societies may find this benefit worth the cost, a developing country is not likely to consider this the best use of scarce resources.

EBM - improving on "common sense"

Taking an evidence-based approach gives a solid framework for evaluating new evidence from the literature, critically appraising existing practices, and effectively using the clinical information gathered from patients. For example, by understanding how much a test result should change the probability of disease, common errors of over-treatment or under-treatment can be avoided. A better understanding of study designs will help understand the importance of randomization and systematic evaluation, rather than relying too much on anecdote and personal experience.

Reducing variation in practice

As more physicians practice in an evidence-based manner, and use the same skills and criteria to answer clinical questions, we should see less variation in practice. Evidence-based practice guidelines take into account differences between populations and allow flexibility for patient and physician decision-making. They are an efficient way to disseminate recommendations based on the best evidence without requiring each physician to do literature searches and validity assessments.

Managing medical information

There are many evidence-based sources of information with "pre-digested" evidence in a “user-friendly” format. In the next section (Information Mastery) we

will discuss a valuable strategy for identifying the most useful new information as you "seek" to keep up with the medical literature every month.

Reversing the decline in medical knowledge

So, is there any evidence for evidence-based medicine? Sackett describes observations of McMaster University graduates (who have been trained in evidence-based medicine) and graduates of other Canadian medical schools (who were not) over a 10 year period. The McMaster graduates maintained their knowledge of the management of hypertension while the other graduates experienced a statistically significant decline. More work is needed in this area, though, and it remains an important area for future research.

Information Mastery

To be useful, medical information should be relevant to everyday practice, correct (valid) and easy to obtain. Slawson and Shaughnessy describe a formula which relates these three factors in a "Usefulness equation":

$$\text{Usefulness of medical information} = \frac{\text{Relevance x validity}}{\text{Work}}$$

VALIDITY: Is the hardest part of information mastery, and it is the element on which the Evidence-Based Medicine Working Group at McMaster University and other pioneers of evidence-based medicine have focused. Their work has focused on evaluating the validity of original research articles in the primary medical literature. However, physicians seek and get information from a wide variety of sources, not just journal articles. In fact, journal articles are used quite infrequently to answer clinical questions. Much more commonly, we ask colleagues, talk to pharmaceutical representatives, read review articles, and consult textbooks and other references. In addition, validity assessment is an acquired skill that can be fairly time-consuming. It may not be realistic or even appropriate to expect all physicians to become experts at validity assessment. An alternate approach is for physicians to identify *YODAs* (Your Own Data Analyzer) in their communities, and high-quality evidence-based resources (such as the *Journal of Family Practice POEMs feature* or the *Cochrane Database*) to do validity assessments for them.

RELEVANCE: Is based on the frequency that we are exposed to the clinical question in our practice and the type of evidence presented. Medical information can be categorized as either disease-oriented (e.g. pathophysiology, pharmacology, etiology) or patient-oriented (e.g. symptoms, mortality, days in the hospital, cost). Relevant information is therefore patient-oriented evidence, and it focuses on medical problems common to our practice.

One of the basic laws of human behavior is that we want to do the least amount of **WORK** we can get away with. That applies to searching for the answers to clinical questions. It is therefore important when considering the usefulness of medical information to balance the relevance and validity with the work needed. The most useful information is therefore very relevant to our practice, has high validity, and doesn't take much work to access. Let's consider apply this equation to some common sources of medical information:

Our Colleagues on the job:

Our colleagues are likely to see the same patients as we do, so his or her relevance is good. Work is low; all I have to do is catch them in the hallway. Validity is likely to be variable, depending on their training in evidence-based medicine and their ability to critically appraise their own practice.

A pharmaceutical representative

While the work is usually low (too low, some might argue!), relevance and validity are questionable. Pharmaceutical reps often present disease-oriented evidence, and they may be more comfortable with a specialty rather than primary care perspective.

An article in the latest New England Journal of Medicine

This source of information generally has pretty good validity. However, the results may have been obtained in a carefully selected sample of patients in a tertiary care setting, and may not have much relevance for our primary care practice. In addition, the work needed to find and read such an article, and assess its validity, is considerable.

Richard Smith, editor of the British Medical Journal, has created a table which classifies different sources of medical information using the "Usefulness equation":

Information source	Relevance	Validity	Work	Usefulness
Evidence-based textbook	High	High	Low	**High**
Systematic review (evidence-based)	High	High	Low	**High**
Portable summary of systematic reviews (InfoRetriever)	High	High	Low	**High**
POEMs based resources: Journal of Family Practice POEMs feature and Evidence-Based Practice newsletter	High	High	Low	**High**
Internet after 10 years	High	High	Low	**High**
Drug reference book (PDR)	High	Mod	Low	**High-mod**
ACP Journal Club, Evidence-Based Medicine	Moderate	High	Low	**High-mod**
Colleagues	High	Mod	Low	**High-mod**
Practice guidelines (evidence-based)	Mod	High	Low	**High-mod**
Cochrane Database of Systematic reviews	Mod-High	High	Mod-High	**High-mod**
Standard textbook	High	Low	Low	**Mod**
Standard journal review	High	Mod	Low	**Mod**
Free medical newspapers	High	Low	Low	**Mod**
CME lectures	Mod	Mod	Low	**Mod**
CME small groups	High	Mod	Mod	**Mod**
Consensus statements	Mod	Mod	Low	**Mod**
Practice guidelines (consensus)	Mod	Mod	Low	**Mod**
Online searching	Mod	High	High	**Mod**
Journal articles	Low	High	High	**Low**
Drug advertising	Mod	Low	Low	**Low**
Drug company representatives	High	Low	Low	**Low**
Mass media	Low	Low	Low	**Low**

Internet now	Low	Low	High	**Low**

Note that some of the best sources of information have yet to be developed (the Internet in 10 years and a regularly updated evidence-based textbook), and some of the most widely available (journal articles, drug advertising, and drug company representatives) are the least useful. Such is the paradox of current medical information management.

By becoming a **"Medical Information Master**", you will learn to emphasize sources of information that are high in relevance and validity, and low in work. It doesn't mean working harder - it means working smarter, and making the best use of your limited time to get the information that matters for you and your patients.

Knowing where to look to answer clinical questions is an important skill. A somewhat different set of skills is needed to keep up with the literature. While "Hunting" every month in the literature, you should look for POEMs and avoid DOEs. The next section will teach you the difference between POEMs and DOEs, and how they lead the way toward information mastery.

POEMs and DOEs

Slawson and Shaughnessy have developed an innovative and extremely useful approach to information mastery. Central to it is the concept of **POEMs** and **DOEs**.

The term **POEM** stands for "***Patient Oriented Evidence that Matters***", and refers to the kind of article that:

- Addresses a clinical problem or clinical question that primary care physicians will encounter in their practice.
- Uses patient-oriented outcomes.
- Has the potential to change our practice if the results are valid and applicable.

An example of a POEM is a randomized controlled trial that demonstrates that not patching corneal abrasions leads to less pain and faster healing than patching. Consider each of the previous elements:

- First, corneal abrasion is a problem encountered fairly often by primary care physicians.
- Second, the article considers pain, healing rate, and complications as the primary outcomes. Clearly, these are outcomes that we and our patients care about.
- Third, since the traditional practice has been to patch corneal abrasions, this article is a "practice-changer".

DOE stands for ***"Disease-Oriented Evidence"***. DOEs are all too common in the medical literature, and they are often brought to our attention by pharmaceutical representatives eager to change our practice. However, this kind of evidence is often misleading and generally should be considered premature.

For example, early studies of the drug *finasteride* showed promise, based on the effect on the urinary flow rate. Many patients were put on the drug, at great cost. However, further randomized trials showed that symptom scores did not improve any more than placebo. What do your patients care about - whether their urinary flow rate improved, or whether they are sleeping through the night?

When POEMs exist, forget the DOEs.

The following table divides POEMs and DOE into "common" and "uncommon" categories. Common conditions are those encountered at least every two weeks in the typical primary care physician's office, while uncommon conditions are those encountered less often:

	POEM	DOE
Common	**Read these!**	**Dangerous**
Uncommon	**Read if you have time.**	**Worthless**

A third kind of article fills the medical literature: **POE**'s. These studies use patient-oriented outcomes, but the findings don't have the potential to change practice. They confirm what we already do and, while important, they are not a priority for our reading.

As a physician caring for patients, you have an obligation to provide the best possible care. You also have a moral obligation to stay up to date. Since time is limited, your efforts should focus on identifying, validating, and applying common POEMs to practice. Actually, focusing your work on POEMs frees you from reading the majority of the medical literature, since over 97% of it is DOEs and other material. This figure comes from a six month survey of 90 journals, which identified 8047 articles and only 213 POEMs (2.6%). This figure includes both common POEMs (encountered at least once every 2 weeks) and uncommon

POEMs (encountered less often than every 2 weeks, but at least once over a six month period).

What about all those DOEs? First, ignore the rare DOEs about conditions that occur so rarely that a typical primary physician may encounter them only once or twice in a career, if at all. When you have such a case, you can always look it up.

Common DOEs, because they do not use patient-oriented outcomes, should not change practice. The medical literature is full of examples where preliminary data were promising, or intermediate results looked good, but disappointing when real patients and real outcomes were measured. Have the confidence to reject these findings as premature. Even "obvious" intermediate outcomes such as blood pressure or cholesterol can be misleading. For example, there has never been a randomized, controlled trial which demonstrated that calcium channel blockers reduce cardiovascular or all cause mortality. Now, some case-control studies suggest that there may even be harm. In fact, among diabetic hypertensives, two randomized trials have now demonstrated increased mortality for calcium channel blockers compared with ACE inhibitors. Similarly, while some anti-hyperlipidemics reduce cardiovascular mortality, they raise all-cause mortality. Patients don't care which category their mortality is in, they just want to avoid it.

QUIZ:

Question 1.

An evidence-based approach to practice:

a. Requires that you base your practice entirely on randomized controlled trials

b. Discourages decisions which consider patient values and beliefs

c. Provides a framework for incorporating new evidence into your practice

d. Is based on strict application of causal thinking.

Question 2.

A study demonstrates that clarithromycin has a greater MIC (minimum inhibitory concentration) in the respiratory epithelium than erythromycin. The authors conclude that this should make clarithromycin a better choice for patients with bacterial respiratory infections to susceptible organisms. This study is a good example of a(n):

a. POEM

b. DOE

c. YODA

d. ACK

Question 3.

A study demonstrates that glyceryl trinatrate reduces healing time, improves symptoms, and reduces the recurrence rate of chronic anal fissure when compared with traditional therapy. This study is a(n):

a. Common POEM

b. Uncommon POEM

c. DOE

d. ACK

Module II

ASKING CLINICAL QUESTIONS

CONTENTS:

- **Objectives of the module.**
- **Introduction.**
- **Questions asked by Clinicians at the Point of Care.**
- **Clinical Questions and Lifelong Learning.**
- **Learning to Ask a Focused Clinical Question.**
- **Exercises.**
- **Steps to improve the ability to ask and answer questions.**
- **Worksheet for Constructing a clinical question**

Objectives

At the end of this module, the participant will be able to:

- Define the importance of clinical questions.
- Create well-formed clinical questions.

Introduction:

Adults learn differently than children. While children have their curriculum and agenda set for them, adults typically take a much more active role in determining their learning needs. A businessman may decide to learn Urdu because he has business in an Urdu speaking country, or a researcher may learn to use a statistics program because their research project requires it.

Applying this to physicians suggests that we learn best when learning:

1. Is in the context of patient care.
2. Answers our questions.
3. Is directly applicable to our work.
4. Does not take too much time.

Thus, asking and answering the clinical questions that arise at the point of care is central to our continuing education as physicians. Continuing medical education in different countries has not been very successful at improving patient outcomes, largely because it is disconnected from patient care.

The ways the practice of physicians change is haphazard, often not driven by the best available evidence. Physicians who want to be successful lifelong learners need highly developed ***critical reflection skills***. Applying these skills to their practice and generating clinical questions at the point of care are central to physician education and lifelong learning. To take an evidence-based approach to practice, these questions must be answered using the most valid, relevant information available.

Questions asked by Clinicians at the Point of Care:

When asked at the end of a half-day of patient care to recall how many questions they had related to patient care, physicians reported 1 question for every 4 patients. Direct observation of primary care physicians, however, has shown that they generate an average of 2 questions for every 3 patient encounters. A physician seeing 25 patients per day will generate approximately 15 questions. These and similar studies have shown there is tremendous variety to the questions physicians ask, and these questions are often complex and patient specific. Approximately 33 percent relate to treatment, 25 percent to diagnosis, and 15 percent to pharmaco-therapeutics.

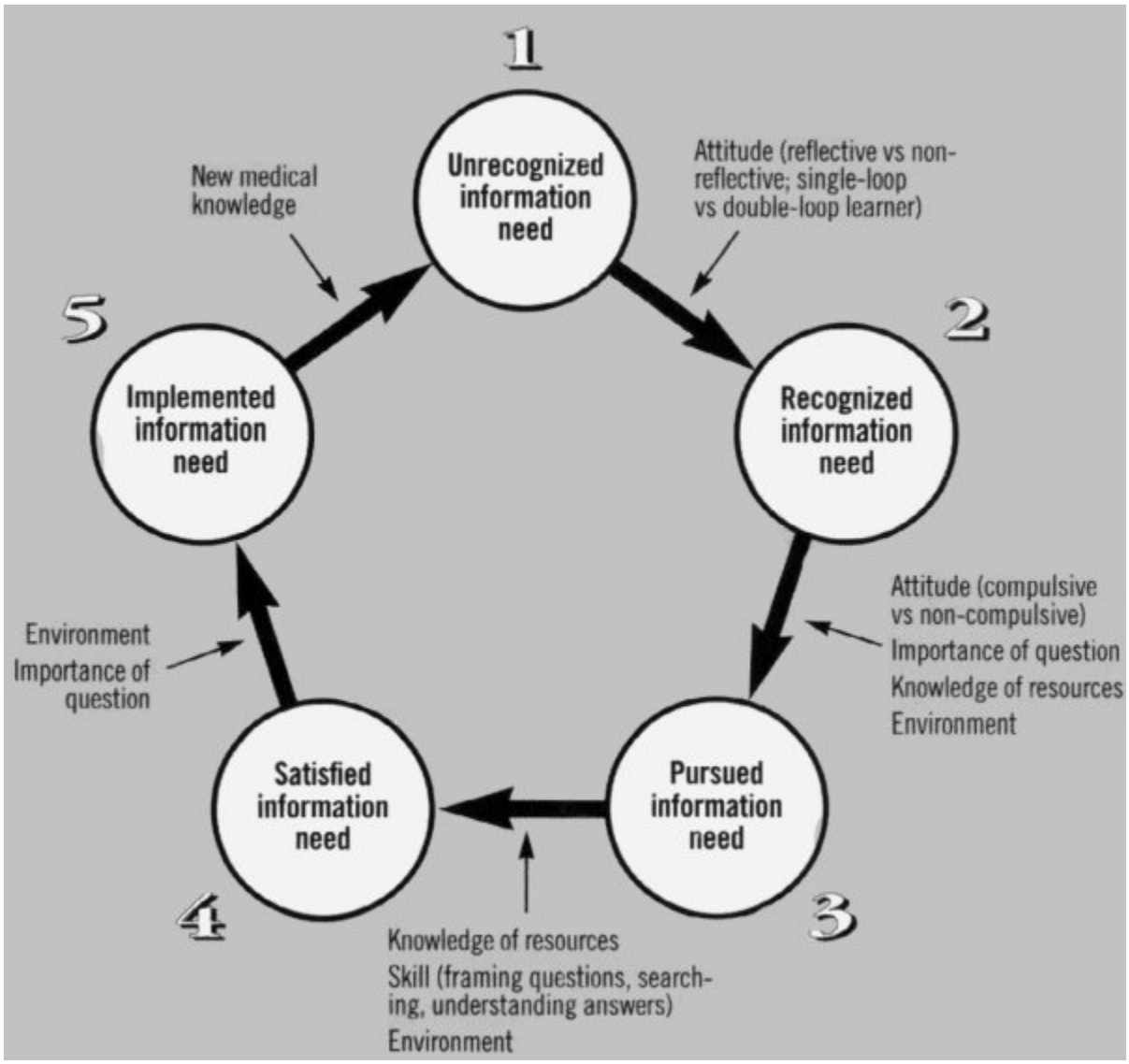
When questions are pursued and answered, more than one half of the answers come from textbooks and human sources, including both office partners and consultants. The *Physician's Desk Reference* (PDR) is perhaps the most commonly named source of answers. Electronic sources of information are rarely used.

Unfortunately, approximately two thirds of the clinical questions generated at the point of care go unanswered. Are these questions important? One study took the unanswered questions and gave them to medical librarians. The authors then gave the answers to the physicians who had asked them and found that approximately one half of the answers would have had a direct impact on patient care.

Why do we not answer more of these questions? Limitations include a lack of convenient access to reference materials at the point of care, the time needed to search for information, and the challenge of formulating an answerable question.

Clinical Questions and Lifelong Learning:

The process of asking and answering questions is summarized in the "Information Pathway" shown below:



Recognized Information Need:

The first step in the information pathway is to recognize that an information need exists. The willingness to recognize that one's knowledge might not be complete would appear to be driven primarily by physician attitude and personality.

Research by educators in the area of ***critical reflection*** suggests that this skill is mature and high-level, but that it might be undermined by an emphasis on ***single-loop rather than double-loop learning***.

Consider the example of a patient who has a cough, which is diagnosed as bronchitis, and who is given erythromycin. When the patient returns 2 days later with persistent symptoms, the physician changes the antibiotic. ***This behavior is single-loop learning, because the physician reacts to the situation without questioning the underlying assumptions***.

A more reflective physician would practice ***double-loop learning***: when the patient returns, both the underlying diagnosis and the treatment strategy are questioned. Maybe the patient does not have bronchitis. Perhaps the patient has reflux or allergies. What if the bronchitis is viral and the antibiotic is not going to help anyway? Do antibiotics help even if the bronchitis is bacterial? The reflective physician, even in this simple clinical situation, generates a number of important questions.

In addition to generating more clinical questions, the reflective physician is more likely to practice patient-centered medicine, tailoring the management strategy to the patient's clinical picture, needs, and situation.

Unfortunately, traditional medical education emphasizes knowing the right answer more than asking the right question.

Pursued Information Need:

The decision to pursue a clinical question is perhaps the most complex step in the information pathway, influenced by:

1. Physician attitude and personality.
2. The characteristics of the question.
3. The practice environment.

Physicians tend to frame questions in relation to specific patients rather than in a generalizable fashion, as is advocated by teachers of EBM

For example, a family physician might ask, “I have this 38-year-old patient with dyspepsia, whose father was diagnosed with pancreatic cancer. He’s really worried that he has pancreatic cancer, and I want to reassure him. What is the best test to rule out pancreatic cancer for him? Should I order a sonogram?”

A clinical epidemiologist or evidence-based medicine proponent might frame this question as, “What is the test with the best negative likelihood ratio for pancreatic cancer among outpatients like this with dyspepsia?”

Research articles might provide information in terms of sensitivity and specificity, but not calculate the likelihood ratio. In addition, the researchers might present data as false-positive and false-negative rates rather than sensitivity or specificity, further confusing the reader.

Implemented Information Need:

Environmental factors as cost, health system constraints, patient acceptance, acceptance of colleagues, and local practice patterns will affect the decision to implement the answer to a clinical question in practice.

Learning to Ask a Focused Clinical Question

Too often clinical questions are formed in terms specific to a particular patient. Asking "What test should I order for this 28-year-old woman with chest pain?" is appropriate when talking to a human consultant but not when searching a medical reference. As phrased, the question requires that you know the patient and the particulars of their case.

You are much more likely to get a helpful answer from a medical reference or a medical librarian if you ask a well-formed, more generalizable question. For example, the above question could be rephrased as: "What is the best test to rule out myocardial infarction in this person with chest pain and a low likelihood of disease?"

PICO

"PICO" is a useful acronym that describes the elements of a well-formed clinical question about therapy. The letters stand for:

- Population
- Intervention
- Comparison
- Outcome

Population is really a description of the group to which the patient belongs, and may include age, gender, race, ethnicity, and stage of disease. The description should be specific enough to be helpful, but not overly specific. You are unlikely to find studies of "Asian 63 year old women" (too specific); describing the population as "Post-menopausal women" is much more likely to be helpful.

Intervention is a description of the test or treatment that you are considering. **Comparison** is the alternative. Not all questions need a comparison, but clinical

questions typically involve a choice between two options, even if the alternative is only watchful waiting.

Outcome should generally be something that not only matters to you, but matters to the patient. You may well want to involve the patient in defining the outcome. In making a decision about anti-depressant drug therapy, some patients may be more concerned about dry mouth and fatigue, while others are more concerned about orgasmic dysfunction.

Exercises

Case #1

Mrs. Haneya presents with a complaint of fatigue and loss of interest in her usual activities. She denies any suicidal ideation, and has a normal history and physical examination. After diagnosing her with minor depression, you are preparing to write a prescription for a Serotonin Specific Reuptake Inhibitor (SSRI). Your patient asks, what about St. John's Wort? You wonder: "Is St. John's Wort a reasonable choice for this patient?"

- Population: Adults with minor depression
- Intervention: St. John's Wort
- Comparison: An SSRI
- Outcome: relief of symptoms

Putting it all together: "In adults with minor depression, is St. John's Wort or an SSRI more effective at relieving symptoms?"

Case #2

Sayed is a 14 year old with seasonal allergies. You wonder: should I prescribe a nasal corticosteroid or non-sedating anti-histamine?

- Population: Patients with seasonal allergies (if you find data for adolescents or children, so much the better, but don't box yourself in at the outset)
- Intervention: nasal corticosteroid
- Comparison: non-sedating antihistamine
- Outcome: relief of symptoms (other important outcomes could include sedation and cost, or you could look at the symptom that is most troublesome to the patient. Coryza, for example.)

Putting it all together: “In patients with seasonal allergies, is an inhaled corticosteroid or a non-sedating antihistamine better at relieving coryza?”

Case #3

Ibrahim has been coughing for 6 weeks. He is a heavy smoker, and you are considering skipping the chest x-ray and going straight to a CT scan.

- Population: Patients with cough
- Intervention: chest x-ray
- Comparison: chest CT
- Outcome: positive and negative predictive value

Putting it all together: “In patients with cough, does chest x-ray or chest CT have a better positive and negative predictive value?”

Comment: This is a tricky one. While we want more accurate tests, that really isn’t the most important thing. The most important thing is whether the increased accuracy improves patient outcomes. Another way to phrase this question is: “In patients with cough, does initial chest x-ray or initial chest CT result in better patient outcomes?” While a better question, in most cases you won’t find evidence that answers this type of question.

Case #4 (therapeutic)

Dawood has had back pain for several days now, after a weekend of chopping wood at his cabin. You wonder whether a muscle relaxant would be helpful.

- Population: ?
- Intervention: ?
- Comparison: ?
- Outcome: ?

Now, put it all together in a single, focused question.

Case #5 (diagnostic)

Bourhan has had persistent heel pain. He wonders if he might have a heel spur, and says "Aren't you going to order an x-ray, doc?"

- Population: ?
- Intervention: ?
- Comparison: ?
- Outcome: ?

Now, put it all together in a single, focused question.

Steps to improve the ability to ask and answer questions.

1. **Become an effective computer user**

One way to reduce the work of answering clinical questions is to make information easily available on a computer.

Excellent sources of evidence-based information for desktop computers include the *Cochrane Library* from the Cochrane Collaboration and the *Best Evidence* reference from the American College of Physicians and the *British Medical Journal* Publishing Group. Both the *Cochrane Library* and Best Evidence are available on CD-ROM.

2. **Create a culture of inquiry**

Medical students, residents, and practitioners build an extensive database of clinical experience and medical knowledge. Like any database, though, it can grow out-of-date. Regularly reflecting on decisions and practices will stimulate clinical questions. Answering these clinical questions using valid, relevant information will keep a database up-to-date. A simple method is to keep an index card in a pocket and write down questions that cannot be answered immediately. Then, make a commitment to answer at least one or two of the most compelling questions each week.

3. **Inquire, Don't Advocate**

Too often, physicians advocate a plan for a patient or a belief about a condition rather than inquire into the best possible approach for the patient.

Advocacy is prefaced by such statements as "I think …," "I believe …," "The facts are …," "Experience says ….," and "My colleagues always …."

Inquiry, on the other hand, is characterized by such statements as, "What do you think?" "I wonder whether there is a better way?" "Should I keep doing this?" and "Why have I always …?" When you find yourself making advocacy

statements, ask yourself whether you could instead inquire. Whereas advocacy tends to uphold the status quo, inquiry leads to new knowledge and new insights.

4. **Feel Good About Not Knowing Everything**

It is impossible for any physician, especially a primary care physician, to know everything. Reading should therefore focus on problems common or important to your patients. When you find an article about such a problem, make sure it uses outcomes that matter to your patients, such as symptom improvement, mortality reduction, cost, or quality of life. An article that is valid and would change your practice is a POEM. Fortunately, only approximately 2 percent of the medical literature is POEMs , so this approach will help get rid of the stack of unread journals.

Because the findings of a POEM have been shown to improve important patient outcomes and because they differ from your current practice, some authors argue that you are ethically obligated to know about POEMs and apply them to your practice.

To learn more about POEMs and search the collected database, see the *Journal of Family Practice* Web site at http://www.jfponline.com/ or visit http://www.medicalinforetriever.com/. Another excellent monthly site is Bandolier, supported by the British National Health Service.

5. **Let Someone Else Do the Heavy Lifting**

There are now several excellent sources of what is called secondary literature. Each has a group of physicians trained in critical appraisal, epidemiology, and research design who select important articles, critically appraise them for validity, and publish the results in brief synopses.

Examples include the ACP Journal Club, ***Journal of Family Practice*** POEMs feature, ***Evidence-Based Practice*** newsletter, and the journal ***Evidence-Based Medicine***.

Construct a Clinical Question Worksheet

Elements of a Good Question	Your Case:
Patient, Population Or Problem How would I describe a group of patients similar to mine?	
Intervention, Treatment, Prognostic Factor, Exposure What main intervention, treatment, prognostic factor, exposure, etc. am I considering?	
Comparison (If appropriate) What is the primary alternative treatment (intervention, prognostic factor, exposure, etc.) to the above?	
Outcome What are you trying to accomplish, improve, or effect?	
The Question is:	
Type of Question How would I categorize this question: Therapy, Prognosis, diagnosis, ?	
Type of Study See below	

Type of Question	Type of Study	Search Strategy
Therapy	Randomized control trial Prospective comparative study Retrospective case-control study	Limit to clinical trial Textword placebo
Prognosis	Prospective Cohort Study Retrospective Case Series	Subject cohort-studies Textword prognosis
Diagnostic	Blinded comparison to a gold standard	Textword sensitivity Subject sensitivity-and-specificity Textword predictive and values
Harm	Randomized control trial Prospective comparative study Retrospective case-control study Case report	Textword risk Subject case-control-studies
Type each line of strategy separately and combine the separate lines with AND or OR at end.		

Module III

SEARCHING FOR THE BEST EVIDENCE

CONTENTS:

- **Objectives of the module.**
- **Characteristics of different types of studies.**
- **Type of study to answer type of question.**
- **Prioritizing your questions.**
- **Search strategy.**
- **Searching Medline.**
- **Using the World Wide Web.**
- **Collection of important sites**

Objectives

- List the characteristics of different study types.
- List the type of study required to answer each clinical situation.
- Define the Priority of different questions according to certain criteria.
- Develop a search strategy.
- Define the terminology and methods of searching the MEDLINE database.
- Perform a web search.
- Identify the web sites where relevant information may be sought.
- Develop skill in searching for relevant literature using the internet.

Characteristics of different types of studies

- **Randomized Controlled Clinical Trial / Controlled Clinical Trial (therapy, diagnosis)**

A clinical trial involving one or more test treatments/diagnostic modality, at least one control treatment/diagnostic modality, specified outcome measures for evaluating the studied intervention, and a bias free method for assigning patients to the test treatment.

The treatment may be drugs, devices, or procedures studied for diagnostic, therapeutic or prophylactic effectiveness. Control measures include placebos, active medicine, no treatment, dosage forms and regiments, historical comparisons, etc.

When randomization using mathematical techniques, such as the use of a random numbers table, is employed to assign patients to test or control treatments, the trial is characterized as a RANDOMIZED CONTROLLED TRIAL.

Trials employing treatment allocation methods such as coin flips, odd-even numbers, patient social security numbers, days of the week, medical record numbers, or other such pseudo- or quasi-random processes are simply designated as CONTROLLED CLINICAL TRIALS.

- **Advantages:**
  - Unbiased distribution of confounders
  - Blinding more likely
  - Randomization facilitates statistical analysis
- **Disadvantages:**
  - Expensive - time and money.

- Volunteer bias.
- Ethically problematic at times.

Multi-center studies:

Controlled studies which are planned and carried out by several cooperating institutions to assess certain variables and outcomes in specific patient populations, for example, a multi-center study of congenital anomalies in children.

▪ Cohort Study (prognosis, harm/etiology, prevention)

Studies in which subsets of a defined population are identified. These groups may or may not be exposed to factors hypothesized to influence the probability of the occurrence of a particular disease or other outcome. Cohorts are defined populations which, as a whole, are followed in an attempt to determine distinguishing subgroup characteristics.

- **Advantages:**
  - Ethically safe
  - Subjects can be matched
  - Can establish timing and directionality of events
  - Eligibility criteria and outcome assessments can be standardized
  - Administratively easier and cheaper than RCT
- **Disadvantages:**
  - Controls may be difficult to identify
  - Exposure may be linked to a hidden confounder
  - Blinding is difficult
  - Randomization not present

- For rare disease, large sample sizes or long follow-up necessary

▪ Case-control Study (prognosis, harm/etiology, prevention)

Studies which start with the identification of persons with a disease of interest and a control (comparison, referent) group without the disease. The relationship of an attribute to the disease is examined by comparing diseased and non-diseased persons with regard to the frequency or levels of the attribute in each group. Involves identifying patients who have the outcome of interest (cases) and control patients without the same outcome, and looking back to see if they had the exposure of interest.

- **Advantages:**
  - Quick and cheap
  - Only feasible method for very rare disorders or those with long lag between exposure and outcome
  - Fewer subjects needed than cross-sectional studies
- **Disadvantages:**
  - Reliance on recall or records to determine exposure status
  - Confounders
  - Selection of control groups is difficult
  - Potential bias: recall, selection

▪ Review Articles:

An article published after examination of published material on a subject. It may be comprehensive to various degrees and the time range of material scrutinized may be broad or narrow, but the reviews most often desired are reviews of the current literature.

The textual material examined may be equally broad and can encompass, in medicine specifically, clinical material as well as experimental research or case reports. State-of-the-art reviews tend to address more current matters. Reviews of the literature must be differentiated from historical reviews on the same subject, but a review of historical literature is also within the scope of this publication type. Specific headings for specific types of reviews (academic, literature, multicase, reported cases, and tutorial) are also available.

- **Systematic Review**

A systematic review usually focuses on a clinical topic and answers a specific question. An extensive review of the evidence is presented after the methodology for identifying the studies in the literature is presented.

- **Meta Analysis**

A quantitative method of combining the results of independent studies (usually drawn for the published literature) and synthesizing summaries and conclusions which may be used to evaluate therapeutic effectiveness, plan new studies, etc. It is often an overview of clinical trials. It is usually called a meta-analysis by the author or sponsoring body and should be differentiated from review of the literature.

Longitudinal Studies

Studies in which variables relating to an individual or group of individuals are assessed over a period of time.

Follow-up Studies: Studies in which individuals or populations are followed to assess the outcome of exposures, procedures, or effects of a characteristic, e.g., occurrence of disease.

Prospective Studies: Observation of a population for a sufficient number of persons over a sufficient number of years to generate incidence or mortality rates subsequent to the selection of the study group.

Retrospective Studies

Studies used to test etiologic hypotheses in which inferences about an exposure to putative causal factors are derived from data relating to characteristics of persons under study or to events or experiences in their past. The essential feature is that some of the persons under study have the disease or outcome of interest and their characteristics are compared with those of unaffected persons.

Cross-Sectional Studies

Studies in which the presence or absence of disease or other health-related variables are determined in each member of the study population or in a representative sample at one particular time. This contrasts with longitudinal studies which are followed over a period of time. The observation of a defined population at a single point in time or time interval. Exposure and outcome are determined simultaneously.

- **Advantages:**
  - Cheap and simple
  - Ethically safe
- **Disadvantages:**
  - Establishes association at most, not causality
  - Recall bias susceptibility
  - Confounders may be unequally distributed
  - Neyman bias
  - Group sizes may be unequal

Type of study to answer the type of question:

Clinical Examination	Prospective, blind comparison to Gold Standard
Diagnostic Testing	Prospective, blind comparison to Gold Standard
Prognosis	Cohort Study>Case Control >Case Series
Therapy	RCT is really the only way we want to answer this question
Etiology / Harm	Cohort Study>Case Control>Case Series
Prevention	RCT>Cohort Study>Case Control>Case Series
Cost	Economic Analysis

Synthesis articles may help you find information for any question:

- Systematic Review / Meta-analysis
- Practice Guideline

Prioritizing your questions:

1. Which question is more important to my patient's well being?
2. Which question is most interesting?
3. Which question is most feasible to answer within the time frame available?
4. Which question are you likely to encounter again in your practice?

Search Strategy:

The following is a guide to help you organize a *Boolean* search strategy. The approach should be used in planning your literature search on the Databases.

STEP 1: **STATE YOUR TOPIC IN ONE OR TWO SENTENCES:**
EX: Information about the Gene that shows susceptibility to breast cancer.

STEP 2: **BREAK YOUR TOPIC INTO KEY IDEAS OR CONCEPTS**:
EX: CONCEPT 1: Breast Cancer AND CONCEPT 2: Gene

STEP 3: **CONSIDER OTHER TERMS TO DESCRIBE YOUR CONCEPTS:**
EX:CONCEPT 1: Breast Neoplasm OR Breast AND Tumor AND CONCEPT 2:Genes OR Genetics OR Hereditary OR Genic OR BRCA1

STEP 4: **COMBINE LIKE OR SYNONYMOUS TERMS USING "OR"**
EX: Gene OR Genes OR BRCA1
EX: Breast Cancer OR Breast Neoplasm OR Breast Tumor

STEP 5: **NARROW YOUR TOPIC BY USING THE "AND" OPERATOR TO COMBINE YOUR CONCEPTS:**
(EX: (Gene or Genes or Hereditary or Genic or BRCA1) AND (Breast Cancer or Breast Neoplasm or Breast Tumor)

Searching MEDLINE

- ***MeSH* (*Medical Subject Headings*)**

The MEDLINE database is indexed by professional reviewers who read each article completely, and then assign index terms to describe (as specifically as possible) the contents of the article. The terms come from a standardized list of vocabulary and definitions called *MeSH* (*Medical Subject Headings*). The idea behind *MeSH* is that all articles about the same topic are assigned the same standard subject heading, regardless of the exact words the author uses. For example, one author may use *whooping cough*, the other *pertussis*, but both articles will be indexed under *whooping cough*. The best way to search MEDLINE is to use the *MeSH* terms whenever possible.

The Ovid MEDLINE system has a unique feature called "mapping" which will automatically match your term with a list of appropriate *MeSH* term(s).

- **Explode**

The hierarchical arrangement of the *MeSH* terms, going from broad to specific, allows for a very powerful type of searching called exploding. Since each article is indexed as specifically as possible, using the *MeSH* term *Heart Failure, Congestive*, for example, would not necessarily retrieve articles indexed with the narrower term *Cardiomyopathy, Congestive*. Exploding *Heart Failure, Congestive* will retrieve all references indexed with that term and all references indexed to all the more specific terms under it. Explode will broaden your search and increase your retrieval.

Example: Tree for Heart Failure, Congestive:

Select	Subject Heading	Explode	Focus	Scope Note
☐	**Heart Diseases (3853)**	☐	☐	
☐	**Heart Failure, Congestive (5653)**	☐	☐	
☐	**Cardiomyopathy, Congestive (1693)**	☐	☐	
☐	**Dyspnea, Paroxysmal (5)**	☐	☐	
☐	**Edema, Cardiac (35)**	☐	☐	

If you check the explode box for *Heart Failure, Congestive* you will retrieve all references indexed to *Heart Failure, Congestive* as well as all references to *Cardimyopathy, Congestive*; *Dyspnea, Paroxysmal*; and *Edema, Cardiac*.

- **Focus**

Using this option will restrict your retrieval to those articles in which your subject heading is considered the focus, or major topic of the article. To produce a more comprehensive search, you should not restrict to focus. Focus is sometimes used to reduce your retrieval when you have a topic that is covered by a single *MeSH* heading which yields thousands of citations.

- **Subheadings**

Subheadings are used to more fully qualify specific aspects of a subject heading. The default is select all subheadings. However, if you are interested in just a particular aspect of a topic such as the diagnosis, you should consider using

appropriate subheadings such as "diagnosis," "pathology," "ultrasonography," or "radiography."

- **Textwords**

In this approach, you are searching for an exact word or phrase that the author has used in the title and/or abstract. You will need to search for all possible variants and synonyms of these words in order to do a comprehensive search. For example, to search for the information on *whooping cough* by textword, you would have to search for the word *pertussis* as well as the phrase *whooping cough*. Note: Textwords do not always indicate that the article is about the word or phrase.

MeSH and textwords are not mutually exclusive search methods. In fact, very effective strategies often combine these approaches. For example you might build a broad search strategy by first using *MeSH* terms and then narrow it with a specific textword search in the title and/or abstract.

- **Truncation**

Textwords can be truncated, which is a way to include word variations. To truncate, type in the first part of the textword that is the same for all variants and add a "$" to the end. For example, *hormon$* will retrieve textwords such as *hormones* and *hormonal*. (The ":" can also be used to truncate.)

- **Combining sets**

After creating sets of references that cover each concept of your topic, you may wish to combine two or more of these sets. Combining sets helps establish a relationship among the concepts:

AND = ALL OF THE SELECTED SETS/TERMS MUST appear in the same reference;

OR = ANY OF THE SELECTED SETS/TERMS MAY APPEAR in any of the references; (usually used to include similar terms or concepts)

NOT =THE SELECTED TERM DOES NOT APPEAR in the reference; (use with caution, eliminating a term may unintentionally eliminate relevant references)

The combine feature will create a new set on the main search screen.

- **Limiting**

You can limit your retrieval by checking in the appropriate boxes on the main search page:

  - Local Holdings (journals held in the Medical Center Library)
  - English language
  - Human subjects
  - Publication year(s)

There are other limits such as Age Groups and Publication Types (such as *Controlled Clinical Trial* or *Meta Analysis*) which can be selected by clicking on the LIMIT icon at the top of the Main Search Page. Consider using the EBM filter for RCT to limit your search to randomized clinical trials.

- **Saving Strategies**

You can save your search strategy so that it can be re-run at a later time. After creating your complete strategy, click on the SAVE SEARCH HISTORY button (Web version). Follow the directions and provide the information to save your search strategy.

To run a saved strategy, use the RUN SAVED SEARCH button on the Main Search Page.

Searching the World Wide Web

According to the results of a study published by Cyveillance in July 2000, the World Wide Web is estimated to contain more than two billion pages of publicly-accessible information. As if the Web's immense size weren't enough to strike fear in the heart of all but the most intrepid surfers, consider that the Web continues to grow at an exponential rate: tripling in size over the past two years, according to one estimate.

Add to this, the fact that the Web lacks the bibliographic control standards we take for granted in the print world: There is no equivalent to the ISBN to uniquely identify a document; no standard system, analogous to those developed by the Library of Congress, of cataloguing or classification; no central catalogue including the Web's holdings. In fact, many, if not most, Web documents lack even the name of the author and the date of publication.

Imagine you are searching for information in the world's largest library, where the books and journals (stripped of their covers and title pages) are shelved in no particular order, and without reference to a central catalogue. A researcher's nightmare? Without question. The World Wide Web defined? Not exactly. Instead of a central catalogue, the Web offers the choice of dozens of different search tools, each with its own database, command language, search capabilities, and method of displaying results.

Given the above, the need is clear to familiarize yourself with a variety of search tools and to develop effective search techniques, if you hope to take advantage of the resources offered by the Web without spending many fruitless hours flailing about, and eventually drowning, in a sea of irrelevant information.

SEARCH ENGINES AND SUBJECT DIRECTORIES

The two basic approaches to searching the Web are **search engines** and **subject directories**.

Search engines:

These sites allow the user to enter keywords that are run against a database (most often created automatically, by "spiders" or "robots"). Based on a combination of criteria (established by the user and/or the search engine), the search engine retrieves WWW documents from its database that match the keywords entered by the searcher. It is important to note that when you are using a search engine you are not searching the Internet "live", as it exists at this very moment. Rather, you are searching a fixed database that has been compiled some time previous to your search.

While all search engines are intended to perform the same task, each goes about this task in a different way, which leads to sometimes amazingly different results. Factors that influence results include the size of the database, the frequency of updating, and the search capabilities. Search engines also differ in their search speed, the design of the search interface, the way in which they display results, and the amount of help they offer.

In most cases, search engines are best used to locate a specific piece of information, such as a known document, an image, or a computer program, rather than a general subject.

Examples of search engines include:

- AltaVista (http://www.altavista.com)
- Excite (http://www.excite.com/search)
- alltheweb (http://www.alltheweb.com)
- Google (http://www.google.com)
- HotBot (http://hotbot.lycos.com)

Multi-threaded search engines:

The growth in the number of search engines has led to the creation of "meta" search tools, often referred to as **multi-threaded search engines**. These search engines allow the user to search multiple databases simultaneously, via a single interface. While they do not offer the same level of control over the search interface and search logic as do individual search engines, most of the multi-threaded engines are very fast. Recently, the capabilities of meta-tools have been improved to include such useful features as the ability to sort results by site, by type of resource, or by domain, the ability to select which search engines to include, and the ability to modify results. These modifications have greatly increased the effectiveness and utility of the meta-tools.

Popular multi-threaded search engines include:

- Metacrawler (http://www.metacrawler.com)
- Ixquick (http://www.ixquick.com)
- SurfWax (http://www.surfwax.com
- Dogpile (http://www.dogpile.com)
- ProFusion (http://www.profusion.com)

Subject-specific search engines:

These engines do not attempt to index the entire Web. Instead, they focus on searching for Web sites or pages within a defined subject area, geographical area, or type of resource. Because these specialized search engines aim for depth of coverage within a single area, rather than breadth of coverage across subjects, they are often able to index documents that are not included even in the largest search engine databases. For this reason, they offer a useful starting point for certain searches. The table below lists some of the subject-specific search engines by category. For a more comprehensive list of subject-specific search engines, see one of the following directories of search tools:

- Beaucoup! (http://www.beaucoup.com)

- Search Engine Colossus (http://www.searchenginecolossus.com)
- Searchengines.com (http://www.searchengines.com/)

Table of selected subject-specific search engines:

Images	**Software**
• The Amazing Picture Machine (http://www.ncrtec.org/picture.htm) • Lycos Multimedia (http://multimedia.lycos.com/) • WebSEEk (http://www.ctr.columbia.edu/webseek/) • Yahoo! Picture Gallery (http://gallery.yahoo.com/)	• Jumbo (http://www.jumbo.com) • Shareware.com (http://shareware.cnet.com/) • ZDNet Downloads (http://www.zdnet.com/downloads/)
Health/Medicine	**Education/Children's Sites**
• Achoo (http://www.achoo.com/) • BioMedNet (http://www.bmn.com/) • Combined Health Information Database (http://chid.nih.gov/) • MayoClinic.com (http://www.mayohealth.org/) • MEDLINEplus (http://www.nlm.nih.gov/medlineplus/)	• AOL NetFind Kids Only (http://www.aol.com/netfind/kids/) • Blue Web'n (http://www.kn.pacbell.com/wired/bluewebn/) • Education World (http://www.education-world.com/) • Kids Domain (http://www.kidsdomain.com) • KidsClick! (http://www.kidsclick.org/) • Yahooligans! (http://www.yahooligans.com)

Subject directories:

Are hierarchically organized indexes of subject categories that allow the Web searcher to browse through lists of Web sites by subject in search of relevant information. They are compiled and maintained by humans and many include a search engine for searching their own database.

Subject directory databases tend to be smaller than those of the search engines, which means that result lists tend to be smaller as well. However, there are other differences between search engines and subject directories that can lead to the latter producing more relevant results. For example, while a search engine typically indexes every page of a given Web site, a subject directory is more likely to provide a link only to the site's home page. Furthermore, because their maintenance includes human intervention, subject directories greatly reduce the probability of retrieving results out of context.

Because subject directories are arranged by category and because they usually return links to the top level of a web site rather than to individual pages, they lend themselves best to searching for information about a general subject, rather than for a specific piece of information.

Examples of subject directories include:

- LookSmart (http://www.looksmart.com)
- Open Directory (http://dmoz.org)
- Yahoo (http://www.yahoo.com)

Specialized subject directories:

Due to the Web's immense size and constant transformation, keeping up with important sites in all subject areas is humanly impossible. Therefore, a guide compiled by a subject specialist to important resources in his or her area of expertise is more likely than a general subject directory to produce relevant information and is usually more comprehensive than a general guide. Such guides exist for virtually every topic.

Just as multi-threaded search engines attempt to provide simultaneous access to a number of different search engines, some web sites act as collections or clearinghouses of specialized subject directories. Many of these sites offer reviews and annotations of the subject directories included and most work on the principle of allowing subject experts to maintain the individual subject directories. Some clearinghouses maintain the specialized guides on their own web site while others link to guides located at various remote sites.

Examples of clearinghouses include:

- Argus Clearinghouse (http://www.clearinghouse.net)
- About.com (http://about.com)
- WWW Virtual Library (http://www.vlib.org)

Complete Collection of EBM websites

Agency for Healthcare Research and Quality--Clinical Information

Directories of EBM-pertinent information

Access is freely available to all.

Alberta Clinical Practice Guidelines

Guidelines from the Alberta Medical Association

Access is freely available to all.

CDC Recommends: The Prevention Guidelines System

Full set of CDC's guidelines

Access is freely available to all.

CenterWatch Clinical Trials Listing Service

Industry-sponsored clinical trials that are actively recruiting patients

Access is freely available to all.

Centre for Evidence-Based Medicine

EBM tools from Oxford University

Access is freely available to all.

Clinical Alerts

Updates of significant findings from current NIH-funded clinical trials

Access is freely available to all.

Clinical Trials.gov

Current information about NIH-sponsored clinical trials. Access is freely available to all.

CMA Infobase

Clinical practice guidelines produced or endorsed in Canada. Access is freely available to all.

Cochrane Database of Systematic Reviews (Abstracts)

Abstracts of Cochrane Collaboration systematic reviews

Access is freely available to all.

Consensus Statements: NIH Consensus Development Program

Evidence-based assessments of medical practice and state-of-the-science. Access is freely available to all.

Critically Appraised Topics

Pediatric topics from the University of Rochester Medical Center. Access is freely available to all.

DARE Metasearch

Allows simultaneous searching of three EBM databases. Access is freely available to all.

Database of Abstracts of Reviews of Effectiveness (DARE)

Abstracts and commentary from UK's National Health Service Centre for Reviews and Dissemination. Access is freely available to all.

Evidence Based Medicine Tool Kit

EBM tools arranged by topic from the University of Alberta. Access is freely available to all.

Evidence Centre Reports from Monash University

CATs in alphabetical order from Monash University. Access is freely available to all.

Evidence-Based Pediatrics

Pediatric topics from the University of Michigan Health System. Access is freely available to all.

Free Online Journals - Lamar Soutter Library

Access is freely available to all.

Free Online Textbooks - Lamar Soutter Library

Perform a Keyword search on the word **freely.** Access is freely available to all.

Full Critical Appraisal Index

From the College of Family Physicians of Canada. Access is freely available to all.

Guide to Clinical Preventive Services, Second Edition

Established reference source for clinicians. Access is freely available to all.

National Guideline Clearinghouse

Extensive collection of evidence-based guidelines. Access is freely available to all.

Netting the Evidence

EBM tools from the University of Sheffield. Access is freely available to all.

PubMed (MEDLINE)

Access to over eleven million MEDLINE records dating back to the mid-1960s. Access is freely available to all.

PubMed Clinical Queries

Click **Clinical Queries** --> **Systematic Reviews** for automatic, targeted searches. Access is freely available to all.

SUMSearch

Access is freely available to all.

Translating Research into Practice (TRIP)

Access is freely available to all.

ACP Journal Club

Abstracts and commentary from the American College of Physicians Access is restricted.

ACP Journal Club (Ovid)

Abstracts and commentary from the American College of Physicians Access is restricted.

MICROMEDEX®

Drug-related information: toxicology, generics, interactions, dosing tools, and patient handouts. Access is restricted.

MD Consult

Access to evidence-based information in various formats. Access is restricted.

MEDLINE (Ovid)

Ovid interface to MEDLINE. Access is restricted.

Online Textbooks - Lamar Soutter Library

Access is restricted.

Online Journals - Lamar Soutter Library

Access is restricted.

Database of Abstracts of Reviews of Effectiveness (Ovid)

Abstracts and commentary from UK's National Health Service Centre for Reviews and Dissemination. Access is restricted.

EBM Reviews Full Text

Searches all Ovid's full-text EBM databases. Access is restricted.

Cochrane Controlled Trials Register (Ovid)

Citations and abstracts of clinical trials, some not listed in MEDLINE Access is restricted.

Cochrane Database of Systematic Reviews (Ovid)

Full text of the Cochrane Collaboration systematic reviews. Access is restricted.

Cochrane Library

Complete collection of Cochrane databases. Access is restricted.

Module IV

Appraising Literature

1. Diagnosis

CONTENTS:

- **Objectives of the module.**
- **The rationale for evidence based diagnosis.**
- **Evaluating an article about diagnosis.**
- **Sensitivity and specificity.**
- **Predictive values.**
- **Likelihood ratios.**
- **Guidelines for interpreting likelihood ratios.**
- **Quiz**
- **Worksheet**

Objectives

At the end of this module, the participant will be able to:

- List the different methods of reasoning used by physicians during the process of diagnosis.
- Identify the steps of evaluating literature dealing with diagnostic techniques.
- Calculate sensitivity and specificity, positive and negative predictive values and likelihood ratios for different diagnostic tests.
- Use the sensitivity and specificity, positive and negative predictive values and likelihood ratios in clinical decisions.
- Distinguish between tests that rule-in and rule-out disease.
- Appraise literature dealing with diagnostic tests using the worksheet.

The rationale for evidence-based diagnosis

Diagnosis can be defined as the process of using the history, physical examination, laboratory, imaging studies, and other tests to identify the disease responsible for the patient's complaint. The advantage of knowing the responsible disease, and assigning a label to it, is that we can now make an informed decision about treatment and give our patients more accurate information about prognosis.

There are several types of reasoning used by physicians during the diagnostic process. They are:

- Algorithmic : using flowcharts and algorithms
- Pattern-recognition :"instant recognition" of a disease
- Exhaustive : gathering every possible piece of data
- Hypothetic-deductive : generating and rejecting hypotheses as more data are collected

Algorithmic approach is useful for conditions where the information we collect from patients is discrete and accurate, such as diagnosis of anemia or an abnormal liver function test.

Pattern-recognition is commonly used by primary care physicians for the diagnosis of common conditions such as urinary tract infection or sinusitis, and is a very efficient style of diagnosis.

The Exhaustive Approach is appropriate for unusual presentations of illness, where the other modes of decision-making have failed. It is commonly used at tertiary and quaternary care centers, where patients have already received a basic evaluation. However, it is generally inappropriate and inefficient in the primary care setting for an initial evaluation.

The Hypothetic-Deductive Style involves proposing a differential diagnosis, asking a question, using the answer to refine the differential diagnosis, asking another question, again refining the differential, and so on until a final working diagnosis is obtained. It is generally held up as the ideal, although each of the first three approaches is perfectly appropriate under certain conditions.

For each of these styles of diagnosis, it is important to accurately use the information which we gather from patients. That means understanding how much each symptom, sign, or test result increases or decreases the likelihood of a given disease. This process is called "revising the probability of disease", and unfortunately, it's something that physicians aren't very good at.

Consider the following example:

Your 45 year old patient has a mammogram. The study is interpreted as "suspicious for malignancy" by your radiologist. Your patient asks you: "Does this mean I have cancer?", and you (correctly) answer "No, we have to do further testing." Your patient then asks, "OK, I understand that the mammogram isn't the final answer, but given what we know now, what are the chances that I have breast cancer?". Assume that the overall risk of breast cancer in any 45 year old woman, regardless of mammogram result, is 1%. Assume also that mammography is 90% sensitive and 95% specific. Then, select your answer below:

1% **15%** **60%** **85%** **95%**

So, how can we make our use of diagnostic test results more rational and accurate? The answer lies in understanding concepts like sensitivity, specificity, predictive value, and likelihood ratios.

Evaluating an article about diagnosis

Articles about diagnosis, like any study, can be of varying quality. Evaluating the quality of an article involves first determining whether the results are relevant to your practice, then whether they are valid. These terms will be explained in the following section.

I. Relevance

First, the test should be one that is feasible for you in your community. PET scans may give us useful information about central nervous system disease, but aren't a practical option some communities. Similarly, while brain biopsy is an accurate test for diagnosing dementia, it's not practical for living patients!

Second, the population should be reasonably similar to those seen in your practice or community. A study which only includes very sick patients (who have more dramatic findings) will make a test look better than it is. For example, a stress thallium is more likely to be abnormal in patients with severe coronary artery disease than in patients with mild disease.

Applying the test only to patients with disease and healthy controls can also make the test look better than it is, because healthy controls are very unlikely to have an abnormal test result.

Ideally, a study should include:

- Patients with and without disease, selected because they have symptoms that would ordinarily cause you to consider ordering the test, and with a spectrum of severity from mild to advanced disease.
- Patients with diseases that have similar manifestations; for example, a study of MI diagnosis should include patients who turned out to have esophageal spasm, gall bladder disease, dyspepsia, and panic disorder.

II. Validity

After determination that the test and the population studied are relevant to your practice, you have to assess the validity of the information. There are only a few key indicators of diagnostic study quality.

First, did the authors use a reasonable reference standard? For example, in a study of the diagnosis of streptococcal pharyngitis, while ASO titers might be the ideal reference standard, for practical purposes we would probably have to accept throat culture.

In some cases we are limited by ethical considerations. For example, doing an invasive test such as a biopsy in a patient with a negative test result, for the sole purposes of doing a study, might be considered unethical.

Let's consider an example:

In our study of ultrasound in appendicitis, we decide to enroll all patients with abdominal pain and suspected appendicitis who present to our community hospital. We perform the ultrasound, but need a reference standard to definitively diagnose appendicitis.

Unfortunately, we doubt that we will be able to convince either the surgeons in our hospital or their patients to operate on every patient with suspected appendicitis, so we can use pathology of the removed appendix as our reference standard. We therefore decide on the following (reasonable) compromise. Surgeons will not be told the results of the ultrasound, so it doesn't bias their decision of whether or not to operate. For patients who go to surgery, pathology will be the reference standard. Patients who don't go to surgery will be followed for 30 days, to make sure their symptoms resolve or an alternative diagnosis is made. If they do not undergo appendectomy in the following 30 days, we assume that they do not have appendicitis.

Second, once you've determined that the authors used a reasonable reference standard, check to see whether the examiners doing the test were "blinded" to the reference standard result.

Third, the decision to perform the reference standard should ideally be independent of the results of the test being studied. As we already learned above, though, that isn't possible. If the reference standard is not applied independently of the test being studied, at least make sure that the authors follow up patients who don't get it, and have a good justification for designing their study this way.

Fourth, data should be collected prospectively whenever possible. This is especially true when studying the history and physical examination, since we all know how inaccurate the medical record is as a source of information about what physicians actually did. For example, if a physician didn't record that a patient with sore throat had an exudate, does that mean they really didn't have one, or that you didn't look, or that they did but you forgot to record it?

Fifth, the diagnostic test should be applied to a reasonable number (at least 100 is a good rule of thumb) of patients with an appropriate "spectrum" of disease. That is, you want patients with mild, moderate, and severe disease, as well as patients with similar symptoms who don't have the disease in question.

III. Summary points

- New tests should be validated by comparison against an established gold standard in an appropriate spectrum of subjects
- Diagnostic tests are seldom 100% accurate (false positives and false negatives will occur)
- A test is valid if it detects most people with the target disorder (high sensitivity) and excludes most people without the disorder (high specificity), and if a positive test usually indicates that the disorder is present (high positive predictive value)

- The best measure of the usefulness of a test is probably the likelihood ratio—how much more likely a positive test is to be found in someone with, as opposed to without, the disorder.

Sensitivity and specificity

Sensitivity and specificity are the most widely statistic used to describe a diagnostic test. Unfortunately, as we learned from the example of interpreting a mammogram above, they are not very helpful to clinicians trying to revise the probability of disease. Reviewing the definitions of prevalence, sensitivity, and specificity will help us understand why:

- **Prevalence =** probability of disease in the entire population at any point in time (i.e. 2% the population has diabetes mellitus)
- **Incidence =** probability that a person without disease develops the disease during an interval (the incidence of diabetes mellitus is 0.2% per year, referring only to new cases)
- **Sensitivity =** probability of a positive test among patients with disease
- **Specificity =** probability of a negative test among persons without disease

As clinicians, though, we don't generally know whether or not the person has disease; that's why we're ordering the test in the first place. Thus, sensitivity and specificity do not give us the information we need to interpret the test results.

What do we want to know? Ideally, we'd like to know what the probability of disease is given a positive or negative test.

Bayes first described a way to do this in the late 1800's, when he developed an equation to relate the probability of disease before ordering the test (the pre-test probability) and the probability of disease given a positive or negative test (the post-test probability). Bayes' equation is shown below:

Probability of disease given a positive test =

$$\frac{\text{(prevalence x sensitivity)}}{\text{((prevalence x sensitivity) + ((1-prevalence) x (1-specificity)))}}$$

Example:

Take a look at the "2 x 2" table below:

	Patients with disease	**Patients without disease**
Test is positive	A	b
Test is negative	C	d

Using the table, the definitions of sensitivity and specificity can be written as:

sensitivity = a / (a+c)

specificity = d / (b+d)

Occasionally, sensitivity and specificity are helpful, a very high sensitivity, when negative, rules out disease. For example, consider the complaint of "dyspnea on exertion" in the diagnosis of congestive heart failure:

	CHF	**no CHF**
Dyspnea on exertion	41	183
No DOE	0	35

The sensitivity of dyspnea on exertion for the diagnosis of CHF is 100% (41/(41+0)), and the specificity 17% (35/(183+35)). If negative (the patient does not complain of dyspnea on exertion), it is very unlikely that they have CHF (0 out of 41 patients with CHF did not have this symptom).

An easy way to remember this rule of thumb is the acronym "**SnNOut**", which is taken from the phrase: "**Sensitive test when Negative rules Out disease**".

Conversely, a very specific test, when positive, rules in disease. Not surprisingly, the acronym for this kind of test is "**SpPIn**". Consider a gallop (S3)

murmur in the diagnosis of congestive heart failure, with data taken from the same study:

	CHF	no CHF
Gallop (S3) murmur	10	3
No gallop murmur	31	215

The sensitivity of gallop for CHF is only 24% (10/41), but the specificity is 99% (215/218). Thus, if a patient has a gallop murmur, they probably have CHF (10 out of 13).

EXAMPLE:

	CHF	no CHF
Displaced apex and JVD, gallop, rales, or edema	18	3
No displaced apex, or displaced apex alone	23	215

So, the sensitivity is:

1% 10% 44% 56% 86% 99%

And the specificity is:

1% 10% 44% 56% 86% 99%

Thus, another SpPIn: patients with a displaced apex and either JVD, gallop, rales, or edema were very likely to have CHF (18 out of 21 or 86%).

Thus, sensitivity and specificity by themselves are only useful when either is very high (over typically, 95% or higher). There is a useful Web site with a collection of SpPIns and SnNOuts. In the next section, we'll learn about predictive values, which are more useful to clinicians than sensitivity and specificity.

Predictive values

Predictive values help us answer the question:

"Given a positive (or negative) test result, what is the new probability of disease?"

Let's fill some numbers in. In his article on the clinical diagnosis of strep, Frank Dobbs took consecutive patients, prospectively did the same history and physical exam maneuvers on all of them, and then did a throat culture.

One of the things our nurses ask patients who call with a sore throat is how long they've had it. If it's only been a couple of days, they are more likely to ask patients to try symptomatic remedies. If the duration is longer, they are more likely to ask them to come in. The presence of fever is another important factor in advising the patients, with febrile patients more likely to be asked to come in for evaluation. Is there any evidence to support these strategies?

Essentially we are asking, "If a patient has fever, what is the likelihood of strep pharyngitis?" and "If a patient has symptoms for 3 or more days, what is the likelihood of strep pharyngitis?" Let's start by creating 2 x 2 tables for each question, using data from Dobbs' article:

	Strep	**No strep**			**Strep**	**No strep**	
Fever	58	80	138	**Duration >= 3 days**	16	62	78
No fever	14	54	68	**Duration < 3 days**	56	72	128
	72	134			72	134	

Using the equations for sensitivity and specificity, we find that for fever:

sensitivity = 58 / (58 + 14) = 0.81

specificity = 54 / (54 + 80) = 0.40

Note that the sensitivity can be written as "0.81" or "81%". One is no better than the other - just be consistent! Similarly, for duration of symptoms >= 3 days,

sensitivity = 16 / (56 + 16) = 0.22

specificity = 72 / (72 + 62) = 0.55

At this point, you should be getting a little uneasy about your triage policy. While most patients with strep had fever, only 22% with symptoms for more than 3 days had the diagnosis.

We still haven't answered our question, though. To do that, we have to calculate predictive values. They are defined as:

Positive predictive value = probability of disease among patients with a positive test

Negative predictive value = probability of no disease among patients with a negative test

- The probability of disease given a positive test can therefore be called the "post-test probability of disease given a positive test", the "positive predictive value", or the "posterior probability of disease given a positive test". These names are interchangeable.
- The probability of disease given a negative test is called the "post-test probability of disease given a negative test" or the "posterior probability of disease given a negative test"; this is equal to one minus the negative predictive value. Note this last point: the negative predictive value does not equal the post-test probability of disease given a negative test. They are the converse of one and other.

	Patients with disease	Patients without disease
Test is positive	**A**	**B**
Test is negative	**C**	**D**

We can now define positive and negative predictive value as follows:

Positive predictive value = a / (a+b)

Negative predictive value = d / (c+d)

Post-test probability of disease given a positive test = a / (a+b)

Post-test probability of disease given a negative test = c / (c+d)

Notice that we are now using the rows instead of columns, as for sensitivity and specificity. What about our original question on the diagnosis of strep throat?

	Strep	No strep			Strep	No strep	
Fever	58	80	138	**Duration >= 3 days**	16	62	78
No fever	14	54	68	**Duration < 3 days**	56	72	128
	72	134			72	134	

We can quickly calculate that for fever:

Positive predictive value = 58 / (58+80) = 0.42

Negative predictive value = 54 / (54+14) = 0.79

And for duration of symptoms of 3 or more days:

Positive predictive value = 16 / (16+62) = 0.20

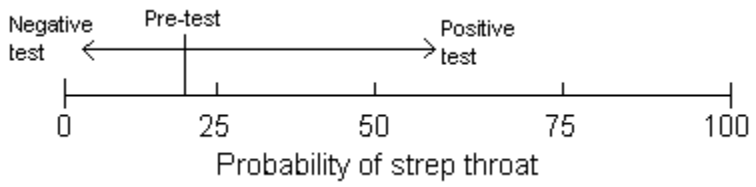
Negative predictive value = 72 / (72+56) = 0.56

If a patient has fever, there is a 42% chance of strep, and if they have symptoms for 3 or more days, only a 20% chance. It appears that it may be appropriate to revise our triage policy.

Likelihood ratios

When we decide to order a diagnostic test, we want to know which test (or tests) will best help us rule-in or rule-out disease in our patient. In the language of clinical epidemiology, we take our initial assessment of the likelihood of disease ("pre-test probability"), do a test to help us shift our suspicion one way or the other, and then determine a final assessment of the likelihood of disease ("post-test probability").

Take a look at the diagram below, which graphically illustrates this process of "revising the probability of disease".



Likelihood ratios tell us how <u>much</u> we should shift our suspicion for a particular test result. Because tests can be positive or negative, there are at least two likelihood ratios for each test. The "positive likelihood ratio" (LR+) tells us how much to increase the probability of disease if the test is positive, while the "negative likelihood ratio" (LR-) tells us how much to decrease it if the test is negative. The formula for calculating the likelihood ratio is:

$$\textbf{LR} = \frac{\textbf{probability of an individual with the condition having the test result}}{\textbf{probability of an individual without the condition having the test result}}$$

Thus, the positive and negative likelihood ratios are:

$$\textbf{LR+} = \frac{\text{probability of an individual } \textbf{with} \text{ the condition having } \textbf{positive} \text{ test result}}{\text{probability of an individual } \textbf{without} \text{ the condition having } \textbf{positive} \text{ test result}}$$

$$\textbf{LR - =} \frac{\text{probability of an individual } \textbf{with} \text{ the condition having } \textbf{negative} \text{ test result}}{\text{probability of an individual } \textbf{without} \text{ the condition having } \textbf{negative} \text{ test result}}$$

You can also define the LR+ and LR- in terms of sensitivity and specificity:

LR+ = sensitivity / (1-specificity)

LR- = (1-sensitivity) / specificity

Example:

In a study of the ability of rapid antigen tests to diagnose strep pharyngitis, 90% of patients with strep pharyngitis have a positive rapid antigen test, while only 5% of those without strep pharyngitis have a positive test. The LR+ for the ability of rapid antigen tests to diagnose strep pharyngitis is:

LR+ = 90% / (100%-95%) = 90% / 5% = 18

Don't get too caught up in the calculations. The important thing is to understand the meaning of a likelihood ratio. They have unique properties that make them particularly relevant to clinicians:

- The LR+ corresponds to the clinical concept of "ruling-in disease"
- The LR- corresponds to the clinical concept of "ruling-out disease"
- The LR+ and LR- don't change as the underlying probability of disease changes (predictive values do change)
- LR's using multiple "levels" of positive (i.e. not just a simple yes/no or positive/negative result) provide much richer, more useful information to the clinician.

Guidelines for interpreting likelihood ratios

The first thing to realize about LR's is that an LR > 1 indicates an increased probability that the target disorder is present, and an LR < 1 indicates a decreased probability that the target disorder is present. The following are general guidelines, which must be correlated with the clinical scenario:

LR	Interpretation
> 10	Large and often conclusive increase in the likelihood of disease
5 - 10	Moderate increase in the likelihood of disease
2 – 5	Small increase in the likelihood of disease
1 – 2	Minimal increase in the likelihood of disease
1	No change in the likelihood of disease
0.5 – 1.0	Minimal decrease in the likelihood of disease
0.2 – 0.5	Small decrease in the likelihood of disease
0.1 – 0.2	Moderate decrease in the likelihood of disease
< 0.1	Large and often conclusive decrease in the likelihood of disease

The decision to order a test is also based on our initial assessment of the likelihood of the target disorder, and how important it is to rule-in or rule-out disease. For example, a chest x-ray might have a good likelihood ratio for pneumonia. But if you believe a patient has a simple cold, this test, no matter how good the LR, probably shouldn't be ordered. It is sometimes helpful to be able to calculate the exact probability of disease given a positive or negative test. We saw that this is next to impossible using sensitivity and specificity at the bedside (unless you can do Bayes' Theorem in your head!). Next, we'll learn how we can use likelihood ratios to quickly estimate the probability of disease in our patients.

Quiz

Question 1.

A very sensitive test, when negative, helps you:

a. Rule-in disease
b. Rule-out disease
c. Confuse medical students
d. Save money

Question 2.

A test which is highly specific, when positive, helps you:

a. Rule-in disease

b. Rule-out disease

c. Confuse medical students

d. Save money

Question 3.

A likelihood ratio of 12.3 provides:

a. Strong evidence to rule-in disease

b. Weak evidence to rule-in disease

c. Weak evidence to rule-out disease

d. Strong evidence to rule-out disease

Worksheet for evaluating an article about diagnosis

Assessing relevance

- The test in question is feasible in my community and practice:

  o Yes **o No**

- Patients are similar to those I see in my practice, and included patients with varying severity of disease as well as other patients with similar presenting symptoms but no disease:

  o Yes **o No**

Assessing validity

- A reasonable reference standard was used:

  o Yes **o No**

- Results of the reference standard were not known to the examiner:

  o Yes **o No**

- Data were collected prospectively:

  o Yes **o No**

- All patients received both clinical examination and reference standard:

  o Yes **o No**

- If no, were patients who didn't get the reference standard followed up closely enough and long enough to convince you that they are free of disease? **o Yes** **o No**

Usefulness

Fill in the 2 x 2 table for the diagnostic test:

	Disease	**No Disease**
_____ **positive**	A	b
_____ **negative**	C	d

sensitivity = a / (a + c) = _____

specificity = d / (b+d) = _____

PV+ = a / (a + b) = _____

PV- = d / (c + d) = _____

LR+ = sens / (1-spec) = _____

LR- = (1-sens) / spec = _____

	Disease	**No Disease**
_____ **positive**	a	b
_____ **negative**	c	d

sensitivity = a / (a + c) = _____

specificity = d / (b+d) = _____

PV+ = a / (a + b) = _____

PV- = d / (c + d) = _____

LR+ = sens / (1-spec) = _____

LR- = (1-sens) / spec = _____

Module V

Appraising Literature

2. Therapy

CONTENTS:

- **Objectives of the module.**
- **Introduction**
- **Critically reading a paper on therapy.**
    - **Evaluating Relevance**
    - **Evaluating Validity**
    - **What are the results**
- **Homework**
- **Summary**
- **Quiz**
- **Worksheet for evaluating an article about therapy**

Objectives

At the end of this module, the participant will be able to:

- Identify pitfalls in common approaches to clinical decision-making.
- Determine if a study about therapy is relevant and valid.
- Analyze study results on an “intention to treat" basis.
- Distinguish between “relative risk reduction” and “absolute risk reduction”.
- Calculate the “number needed to treat” (NNT) in a study.

Required Materials

- A copy of the following paper: SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: final results of the Systolic Hypertension in the Elderly Program (SHEP). JAMA 1991 Jun 26; 265(24): 3255-3264.
- A calculator.

Introduction

Clinicians make therapeutic decisions all the time; many of them are made with little thought. And though the same decisions are made all the time, when confronted with difficult questions, we usually ask for advice from a colleague or from a consultant.

Evidence-based medicine as a new paradigm, does not accept the above-mentioned approaches to therapeutic decision-making. The following is a review of common approaches to therapeutic decision-making and their pitfalls.

Pathophysiologic approach

This is a reductionism approach: if we understand the parts, we know about the whole. In theory, this makes sense. In the absence of evidence based on true patient-oriented outcomes, this may be all we have. Many examples of where this approach has led to dangerous outcomes.

- In the 1980s, ventricular arrhythmia was the most common cause of death among patients with acute myocardial infarction. Based on this, patients were routinely started on lidocaine and if they continued to have lots of premature ventricular contractions, medications such as tocainide, encanide, or flecanide were started. It was later found out that these medications **increased** mortality.
- Similarly, while reducing cholesterol should improve mortality, studies have shown that Lopid may **increase** overall mortality.
- Two recent randomized trials have shown that calcium channel blockers, which do a fine job of lowering blood pressure, also **increase** the risk of MI compared with ACE inhibitors.

Human beings are complex, non-linear beings, and simple, linear reasoning doesn't always (or even often) work.

Experience

This is a major component of the "art" of medicine. Generally, the experience of the clinician is an important and underappreciated part of practice. However, we have to be careful. Let's look at a couple elements that affect our experience.

We observe a number of children with bronchitis who failed to respond to amoxicillin, but after putting them on azithromycin, they improve. Because of this experience, we decide to start treating all our children with bronchitis with azithromycin. There are at LEAST two explanations for this phenomenon. Unless we systematically follow-up all our patients, we forget how many children treated with amoxicillin did not come back because they got better. Additionally, bronchitis is generally a self-limited illness that gets better regardless of how it's treated. The subsequent improvement with the second antibiotic (ANY antibiotic) may merely reflect the natural history of bronchitis!

Expert opinion

We generally use consultants for a few limited things: technical expertise (a procedure) or advice on diagnosis or treatment. How many times have the recommendations of a consultant left you scratching your head? Quite often we choose to ignore their suggestions, most often because we believe that the consultant's perspective is inconsistent with the needs of the patient, their recommendations are not feasible, or the diagnosis is such a zebra that it makes no sense to pursue. Many of our consultants live in a world very different from our own. Additionally, some of our consultants may not be up to date on the literature in their own field.

Enterprise

In the good old days of medicine, we all knew of physicians who did tests or recommended procedures because they needed to make a boat payment.

Critically reading a paper on therapy

Being able to critically read an article on therapy puts the power of decision making back in your hands, freeing you from an over-reliance on "experts".

Reading a paper evaluating a therapeutic intervention requires addressing the same three basic issues as in other studies: relevance, validity and results.

Evaluating relevance

Evaluating the relevance of a study is the first step in critically reading a paper. This directly relates to whether you can (or should) apply the results of the paper to the care of your patients.

- The first question you should ask is whether the treatment is feasible. If you can't do it or refer a patient for it, the treatment is irrelevant to your practice, and you can ignore the paper.
- The study should be generalizeable to your patients. A common criticism of clinical trials is they include a narrow spectrum of disease and patients.
- You should also see if the authors included all the important clinical outcomes. Some exceptions may be needed , for instance, a study of a drug given to twenty year olds to prevent cardiovascular disease might need several decades of follow up before deaths are likely to occur.
- The author should also provide enough information so you can weigh the benefits of the treatment against its risks.

Evaluating validity

Before we get to search the literature, we probably need to talk about one or two issues related to therapy. Remember that we are looking for the best

available evidence to guide us. In other words we want to find articles that are most likely going to provide us with valid results. The most valid study designs in considering a therapy are randomized clinical trials (RCTs). What elements of RCTs make them more valid than other study designs?

Random Assignment

Are subjects randomly assigned to the different treatment groups? With random assignment, we have the best opportunity to make sure that potentially important factors like age, sex, co-morbid conditions, etc. are equally distributed among the treatment groups. It also eliminates another important source of bias: the investigator's enthusiasm for the treatment. For example, if we think that gorillacillin is better than cephakillitall in treating otitis media in children, we could be influenced by this and assign more severe cases or kids with recurrent otitis media to the cephakillitall group and the milder ones to the gorillacillin group. This is called of selection bias.

Unfortunately, sometimes randomization doesn't work perfectly and our treatment groups may differ slightly. One group might be older, have fewer minority subjects, or have more smokers. The larger the study the less likely this is going to happen. The researcher (and the reader) have to decide if these differences are important and how to interpret the results in light of these differences. In most papers, Table 1 is where you'll find a comparison of baseline characteristics of the study groups.

When no randomized studies exist, you have to resort to other study designs and make the best of things. After all, you still have to make a decision!

Blinding

Do the subjects, researchers, and data analyzers know which treatment was given to a particular individual? This can be done in other study designs and isn't unique to RCTs, but it is a crucial element. In our otitis media example, you could imagine that if you had a personal stake in a particular

treatment, if you knew which treatment was given, your judgment of improvement might be influenced. With some outcomes, blinding is not very critical.

In a clinical trial, we compare the effects of two or more treatments. Another aspect of blinding is the nature of the comparison treatment. When a new drug is evaluated, we usually have a placebo that is identical to the real drug except it lacks the active ingredient. Once you have an effective therapy, it may be considered unethical to use a placebo in subsequent studies. In this instance, a researcher will compare a new drug to the old one. The two drugs should again look identical. Sometimes you have a situation where blinding is tough.

Some solutions to these dilemmas include the use of objective and reproducible outcome measures (like death, cholesterol levels, etc.) and having a separate data collector who is blind to the treatment group. In other words, the members of the research team who do the outcomes assessments should be blind to group assignment, thereby reducing systematic bias.

Follow up

Were all the study subjects accounted for at the end of the study? Again, this is not a unique issue for RCTs. You should look for important differences in drop-out rates. Additionally, you should see how the author handles them. In the most conservative approach, you would assign the worst outcome to the drop-outs from the intervention group and the best outcome to the drop-outs from the control group. This way if the experimental treatment is still beneficial, it is in spite of these potentially bad outcomes.

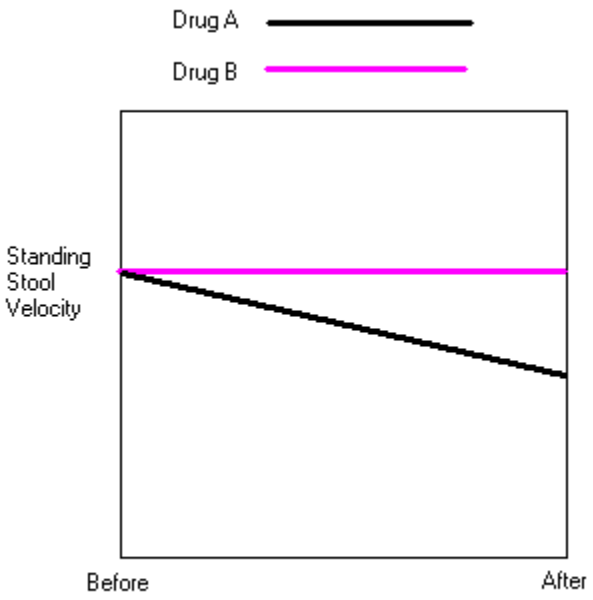
If the dropout rate is similar between intervention and treatment groups, this is somewhat reassuring.

Intention to Treat (also known as "analyzed as randomized")

Were the patients analyzed in the groups to which they were originally assigned? This addresses what happens to subjects in a study. Some

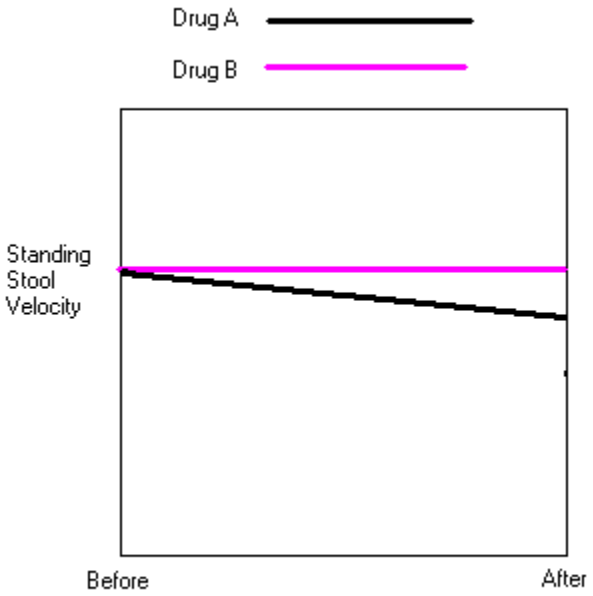
subjects might drop out, have a change of therapy, move out of town, get mad and leave the study, or even die.

The "drop outs" are dealt with above, but what about the "crossovers"? If we have a study comparing medical therapy and surgery, what happens if a subject starts with medication and at a later point in the study gets surgery? To minimize the possibility of bias in favor of either treatment, researchers will analyze subjects based upon their original treatment assignment **regardless of what happens afterwards.** I know, this sounds crazy, but lets look at the ramifications of this. Let's say that the truth is that drug A lowers standing stool velocity and drug B has no effect. As researchers, we don't know this yet (why else would we do the study?), so we set up a randomized double blinded study comparing drug A and drug B. A perfect study will show the following:

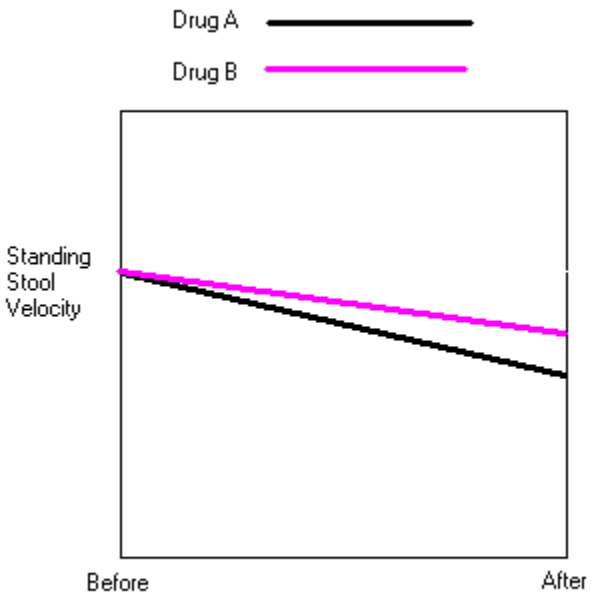


What happens if a number of subjects receiving drug A stop taking it because of side effects, and instead begin taking drug B? At the end of the study, they will have higher standing stool velocities (because, remember, drug B is

ineffective). In the "intention to treat" approach, it would look like drug A is less effective than it really is:



OK, so now what would happen if patients discontinued drug B and started taking drug A? You got it, it would look like drug B has some effect on standing stool velocity (when it really doesn't):



One last possibility to consider? You guessed it, what happens if there is some crossover between both groups? It would look like drug A is less effective than it really is and drug B has some effect when it really has none. Depending on how many cross over, it is possible that

- drug A may look less effective than drug B
- there may be no difference between the two
- drug A may still look more effective than drug B

The intention to treat approach is very conservative. If you still see a difference, the result is more robust.

What are the results?

So far, we have talked about the design of a study. Clearly, this is important in deciding if the study is valid. What about the results? Research papers have two broad types of analysis: statistical and clinical. The statistical results are the p-values and confidence intervals and stuff. In one of your on campus sessions, you will have (or already have had) a series of lectures on basic statistics (mean, median, confidence intervals, p-values, statistical tests,

etc.). In this module, we will focus primarily on the clinical results and will not discuss the statistical stuff. The main reason for making this decision is twofold. The first is a pragmatic one. This module could be **very** long if we went into all the discussion about descriptive statistics and comparative statistics, hypothesis testing, and so on. The second and more important reason is that when you read a paper, the statistical stuff is rarely critical; you base your decisions on the clinical stuff. We want you to emphasize the clinical importance. A p-value only tells you the likelihood that the results were due to chance. If a study showed that drug A lowered blood pressure an average of 1 mm Hg and that this was highly statistically significant, you'd still say "So what?" because the difference is not important clinically! The literature is full of statistically significant but clinically irrelevant results.

When we read a therapeutic paper, we are usually interested in seeing if one treatment group does better than another. We are also interested in which outcomes are improved and whether patients were harmed. To make the most sense of the outcomes, we need to know how often patients experience good outcomes and bad outcomes. When we compare their frequency, we are comparing rates or risks.

Example

A researcher develops a new drug to prevent the common cold in children over the age of two. In preliminary uncontrolled studies, when the drug was given as a single dose in September, it reduced the frequency of colds for six months. In the current study, 1000 children received placebo and 1000 received the drug. During the six month follow up the researchers diagnosed colds in 650 of the kids on placebo and in 500 of the active treatment. None of the children were hospitalized for complications of the cold. Otitis media was diagnosed in 300 of the children with colds on active treatment and in 299 of the children with colds on placebo. There were no deaths. Oh yeah, the mean number of days of school missed in the placebo group was 2.7 (95% confidence interval 0.8, 4.9) and in the treatment group was 2.2 (95%

CI 0.3, 3.1). As you might have guessed, researchers have several ways of expressing the differences between groups. Before discussing them, do the following for each group:

Task: Calculate the rate of colds and the rate of ear infections for the children that got colds.

Absolute risk reduction

Calculating absolute risk reduction (ARR) is simple. It is merely the difference in the proportion of subjects with the outcome of interest in each group. If the event rate in the placebo group is X and the event rate in the treatment group is Y, the ARR is X-Y.

In a study of the management of the third stage of labor (this is the time spent waiting for the placenta after the baby is delivered), 16.5% of women managed expectantly had post partum hemorrhage compared to 6.8% in the actively managed group. This represents an ARR of (0.165-0.068) = .097 or 9.7%. In English, this means that active management of the third labor reduced the rate of postpartum hemorrhage 9.7%

Task: Calculate the ARR for colds and for otitis media from the above example.

Relative risk reduction

The most commonly used measure of risk is relative risk reduction (RRR). It is expressed as a percent difference and is calculated as 100*(X-Y)/X. In the postpartum hemorrhage study the RRR is 100*(16.5-6.8)/16.5 = 59%. In English, this means that active management of labor reduced the rate of postpartum hemorrhage by 59%.

Task: Calculate the RRR for developing colds in the hypothetical study in our example.

RRR is the most commonly reported form of risk reduction. This is probably no accident. Why do you think this is the case?

Number needed to treat

The number needed to treat (NNT) is the number of patients you need to treat to prevent one additional bad outcome. It is calculated as the reciprocal of ARR or (1/ARR). If we have to treat 1000 patients to prevent a single bad outcome, we might be less enthusiastic about the treatment, especially if we only need to treat 15 with an alternative therapy to prevent the same bad outcome (assuming relatively equal adverse effects). The NNT is also useful because if we also know the rate of adverse events, we can balance risk and benefit. If, for example, the NNT for a drug to prevent cancer is 300, but the rate of fatal pulmonary embolism is 2%, we know that for every cancer we prevent, we cause 6 fatal pulmonary emboli.

Unfortunately, the NNT is not reported as frequently as EBM proponents would like. This is slowly changing.

Task: Calculate the NNT for developing colds in our hypothetical study.

Confidence intervals

The results can tell you both the magnitude of effect and the precision. A confidence interval combines the statistical stuff and the clinical stuff to provide you with an idea of how precise the treatment effect is. Data is usually presented as a 95% confidence interval, meaning that if the study is repeated multiple times, 95% of the studies will have result within that range.

When you compare the confidence intervals of different treatment groups, look for overlap. If the confidence intervals don’t overlap, the difference is statistically significant. To see if the difference is clinically meaningful, compare the closest extremes. For instance, if the upper limit of the treatment effect for the control group is very close to the lower limit of the treatment group, it’s possible the results aren’t clinically important.

When you look at confidence intervals, see how wide they are. A narrow or tight confidence interval represents a precise estimate. These are usually found in studies with a large number of participants.

No difference

Occasionally you'll read a study that shows no difference between treatment groups. It doesn't happen often, though. These are called negative studies. You don't see as many of these because researchers are less likely to submit them for publication. When they are submitted, they take longer to get published and are less likely to get published.

A negative study manifests itself either as insignificant p-values, p-values over 0.05, or the confidence intervals overlap. How might you interpret the results of a negative study? There are only **two** possible explanations.

The first explanation is that maybe ***there really is no difference***. The second is that ***the study did not have enough power*** (not enough subjects) to find a difference if there really was one. This is called a beta-error or a type II error and the study is referred to as being underpowered. When you read a negative paper, you need to go back to the methods section. The author should discuss how they calculated the size of the study.

If the author reported a sample size estimation, it will include some assumptions about the data (remember that a sample size estimation is done **before** the study is started, so they had to make their best guess about the mean treatment effect, standard deviation, and magnitude of difference). Compare their assumptions with the actual results. How did they do? If they came pretty close, then the likelihood of a type II error is minimized (keep in mind that most studies use 80-90 percent power meaning that 10-20 percent of the time an adequately powered study may still not be able to find a difference). If not, it is still possible that the study lacked power.

Homework

You have a 72 year-old woman in your office for a routine checkup. She is generally healthy, except for some mild osteoarthritis for which she takes acetaminophen. Her blood pressure is 190/84. Your office nurse teaches her to take her own blood pressure. Her blood pressure readings correlate well with those of your nurse.

She returns two weeks later with her readings. The range is 160 to 210 systolic and 72 to 96 diastolic. Her average blood pressure is 193/88. You confirm that she is not taking any anti-inflammatory agents, decongestants or other sympathomimetic drugs.

We are faced with a number of potential clinical issues:

- Could this patient have reno-vascular hypertension?
- Does she need to be treated?
- Will she be better off being treated?

Since this is not the diagnostic module or the module on prognosis, we'll ignore the first two issues! Before we can proceed, we really need to have a better question to answer. Before moving on write down on a separate piece of paper 2 or 3 questions related to therapy for this patient scenario.

A global question, such as "How should I treat systolic hypertension in the elderly" is a good starting point but still needs more focus. A more focused question might be "Are patients better off if we treat isolated systolic hypertension?" An even more focused question might be "Is active therapy better than placebo in reducing all-cause mortality, cardiovascular mortality, and cardiovascular end points in isolated systolic hypertension (ISH) in the elderly?" This last question has four parts to it:

1. The patient and problem (ISH in an elderly woman)
2. The intervention (active therapy)

3. The comparison intervention (placebo)
4. The outcomes of interest (mortality and morbidity).

To answer this question, we need to search for it! A common source of answers comes from MEDLINE. The PUBMED website is a nice free resource

Use the following terms:

"hypertension"[MeSH Major Topic] AND ("aged"[MeSH Terms] OR "aged, 80 and over"[MeSH Terms]) AND "randomized controlled trial"[Publication Type] AND "mortality"[MeSH Terms]

You should get about 41 "hits". You can scan the titles quickly and see that only a few look particularly relevant. For this exercise, we will use the following article:

SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: final results of the Systolic Hypertension in the Elderly Program (SHEP). JAMA 1991 Jun 26;265(24):3255-3264

Get a copy of it

Print the worksheet, and use it as a guide to critically reading the paper. Once you have done that (and not before!) take a look at the critical appraisal of this article.

Be sure you calculate NNTs for the key outcomes.

A separate table lists the ARR and NNT for the major outcomes in this article.

Summary

You have just gone through critically reading a therapeutic paper. How are you going to manage your patient? You may determine that the study design is valid and the results meaningful. Using the explicit guides described above, most readers would reach the same conclusions. The issue of applicability is where the greatest variability is likely to occur.

Now that you have gone through the painful and lengthy process, you must be asking yourself, "Surely there has to be a better (less painful) way?"

There are a number of **secondary sources** where others use the same appraisal skills we described to assess the validity and summarize the important results, and you are left to determine the relevance! Some of these sources include the Cochrane Library, Best Evidence, ACP Journal Club, Evidence-Based Practice, Evidence-Based Medicine, Journal of Family Practice POEMs, etc.

If you go to the ACP Journal Club Nov-Dec 1991 issue (revised November 1996), you will see their favorable review of this paper. You will also find other synopses of clinical trials of systolic hypertension in the elderly. If you search the Cochrane library, you will also find a meta-analysis of clinical trials of treating hypertension in the elderly that reaches the following conclusion:

"Randomized controlled trials establish that treating healthy older persons with hypertension is highly efficacious. Benefits of treatment with low dose diuretics or beta-blockers are clear for persons in their 60s to 70s with either diastolic or systolic hypertension. Differential treatment effects based on general risk factors, pre-existing cardiovascular disease and competing comorbidities could not be established from the published trial data."

When you are in clinical settings, go for the quick source when it is available.

Quiz

1. The preferred approach when analyzing data is to:

a. Analyze data only for those patients who actually completed the study

b. Analyze data for all patients as randomized, whether or not they finished the study or adhered to the protocol

c. Analyze data for patients who actually completed the study AND adhered to the protocol.

2. Patients with acute MI receiving *Clotbustase* have a 9% mortality rate at 3 months, compared with a 10% mortality rate for Streptokinase. The relative risk reduction is:

a. 1%

b. 9%

c. 10%

d. 90%

Worksheet for evaluating an article about Therapy

Guide	Comments
I. Will the results help me in caring for my patients?	
1. Can the results be applied to my patient care?	
2. Were all clinically important outcomes considered?	
3. Are the likely benefits worth the potential harms and costs?	
II. Are the Results Valid?	
1. Was the assignment of patients to treatments randomized?	
2. Were all patients who entered the trial properly accounted for and attributed at its conclusion?	
a) Was follow-up complete?	
b) Were patients analyzed in the groups to which they were randomized	
3. Were patients, health workers, and study personnel "blind" to treatment?	
4. Were the groups similar at the start of the trial?	
5. Aside from the experimental intervention, were the groups treated equally?	
III. What are the Results?	
1. How large was the treatment effect?	
2. How precise was the treatment effect?	

Module VI

Appraising Literature

3. Prognosis

CONTENTS:

- **Objectives of the module.**
- **Introduction.**
- **Cohort and case control studies.**
- **Determining the relevance.**
- **Evaluating validity.**
- **Exercise.**
- **Interpreting the results of a study on prognosis**
- **Quiz.**

Objectives:

By the end of this module, the participant will be able to:

1. Formulate a precise question relevant to a prognosis study.
2. Understand what a cohort study is.
3. Understand what a case-control study is.
4. Determine if the study is relevant.
5. Determine if the study is valid.
6. Appraise an article dealing with prognosis.

Introduction to articles about prognosis

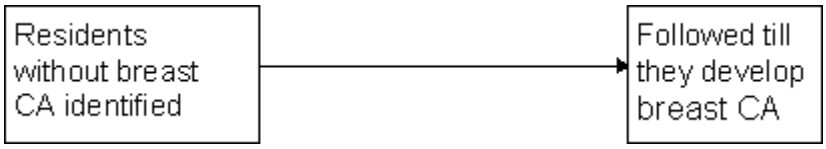
Prognosis is an important factor in diagnostic and treatment decisions. It can affect whether a test is ordered at all, the choice of treatment, and when a patient will return for follow-up. For example, knowing that the survival rate for 20 year survivors of breast cancer is the same as that of the general population would help you decide how often they need mammography and clinical breast examination. Similarly, knowing that the prognosis for men over 70 with a Gleason stage 1 or 2 prostate cancer without metastasis is the same as that for the general population would discourage you from recommending surgery or radiation for that patient.

Studies of prognosis are often (and ideally) cohort studies. In a cohort study, a large group of individuals with (and sometimes without) the condition in question are followed over time. Case-control studies are also used to identify prognostic factors, but they are subject to additional biases. In this module, we will first learn how to distinguish these two important study designs, and then move on to evaluating the relevance, validity, and results of an article on prognosis.

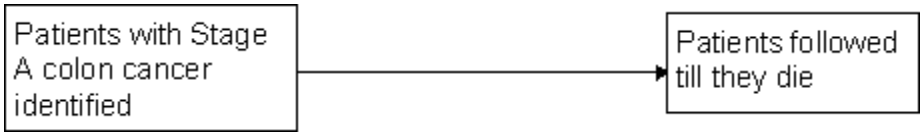
Cohort and case-control studies

A **cohort study** assembles a group of patients and follows them over time. An example is the Nurses' Health Study, in which over 20,000 nurses were identified and followed-up annually with tests and surveys for over 25 years (this study is still ongoing). These studies provide very valuable information, but are obviously very expensive and time-consuming.

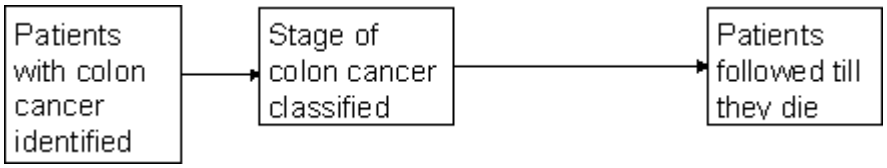
In most cohort studies, you want to assemble patients without the disease in question, and then follow them until they develop the disease. By comparing the characteristics of patients with and without disease, you can identify risk factors. A study to identify risk factors for breast cancer is diagrammed below:



In a cohort study of prognosis, you would continue to follow patients after they develop the disease in question to see how they do regarding mortality, disease progression, and other important outcomes. Therefore, cohort studies of prognosis don't necessarily begin with a group of disease-free patients. Because they don't, it is important that patients either be at the same stage of the disease at the start of the study, or that you are able to precisely characterize the stage of the disease. A study to determine the prognosis for patients with colon cancer is shown below:

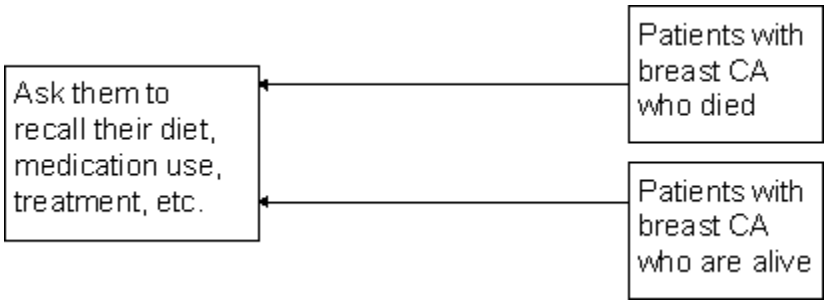


Alternately, you can enroll all patients with colon cancer, being careful to classify them at the beginning of the study:



When it is important to compare the prognosis for patients with disease to those without, a group of patients without the disease in question is sometimes followed in a cohort study. Consider the example of well differentiated prostate cancer described earlier: the best way to know that these patients have the same prognosis as members of the population without prostate cancer is to follow a similar group of patients without prostate cancer for a long period, perhaps 10 or 20 years.

While cohort studies are considered the best way do a study about prognosis, you can also use a case-control design. While cohort studies are prospective (patients are followed forward in time), case-control studies are retrospective (looking back in time). Patients with a disease are identified who have suffered a bad outcome such as death or recurrence, and compared with patients who have the disease but haven't suffered the bad outcome. For example, a researcher might identify a group of breast cancer patients who have died from a cancer registry, and compare them with a similar group of patients with breast cancer who are still living. This is diagrammed below:



Note that in a case-control study there are two major potential biases which don't exist to the same extent in a cohort study. First, you have to pick controls. If the controls are different then the cases (i.e. older, larger, different lifestyle, different habits), that introduces bias. Second, you are looking backwards in time (retrospective design) to determine prognostic factors. While

in some cases you may have good records and little bias (for example, descriptions of the surgery performed and the amount of radiation given), other variables may be subject to significant "recall bias". For example, knowing that breast cancer is now thought to be related to fatty food and alcohol could bias women's recall of their diet. Another important bias should be obvious from the diagram above: you will have to rely on the families, friends, and caregivers of women who have died to tell you that patient's diet and habits.

Determining relevance

Before bothering to evaluate the validity of an article, it is important to determine its relevance to you and your patients. There are three major **relevance criteria**, true of any type of article:

1. Will this information, if true, have a *direct* bearing on the health of my patients and is it something *they* will care about?
2. Is this problem *common* to my practice?
3. Will this information, if true, require me to *change* my current practice?

To answer the first question, look at the outcomes being prognosticated. In a study of congestive heart failure, are we predicting mortality, or wall motion? Quality of life? or hemoglobin A1C?

Articles which tell us about the prognosis for important patient-oriented outcomes are probably worth reading, while those which only describe intermediate or surrogate outcomes such as lab tests or imaging abnormalities are generally not.

The second and third questions can only be answered by the individual practitioner. A rural physician in a small southern town may have a very different spectrum of patients and problems than a suburban HMO physician. In some cases, a rare but severe problem such as pheochromocytoma may prompt a search for an article about prognosis, even if the condition is not common to your practice.

Once you have determined that an article is **relevant**, you have to determine the **validity**.

Evaluating validity

Laupacis and colleagues of the Evidence-Based Medicine Working Group have described several important questions regarding validity of an article about prognosis. They are paraphrased below:

1. Were the study subjects **representative** of patients with the disease in question?

Many studies are done in referral settings, where the sickest patients end up. The study may describe the prognosis of more severely ill patients, rather than the patients typically seen in the primary care setting.

2. Were the patients **at a similar point** in the course of their disease?

In a study about breast cancer prognosis, patients should be at a similar stage in the disease. Otherwise, it is difficult to interpret the results.

3. Were the **outcomes well-defined**, and were the people recording the outcomes blinded to the prognostic factors?

Mortality is a clearly defined outcome. Some studies, though, use more subjective outcomes. For example, a study of the decline in patients with Alzheimer's may have institutionalization as a primary outcome. Clearly, the decision to institutionalize is social and behavioral as well as medical.

4. Were patients **followed long enough** for outcomes to occur?

It is important that patients are followed long enough for the outcome in question to occur. Five years is a minimum for most chronic illnesses, and ten years is preferable.

5. Was the **dropout rate** excessive?

If too many patients are lost to follow-up, a serious bias may be introduced. Patients lost to follow-up may differ significantly in terms of severity

and outcome of their disease. For a study to be valid, at least 70% to 80% of patients should be followed for the duration of the study.

7. Did the authors **adjust for differences between groups**?

Because prognostic trials are usually cohort or case-control studies, and patients are not randomized to various prognostic factors, it is important that the researchers adjust for differences between groups. Common confounders include age, stage of disease, and co-morbidities.

Exercise

You have a 64 year old patient with new onset congestive heart failure, NYHA Class I. He asks you what he can expect as his prognosis. You remember the article by ***Ho and colleagues***, who look at the prognosis of patients with congestive heart failure (CHF) from the Framingham Heart Study, and you promise to consult it before his next visit.

Before evaluating the **validity**, let's consider the **relevance**. Do the patients in this study look like our gentleman? On page 107, under "Study Population", we learn that over 5000 residents of Framingham aged 28 to 62 years were enrolled in 1948, and in 1971 5000 of their offspring were enrolled in the Framingham Offspring Study. Patients with CHF at the beginning of the study were excluded from the analysis. These patients were then followed longitudinally with regular follow-up. Are these patients generally similar to our patient?

Yes No

Now, let's talk about the validity. Is this a:

Cohort study

Case-control study

Randomized controlled trial

This study design is a good choice for a study about prognosis, so we'll continue with our evaluation of validity. We will consider each of the criteria described above in turn:

<u>1. Were the study subjects **representative** of patients with the disease in question?</u>

Yes No

2. Were the patients **at a similar point** in the course of their disease?

Yes No

3. Were the **outcomes well-defined**, and were the people recording the outcomes blinded to the prognostic factors?

Yes No

4. Were patients **followed long enough** for outcomes to occur?

Yes No

5. Was the **dropout rate excessive**?

Yes No

6. Did the authors **adjust for differences between groups**?

Yes No

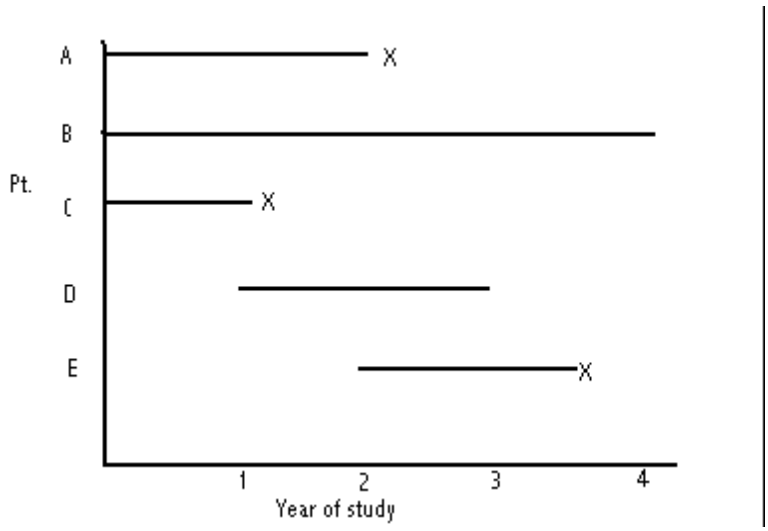
You've just examined this article, and found it to be a **valid** study of prognosis among patients with CHF.

Interpreting the results of a study on prognosis

The results in an article about prognosis can be presented in different ways. For example, you can simply state the probability of the outcome at a given point in time:

- 23% of patients with Grade A colon cancer have died at 5 years of follow-up
- 30% of patients with Alzheimer's dementia are institutionalized by 3 years following diagnosis

A somewhat richer way to present this information is in a survival curve. Before discussing survival curves, let's look at some sample data from a prognosis study:



Clearly, at any point in the study, patients have been followed for different periods of time, and some patients have died. By the end of the study, one patient has been lost to follow-up (D) and one is still alive (B). How do we make sense of these data?

These kinds of observations are called "**censored observations**"; when patients are not all enrolled at the same time (as in the graph above), they are

called "**progressively censored observations**". The fundamental principle of analyzing these data is that for each point in time, we look at how many patients are eligible to be counted, and how many are still alive at that point.

The table below shows some censored data for a study of kidney transplant rejection:

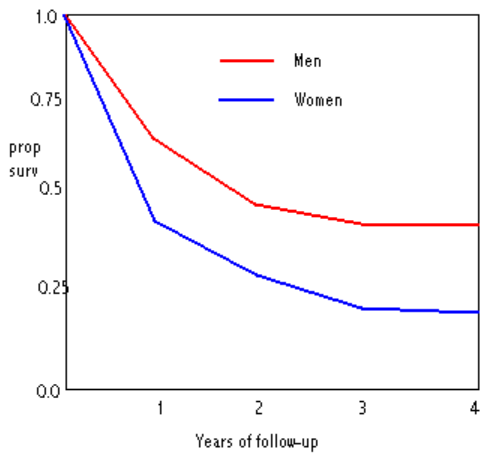
1	2	3	4	5	6	7
Months since entry into study	**Alive at beginning of interval (n_i)**	**Rejections during interval (d_i)**	**Withdrawn or lost to follow-up (w_i)**	**P [rejection or death]**	**P [Kidney retention]**	**Cumulative P [kidney retention]**
0 up to 2	31	3	2	3/[31-(2/2)] = 0.10	0.90	0.90
2 up to 4	26	3	2	3/[26-(2/2)] = 0.12	0.88	0.79
4 up to 6	21	1	3	1/[21-(3/2)] = 0.05	0.95	0.75
6 up to 9	17	0	3	0/[17-(3/2)] = 0.0	1.0	0.75
9 up to 12	14	0	2	0/[14-(2/2)] = 0.0	1.0	0.75
12 up to 15	12	0	4	0/[12-(4/2)] = 0.0	1.0	0.75
15 up to 18	8	1	1	1/[8-(1/2)] = 0.13	0.87	0.65
18 up to 21	6	0	4	0/[6-(4/2)] = 0.0	1.0	0.65
21 up to 24	2	0	2	0/[2-(2/2)] = 0.0	1.0	0.65

Where do those numbers come from? Let's review:

- **Column 1**: The number of months since the start of the study. In this case, all 31 of the patients were enrolled at the same time.

- C**olumn 2**: How many were alive at the beginning of the interval? Obviously, this number gets smaller as the study goes on and patients die.
- C**olumn 3**: How many rejections occur during this interval? The number of rejections seems higher earlier in the study, which is biologically plausible.
- **Column 4:** How many patients withdrew from the study, or were lost to follow-up? If a patient is lost to follow-up, we would like to make use of the information gained during the time they were in the study.
- **Column 5**: Be careful! Now it gets a little tricky. The probability of rejection is given by the formula q = d/[n-(w/2)]. The numerator (d) is the number of rejections during the interval - that's the easy part. In the denominator (n - (w/2)), we have to account for the patients lost to follow-up (w). We divide by 2 because we don't know when in the interval they were lost, and therefore assume that they were lost half-way through.
- **Column 6**: The probably of retaining a kidney during the interval is simply 1 minus the probability of losing one. For example, if the probability of losing one is 0.1, the probability of retaining one is 1 - 0.1 = 0.9.
- **Column 7**: Finally, the cumulative probability of retaining a kidney is the product of each interval probability. Thus, for the 2nd interval it is 0.9 x 0.88 = 0.79, and for the 3rd interval is 0.9 x 0.88 x 0.95 = 0.75.

These data can be presented graphically in a survival curve. A survival curve for hypothetical data comparing men with women over a four year period of follow-up is shown below:



In this graph, the proportion of survivors is shown along the y-axis, and the years of follow-up along the x-axis. What proportion of men are still alive at 3 years of follow-up?

0.2 0.4 0.6 0.8

Let's answer our original question. Recall, we wanted to know the prognosis for a 64 year old man with newly diagnosed congestive heart failure. Refer to the Framingham article, which we have already decided is valid. Sure enough, there is a group of four survival curves on the bottom right corner of page 111, in Figures 1 and 2. Since our patient is a man, we are interested in the two left-hand curves. In particular, Figure 1 gives us the survival rates for the most recent years of the study (1970 to 1988) stratified by age. What is the approximate 5 year survival rate for our patient?

5% 20% 30% 70%

We also notice in Figure 2 that the cause of CHF is an important predictor of mortality, particularly for men. Patients with CHF caused by valvular heart disease do:

Worse than average Better than average

Since our patient does not have valvular heart disease, we find this somewhat reassuring.

Quiz

Question 1.

What is the best study design for an article about prognosis?

a. Cohort (retrospective)

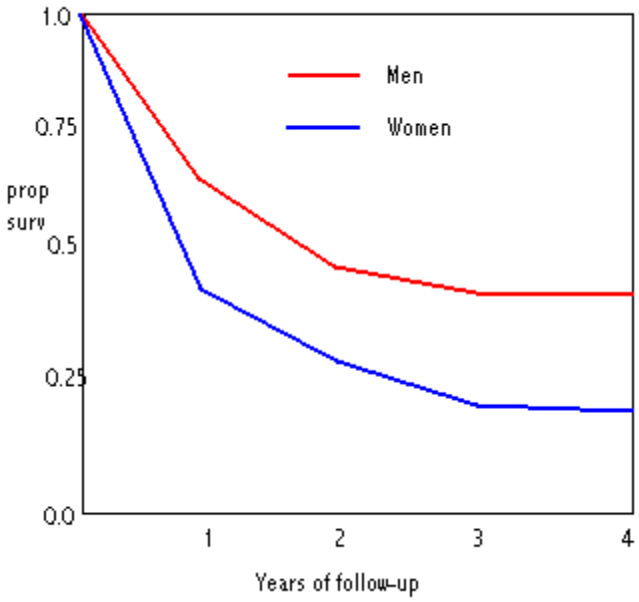
b. Cohort (prospective)

c. Case-control

d. Case series

Question 2.

Examine the following diagram:



The 4 year survival rate for women is how much higher or lower than that for men (percentage given are absolute differences, not relative)?

a. 25% higher

b. 25% lower

c. 50% lower

d. 50% higher

Question 3.

A study enrolls 5400 patients with hypertension, and then follows them for 20 years to see which ones develop a stroke. This is an example of a:

a. Cohort study (prospective)

b. Cohort study (retrospective)

c. Case-control study

d. Randomized controlled trial

Module VII

Appraising Literature

3. Integrative studies

CONTENTS:

- **Objectives of the module.**
- **Introduction.**
- **The need for meta-analyses.**
- **Steps of meta-analyses.**
- **Evaluating the relevance of a meta-analysis.**
- **Evaluating the validity of a meta-analysis**
- **Using the results of a meta-analysis.**
- **Quiz.**

Objectives:

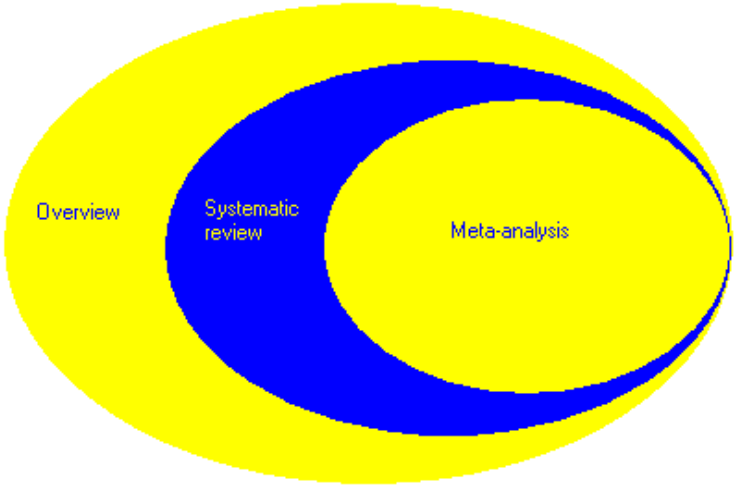
At the end of this module, the participant will be able to:

- Explain the steps of performing a meta-analysis: identification, selection, abstraction, and analysis.
- Distinguish between random and fixed effects models when reading a meta-analysis
- Evaluate the relevance, validity and usefulness of a meta-analysis.
- Critically appraise a simple meta-analysis.

Introduction

The fourth, fifth and sixth modules have helped you learn how to read articles about diagnosis, therapy and prognosis respectively. Now, it's time to learn about how to read an article about meta-analysis. While on the surface more complex and involved than most studies, by applying the principles learned in the first few modules it will be easy to interpret these types of studies.

One area of confusion is the distinction between review articles, systematic reviews, and meta-analyses. The diagram below illustrates their relationship using a Venn diagram:



Review articles (also called overviews) are the broadest category. Most review articles are unsystematic, because the author does not look at all of the evidence. A systematic review has a formal approach to gathering, evaluating, and presenting the evidence. Some systematic reviews are meta-analyses; a meta-analysis goes the final step by using formal statistical methods to calculate a summary result or results.

The need for a meta-analysis:

Meta-analyses is needed for the following two major reasons:

1. To quantitatively combine the results of previous studies to arrive at a summary estimate.
2. As a "study of studies", to help guide further research and identify reasons for heterogeneity between studies.

Meta-analyses of the first kind can help resolve medical controversies caused by conflicting studies, are an inexpensive alternative to very large randomized trials, and can in this way shape health policy.

The second kind of meta-analysis is particularly useful for designing future studies, by systematically identifying key patient and study characteristics from previous work.

Meta-analyses were initially used in the social sciences in the mid-1970's, and were adapted to medical data sets in the early to mid-1980's. The number of meta-analyses is growing rapidly, based on the results of this Medline search for the keyword "meta-analysis":

Years	Number of meta-analyses in Annals, JAMA, Lancet, NEJM
1971-75	0
1976-80	0
1981-85	1
1986-90	32

1991-95 175

Meta-analyses often form the initial step of a cost-effectiveness analysis, decision analysis, or grant application. While the results usually correspond to later randomized trials, they do not always (LeLorier, 1997):

	Results of RCT	
Results of meta-analysis	**Positive**	**Negative**
Positive	13	6
Negative	7	14

Steps of a meta-analysis

There are four basic steps to any good meta-analysis:

1. Identification
2. Selection
3. Abstraction
4. Analysis

1. Identification

The first step in a meta-analysis is to find all of the pertinent articles on your topic. Important sources of information for a meta-analysis include:

- MEDLINE
- EMBASE
- CancerLit, AIDSLine, and ToxLine
- Index Medicus

While MEDLINE, the database of the National Library of Medicine, is a good starting point, it is not the only source of information. MEDLINE indexes approximately 4100 journals, dating from 1966 to the present. It now offers an excellent Web-based interface for free searching, including a special page for clinical queries.

There are also CD-ROM based search engines from BRS Colleague, WinSPIRS, and others which offer different search options, but use essentially the same underlying database. The European version of MEDLINE is called EMBASE, and is a Dutch/English collaboration. Depending on the topic, it may be appropriate to search the more specialized National Library of Medicine databases, such as CancerLit, AIDSLine, and ToxLine.

The Cochrane Collaboration Controlled Trials Register, established in 1993, is an important source of studies for a meta-analysis. Available by subscription on CD-ROM or Web interface, the Register includes abstracts of

over 160,000 trials. It includes all controlled trials in the MEDLINE and EMBASE, as well as the results of hand searches by Cochrane Collaboration volunteers of thousands of journals not indexed by MEDLINE or EMBASE. The Collaboration is even negotiating with drug companies to get abstracts of the results of unpublished Phase III drug trials.

The ***Index Medicus*** can still be useful when it is important to search for articles published before 1966, when MEDLINE and the other electronic databases were established.

Finally, there are other sources of "fugitive literature" that may be important for the author of a meta-analysis (some of which may be found in the Cochrane Controlled Trials Register):

- Unpublished studies – by contacting the authors themselves
- Dissertations - the national indexes of dissertations at university libraries
- Drug company studies - by contacting the company directly
- Non-indexed studies – by searching the bibliographies and Cochrane
- pre-MEDLINE (1966) – by using Index Medicus

It's important to know that different search strategies can result in different results (Table 4.5, Petitti):

Topic	**Cochrane CTR**	**MEDLINE (expert searcher)**	**MEDLINE (amateur searcher)**
Neonatal hyperbilirubinemia	88	28	17
Intraventricular hemorrhage	29	19	11

Note: CTR = Controlled Trials Register

If you are thinking about doing a meta-analysis, it is important to enlist the aid of an expert Medline searcher such as a medical librarian. The above table also highlights the importance of using the Cochrane Controlled Trials Register.

2. Selection

Once the author of a meta-analysis has assembled a large number of studies, it is important to select the right ones. There are a variety of possible inclusion (also called eligibility) criteria:

- Whether the study includes enough information for analysis (i.e. standard deviation or standard error in addition to point estimate)
- The study design (i.e. controlled trials only vs randomized controlled trials only, especially for studies of therapy)
- The year of study, if technology or typical dosing changes (for example, only include studies since 1984 on dyspepsia if you're interested in helicobacter pylori)
- The dosage used in the study (to assure that an effective dose was used)
- The language of the article; you or a colleague have to be able to read it.
- The minimum sample size; very small studies may be unrepresentative and/or not worth the effort
- The patient age (adults only, > 60 only, etc)
- The study setting (emergency department, outpatient, inpatient)

3. Abstraction

Once an appropriate group of studies has been identified, the author(s) have to abstract the relevant data from each study. There are many sources of potential error in data abstraction:

- The article may be wrong due to typographical or copyediting errors
- Tables can be misinterpreted
- Errors can occur during you own data entry or abstraction process

A good meta-analysis will take some or all of the following steps to minimize errors:

- Use 2 independent reviewers
- Use a 3rd reviewer or consensus meeting to resolve conflicts
- Train reviewers by practicing with several articles to "calibrate"
- Compare abstract and text to look for inconsistencies
- Use a standard form or database which constrains entries to the expected range
- Report the results of the data abstraction, including the percentage concordance or even a kappa statistic

Bias can also creep into a meta-analysis. For example, the authors may be biased in favor of (or against!) well known researchers. Also, prominent journals may be given greater weight or authority (rightly or wrongly). It is therefore best (although not often done) to have identifiers eliminated from articles.

Finally, part of the data abstraction phase is an assessment of study quality. Chalmers has proposed a fairly complex set of criteria which apply well to randomized controlled trials. Simpler criteria may be sufficient. For example, in a diagnostic meta-analysis, simply assuring a high quality gold standard, independent assessment of reference and study tests, and blinding may be adequate. Too often, the quality assessment is done and ignored.

Ideally, the results of the quality assessment should inform the analysis and interpretation of results.

4. Analysis

There are many issues and controversies in the analysis of meta-analytic data. First, let's define some important terms:

- **Homogeneity** and **heterogeneity** describe the degree of between-study variability in a group of studies. It is probably appropriate to combine the

results from a homogenous set of studies, but many would argue that results from heterogeneous studies should not be combined. The Q statistic, interpreted using a chi-square distribution, is often used as a test of homogeneity.

- **Fixed effects models** consider only within-study variability. The assumption is that studies use identical methods, patients, and measurements; that they should produce identical results; and that differences are only due to within-study variation. By using a fixed effects model, the researcher answers the question: "Did the treatment produce benefit on average in the studies at hand?" The Peto and Mantel-Haenszel odds ratios are both based on a fixed effects model.
- **Random effects models** consider both between-study and within-study variability. The assumption is that studies are a random sample from the universe of all possible studies. With a random effects model, the researcher answers the question: "Will the treatment produce benefit ‘on average’?" The DerSimonian Laird statistic is based on a random effects model.

Usually, the answers provided by these different modeling assumptions are similar. Differences only arise when studies are not homogenous. In a comparison of 22 meta-analyses, fixed and random effects models gave the same answer in 19 out of 22. In 3 cases, fixed effects models were significant while random effects models were not.

When there is significant heterogeneity, the between-study variance becomes much larger than the within, and studies of different sample size receive relatively similar weight. When there is homogeneity, sample size dominates, and both models give similar results. Random effects models are therefore more "conservative" and generate a wider confidence interval. Put another way, a random effects model is less likely to show a significant treatment effect than a fixed effects model.

In general, if the studies are homogenous, the researchers should use a fixed effects model. If the studies are heterogeneous, the researchers (and the reader) should first ask why! While it may be appropriate to do a random effects analysis on all of the studies, it may be better to identify an important subgroup difference (i.e. studies using one dose showed significant effect, while lower dose did not) and then do a fixed effects analysis of each and report all of the results.

A term you will encounter in many meta-analyses is "sensitivity analysis". A sensitivity analysis is a way of looking at only certain studies, certain groups of patients, or certain interventions. For example, a meta-analysis of aspirin in prevention of acute MI might first analyze all studies, but then also look separately at only studies of men and studies of women.

However, details of calculations for homogeneity, fixed effects models, and random effects models are beyond the scope of this course.

Evaluating the relevance of a meta-analysis

After reviewing the steps in a meta-analysis, the three basic steps for formal critical appraisal are achieved by answering the following three questions:

1. Would the results change my practice if valid?
2. Are the outcomes important to my patients?
3. Are the results valid?

A study which meets all three of these criteria is a POEM (Patient-Oriented Evidence that Matters). The first two criteria have to do with relevance, while the third is concerned with validity.

Would the results change my practice if valid?

This question must be answered by the individual clinician. A meta-analysis which confirms existing practice need not be subjected to rigorous evaluation, since they results don't change what you are already doing. On the other hand, if the results would change what you do for your patients, it is incumbent on you to take the next step.

Are the outcomes important to my patients?

Most meta-analyses use patient-oriented outcomes such as morbidity, mortality, and symptoms. However, if a study uses surrogate or intermediate outcomes such as FEV1 or hemoglobinA1C, the results should be interpreted cautiously. Proponents of the use of POEMs would argue that if a meta-analysis does not use patient-oriented outcomes, you don't have to proceed to the next step.

Evaluating the validity of a meta-analysis

If a study would change practice, and uses patient-oriented outcomes, you are obligated to evaluate its validity. Key criteria include:

- Did the authors ask a focused clinical question?
- Were the criteria used to select articles for inclusion appropriate?
- Is it unlikely that important, relevant studies were missed?
- Was the validity of the included studies appraised (study quality)?
- Were assessments of studies reproducible (data abstraction)?
- Were the results similar from study to study (homogeneity)?

Using the results of a meta-analysis

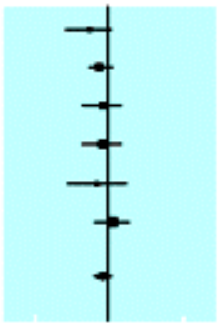
Most meta-analyses will initially give you a table of included studies. In our study, that table is published on the Web only. We see that most studies used doxycycline, sulfa/trimethoprim, or erythromycin. This is a possible limitation - maybe the pharmaceutical reps are right, and Biaxin or Zithromax are really better, although I doubt it!

Let's look at the results now. The three key figures are reproduced below. In Figure 2, the first column has the study name; the second column the proportion receiving antibiotic with productive cough at follow-up; the third column the proportion receiving placebo with productive cough at follow-up. Let's stop and think here. If antibiotics work, patients receiving them should be less likely to have cough at follow-up. That would make the relative risk of cough less than 1.0. In the light blue "relative risk diagram" below, the vertical line indicates a relative risk of 1.0. The horizontal dots and bars represent the relative risk and 95% confidence interval for each study. The weight is proportional to the inverse of the variance of the study - high variance, associated with a small study, means less weight given to that study, and vice versa. The relative risk is shown numerically in the final column of each table. Finally, the summary estimate of relative risk with its confidence interval is shown in the last row, along with an estimate of homogeneity (the chi-square).

Figure 2. Productive cough at 7 to 11 days follow-up.

Proportion of subjects with productive cough at follow up

Study	Antibiotic	Placebo	Relative risk (95% CI) random effects model	Weight (%)	Relative risk (95% CI) random effects model
Dunlay et al[25]	10/21	17/24		9.6	0.67 (0.40 to 1.13)
King et al[27]	28/41	27/31		41.3	0.78 (0.61 to 1.01)
Stephenson (unpublished)	24/81	27/82		12.3	0.90 (0.57 to 1.42)
Stott and West[21]	30/104	32/103		14.7	0.93 (0.61 to 1.41)
Verheij et al[26]	13/72	16/72		6.0	0.81 (0.42 to 1.56)
Williamson[23]	23/37	18/32		16.3	1.11 (0.74 to 1.64)
Total (95% CI)	128/356	137/344		100.0	0.85 (0.73 to 1.00)
χ^2=3.21, df=5, Z=1.94					

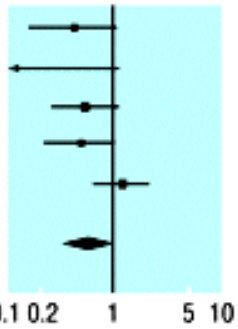


Looking at this table, you see that the summary estimate of relative risk just barely includes one, suggesting that the association is NOT statistically significant. Look at the next figure below:

Figure 3. Failure to clinically improve at 7 to 11 days follow-up**.**

Proportion of subjects who had not improved clinically at follow up

Study	Antibiotic	Placebo	Relative risk (95% CI) random effects model	Weight (%)	Relative risk (95% CI) random effects model
Brickfield et al[24]	5/26	10/24		19.6	0.46 (0.18 to 1.16)
Dunlay et al[25]	0/23	6/21		3.6	0.07 (0.00 to 1.18)
Stott and West[21]	10/104	17/103		24.4	0.58 (0.28 to 1.21)
Verheij et al[26]	9/73	17/72		24.2	0.52 (0.25 to 1.09)
Williamson[23]	16/37	11/32		28.2	1.26 (0.69 to 2.30)
Total (95% CI)	40/263	61/252		100.0	0.62 (0.36 to 1.09)
χ^2=8.21, df=4, Z=1.66					

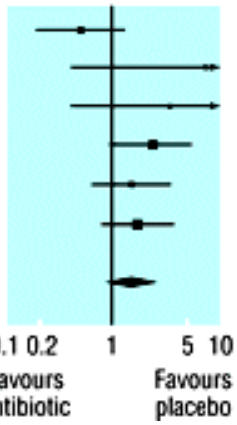


In this case, the confidence interval of the summary relative risk also includes 1.0 (it is 0.36 - 1.09).

Figure 4. Adverse drug effects.

Proportion of subjects with side effects

Study	Antibiotic	Placebo	Relative risk (95% CI) random effects model	Weight (%)	Relative risk (95% CI) random effects model
Brickfield et al[24]	5/26	9/24		20.6	0.51 (0.20 to 1.32)
Dunlay et al[25]	4/31	0/26		3.5	7.59 (0.43 to 134.81)
Franks and Gleiner[22]	3/25	1/29		5.7	3.48 (0.39 to 31.38)
King et al[27]	14/39	5/32		21.5	2.30 (0.93 to 5.70)
Stott and West[21]	12/104	8/103		23.0	1.49 (0.63 to 3.48)
Verheij et al[26]	15/78	9/80		25.7	1.71 (0.80 to 3.67)
Total (95% CI)	53/303	32/294		100.0	1.51 (0.86 to 2.64)
χ^2=7.80, df=5, Z=1.45					



Finally, while there appears to be more adverse drug effects with antibiotic use, this association is also not statistically significant.

So, have we learned nothing? No. Even meta-analyses can suffer from Type II error, i.e. too small a sample size. The total number of patients studied was only about 600 in the 6 studies, and it is interesting to note that while the results were not significant, they were consistent. That is, 5 out of 6 studies fell either to the left or right of the vertical bar for each outcome.

Quiz

Question 1.

The most rigorous and methodologically complex kind of review article is an:

a. Overview

b. Systematic Review

c. Meta-analysis

Question 2.

Steps in a meta-analysis include all of the following except:

a. analysis

b. abstraction

c. selection

d. randomization

Question 3.

A fixed effects model is most appropriate in a meta-analysis when study findings:

a. Are homogenous

b. Are heterogeneous

c. Either homogenous or heterogeneous

d. Neither homogenous or heterogeneous

Question 4.

Good sources for meta-analyses include:

a. The Cochrane Collaboration Controlled Trials Register

b. MEDLINE

c. EMBASE

d. SEARCHBOY

Student _____

Pre-test/Post-test

EBM Principles

Multiple Matching

Match the type of clinical/research question with the type(s) of research design that can be used to answer the question. Research designs can be used more than once.

1. Diagnosis (1) _____ 2. Harm (4) _____ _____ _____ _____ 3. Prognosis (2) _____ _____ 4. Therapy (3) _____ _____ _____	**Research Designs** A. Blinded comparison to a gold standard B. Case report C. Prospective cohort study D. Prospective comparative study E. Randomized control trial F. Retrospective case series G. Retrospective case-control study

Levels of Evidence

For each of the types of clinical questions, below, number the level/types of evidence from highest (1) to lowest (5)

5. Therapy/Prevention/Etiology/Harm:

Student _____

Level

_____ Case-series (and poor quality cohort and case-control studies)

_____ Expert opinion without explicit critical appraisal, or based on physiology, bench research or 'first principles'

_____ Systematic review (with homogeneity) of case-control studies; or Individual case-control study

_____ Systematic review (with homogeneity) of cohort studies; or Individual cohort study (including low quality randomized control trial; e.g., <80% follow-up); 'Outcomes' Research

_____ Systematic review (with homogeneity) of randomized control trials; or Individual randomized control trial (with narrow confidence interval)

6. Diagnosis:

Level

_____ Any of: 1) reference standard was unobjective, unblinded or not independent; 2) positive and negative tests were verified using separate reference standards; 3) study was performed in an inappropriate spectrum of patients.

_____ Expert opinion without explicit critical appraisal, or based on physiology, bench research or 'first principles'

_____ Independent blind comparison of an appropriate spectrum, but the reference standard was not applied to all study patients

_____ Systematic review (with homogeneity) of Level 1 diagnostic studies; or a clinical rule validated on a test set; or Independent blind comparison of an appropriate spectrum of consecutive patients, all of whom have undergone both the diagnostic test and the reference standard.

Student _____

_____ Systematic review (with homogeneity) of Level >2 diagnostic studies; or Any of: 1)independent blind or objective comparison; 2)study performed in a set of non-consecutive patients, or confined to a narrow spectrum of study individuals (or both) all of whom have undergone both the diagnostic test and the reference standard; 3) a diagnostic clinical rule not validated in a test set.

7. Prognosis: (There is no Level 3 in prognosis, so number these 1, 2, 4 and 5)

Level

_____ Case-series (and poor quality prognostic cohort studies)

_____ Expert opinion without explicit critical appraisal, or based on physiology, bench research or 'first principles'

_____ Systematic review (with homogeneity) of either retrospective cohort studies or untreated control groups in randomized control trials; or Retrospective cohort study or follow-up of untreated control patients in an randomized control trial; or clinical rule not validated in a test set; or 'Outcomes' research

_____ Systematic review (with homogeneity) of inception cohort studies; or a clinical rule validated on a test set. or Individual inception cohort study with > 80% follow-up; or All or none case-series

EBM Resources

Matching

_____8. EBM resources that is a collection of systematic reviews of randomized clinical trials of the highest methodical rigor by panels of experts on specific topics; and is updated regularly.	A. ACP Journal Club B. Cochrane Database C. DARE D. MEDLINE

Student _____

_____9. Structured index of all medical peer reviewed journal articles. Currently covers 4600 journals and contains over 12M articles. _____10. EBM resource prepared by the National Health Service in th UK that is a database of abstracts of quality assessed systematic reviews and metaanalysis. _____11. EBM resource that systematically surveys 100 of the major journals and prepares structured abstracts of the best original and review articles on the cause, course, diagnosis, clinical prediction, prevention, treatment, or economics of medical disorders managed by internists and related subspecialists, and on quality improvement and continuing medical education intervention trials in internal medicine.	

Types of Clinical Questions

What type of clinical question are each of the following?

A. Diagnosis

B. Harm

C. Therapy

D. Prognosis

Student _____

12. _____You have a 28 year old male non-smoker patient who has been exposed to asbestos during a home renovation. What is the probability that he will suffer some bad outcomes?

13. _____A 18 year old female presents to the emergency room where you are working with a sprained ankle. Should you x-ray it to determine if it is fractured?

14. _____A 19-year-old woman comes home from college with an acute maxillary sinusitis. You just heard about treating this with a shorter 3-day course of antibiotics rather than the usual 10 days of prescribed treatment. You wonder whether you should try the shorter course with this patient.

15. _____A 44-year-old woman with recently diagnosed ovarian cancer presents to the emergency room with dyspnea and inspiratory chest discomfort. The ventilation-perfusion scan is read as 'indeterminate.' The emergency room doctor asks your advice 'now that embolism has been ruled out.'

16. _____A 69-year-old retired teacher returns for follow-up of congestive heart failure that first presented 1 month before. After you review her progress, she asks you about her life expectancy.

Student _____

ANSWERS TO Pretest

EBM Principles

Multiple Matching

Match the type of clinical/research question with the type(s) of research design that can be used to answer the question. Research designs can be used more than once.

5. Diagnosis (1) **A** **6.** Harm (4) **E, D, G, and B** 7. Prognosis (2) **C and F** 8. Therapy (3) **E, D and G**	**Research Designs** H. Blinded comparison to a gold standard I. Case report J. Prospective cohort study K. Prospective comparative study L. Randomized control trial M. Retrospective case series N. Retrospective case-control study

Levels of Evidence

For each of the types of clinical questions, below, number the level/types of evidence from highest (1) to lowest (5)

5. Therapy/Prevention/Etiology/Harm:

Level

Student _____

4__ Case-series (and poor quality cohort and case-control studies)

5__ Expert opinion without explicit critical appraisal, or based on physiology, bench research or 'first principles'

3__ Systematic review (with homogeneity) of case-control studies; or Individual case-control study

2__ Systematic review (with homogeneity) of cohort studies; or Individual cohort study (including low quality randomized control trial; e.g., <80% follow-up); 'Outcomes' Research

1__ Systematic review (with homogeneity) of randomized control trials; or Individual randomized control trial (with narrow confidence interval)

6. Diagnosis:

Level

4__ Any of: 1) reference standard was unobjective, unblinded or not independent; 2) positive and negative tests were verified using separate reference standards; 3) study was performed in an inappropriate spectrum of patients.

5__ Expert opinion without explicit critical appraisal, or based on physiology, bench research or 'first principles'

3___Independent blind comparison of an appropriate spectrum, but the reference standard was not applied to all study patients

1__ Systematic review (with homogeneity) of Level 1 diagnostic studies; or a clinical rule validated on a test set; or Independent blind comparison of an appropriate spectrum of consecutive patients, all of whom have undergone both the diagnostic test and the reference standard.

2_ Systematic review (with homogeneity) of Level >2 diagnostic studies; or Any of: 1)independent blind or objective comparison; 2)study performed in a set of

Student _____

non-consecutive patients, or confined to a narrow spectrum of study individuals (or both) all of whom have undergone both the diagnostic test and the reference standard; 3) a diagnostic clinical rule not validated in a test set.

7. Prognosis: (There is no Level 3 in prognosis, so number these 1, 2, 4 and 5)

Level

4_ Case-series (and poor quality prognostic cohort studies)

5_ Expert opinion without explicit critical appraisal, or based on physiology, bench research or 'first principles'

2_ Systematic review (with homogeneity) of either retrospective cohort studies or untreated control groups in randomized control trials; or Retrospective cohort study or follow-up of untreated control patients in an randomized control trial; or clinical rule not validated in a test set; or 'Outcomes' research

1_ Systematic review (with homogeneity) of inception cohort studies; or a clinical rule validated on a test set. or Individual inception cohort study with > 80% follow-up; or All or none case-series

EBM Resources

Matching

Student _____

B _____12. EBM resources that is a collection of systematic reviews of randomized clinical trials of the highest methodical rigor by panels of experts on specific topics; and is updated regularly. D _____13. Structured index of all medical peer reviewed journal articles. Currently covers 4600 journals and contains over 12M articles. C _____14. EBM resource prepared by the National Health Service in th UK that is a database of abstracts of quality assessed systematic reviews and metaanalysis. A _____15. EBM resource that systematically surveys 100 of the major journals and prepares structured abstracts of the best original and review articles on the cause, course, diagnosis, clinical prediction, prevention, treatment, or economics of medical disorders managed by internists and related subspecialists, and on quality improvement and continuing medical education intervention trials in internal medicine.	A. ACP Journal Club B. Cochrane Database C. DARE D. MEDLINE

Types of Clinical Questions

What type of clinical question are each of the following?

A. Diagnosis

B. Harm

Student _____

C. Therapy

D. Prognosis

17. B _____You have a 28 year old male non-smoker patient who has been exposed to asbestos during a home renovation. What is the probability that he will suffer some bad outcomes?

18. A _____A 18 year old female presents to the emergency room where you are working with a sprained ankle. Should you x-ray it to determine if it is fractured?

19. __C__A 19-year-old woman comes home from college with an acute maxillary sinusitis. You just heard about treating this with a shorter 3-day course of antibiotics rather than the usual 10 days of prescribed treatment. You wonder whether you should try the shorter course with this patient.

20. A _____A 44-year-old woman with recently diagnosed ovarian cancer presents to the emergency room with dyspnea and inspiratory chest discomfort. The ventilation-perfusion scan is read as 'indeterminate.' The emergency room doctor asks your advice 'now that embolism has been ruled out.'

21. D _____A 69-year-old retired teacher returns for follow-up of congestive heart failure that first presented 1 month before. After you review her progress, she asks you about her life expectancy.