

Continuing Education Lesson



Evidence-Based Pharmacotherapy

A practical guide for pharmacists

Peter Loewen, BSc(Pharm), PharmD

This program has been approved
for **3.0 CEUs**
by the Canadian Council on
Continuing Education in Pharmacy
CCCEP #915-1202



This program was developed with an unrestricted
educational grant by Merck Frosst Canada



*Découvrir toujours plus. Vivre toujours mieux.
Discovering today for a better tomorrow.*

Merck Frosst Canada Ltée, Kirkland Québec
Merck Frosst Canada Ltd., Kirkland Quebec

Correspondence

Editorial correspondence may be sent to

Virtual Learning Inc.
CCCEP Home Study Program
18 Wynford Drive, Suite 507
Toronto, Ontario M3C 3S2

Queries regarding subscriptions or receipt of marks should be sent to your provincial Continuing Education Coordinator.

Disclaimer

We have done our best to produce an accurate, timely, and educational Home Study Program. However, Virtual Learning Inc., the Authors, the Reviewers, the Editors, the Provider Committee, and the Canadian Council on Continuing Education in Pharmacy assume no responsibility for any errors or consequences arising from the use of information contained within this program. With the constant changes in practice and regional differences, it remains the responsibility of the readers as professionals to interpret and apply this lesson's information to their own practices. All rights reserved.

Learning objectives

- To provide learners with a thorough understanding of the origins and rationale for adopting an evidence-based perspective on drug therapy.
- To provide learners with an understanding of the principles and processes of EBP.
- To provide learners with a working knowledge of the terminology and interpretative skills required to effectively evaluate, interpret, and apply the results of randomized controlled trials.
- To provide learners with examples of the applicability of these principles and skills to their own practice.
- To convince learners of the importance, for their own professional advancement, of incorporating an evidence-based perspective into their practice.

Author

Peter Loewen, BSc(Pharm), PharmD

- Clinical Assistant Professor of Pharmacy, Faculty of Pharmaceutical Sciences, University of British Columbia
- Pharmacotherapeutic Specialist—Internal Medicine, Pharmaceutical Sciences Clinical Service Unit, Vancouver Hospital & Health Sciences Center
- Coordinator – Advanced Pharmacotherapeutics, Doctor of Pharmacy Program, UBC
- Chair, UBC Clinical Research Ethics Board
- Publishing Editor—The Journal of Informed Pharmacotherapy

Expert reviewers

David Williamson, BPharm, MSc

Mr. Williamson is an assistant clinical professor of the Faculté de Pharmacie at l'Université de Montréal since 2002 and a critical care pharmacist at l'Hôpital Sacré-Cœur de Montréal since 1999. He has also completed a specialized residency in cardiovascular pharmacotherapy. Mr. Williamson has participated in several educational events for community and hospital pharmacists in Canada on acute coronary syndromes, hypertension, cholesterol management, coagulation and sepsis. His research involves impact of renal failure on the effects of antiplatelet agents, impact of antibiotic usage in the intensive care unit and the treatment of sepsis. His responsibilities include pharmacotherapy education in the intensive care unit and teaching of nosocomial infections and cardiovascular diseases modules at the Faculté de Pharmacie of the Université de Montréal.

Daniel Thirion, BPharm, MSc, PharmD, BCPS

Dr. Daniel Thirion is an assistant clinical professor of the Faculté de Pharmacie at l'Université de Montréal and pharmacist at l'Hôpital Sacré-Cœur de Montréal since 2001. Responsibilities include infectious disease education for l'Université de Montréal and pharmacotherapy education on the Internal Medicine Teaching Unit. His research involves impact of antibiotic management in a hospital setting on resistance, side effects, and cost. He chairs the committee on Evidence Based Medicine for the Faculté de Pharmacie. He also serves as consultant for hospital and provincial drug utilization committees. He has participated in several educational events for community and hospital pharmacists in Canada and the US.

Contents

1	1. Introduction
1	2. Definitions of EBM & Evidence-Based Pharmacotherapy (EBP)
2	3. Principles of EBM
2	<i>Table 1: Oxford Centre for Evidence-Based Medicine Levels of Evidence (May 2001)</i>
3	<i>Table 2: Oxford Center for Evidence-Based Medicine “Grades of Recommendations”</i>
3	4. Putting EBP into Practice
15	5. Beyond the “Method” of EBP: EBP as a Professional Lifestyle
15	6. Limitations, Pitfalls, and Challenges of Using an Evidence-Based Perspective
16	7. EBM/EBP Resources & Further Reading
16	8. Glossary of Selected Terms
18	References
22	Questions

1. Introduction

Terms such as “Evidence-Based Medicine” and “Evidence-Based Health Care” have been rapidly integrated into the vocabulary and philosophy of many clinicians over the past decade. The care that pharmacists provide to patients and the advice they give to other professionals can be strongly influenced by the perspective and principles of evidence-based medicine and, in particular, the evidence surrounding pharmacotherapy—hence the term “Evidence-Based Pharmacotherapy” (EBP). Thus, the importance of pharmacists in understanding and embracing the principles of evidence-based pharmacotherapy has never been greater.

This module will discuss the history of EBP, its place in contemporary pharmacy practice, the principles and processes involved in EBP, the interpretative skills which are required, and the pitfalls and limitations of EBP. Numerous practical examples to illustrate these points will be provided for pharmacists who wish to incorporate the principles of EBP into their practice.

2. Definitions of EBM & Evidence-Based Pharmacotherapy (EBP)

Evidence-based medicine has been defined by one of its founding fathers, David Sackett, as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.”¹ More colloquially, EBM asserts that clinicians “know the evidence to support what you do and apply it appropriately to the care of your patients.” My preferred definition, “EBM is the integration of *best research evidence* with *clinical expertise* and *patient values*,”² embodies the critical elements of **best evidence**, **clinical judgment**, and the **patient’s perspective**. This view of EBM, as we will see, is most practical for integration into clinical practice.

A host of variations on the term “EBM” have been introduced to embrace the values of various professions (Evidence-Based Nursing, Evidence-Based Education), perspectives (Evidence-Based Healthcare, Evidence-Based Clinical Practice), and even specialities (Evidence-Based Cardiology). Since pharmacists are in many ways linked to the spectrum of medicine, which includes disease prevention, detection, diagnosis, and treatment, refining the terminology and definitions used while retaining the spirit of EBM’s ideas may be helpful. Pharmacists focus their expertise on the therapeutic end of the care spectrum, and on pharmacotherapeutics in particular. Hence, in this module we will refer to **Evidence-Based Pharmacotherapy (EBP)**, recog-

nizing that this is a subset of EBM in the same way that Evidence-Based Diagnosis and Evidence-Based Nursing embody their own specialized perspectives, knowledge and practices.

2.1. A Brief History of EBM

To understand the relevance and importance of EBM/EBP in contemporary pharmacy practice, it is useful to reflect on the context in which its principles and ideas arose. Some authors trace EBM’s philosophical roots to ancient Chinese medicine (the practice of “kaozheng” meant “practicing evidential research”)² or the French physician Pierre Louis, who rejected therapies based on expert opinion and preferred careful observation of patients. However, neither of these movements resulted in profound changes in the practice of medicine.

In 1972, Archie Cochrane, a British epidemiologist, published a short book entitled *Effectiveness and Efficiency: Random Reflections on Health Services*³ in which he identified inefficiencies in Britain’s National Health Service and drew attention to the collective ignorance of health care providers about the effects of the care they provide. At that time, the practice of medicine was based, as it had been for centuries, primarily on personal experience, intuition, and the opinions of other “experts.” Cochrane exposed the limitations of this perspective, particularly on the basis that in a public health system producing measurable health benefits most efficiently was critical to the system’s survival.

One of the fundamental messages that emanated from Cochrane’s work was the need for randomized controlled trials (RCTs, which were in short supply at the time) to assess the efficacy of treatments. Over the ensuing 13 years, Cochrane’s group identified and catalogued 3,500 controlled trials in perinatal medicine published between 1940 and 1984. The impact of Cochrane’s ideas is evident when observing that tens of thousands of controlled trials are now published every year.

The term “EBM” is attributed to Gordon Guyatt at McMaster University in Canada, when, in 1992, his group published a seminal paper outlining the rationale and principles for a “new” philosophy of medical practice.⁴

Over 20 years after Cochrane’s original publication and after observing the importance of the databases accumulated by Cochrane’s group in perinatology, the British National Health Service provided funding for a “Cochrane Center” “to facilitate the preparation of systematic reviews of randomized controlled trials of health care.” Interest in basing clinical decision-making on the results of well-conducted trials grew rapidly during the 1990s as clinicians gained the necessary skills and incorporated EBM principles into their practices. Other health pro-

fessions began to adapt EBM's principles for their own use and incorporation of EBM into academic curricula proliferated between 1996 and 2000.

As of 2002, 15 Cochrane Centers have been established, providing services to 196 countries worldwide⁵ and nearly every health professional faculty in North America formally teaches or espouses EBM principles.

Clearly, the EBM movement was founded upon the need to understand the effects of treatment. From this viewpoint, EBM's principles are directly applicable to practice of pharmacy.

3. Principles of EBM

3.1. The Importance of an Evidence-Based Perspective on Pharmacotherapy

The rapid spread and acceptance of EBM principles in health care can be largely explained by what Malcolm Gladwell would call the "stickiness factor" of the ideas it embodies⁶: namely, that knowing what evidence is available for treatments, sorting out which evidence is the most reliable, and carefully applying that information to patient care situations seems an inherently sensible way to provide health care. Once clinicians are introduced to these ideas, they have a tendency to "stick."

More specifically, the appeal of EBM/EBP may be related to (1) the daily need for valid information about treatment and prevention; (2) the inadequacy of traditional information sources which are either out of date (e.g., textbooks), wrong (e.g., expert opinion⁷), ineffective (e.g., continuing education⁸), or too overwhelming to use (e.g., medical journals); (3) the erosion of clinical knowledge and performance as experience increases²; and (4) the limited time available for reading and study.

EBP encourages critical evaluation of all the available evidence and putting weight on the "best" evidence. This approach accomplishes several things: (1) it fosters awareness of how strongly a therapeutic approach or belief is supported by scientific evidence (if at all); (2) it promotes thorough analysis of the *magnitude* of treatment effects seen in clinical trials and thoughtful reflection about whether these differences are clinically meaningful in the situation at hand; (3) it reduces the potential for caregivers to harm patients by applying out-of-date or unproven treatment strategies; (4) it has the potential to improve the chances of successful treatment by selectively using approaches supported by the best evidence; and (5) it can foster preferential use of the most *efficient* (measured by amount of benefit produced per amount of resources consumed) therapeutic approaches, thereby providing a balance to the forces which promote newer, more expensive

therapies which may not be any more efficient than their predecessors.

3.2. What Is Evidence?

Evidence is everywhere. It is a personal experience with a patient, an adverse drug reaction reported on the TV news, a patient's testimonial about how helpful or useless a drug was for them, a poorly conducted study published in an obscure journal, a well-conducted study published in a reputable journal, an abstract presented at a conference, and an anecdote about an expert's opinion related by a pharmaceutical sales representative. All of these bits of data are evidence. It is obvious, however, that they are all not equally reliable. More to the point, they are all susceptible to different types and degrees of *bias*. Detecting and avoiding bias is a fundamental skill demanded by EBP.

Although this module will mention some specific biases, an illuminating catalogue of biases in analytic research is worth consulting.⁹

3.3. Levels of Evidence

Through identification of the types and degrees of bias inherent in the various research methodologies commonly employed and reported in the medical literature, a hierarchy of these methods based on their validity (relative lack of susceptibility to bias) can be produced.

A commonly accepted scheme pertaining mainly to therapeutic and prevention studies is depicted in

Table 1
Oxford Centre for Evidence-Based Medicine Levels of Evidence (May 2001)¹⁰

Level 1 is considered the most reliable form of evidence and the levels proceed in descending order of validity through 5.

Level	Type of study (of Therapy/Prevention or Aetiology/Harm)
1a	Systematic Review of RCTs
1b	Individual RCT
1c	"All or none" data
2a	Systematic review of cohort studies
2b	Individual cohort study/Low-quality RCT (e.g., <80% follow-up)
2c	"Outcomes" research
3a	Systematic review of case-control studies
3b	Individual case-control study
4	Case series & poor-quality epidemiologic studies
5	Expert opinion based on physiology, bench research, or "first principles"

Table 2
Oxford Center for Evidence-Based
Medicine “Grades of Recommendations”¹⁰

Grade A recommendations are the most forceful.

Grade	Associated Level of Evidence
A	consistent level 1 studies
B	consistent level 2 or 3 studies or extrapolations from level 1 studies
C	level 4 studies or extrapolations from level 2 or 3 studies
D	level 5 evidence or troublingly inconsistent or inconclusive studies of any level

Table 1. This system can be used to weigh different forms of evidence in order to decide which is the “best” evidence. In fact, it is becoming increasingly common to link the strength or “grade” of therapeutic recommendations to the levels of evidence which support them.^{11,12} The Oxford Center for Evidence-Based Medicine promotes the system outlined in Table 2.

3.4. EBM Fallacies and Misconceptions

Sackett was prolific in identifying some misconceptions that EBM’s detractors (and there were/are many) asserted. This resulted in a paper in which he described some things, which EBM is *not*.¹ To adapt and summarize:

- **EBP is not “cookbook medicine”:** Since, by its definition, clinical expertise and patient values must be integrated with external evidence, slavish devotion to algorithmic approaches to therapy are not the ultimate fulfillment of EBM’s promise. Clinical judgement and patient values must dictate whether the evidence is applicable or not in every situation.
- **EBP is not “cost-cutting medicine”:** Applying the best available evidence to pharmacotherapy practice may raise or lower the costs of care. Neither one supersedes the goals of maximizing benefit, minimizing harm, and therefore maximizing efficiency of therapy.
- **EBP is not restricted to areas where randomized controlled trials and meta-analyses exist:** Although these tools are invaluable, the importance of knowing and applying the best available evidence in a given area of therapeutics prevails. Sometimes the best available evidence will be expert opinion. It is critical to recognize what the evidence is, and act accordingly.

4. Putting EBP into Practice

So far we have discussed EBM and EBP as philosophies of practice. The balance of this module will attempt to translate these philosophical tenets into a process that can be implemented by clinicians.

4.1. Formulating Clinical Questions about Pharmacotherapy

The first step in taking an evidence-based approach to a practical problem is to define exactly what the clinical question is that requires answering. It is critical that the questions asked be *answerable*. This is sometimes called “formulating a focused clinical question.” The ability to ask well-formulated questions influences what sources of information you will consult, what types of data you would seek therein, whether you will find a satisfactory answer, and how much time will be consumed in the process.

Clinical questions may be categorized as “background” and “foreground” questions.² Both are necessary at certain times. Background questions are aimed at helping to understand the patient’s condition better and often begin with “What causes...” or “How does...” and end with some aspect of a clinical problem or disorder. Examples include “What causes dry cough in patients on ACE-inhibitor therapy?” or “How does this patient’s dizziness relate to their furosemide therapy?”

Foreground questions tend to ask specifically about how to manage or treat a condition. Examples include “Is a one-day course of antibiotics as efficacious as a three-day course in uncomplicated UTIs?” or “Is an angiotensin receptor antagonist superior to ACE-inhibitor in the prevention of diabetic nephropathy?”

4.2. Background vs. Foreground Questions

As knowledge and experience in a given area grow, clinicians tend to ask fewer background questions and more foreground questions. Classical focused clinical questions asked by pharmacotherapy professionals fall into the following general types:

1. Is drug X effective for treating condition Y? (e.g., Is sotalol effective for converting atrial fibrillation to normal sinus rhythm?)
2. Is the efficacy of drug X worth the harm (adverse effects) it causes in treating condition Y? (e.g., Is cyclophosphamide worth the bone marrow suppression it causes in treating lupus nephritis?)
3. Is the efficacy of drug X worth the cost of treatment for condition Y? (i.e., Is drug X an *efficient* way to treat condition Y?) (e.g., Is azithromycin an efficient therapy for community-acquired pneumonia?)

4. How exactly should drug X be used in treating condition Y? (i.e., What is the actual administration regime?) (e.g., Should pantoprazole be given as intermittent boluses or as a continuous infusion when treating acute upper GI bleeding due to NSAIDs?)
5. What is the place in the therapy of drug X for condition Y? (i.e., is it a first-line or last-line option?) (e.g., Should “triptans” be first-line therapy in patients with migraine?)
6. Can drug X cause problem Y? (i.e., What are its adverse effects, drug interactions, etc.?) (e.g., Can metformin cause Vitamin B₁₂ deficiency?)
7. What other drugs may be helpful in a patient who has not responded to drugs X, Y, and Z for condition B? (e.g., What other drugs may be useful in a patient who still has nausea despite trying dimenhydrinate and metoclopramide?)
8. What is the expected outcome of drug X? (i.e., is drug X having an impact, and how will I know it is working”) (e.g., What is the best way to monitor whether irbesartan is helping my patient with diabetic nephropathy?)

4.3. Locating the Best Available Evidence

It is impossible in the space available to cover even the most important elements of effective search strategies and the depth and breadth of currently available pharmacotherapy information sources. Much of effective searching is specific to the individual data sources, which are constantly changing in name, scope, location, and ease of accessibility. For tutorials on some prominent indexes of pharmacotherapeutic data (e.g., MEDLINE, PubMed, EMBASE, Cochrane Library), readers are referred to more specialized resources.¹³⁻¹⁶

In EBP, information sources can be divided into four general categories:

1. **Primary literature.** This refers to articles of all types published in the medical literature, including editorials, letters, RCTs, narrative reviews, meta-analyses, epidemiologic studies, and guidelines. The primary modes of accessing this data is by searching indexes such as MEDLINE, EMBASE, Cochrane Controlled Trials Register, and more specialized indexes such as CancerLit or AIDSline. This immense and rapidly expanding body of data represents the main “target” for EBP practitioners to be aware of.
2. **Secondary collections of the primary literature.** These sources mainly summarize and comment on prominent studies published in the primary literature. They tend to take the form of periodical digests (e.g., Drug & Therapy Perspectives, InPharma, Evidence-Based Medicine, ACP Journal Club) or more fluid online publications with

daily updates (e.g., Medscape, theHeart.org’s HeartWire). These services greatly assist clinicians in surveying the most prominent developments in the primary literature.

3. **Tertiary sources.** This mainly includes traditional textbooks, an information source which has rapidly become outmoded due to limitations on how frequently they can be updated, the tendency not to be thoroughly referenced, and the tendency not to embrace EBM principles in making therapeutic recommendations. A notable exception to the erosion of textbooks is *Harrison’s Internal Medicine*,¹⁷ a venerable textbook of medicine which is energetically trying to remain relevant through porting its content online, updating it regularly, rapidly integrating clinical trial data into the body of the text, and producing versions for handheld computers. Another excellent online resource of this type is UpToDate (www.uptodate.com).
4. **Other Sources.** The Internet has given rise to a multitude of “niche” pharmacotherapy information sources that do not fit easily into the traditional primary/secondary/tertiary categories. Some of these include Motherisk (www.motherisk.org), Pneumotox (www.pneumotox.com), guidelines.gov, controlled-trials.com, and stroketrials.com. Once discovered (by chance or through systematic internet searching), these sites may dramatically improve the efficiency of access to certain types of information. Unfortunately, they may be difficult to find, change their access policies, move or disappear altogether.

A more complete list of EBM/EBP resources is offered later in this module.

4.4. Critically Evaluating the Evidence

Much of what is discussed in this section has been described in less pharmacotherapeutic-specific terms in a series of articles entitled “User’s Guides to the Medical Literature,”¹⁸⁻²² which have become a generally accepted curriculum for those wanting to learn the basics of critical literature appraisal. However, I have added numerous other issues of particular interest to pharmacotherapy in the following sections.

Furthermore, it is accepted that there are numerous types of pharmacotherapy information (from editorial opinion to meta-analyses of RCTs) and each requires a specialized body of knowledge, skill, and judgment in properly interpreting their findings. Since RCTs represent the “gold standard” of reliable types of information, critically appraising them will form the focus for this module and guidelines for evaluating the quality of other types of data such as “nebulous information found on the inter-

net," though an important skill, will be left for a more appropriate future discussion. Furthermore, for a detailed discussion about the statistical methods forming the underpinnings of RCTs and meta-analyses, readers should consult an appropriate clinical epidemiology reference.²³

4.4.1. Research about Pharmacotherapy for Treatment or Prevention

4.4.1.1. Clinical Trials

Though now generally accepted as the most reliable way to test a clinical hypothesis (e.g., that Drug X is better than placebo for Condition Y), the randomized controlled trial (RCT or clinical trial, for short) has only relatively recently been appreciated as critical to making health care decisions. One of the first RCT reports involved, not surprisingly, drug therapy for tuberculosis,²⁴ and the number of RCTs published annually since 1952 has grown exponentially.

Here I will present a systematic approach to evaluating and interpreting RCTs. This can be accomplished by asking a series of key questions. Some comments on why these questions are important are provided. In some cases, a more thorough discussion of statistics than can be provided here is required to fully appreciate the implications of a "yes" or "no" answer to some questions, and readers are encouraged to satisfy their curiosity by consulting a health statistics reference.²⁵

The questions are divided into 4 categories:

1. What were the aims of the study?
2. Are the results of the study valid?
3. What were the results?
4. Can the results help me in caring for my patients?

4.4.1.1.1. What were the aims of the study?

The answers to these questions should always be found in the METHODS section of a RCT report.

4.4.1.1.1.1. Was it designed to show superiority or non-inferiority (sometimes called equivalence) of one treatment compared to another?

It is important to consider what the study was designed to test. If the study was designed to show that A is better than B (a *superiority* trial) then the methods used must reflect this intent. This generally means that conservative methodologic approaches (e.g., blinding) and statistical approaches (e.g., intention-to-treat principles) should be used. If the intent was to show that "A is equal to B" or, more commonly, "A is not worse than B," then special "equivalence trial" methods are expected. Some key methods to look for here are specifying the "minimally clinically important difference" (MCID),

which the trial is designed to detect, and an indication that the size of the sample was calculated with this MCID in mind. If the size of the difference between A and B is less than some appropriately selected amount for the condition being studied (e.g., less than a 1% absolute difference in rate of death between streptokinase and reteplase), then the two drugs can be considered "equally" or "similarly" effective. Note that this says nothing about whether they are similarly safe, unless that is specifically evaluated in the study. For more details, an excellent summary of this topic has been published by Massel.²⁶

What readers should be most cautious about here is a superiority trial which failed to show a difference between A and B being interpreted by its authors and others as demonstrating that A is *equal* to B. This is a common fallacy and has been criticized on clinical and ethical grounds.²⁷

4.4.1.1.1.2. Were the research questions clearly defined?

Keeping a clear focus on what the primary research question is and distinguishing it from the secondary research questions greatly assists in understanding the results and their credibility. Generally speaking, the primary research question (e.g., Does Drug X reduce mortality in Condition Y?) is what the primary outcome variable (e.g., death) should be based on. Secondary research questions should each have a secondary outcome variable associated with them (e.g., rate of hospitalization, reduction in symptoms). The research questions are usually clearly stated at the end of the introduction section and the primary and secondary outcome variables are usually stated in the "statistics" subsection of the "methods" section.

4.4.1.1.1.3. Were appropriate outcome measures used?

This is where readers must decide whether the primary outcome variable is actually appropriate. This depends mostly on what the research question associated with it is. For example, if the primary research question is "Does drug X prevent ischemic strokes in patients with previous stroke?," then an appropriate primary outcome variable would be "incidence of ischemic stroke." An *inappropriate* outcome variable for that question would be "incidence of hospitalization for any cerebrovascular event." That outcome variable might be appropriate for a secondary research question such as "Does drug X reduce morbidity related to all cerebrovascular events?"

A classic pitfall here is when less clinically meaningful research questions and outcome measures are used and promoted as being more meaningful than they are. Imagine the headline "Drug X is superior to Drug Y in patients with hypertension." This could mean a variety of things, but in hypertension

research, it could be argued that the most clinically meaningful research question and outcome variable is mortality, or even “heart attack or stroke.” Frequently, however, the headline above results from a study which showed that Drug A lowered systolic blood pressure more than Drug B. Whether this is clinically meaningful is debatable. Examples of using these “surrogate outcomes” are plentiful in nearly every area of pharmacotherapeutics and readers are encouraged to interpret them appropriately.

4.4.1.1.1.4. *Were the comparator interventions appropriate, including duration?*

Readers must decide whether a study comparing Drug X to placebo really helps very much in determining whether Drug X is useful or not. Sometimes what is really needed is a comparison between Drug X and Drug Y. On the other hand, what if Drug Y has never been shown to be superior to placebo? In that case, how helpful is a trial showing that Drug X is equal to Drug Y? Perhaps not very. The same may be true if Drug X is compared to Drug Z, when Drug Z is not used any more because it is unsafe or not available.

No generalizations can be made about the usefulness of placebo controlled versus active-controlled (also known as “head-to-head”) trials here, since each area of drug therapy is unique in this regard.

Readers must also critically evaluate whether the *duration* of the comparison between Drugs A and B (or placebo) is appropriate. Clinical trialists generally want to make trials as short as necessary to answer the research question. Sometimes, though, they are not long enough. Consider a 4-week comparison of Drug X versus placebo for Alzheimer’s dementia. This may not be long enough to see any effect of the drug. On the other hand, if an effect was shown, such a short trial would not help to detect whether the effect was short-lived or persistent. Hence, the duration of RCTs must be appropriate given the chronicity or acuity of the condition being studied. For this reason, studies of prevention of heart disease must last for several years, since there is virtually no chance a drug can prove better than placebo or another drug in a 1-month period when the incidence of heart attacks is very low during such a short time period.

Choosing the right duration of the experiment therefore protects against mistakenly saying the drug is not effective when it really is (called a “Type II error”) as well as mistakenly saying a drug is effective when it is not (a “Type I error”).

4.4.1.1.1.5. *Was the sample size appropriately determined?*

This is an area that is difficult for most clinicians to fully appreciate because of its direct relation to the

inferential statistical techniques used in most clinical trials.

In short, it is important to discern whether the investigators formally calculated the size of the sample population they would require in order to answer the primary research question *before* they actually started doing the trial. This information, if it is reported, will be contained in the “methods” section of the RCT report.

The principle here is that trials must have enough *power* to detect a meaningful difference between treatment A and B, if there really is a difference between the two. That way, if *no difference* is found, readers can confidently conclude either that the two treatments are similar (sometimes even “equivalent”) or at least that the difference between them is smaller than some amount specified in the sample size calculation the authors describe. Readers must then decide if this potential small difference is less than a clinically meaningful difference. The *power* of a trial is the probability that it will detect a difference between A and B, if there really is a difference between them. Power is increased by larger sample sizes (and decreased by smaller ones) and increased by smaller *variance* in the outcome variable (and decreased when there is larger variance). Power is, by convention, set at 80 or 90 percent when the investigator does a sample size calculation. For example, in a study designed to have 80% power to detect a difference between Drug A and placebo, the authors have a 20% chance of not finding a difference between them even though a difference may really exist (a Type II error).

4.4.1.1.1.6. *An important nuance*

There is an important nuance here which is frequently misunderstood by clinicians: When a superiority trial successfully demonstrates a difference between Drug A and B, criticizing the investigators about having too small a sample size or not providing a sample size calculation is nonsensical. The simple fact is that, because a difference was shown, they clearly had a large enough sample to detect a difference. The “power argument” and speculation about a Type II error ends there (by definition, they’ve not “failed to detect a difference when one really exists”). Criticisms related to power are generally only meaningful when a trial *fails* to show a difference. In that case, readers can question whether there may really be a difference and the trial simply failed to detect it because it didn’t have enough power.

4.4.1.1.1.7. *Who funded the study?*

When evaluating an RCT report I always look for an indication about who sponsored the study. The primary distinction I am trying to make is whether it was funded by the manufacturer of the experimen-

tal drug (extremely common), the “control” drug (extremely rare), or a more impartial third party, such as a public funding agency like the Canadian Institutes of Health Research the National Institutes of Health, or the British Medical Research Council.

At issue here is whether the sponsor might have any reason to be inherently biased toward the study producing one result or another. The study must still absolutely be judged on its own merits and the answers to all the other questions posed here. However, when weaknesses are detected, one should carefully consider whether those weaknesses might favour the sponsor’s bias (the usual case), the opposite view, or neither. This often influences how enthusiastic you are about the results of the trial.

Publication bias asserts that studies showing positive results are far more likely to be published than those showing negative or neutral results.²⁸ Furthermore (and possibly related to publication bias), published studies comparing one drug to another tend to favour the sponsor’s drug more often than the comparator agent.

The study’s authors may also have significant conflicts of interest. An extreme (and common) example of this is when investigators are employees of the sponsoring company. Journals are becoming increasingly rigorous about requiring authors to declare actual or potential financial conflicts of interests and publishing them when they are present.

Thus, readers are encouraged to exercise healthy skepticism in light of the identity of the sponsor and author conflicts of interest.

4.4.1.1.2. Are the results of the study valid?

4.4.1.1.2.1. Was assignment to treatment randomized?

Since we are discussing randomized controlled trials, it is expected that subjects were randomized. The reason why this is important is often overlooked, however, so we will specifically mention why randomization is important.

Randomization (assigning subjects to receive one treatment or another based on a factor such as a coin toss or a random number which is totally unrelated to themselves or their condition) primarily protects against bias resulting from a researcher deciding which treatment a subject should receive. This opens the possibility that the outcomes of the treatment were not related directly to the treatment itself, but to some other factor such as disease severity, age, other treatments, etc. This is called *procedure selection bias*.⁹ Randomization ensures that all the patient factors which might influence outcome get equally distributed between the two groups being studied, so if differences in the effect of the treatments are seen, they can confidently be attributed to the treatments and not some other known or

unknown factor (a “confounder”).

It has also been shown that lack of randomization results, in general, in larger treatment effects.²⁹ Hence, with only a basic understanding of the underlying principles, an astute reader could suspect that the treatment difference shown in a non-randomized trial may be overestimated.

4.4.1.1.2.2. Were all patients assigned to treatment properly accounted for at the conclusion?

Readers should read the “results” section carefully for a description of how many patients were randomized and how many were later excluded (and for what reasons) and ensure that the number of patients eventually included in the outcome analysis adds up appropriately. Of concern here would be (1) disappearing patients who are not accounted for, and (2) disproportionate dropouts between the two groups.

Recent guidelines for how RCTs should be reported are intended to eliminate problem 1.^{30,31} One way of dealing with unaccounted-for patients is to assume that they had the worst possible or “most conservative” outcome. For the experimental treatment group, this usually means assuming that they got no benefit, and for the control arm, this usually means assuming that they did benefit. The courageous can try recalculating what the results would have been under such circumstances.

4.4.1.1.2.3. Were patients analyzed in the groups to which they were randomized?

This question is meant to prompt an assessment of whether *intention-to-treat* (ITT) analysis principles were used in the trial. One can simply look for this statement in the methods section and be satisfied if it appears. More critical readers will want to understand why and when ITT is important.

ITT simply means that all subjects who are randomized to one treatment or another get analyzed at the end of trial according to the treatment to which they were *originally assigned*, regardless of whether they dropped out, were noncompliant, or started taking the other group’s treatment (or one like it). Although this may seem unrealistic, there are two primary reasons for insisting on this methodology in superiority trials (it is less important or even undesirable in equivalence trials and safety studies):

1. Since it is accepted that compliance with treatment is closely linked to prognosis in some conditions,^{19,32} *not* using ITT may negate the effects of randomization! In other words, you randomized the subjects because you wanted to eliminate those confounding factors which influence which treatment they take and their chances of an outcome event, so why would you let the patients

themselves later decide which treatment to take (e.g., to stop taking the experimental treatment or stop taking placebo and start taking some other treatment similar to the experimental treatment) and analyze them according to their choice (what they *actually* took) rather than what you assigned them to take? Such a non-ITT methodology seems invalid in that light.

2. ITT is a more *conservative* way of analyzing the data in that it decreases the chances of finding a difference between the two groups, even if one exists. In this way, it provides desirable protection against mistakenly saying there is a difference between A and B when there really isn't one (Type I error). Similarly, it increases the chance of a Type II error (finding no difference when one truly exists) because it tends to increase the variance in the outcome variable, and hence decrease power. Variance tends to increase in ITT analyses because in, for example, the experimental arm, there is a mixture of those who are still receiving the experimental treatment and those who are not. Both of these effects occur because as subjects "drop out" of the experimental group (i.e., start being treated like placebo or control group patients) and/or "drop in" to the experimental group or one similar to it (e.g., start taking anti-hypertensive therapy, even if it is a different drug than the one being studied) the two groups become more similar to each other in terms of their chance of experiencing whatever the outcome variable is (e.g., hypertension-related stroke). This is the conservative and generally desirable bias referred to above.

Readers are challenged to consider, using the two reasons above, why an ITT methodology might be *undesirable* in equivalence and safety trials.

4.4.1.1.2.4. Was blinding to treatment used?

It is commonly accepted that it is important for subjects to not know which study treatment they are receiving. We inherently understand that this reduces the chances of their own beliefs about whether the treatment is helpful or harmful interfering with the measurements in the various outcome variables.

Researchers and clinicians are susceptible to a similar set of biases related to their own beliefs and expectations about treatments (therapeutic personality bias). They may be influenced by these expectations in how they ask questions of subjects (recall bias) or how intensively they monitor them (expectation bias).

Thus, blinding of subjects and/or investigators may eliminate several biases to which both groups are prone. These include recall bias, rumination bias,

therapeutic personality bias, expectation bias, and obsequiousness bias.

It is commonly believed that when so-called "hard" endpoints such as mortality or cholesterol levels are used, blinding is less important because subjects and researchers are unlikely to be able to influence the occurrence or magnitude of these events or results. This is only partly true, since numerous influences are still possible, including the frequency with which subjects are monitored or queried about their condition, the propensity to interpret a finding as "significant" depending on personal bias, a subject's or researcher's sense of hopefulness or hopelessness about an outcome (which could even affect mortality), or stress levels (worry about knowing you are receiving placebo, even though you have a disease). Importantly, the adverse effect profile of a drug can be strongly influenced when blinding of both parties is not employed, since most adverse effects are subjective in nature and dependent upon the subject's reporting them and the researcher's enthusiasm in "extracting" such information from them.

When blinding is not used, readers must judge for themselves whether any of these biases might have come into play, and in what direction they may have influenced the results. Furthermore, readers should consider whether blinding could have been used. In some situations it is difficult, but not necessarily impossible, to properly blind subjects. Examples include comparing traditional antipsychotics, which have obvious and characteristic side effects, to atypical ones, which do not; or comparing warfarin, which requires regular blood tests and dosage adjustments, to aspirin, which does not.

Another important aspect of blinding that is frequently overlooked is whether the blinding was actually maintained throughout the trial. Just because the researchers *tried* to blind themselves and the subjects doesn't mean they were successful in doing so. If they were not, all of the biases mentioned above are in play. In some cases, such as in the landmark Women's Health Initiative study, 47% of the hormone-replacement recipients had to be unblinded *per protocol* because of persistent vaginal bleeding. It could be argued that this may have influenced the efficacy and safety results in either direction. Formal assessments of blinding are infrequently reported in RCTs but may be very important in some cases.

4.4.1.1.2.5. Were the groups similar at the start of the trial?

Since a multitude of patient characteristics may affect the likelihood of an outcome event in an RCT, it is important to confirm that the important factors were similarly distributed between the two groups

during the randomization process. This is usually easily done by looking at the table in most RCTs, which describes the baseline characteristics of the study patients. If randomization failed to equally distribute, for example, patients who smoked between the two groups, readers may speculate on how that may have affected the results, if at all. Similarly, if the investigators do not report on a factor that the reader knows may be important to prognosis, the reader may have an appropriate amount of skepticism about the results (although this is usually a *small* amount of skepticism since whatever the factor is, because of randomization it is very likely to be distributed evenly between the groups).

4.4.1.1.2.6. Were the groups treated equally, aside from the experimental intervention?

In RCTs one expects and requires that, aside from the experimental intervention, the patients were treated and evaluated similarly in every respect. This is what gives confidence that differences in the effects seen are, in fact, due to the experimental treatment and not some other factor.

Some scenarios to watch out for are: In studies involving antibiotics, were other antibiotics besides the study ones allowed at the discretion of the physician? In arthritis studies, were other medications besides the study one(s) being used and dose-adjusted at the same time? In cardiology studies, were patients who didn't respond to the study treatment getting angioplasty or bypass surgery, and therefore having their outcomes influenced by something other than the study drugs? All of these examples include an element of *cointervention*, which can obscure or exaggerate the real effects of the study treatments.

4.4.1.1.3. What were the results?

4.4.1.1.3.1. How large was the treatment effect?

EBM's forefathers placed the bulk of the justification for EBM's philosophy on the importance of understanding the effects of therapies, their magnitude, and their clinical relevance.

In this section, we will briefly describe how to extract maximum meaning from the most common type of data presented in RCTs. The primary outcome variable in most RCTs is a *binary* variable (also known as a *nominal* variable). In other words, each research subject either experiences or does not experience an event, regardless of whether it is a death, hospitalization, successful treatment outcome, adverse event, or any other event. Thus, a percentage of patients experiencing the event in each treatment arm ends up being the main analysis of this outcome variable.

The data are usually presented in a manner sim-

ilar to this, where an event is undesirable (say, a gastrointestinal bleed) and Drug X is designed to prevent it:

	Drug X (n=2234)	Placebo (n=2105)
Number of patients with an event (%)	352 (15.8%)	625 (29.7%)

Believe it or not, a wealth of information and different ways of representing and describing this result can be extracted from this simple table. These different "views" of the data can be helpful in deciding whether the treatment effect is meaningful to you or to your patient.

The best way to become comfortable with these simple number manipulations is to practice doing it yourself with every study you read. Forget about trying to memorize what look like formulas here, but use them as a template you can apply to any kind of study that takes this form.

To make the simple calculations involved as clear as possible, it may help at first to think of the data in a slightly modified form like that below:

	Drug X (experimental group)	Placebo (control group)
Number of patients with an event	A	B
Number of patients without an event	C	D

It should be readily apparent that the "n" for Drug X (the total number of people in the Drug X arm) is equal to A+C. For placebo, this would be B+D.

What you can calculate:

- Percentage of subjects with an event in the Drug X group = $A/(A+C)$. In this case, 15.8%.
- Percentage of subjects with an event in the Placebo group = $B/(B+D)$. In this case, 29.7%.

Absolute Risk Reduction (ARR) = The percentage of subjects who had an event in the Placebo (control) group MINUS the percentage of subjects who had an event in the Drug X (experimental) group.

- $ARR = (\% \text{ events in Placebo}) - (\% \text{ events in Drug X}) = (B/(B+D)) - (A/(A+C))$

In our example, the ARR is $29.7\% - 15.8\% = 13.9\%$

Number Needed to Treat (NNT) = $1/ARR$. This signifies the number of patients who must be treated with the experimental treatment in order for one

patient to receive a benefit (e.g., to avoid a bad outcome event or experience a good outcome). Note that the ARR must be expressed in *decimal* form (i.e., the %/100, in this case 0.139) when using this equation. In this example, the NNT would be 7.19. It is conventional to round fractional NNTs to the next highest integer, since a part of a person cannot be treated. Hence, in this case the NNT would be 8. This would be appropriately phrased as “8 patients must be treated for Z years [insert duration of the trial here] with Drug X instead of placebo in order for one GI bleed to be avoided.” Where the experimental treatment produces *harm*, the NNT becomes a **number needed to harm (NNH)**, and the statement would be phrased as “8 patients must be treated for Z years [insert duration of the trial here] with Drug X instead of placebo in order for one event to be *caused*.” As illustrated in these examples, it is critical to include the *duration* of treatment in any statement involving an NNT since the ARR (and hence, the NNT) is highly dependent on the duration of therapy.

Relative Risk (RR) = The percentage of subjects who had an event in the Drug X group *divided by* the percentage of subjects who had an event in the Placebo group.

$$\bullet \text{ RR} = \frac{\% \text{ events Drug X}}{\% \text{ events Placebo}} = \frac{A/(A+C)}{B/(B+D)} = \frac{15.8}{29.7} = 0.53$$

Relative Risk Reduction (RRR) = 1-RR. In this case the RRR would be 0.47, which is usually phrased as “Drug X produced a 47% relative risk reduction compared to placebo.”

Odds Ratios (OR) are sometimes used in reporting of RCTs, although they are more commonly used in case-control epidemiologic studies and some types of meta-analyses. They are included here for completeness

- OR = odds of an event in the experimental arm *divided by* the odds of an event in the control arm.

$$\bullet \text{ OR} = \frac{A/C}{B/D} = \frac{0.187}{0.422} = 0.44$$

It is common in the medical literature for authors to phrase treatment effects only in terms of relative risk reductions. It is also known that using RRR values in advertisements and presentations is more likely to influence prescribing habits (among physicians and pharmacists) and patient acceptance of treatment,³⁵⁻³⁷ likely due to the more “dramatic” RRR figure compared to the ARR. When treatment

effects are presented as NNTs, they reduce the likelihood of prescribing or taking a drug, compared to RRRs.

Example of how different ARRs can result in the same RRR, with possibly different conclusions about the “importance” of the effect:

Event rate (treatment vs. placebo)	RRR	ARR	NNT
1% vs. 2%	50%	1%	100
20% vs. 10%	50%	10%	10
80% vs. 40%	50%	40%	2.5

Using these simple manipulations, clinicians can better understand the magnitude of treatment effects and the likelihood that an individual patient will benefit or be harmed.

For example, let’s say the event in our example above is stroke and the trial was conducted over a 1-year period. You have a patient who you know is at risk for stroke so you are considering using Drug X for prevention. If you believe (based on your clinical experience, patient’s history, and lab tests, combined with any other available risk assessment tools) that your patient’s risk of stroke is a little lower than the trial population’s (the placebo group) over the next year (say, 20%), then you can make a number of patient-specific estimations based on the study above:

- If you don’t treat your patient at all, they have a 20% chance of having a stroke in the next year (your own informed estimate).
- Using Drug X reduces the chance of stroke in people similar to your patient by 47% (the RRR from the trial).
- If you treat your patient with Drug X, their chance of a stroke in the next year will be reduced to 10.6% (20% × the RR (0.53) = 10.6%).
- Thus, in your patient, the absolute risk reduction (ARR) would be 9.4% (20%–10.6%).
- The number of patients *exactly like your patient* who would have to be treated for 1 year with Drug X in order for one stroke to be prevented would be 11 (the NNT, which is 1/ARR (0.094)).
- Perhaps more meaningfully, the chance that your patient will benefit in the form of avoiding a stroke by taking Drug X for 1 year is 1 in 11.
- In other words, there is a 10 in 11 (~90%) chance that they will either not have a stroke or have a stroke despite taking Drug X during the next year.

Armed with this **best available evidence** (and the various “views” of the data), combined with your **clinical expertise** (what you used to identify

that your patient was at risk and how much) and the **values of the patient** (e.g., they can't afford Drug X, Drug X has annoying side effects), a decision can be made as to whether to treat your patient with Drug X or not. This integrative process is the essence of Evidence Based Pharmacotherapy.

A note about non-nominal variables:

When the outcome variable is something like blood pressure, cholesterol level, or score on a Quality of Life questionnaire, it is called an *ordinal* or *continuous* variable and is subject to a whole other realm of statistical tests and interpretations, which are beyond the scope of this module.

Other outcome measures such as Hazard Ratios are sometimes used in RCTs. Understanding and interpreting these requires some background in proportional hazard models and Kaplan-Meier plots and is best tackled by more advanced or curious readers.

4.4.1.1.3.2. *How precise was the estimate of the treatment effect?*

Now suitably armed with the tools to fully understand the *magnitude* of treatment effects in RCTs, we must have some appreciation for the *precision* of these measurements. In other words, it is fine to say that Drug X produces a 47% relative risk reduction, but how confident are we that 47% is an accurate estimate of the "real" effect of Drug X? To appreciate this, we must have some knowledge of *confidence intervals*.

The common confidence interval (CI) presented in RCTs is the 95%CI. A 95%CI can be calculated around any of the ARR, RRR, RR, or NNT/NNH. The 95%CI is defined as the range of values within which we are 95% sure the true population value lies.² It is important to realize, however, that not all results within this interval are equally likely. The further away (in either direction) from the mean you go, the less likely that result is. Smaller confidence intervals imply more precise estimates and vice-versa. CI's are made smaller by larger sample sizes and smaller amounts of variance in the outcome variable. They get larger as the sample size shrinks and variance increases.

It is not mathematically difficult to calculate CI's, although the formulae involved are long enough to be difficult to memorize. Readers are referred to other sources for this information.^{2,26,38}

When the 95%CI range around an ARR includes zero (i.e., no difference between the two groups), this can be taken to mean that there was no statistically significant difference between the groups in this parameter. For OR's and RR's, lack of statistical significance is implied when the 95%CI's around the OR or RR include 1.0.

In practice, some simple questions can be asked: Is the lower limit of the confidence interval around the ARR or RRR clinically meaningful? If not, your confidence that the treatment effect is clinically meaningful may be decreased. For an RR, if the upper limit of the 95%CI approaches or includes 1.0 in an efficacy study, your confidence that this is a meaningful treatment effect may be diminished. In an equivalence trial, if the 95%CI's around the ARR or RRR go beyond your self-designated minimally clinically important difference (MCID), you may conclude that the two treatments are not similar at all.

For example, consider that an RCT reported an absolute difference between amiodarone and sotalol for the incidence of atrial fibrillation at 1 year of 20% favouring amiodarone and the 95%CI around this difference is reported as "(2-38%)." This can be interpreted to mean that the difference between amiodarone and sotalol is statistically significant (because the 95%CI does not include zero); however, you may decide that the difference is of questionable clinical significance since the difference between the two treatments may be as small as 2%. A difference this small is much less likely than the 20% difference reported, but it is possible based on the data. The same is true of a 38% difference, which may cause you draw the opposite conclusion and to heavily favour using amiodarone over sotalol in your patients.

The best way to appreciate CI's is to practice by thinking about these ideas every time you read an RCT.

4.4.1.1.3.3. *For composite endpoints, which endpoints "drove" the result?*

Many trials, particularly in cardiovascular medicine, use a *composite endpoint* as the primary outcome variable. This means that the primary outcome variable is something like "death + recurrent MI + stroke." There are various methodologic reasons why composite endpoints are used, the most compelling being that a smaller RCT can be conducted since the event rates in both groups being studied will be higher than if, for example, only death was used.

When a trial reports that there was a beneficial effect of Drug X on the composite endpoint of death + recurrent MI + stroke, astute readers should dig further, usually by consulting the table in the report entitled something like "all endpoints" or "secondary endpoints" to see whether there appear to be any important differences in the individual endpoints of death, recurrent MI, or stroke. It is common to find that the difference in the composite endpoint is driven not at all by any difference in death, but by the far more common events like recurrent MI or

stroke. Appreciating this avoids the misconception that the therapy does anything to reduce mortality, and makes it clear that it mainly prevents the other events named. These other events may still be very important, but not as important as prolonging life.

4.4.1.1.4. Can the results help me in caring for my patients?

The following questions address the most critical step in the EBM process, wherein clinicians decide whether the available evidence (thoroughly analyzed and critiqued) is applicable to the patient at hand.

These questions apply to almost all the different research methodologies and not just to RCTs.

This portion of the process is considerably more subjective than the other more analytical steps just discussed.

4.4.1.1.4.1. What was the patient population?

It is critical to have a thorough understanding of the patient population studied in an RCT so the similarities and differences compared to the patients you are trying to make therapeutic decisions about are taken into consideration.

Properly reported RCTs should clearly describe the patient population they *planned* to enrol in the “population” section of the “Methods.” The plan is described in terms of *inclusion* and *exclusion criteria*, which should give a fairly clear sense of who should and should not have been enrolled in the trial. The first portion of the Results section will generally describe who was actually enrolled in the trial.

Giving some thought to this information will help readers answer questions like: Compared to the kinds of patients with the same condition which I care for, were the RCT patients similar in age? In disease severity? In risk factor profile? In race? In the kinds of medications they were taking at the same time as the study medication(s)?

A couple of tips for astute readers, which are often not fully appreciated, are:

1. In a *prevention* study (e.g., stroke prevention), carefully discern whether the RCT is a *primary* prevention (i.e., the subjects had never had a stroke before) or *secondary* prevention study (i.e., all subjects had already had a stroke previously). In general, it is less appropriate to extrapolate positive results from a secondary prevention trial to patients who have never had the event before (e.g., giving aspirin to patients at risk of MI, but who’ve never had one before, on the basis of post-MI aspirin studies).
2. In a *treatment* study, pay close attention to whether the patients studied were considered refractory (i.e., had not responded) to other treatments. This is often not clearly stated, but is criti-

cal in deciding whether the results apply to your patient, who may or may not have tried anything else for their condition. It is generally less appropriate to extrapolate results from a study involving patients who had never tried anything else before (e.g., newly diagnosed depression) to patients who have not responded to several other agents (e.g., refractory depression). These can generally be considered two different patient populations, and readers are cautioned against extrapolating from one to the other.

4.4.1.1.4.2. Were all clinically important outcomes considered?

We asked a question similar to this under “What were the aims of the study.” However, at this point we are expanding our scope and wondering whether, besides the endpoints which were studied in the RCT (which may be perfectly appropriate and the most important), there are any other endpoints which the investigators might have considered that would help you in deciding whether to apply the results of the study to your patient.

Commonly neglected endpoints include, depending on the condition being studied, Quality of Life measurements, length of hospital stay, number of hospital admissions for relapses of the condition, and number of patients who withdrew due to adverse effects of the medication(s). Each condition studied has its own set of relevant outcome variables, so this list is offered only as food for thought.

Readers are cautioned against being too critical of investigators who do not include an outcome variable, which is probably of only marginal importance given the research question(s) posed in the study. For example, readers could criticize the Women’s Health Initiative investigators for not including “cognition” as an outcome variable in their initial report.³⁴ However, it could be debated whether this is an important parameter in a study aimed at detecting cardiovascular and breast cancer risks or benefits of hormone replacement therapy. It may be outside the scope of the study, or it may have been a missed opportunity, depending on your viewpoint. Consider, also that the investigators may be planning to publish the results about a particular outcome as a separate publication in the future.

4.4.1.1.4.3. Was subgroup analysis used?

Subgroup analysis is, as its name implies, an attempt by the investigators to analyze subpopulations within the study population for a particular outcome. For example, in a study that included patients from 40 to 80 years old, the investigators may do a subgroup analysis on the >70-year-olds to see whether there was any effect of the intervention or to compare the size of the effect with that in the younger patients.

Much is said about the limitations of doing such analyses, but they can be summed up in the following guideline: Unless the authors state that they planned to do a specific subgroup analysis (in the methods section), then any such analyses reported in the results should be considered “hypothesis generating” rather than conclusive.

There are numerous statistical arguments why such analyses are less reliable than the primary analyses.³⁹ Without in-depth knowledge of these arguments, readers can apply the following practical tips when deciding how much weight to place on subgroup analysis results. Subgroup analysis results are more likely to be “real” when all or many of the following are answered in the affirmative:

1. The difference found is very large.
2. The difference is unlikely to occur by chance.*
3. The difference was a pre-specified hypothesis.
4. It is biologically plausible that the effect could occur.
5. It is replicated in other trials.
6. It is one of a small number of subgroup analyses conducted.†

4.4.1.1.4.4. Assessing clinical vs. statistical significance
Readers are likely comfortable with the idea of statistical significance. We are accustomed to checking to see if, for example, the “p-value” attached to a particular result is “less than 0.05” or “that the 95% CI around the relative risk does not include 1.0.” Practically speaking (though not quite statistically true[‡]), this provides us with greater than 95% assurance that the result seen did not occur by chance.

Equally important is for readers to decide whether the result seen is clinically significant. In other words, “is an effect of this size likely to be meaningful to my patient?”

Very often, statistically significant results are generally accepted as clinically meaningful as well. However, a classic example where this could be questioned is the CAPRIE trial comparing aspirin to clopidogrel for the secondary prevention of stroke, MI, or amputation (due to peripheral vascular disease).⁴⁰ This study found that after 3 years of treat-

ment, the incidence of this composite endpoint in the clopidogrel arm was 5.32% and was 5.83% in the aspirin arm. This result was statistically significant (there were over 19,000 patients in the trial) and translated into an 8.7% relative risk reduction favouring clopidogrel (you can calculate this yourself). Deriving more “views” of the data, this is a 0.51% ARR, which translates into requiring that 196 patients be treated with clopidogrel instead of aspirin for 3 years in order for a single stroke/MI/amputation to be prevented. This may cause readers to conclude that although statistically significant, the difference between aspirin and clopidogrel for this efficacy parameter is not very *clinically* significant. Deciding on the role of clopidogrel, then, requires answering the next question...

4.4.1.1.4.5. Do the potential benefits outweigh the potential harms and costs and/or is the treatment feasible in our setting?

This is the final question evidence-based practitioners must ask themselves when deciding whether to apply the results of a RCT to their patient’s care. The answer to this question requires that the effect, which we have spent the majority of our time focusing on, be weighed against the harms, costs, and other practical implications of the therapy.

Clinicians are generally well versed in assessing adverse effects, considering their seriousness and frequency, and balancing them against the nature and size of the effects of the therapy. This is the essence of evaluating the usefulness and applicability of a therapy.

More recently, methods for formal assessment of the economic aspects of pharmacotherapy have emerged. Pharmacoeconomic research has become a cottage industry of sorts and it allows the effects of drugs to be assessed from the viewpoint of resources consumed (cost of drugs, cost of medical care and monitoring, treatment of adverse effects) versus outcomes improved or avoided (prevention of medical expenses such as hospitalization and physician visits, increased productivity or quality of life). Pharmacoeconomics includes three main methodologies: cost-minimization analysis, cost-utility analysis, and cost-benefit analysis. All of these methodologies are commonly referred to as *cost-effectiveness* analyses. This allows conclusions to be drawn such as “It costs the provincial Ministry of Health X dollars to prevent one gastrointestinal bleed by prescribing omeprazole to all high-risk patients on NSAIDs.” Clinicians, politicians, policy analysts, and taxpayers can decide whether that dollar figure is worth spending or not. Sometimes, pharmacoeconomic analyses allow the conclusion that “Using drug X results in an overall cost *savings* of X dollars per patient treated.”

* This can be further assessed by determining whether appropriate statistical adjustment was done. These adjustments are called adjustments for *multiple comparisons*, of which Bonferroni’s Correction is one example.

† The general rule here is that the more subgroup analyses you do, the greater the chance that one of them will turn out to be statistically significant, regardless of whether it is a “real” effect or not. Hence, fewer subgroup analyses are better.

‡ The statistically correct meaning of the p-value is the likelihood of a result (i.e., a difference) as extreme as or more extreme than the one seen occurring, given that the null-hypothesis is true (i.e., assuming that there is no difference between the two treatments).

Some practical issues that may limit the ability to employ a therapy include an onerous administration schedule, impractical route of administration, or impractical monitoring requirements.

4.4.2. Special Considerations in Interpreting Research about Harms of Drug Therapy

The discussion so far has focused mainly on studies designed to assess the *efficacy* of pharmacotherapy. Sometimes studies are specifically designed to evaluate differences in *safety* between one therapy and another, or between a therapy versus no therapy.

These kinds of studies turn some of the methods we've described as desirable for efficacy studies upside down. In efficacy studies we want the methods to be rigged *against* showing a difference (so that if a difference is found, we can more strongly interpret it as being "real", e.g., intention-to-treat methodology) and we call this approach *conservative*. In safety studies, investigators and astute readers take the opposite approach, which says "if there really is any difference in safety, I want this study to be able to detect it." This means you will be more likely to accept a Type I error (saying there is a difference when there really isn't one) and less likely to accept a Type II error (saying there is no difference when there really is one). Generally, if there are safety issues at stake, we want to know about it and are more willing to err on the side of *overestimating* the risk than on underestimating it. You can see, therefore, that *conservatism* in safety studies has the opposite definition to that in efficacy studies.

Practically speaking, this means that readers should look for results of "on-treatment" analyses (what happened to the patients who actually took the medication properly, which is the opposite of intention-to-treat analysis). It also means that the statistical corrections for multiple comparisons are less desirable since they are used to guard against "random" detection of associations, which you may actually want in safety studies. In some cases, using p-value thresholds for significance which are greater than the usual 0.05 (say, 0.10) may be appropriate, since even a less impressive *trend* toward a safety concern may be clinically interesting.

4.4.3. Critical Appraisal of Systematic Reviews

Another commonly encountered type of research in the medical literature is the systematic review. Systematic reviews come in two primary forms, qualitative (or narrative) and quantitative (or meta-analysis). In general, systematic reviews aim to synthesize the available data related to a specific research question. Depending on the question and the quality and quantity of data available, either a qualitative or quantitative approach to synthesizing it may be appropriate.

Several excellent articles to assist readers in understanding the methods involved and critical evaluation of qualitative and quantitative systematic reviews are available.⁴¹⁻⁴⁶

4.4.4. Practice Guidelines

Practice guidelines are usually the result of a gathering of experts in which they attempt to produce specific recommendations about how to diagnose and/or treat a particular condition. These guidelines may be based on an exhaustive review all available evidence on the topic combined with the experts' clinical opinions. When done properly, guidelines may represent a very useful tool for clinicians to understand and apply the best available evidence to their practice. When not done properly, they can be used as a mechanism for pharmaceutical companies to influence prescribing of particular products.⁴⁷ Readers are advised to pay close attention to who sponsored the gathering of experts as a possible indication of bias in the recommendations, which resulted.

Because of the tremendous variability in the quality and transparency of the process which leads to the recommendations in practice guidelines, readers are advised to educate themselves about how to critically appraise them.⁴⁸⁻⁵⁰

4.5. Bringing It All Together

As the definition of EBM implies, the detailed critical evaluation of the evidence we've just discussed is only one-third of the process of practicing EBP. Before any patient care decision can be made, the implicit process of bringing your *clinical skills* and experiences to bear and discovering and considering your *patient's values* must be invoked.

As our example has shown, the *clinical skill* element involves many things, including assessing what condition your patient has, assessing its severity, how urgent the need is for therapy, whether the available evidence applies to this patient, what other relevant conditions the patient has, what their individualized level of risk is, and therefore what their individual probability is that they will benefit or be harmed by the therapy.

The *patient values* element involves assessing, among many other things, whether the patient is likely to accept drug therapy or feel uncomfortable with it, whether there are cultural factors which would make them more or less likely to want to treat the condition, whether they can afford a therapy, their stage in life and sense of well-being, and even their cognitive ability to be involved in the decision-making process.

5. Beyond the “Method” of EBP: EBP as a Professional Lifestyle

Considering the apparent complexity of the process of evaluating the evidence, evidence-based practitioners quickly learn that it is not practical to answer every clinical question at the time it arises using the approach described above. This approach may be called a “just in time” approach to EBP, and although often necessary, it is simply too time-consuming.

This leads to the daunting conclusion that clinicians must be aware of and have evaluated the best available evidence in the necessary areas *before* they actually need to apply it. This can be called a “just in case” approach. This requires a professional lifestyle of keeping abreast of developments in the relevant literature and casts pharmacists in the light of “information managers.”

An entire module (or several) could be devoted to the topic of pharmacotherapy information management. This would include deciding what literature and resources to survey, how to assess their reliability, how to process and store the information for later retrieval, and tools to facilitate retrieval at the “point of curiosity” or point of care. For our purposes, we will direct readers to several useful online resources that can be used to more efficiently facilitate the information management (or “just in case”) process. Readers are referred to the individual services for instructions on how to use them.

Surveying the Primary Literature and Conference Proceedings:

Medscape (Reuters Health) News
(www.medscape.com)
MDLinx Newsletter (www.mdlinx.com)
Journal Watch (www.jwatch.org)
theHeart.org Heartwire (www.theheart.org)
ACP Journal Club (www.acpjournals.org)
The Journal of Informed Pharmacotherapy
(www.informedpharmacotherapy.com)
Ingenta.com’s TOC Alerts and Research Alerts
(www.ingenta.com)
Amedeo.com’s topic alerts (www.amedeo.com)
PubMed Cubby (www.ncbi.nlm.nih.gov/entrez/query.fcgi)
Highwire (highwire.stanford.edu)

PDA Resources:

The Journal of Informed Pharmacotherapy
(www.informedpharmacotherapy.com)
Cogniq (www.cogniq.com)
Journal To Go (www.journalstogo.com)

Individual Journals:

Thousands of journals have either their table of contents or the full text of the journal available online. Costs vary from free to hundreds of dollars, depending on whether you have a subscription to the hardcopy of the journal or have access through a university library. Some pharmacotherapeutically relevant journals are listed below:

Annals of Pharmacotherapy (www.theannals.com)
Pharmacotherapy (www.accp.com/pharmacotherapy.html)
American Journal of Health-System Pharmacy (via www.ashp.com)
New England Journal of Medicine
(www.nejm.com)
Journal of the American Medical Association
(www.jama.com)
British Medical Journal (www.bmj.com)
The Lancet (www.thelancet.com)
Annals of Internal Medicine (www.acponline.org)
Archives of Internal Medicine
(www.archinternmed.com)
Evidence-Based Medicine (www.evidence-basedmedicine.com)

Thousands of other journals via
www.freemedicaljournals.com

Finding and Keeping Up with Practice Guidelines:

National Guidelines Clearinghouse
(www.guidelines.gov)
MDConsult (www.mdconsult.com)
Canadian Coordinating Office for Health Technology Assessment (www.ccohta.ca)
Health Services/Technology Assessment Text
(hstat.nlm.nih.gov)

6. Limitations, Pitfalls, and Challenges of Using an Evidence-Based Perspective

There are several important factors that make it difficult for some clinicians, no matter how enthusiastic they are about the principles EBP, to implement the ideas we’ve discussed. A brief discussion is offered here.

6.1. Information Access

Clearly, access to internet-based resources has played a decisive role in the spread of EBP. It is difficult to imagine how clinicians could effectively practice from an EBP perspective prior their availability. There is, however, some disparity among pharmacists in terms of how much information they can access, and in particular the ability to lay hands on the actual articles in the primary literature. Of

course, unlimited funds would allow subscription access to all the resources mentioned here. Barring availability of those funds, the next best solution is to leverage the status that many pharmacists have with a university. Many pharmacists have “clinical instructor” status with a university which they precept students from, provide internships for, teach at, etc. This is ideal, since this status usually entitles you to a library card from that university and in turn allows free access to a plethora of online resources. If you don’t have status at a university, or its library is not helpful, some employers are willing to provide access to resources on the basis that they are essential for professional development and quality patient care. Finally, a group of interested pharmacists may wish to pool their resources to buy access to the most useful online resources or subscriptions.

6.2. Information Overload

More common than lack of access to information is “information anxiety” associated with its ready availability. This is an acknowledged clinical entity associated with anger, aggression, fear, anxiety, apathy, depression, and helplessness to which health care providers are prone.

Coping with information overload is a discipline in itself which we cannot possibly address here, except to say that efficient means of identifying what information must be reviewed (and what need not be), rapid methods for reviewing it in priority sequence, and efficient means for storing and retrieving it are the tools for avoiding being paralyzed by information overload.

6.3. Lack of Evidence

EBP’s supporters may be prone to not being able to cope with situations in which there is no evidence to rely upon, or where the “best available” evidence is deemed inadequate. In fact, in some areas of drug therapy (e.g., psychiatry) the majority of clinical decisions must be made in the absence of such evidence. This is a problem with the EBP perspective. The solution is to accept that often decisions must be made using only two of the three aspects of the EBM process (clinical judgement and patient values) and that knowing where evidence does not exist may be as important as knowing where it does.

6.4. Lack of Receptivity by Patients

Some of EBM’s proponents have suggested that, ideally, patients themselves should be presented with the best available evidence (in a readily understandable form) and invited to participate in decision-making about their drug therapy.^{51,52} To some degree, this seems like an obvious extension of EBM’s philosophy. Research in this area varies in

how useful this approach is in terms of how much it affects the ultimate decision about drug therapy.^{53,54} The patient population and condition studied is the most influential factor in whether patients accept being involved in the process or not. Our own research in this area indicates that elderly patients with atrial fibrillation prefer to trust their physician’s judgement rather than be presented with the evidence for stroke prevention therapies and participate in the decision of which therapy to employ.⁵⁵

7. EBM/EBP Resources & Further Reading

Centres for Evidence-Based Medicine:
minerva.minervation.com/cebm,
www.cebm.utoronto.ca, www.ebmny.org,
www.cochrane.org
 E-Bandolier: www.jr2.ox.ac.uk/bandolier/
 Ontario Program for Optimal Therapy:
www.opot.org
 Therapeutics Initiative: www.ti.ubc.ca

8. Glossary of Selected Terms

Absolute Risk Reduction (ARR): The percentage of subjects who had an event in the placebo (control) group *minus* the percentage of subjects who had an event in the Drug X (experimental) group.
 $ARR = [\% \text{ events in Placebo}] - [\% \text{ events in Drug X}]$
 $= [B/(B+D)] - [A/(A+C)]$

Expectation Bias: Observers may systematically err in measuring and recording observations so that they concur with prior expectations.

Intention-to-Treat (ITT): An approach to analyzing study results in which all subjects who are randomized are eventually analyzed for outcomes as though they had continued the treatment to which they were *originally assigned*, regardless of whether they actually received the treatment, were compliant with it, or were switched to another treatment during the trial.

Number Needed to Treat (NNT): $NNT = 1/ARR$. This signifies the number of patients who must be treated with the experimental treatment in order for one patient to receive a benefit (e.g., to avoid a bad outcome event or experience a good outcome). Note that the ARR must be expressed in *decimal* form (i.e., the %/100) when using this equation.

Obsequiousness Bias: Subjects may systematically alter questionnaire responses in the direction they perceive desired by the investigator.

Odds Ratios (OR): Odds ratios are sometimes used in reporting of RCTs, although they are more commonly used in case-control epidemio-

logic studies and some types of meta-analyses. $OR = \text{odds of an event in the experimental arm divided by the odds of an event in the control arm.}$

It is common in the medical literature for authors to phrase treatment effects only in terms of relative risk reductions. It is also known that using RRR (see below) values in advertisements and presentations is more likely to influence prescribing habits (among physicians and pharmacists) and patient acceptance of treatment,^{56,57,58} likely due to the more “dramatic” RRR figure compared to the ARR. When treatment effects are presented as NNTs, they reduce the likelihood of prescribing or taking a drug, compared to RRRs.

P-value: The statistically correct meaning of the p-value is the likelihood of a result (i.e., a difference) as extreme as or more extreme than the one seen occurring, given that the null-hypothesis is true (i.e., assuming that there is no difference between the two treatments). This value is commonly set at 0.05 as the threshold for concluding “statistical significance.” In some cases, using p-value thresholds for significance that are greater than the usual 0.05 (e.g., 0.10) may be appropriate, since even a less impressive trend toward a safety concern may be clinically interesting.

Power: Directly related to the “beta level” (see Type II error above) by the formula: $\text{Power} = 1 - \beta$. If β is set at 20% (0.2), then the *power* of the study to detect a difference between two treatments being compared is 80% (0.8).

Relative Risk (RR): The percentage of subjects who had an event in the Drug X group *divided by* the percentage of subjects who had an event in the placebo group.

Relative Risk Reduction (RRR): $1 - RR$. This value is usually phrased as “Drug X produced a Z% relative risk reduction compared to placebo.”

Recall Bias: Questions about specific exposures (or effects or adverse effects) may be asked several times of cases (or experimental treatment subjects) but not of control subjects.

Rumination Bias: The possibility that subjects on one experimental treatment versus another will try harder or spend more time thinking about their condition, side effects of the treatment or the effects of the treatment, by virtue of knowing which treatment they are receiving.

Therapeutic Personality Bias: The investigators’ convictions about efficacy may systematically influence outcomes (positive personality) and their measurement (desire for positive results) in the ways that they interact with study subjects.

Type I error: Mistakenly concluding that there is a difference between two treatments when, in fact, there is no difference. The allowable probability of making a Type I error (called the “alpha

level”) is usually set at less than 5% (by convention) when determining how many patients are required in a particular trial. This is not the same as the “p-value,” which, coincidentally, is commonly set at 0.05 as the threshold for concluding “statistical significance.”⁵⁶

Type II error: Mistakenly concluding that there is no difference between two treatments when, in fact, there is a difference. The allowable probability of make a Type II error (called the “beta level”) is usually set at less than 20% (by convention) when determining how many patients are required in a particular trial.

Variance: A mathematical measure of the distribution (or spread) among a series of numbers. In clinical trials, it can be thought of as a measure of the variability between patients in the measurement of a parameter (e.g., blood pressure, cholesterol). Mathematically, the variance is equal to the standard deviation squared. For example, a large variance in the outcome variable “blood pressure” might exist when subjects with blood pressures ranging from 120/80 to 180/110 are enrolled in a clinical trial. If only subjects with systolic blood pressures between 140 and 160 were enrolled, the variance in this variable would likely be smaller.

References

1. Sackett, D.L. et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312 (7023):71-72.
2. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based Medicine: How to Practice and Teach EBM. 2nd Edition. Churchill Livingstone, 2000.
3. Cochrane AL. Effectiveness and Efficiency: Random Reflections on Health Services. Burgess & Son: Abington, 1972.
4. Evidence-Based Medicine Working Group. Evidence-based medicine: A new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-2425.
5. <http://www.cochrane.org/cochrane/country.htm>. Accessed 7 November 2002.
6. Gladwell M. *The Tipping Point*. New York: Back Bay Books, 2002.
7. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TCA. A comparison of results of meta-analyses of randomized controlled trials and recommendations of clinical experts. *JAMA* 1992;268:240-248.
8. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance: a systematic review of the effect of continuing medical education strategies. *JAMA* 1997;274:700-705.
9. Sackett DL. Bias in analytic research. *J Chron Dis* 1979; 32:51-63.
10. <http://minerva.minervation.com/cebmb/docs/levels.html>. Accessed 7 November 2002.
11. Cook DJ, Guyatt GH, Laupacis A, Sackett DL, Goldberg RJ. Clinical recommendations using levels of evidence for antithrombotic agents. *Chest* 1995;108(4 Suppl):227S-230S.
12. Bach PB, Brown C, Gelfand SE, McCrory DC. Management of acute exacerbations of chronic obstructive pulmonary disease: A summary and appraisal of published evidence. *Ann Intern Med* 2001;134:600-620.
13. PubMed Tutorial. http://www.nlm.nih.gov/bsd/pubmed_tutorial/m1001.html. Accessed 18 November 2002.
14. MEDLINE Tutorial. <http://www.biomed.lib.umn.edu/ovidtut/>. Accessed 18 November 2002.
15. Cochrane Library Tutorial. <http://www.glenlib.demon.co.uk/clibonline/>. Accessed 18 November 2002.
16. EMBASE Tutorial. <http://www.library.usyd.edu.au/subjects/pharmacy/embase/>. Accessed 18 November 2002.
17. Harrison's Online. <http://www.harrisononline.com/>. Accessed 18 November 2002.
18. Oxman AD, Sackett DL, Guyatt GH, et al. User's Guides to the Medical Literature: I. How to get started. *JAMA* 1993;270:2093-2095.
19. Guyatt GH, Sackett DL, Cook DJ, et al. User's Guides to the Medical Literature: II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA* 1993;270:2598-2601.
20. Guyatt GH, Sackett DL, Cook DJ, et al. User's Guides to the Medical Literature: II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA* 1993;271:59-63.
21. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V, et al. User's Guides to the Medical Literature: IV. How to use an article about harm. *JAMA* 1994;271:1615-1619.
22. Dans AL, Dans LF, Guyatt GH, Richardson S, et al. User's Guides to the Medical Literature: XIV. How to decide on the applicability of clinical trial results to your patient. *JAMA* 1998;279:545-549.
23. Sackett DL, et al. (editor). *Clinical epidemiology: a basic science for clinical medicine*, 2nd Edition. Boston: Little, Brown, 1991.
24. Daniels M, Hill AB. Chemotherapy of pulmonary tuberculosis in young adults. An analysis of the combined results of three Medical Research Council trials. *Br Med J* 1952;1:1162.
25. Daniel WW, Wiley B, eds. *Biostatistics: A Foundation for Analysis in the Health Sciences*. 7th Edition. John Wiley & Sons, 1998.
26. Massel D. Similar, the same, or just not different: A guide for deciding whether treatments are clinically equivalent. *Can J Cardiol* 1999;15(5):556-562.
27. Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA* 2002;288:358-362.
28. Dickerson K, Min YI. Publication bias: the problem that won't go away. *N Y Acad Sci* 1993;703:135-146.
29. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy, I: *Stat Med* 1989;8:441-454.
30. Moher D, Schultz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendation for improving the quality of reports of parallel group randomized trials. *Lancet* 2001;357:1191-1194.
31. Pakes GE. Writing manuscripts describing clinical trials: A guide for pharmacotherapeutic researchers. *Ann Pharmacother* 2001;35:770-779.
32. Horwitz RI, Viscoli CM, Berkman L, et al. Treatment adherence and risk of death after myocardial infarction. *Lancet* 1990;336:542-545.
33. Horwitz RI, Viscoli CM, Berkman L, et al. Treatment adherence and risk of death after myocardial infarction. *Lancet* 1990;336:542-545.
34. Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women. *JAMA* 2002;288:321-333.
35. Hux JE, Naylor CD. Does the format of efficacy data determine patients acceptance of treatment? *Med Decis Making* 1995;15:152-157.
36. Forrow L, Taylor WC, Arnold RM. Absolutely relative: How research results are summarized can affect

- treatment decisions. *Am J Med* 1992;92:121-4.
37. Loewen P, Marra C, Marra F. Influence of presentation of clinical trial data on pharmacists willingness to recommend drug therapy. *Can J Hosp Pharm* 1999;52:145-149.
 38. Altman DG. Confidence intervals for the number needed to treat. *Br Med J* 1998;1309-1312.
 39. Schneider B. Analysis of clinical trial outcomes: alternative approaches to subgroup analysis. *Controlled Clinical Trials*. 1989;10(4 Suppl):176S-186S.
 40. CAPRIE Steering Committee. A randomized blinded trial of clopidogrel versus aspirin in patients at risk of ischemic events. *Lancet* 1996;348:1329-1339.
 41. Hunt DL, McKibbin KA. Locating and appraising systematic reviews. *Ann Intern Med* 1997;126:532-538.
 42. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med* 1997;127:380-387.
 43. Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. *Ann Intern Med* 1997;127:531-537.
 44. Lau J, Ioannidis PJA, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997;127:820-826.
 45. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, for the QUORUM Group. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement. *Lancet* 1999;354:1896-1900.
 46. Jada AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *Can Med Assoc J* 1997;156:1411-1416.
 47. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for critical appraisal. *Lancet* 2000;355:103-106.
 48. Cook DJ, Greengold NL, Ellrodt AG, Weingarten SR. The relation between systematic reviews and practice guidelines. *Ann Intern Med* 1997;127:210-216.
 49. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G, et al. User's Guides to the Medical Literature: VIII. How to use Clinical Practice Guidelines. A. Are the recommendations valid? *JAMA* 1995;274:570-574.
 50. Wilson MC, Hayward RS, Tunis SR, Bass EB, Guyatt G, et al. User's Guides to the Medical Literature: VIII. How to use Clinical Practice Guidelines. B. What are the recommendations and will they help you in caring for your patients? *JAMA* 1995;274:1630-1632.
 51. Edwards A, Bastian H. Risk communication—making evidence part of patient choice. In: Evidence based patient choice: inevitable or impossible? Edwards A, Elwyn G, eds. Toronto: Oxford University Press, 2001. pp. 144-160.
 52. Charles C, Gafni A, Whelan T. Shared decision making in the clinical encounter: what does it mean? *Soc Sci Med* 1997; 44: 681-692.
 53. Devereaux PJ, Anderson DR, Gardner MJ, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ* 2001; 323:1-7.
 54. Barry MJ, Cherkin DC, Yuchhiao C, et al. A randomized trial of a multimedia shared decision making program for men facing a treatment decision for benign prostatic hyperplasia. *Dis Manage Clin Outcome* 1997; 1:1-5.
 55. Loewen P, Sprague DS. Feasibility of implementing an individualized, evidence-based stroke prevention counseling program for patients with atrial fibrillation. Unpublished results.
 56. Goodman, SN. Toward evidence-based medical statistics. 1: the P value fallacy. *Ann Intern Med*. 1999;130:995-1004.

Questions

1. The Evidence-Based Medicine process can be most accurately summarized as
 - a. using guidelines and algorithms to make treatment decisions
 - b. using the “best evidence” to choose therapies which are most cost-effective
 - c. knowing the “best evidence” and combining clinical skill and the patient’s values in order to make therapeutic decisions
 - d. critical evaluation of the different types of evidence published in the medical literature
2. One of the important fundamental principles which Archie Cochrane promoted in his 1972 book was:
 - a. Evidence-based medicine is needed in order to restore the public’s trust in the medical profession.
 - b. Clinicians are threatened by studies which might prove their preferred treatments to be ineffective.
 - c. Industrial sponsorship of research threatens the integrity of the results produced.
 - d. Randomized controlled trials are needed to understand the effects of widely used treatments.
3. The term “Evidence-Based Medicine” is most appropriately attributed to:
 - a. Archie Cochrane and the British National Health Service
 - b. Gordon Guyatt and the Evidence Based Medicine Working Group at McMaster University in Hamilton, Ontario
 - c. David Sackett, from his book *How to Practice & Teach Evidence-Based Medicine*
4. Examples of “evidence” include
 - a. a clinician’s observation of an adverse effect in a patient
 - b. a randomized clinical trial
 - c. a case-control epidemiologic study
 - d. a cohort epidemiologic study
 - e. all of the above
5. On the basis of the Oxford Center for Evidence-based Medicine’s Levels of Evidence, which of the following types of evidence is considered the most reliable?
 - a. A case series
 - b. A systematic review (meta analysis) of cohort studies
 - c. An individual randomized controlled trial
 - d. A case-control study
6. Which of those below is the most focused clinical question?
 - a. “Is valsartan superior to ramipril for reducing mortality in patients with congestive heart failure?”
 - b. “What is the role of baclofen in managing spasticity associated with multiple sclerosis?”
 - c. “What effects do non-steroidal anti-inflammatory drugs have on hypertension?”
 - d. “What is the best way to manage primary pulmonary hypertension?”
7. Is the following question considered a “foreground” or “background” question: “How does congestive heart failure result in hyponatremia?”
 - a. Foreground
 - b. Background
8. Pharmacotherapy references such as DiPiro’s *Pharmacotherapy* or Koda-Kimble’s *Applied Therapeutics* are classified as which type of information source?
 - a. Primary
 - b. Secondary
 - c. Tertiary
 - d. Other
9. In trying to decide whether an RCT which did not show a difference between enalapril and doxazosin proves that enalapril is equivalent to doxazosin, which of the following factors would be important to consider?
 - a. Was the trial set up to demonstrate equivalence of the two treatments?
 - b. Were the results analyzed such that if there really were differences between the two drugs, they would be likely to show up in the results?
 - c. Were there enough patients enrolled so that the study had the power to show a difference, if one really existed?
 - d. All of the above
10. An example of a surrogate outcome used in a clinical trial of a drug in patients with rheumatoid arthritis would be
 - a. patients’ functional ability
 - b. mortality
 - c. synovial fluid eosinophil count
 - d. time to useful symptom reduction

11. A “Type II Error” in research is most precisely defined as
- mistakenly concluding that there is a difference between two treatments when there really is none.
 - mistakenly concluding that there is no difference between two treatments when there really is a difference
 - mistakenly concluding that Drug A is superior to Drug B when the trial actually showed that Drug B was superior to Drug A
 - mistakenly concluding that Drug A and B are equivalent when the trial actually showed one was superior to the other
12. Power in an RCT is influenced directly by which of the following?
- Sample size and variance
 - Sample size only
 - Variance only
 - The p-value found when a statistical test is done
13. In an RCT which shows that cefuroxime is superior to amoxicillin + clavulanic acid for “curing” pneumonia (p-value for this primary outcome = 0.01), which of the following conclusions is most appropriate?
- There is less than a 1% chance that the trial would have produced the results seen if, in fact, there was no difference between cefuroxime and amoxicillin + clavulanic acid.
 - There is a 1% chance that we have made a “type II error” in concluding that cefuroxime is superior to amoxicillin + clavulanic acid.
 - It is impossible to conclude that cefuroxime is superior to amoxicillin + clavulanic acid without knowing what power the study had to demonstrate such a difference.
14. Not using an Intention-to-Treat analysis in a RCT could reasonably cause a reader to conclude which of the following?
- The effects of randomizing the patients in the first place may be negated and the results presented may be due to a confounding factor rather than to the treatments themselves.
 - Patients who did not stay on the study treatment may have had bad outcomes, which isn’t reflected in the way the results are presented, hence the reported benefits may be overestimated.
 - Since a less conservative (non-ITT) analysis method was used, the chance of a Type I error may be increased.
 - All of the above.
15. When so-called “hard” endpoints such as death are used, it is not necessary to use blinding in a clinical trial.
- True
 - False
16. In an RCT comparing sertraline to placebo without blinding, which of the following biases might influence the results?
- Over-estimation of the adverse effects of sertraline because of “rumination” by subjects.
 - Under-estimation of the efficacy of sertraline because of low expectations of positive results among subjects because of media reports.
 - Over-estimation of the efficacy of sertraline because of positive expectations of the investigators, which influences how they ask questions on questionnaires.
 - All of the above.
17. “Cointervention” refers to bias which may be introduced in a study because of
- adjustment during the study of non-study medications used to treat the same condition as the study medication
 - subjects seeking opinions from other specialists while participating in a clinical trial
 - subjects reading media reports about the study medications, possibly influencing their feelings about the study
 - the study being too long and therefore increasing the possibility that subjects won’t remain on the study treatment

For Questions 18–21, use the following table of results from an RCT:

	Simvastatin (n=1825)	Placebo (n=1765)	p-value
Cardiovascular deaths at 3 years of treatment (n)	157	237	<0.05

18. The absolute risk reduction for simvastatin compared to placebo is

- a. 36%
- b. 5.5%
- c. 64%
- d. 4.8%

19. The relative risk (RR) for simvastatin compared to placebo is

- a. 0.64
- b. 1.56
- c. 64%
- d. 4.8%

20. The relative risk reduction (RRR) for simvastatin compared to placebo is

- a. 8.6%
- b. 64%
- c. 36%
- d. 4.8%

21. The most appropriate way to state the NNT for the cardiovascular death results from this trial would be:

- a. 21 patients must be treated with simvastatin instead of placebo for 3 years in order to prevent one cardiovascular death.
- b. The number needed to treat from this trial is 20.7.
- c. The number needed to treat with simvastatin from this trial is 21.
- d. 12 patients must be treated with simvastatin instead of placebo for 3 years in order to prevent one cardiovascular death.

For Questions 22 & 23, use the following table of results from an RCT:

	Nadolol (n=55) (experimental treatment)	Isosorbide dinitrate (n=57) (control treatment)	p-value
Alcoholic subjects with recurrent esophageal bleeding at 1 year (n)	18	22	p=0.65

Alcoholic subjects with recurrent esophageal bleeding at 1 year (n)

22. The absolute risk reduction for nadolol compared to isosorbide dinitrate in this trial is

- a. 36%
- b. 5.9%
- c. 64%
- d. 4.8%

23. The most appropriate way to state the NNT for the esophageal bleeding results from this trial would be:

- a. 17 patients must be treated with nadolol instead of isosorbide dinitrate for 1 year in order to prevent one recurrent esophageal bleed.
- b. 17 patients must be treated with nadolol instead of placebo for 1 year in order to prevent one recurrent esophageal bleed.
- c. This trial does not provide evidence that there is a difference between nadolol and isosorbide dinitrate in preventing esophageal bleeding over a 1 year period.

24. The format for reporting RCT results which is shown to have the most significant positive effect on clinicians' decision to use a drug therapy is

- a. ARR
- b. RR
- c. RRR
- d. NNT

25. Which of the following statements is correct in the context of a clinical trial?

- a. A treatment which reduces the incidence of an event by HALF could produce a large RRR and a small ARR.
- b. A treatment which reduces the incidence of an event by HALF could produce a large RR and a large NNT.
- c. A treatment which reduces the incidence of an event by HALF should produce an identical RRR and RR.
- d. All of the above.

26. In a clinical trial, gabapentin produces a reduction in pain symptoms compared to placebo such that the relative risk is 0.85 (95%CI 0.69-1.01). Which of the following statements is most correct?

- a. This trial did not detect a statistically significant difference between gabapentin and placebo for pain symptom reduction.
- b. Gabapentin produces a 15% RRR in pain symptoms compared to placebo.
- c. The probability that gabapentin is superior to placebo for reducing pain symptoms is very high.
- d. The NNT for gabapentin to reduce pain symptoms is 7.

27. True or False: In the trial in Question #26, it is equally likely that the difference between gabapentin and placebo, in RR terms, is 0.86 and 0.70.

- a. True
- b. False

Use the following table of results from a study aimed at preventing cardiovascular events in patients with a previous heart attack to answer questions 28-30.

	Simvastatin (n=1825)	Placebo (n=1765)	p-value
Composite endpoint of cardiovascular death, recurrent ischemia, urgent revascularization, bypass surgery at 2 years	165	250	0.001
Cardiovascular Death	25	26	>0.05
Recurrent ischemia	93	182	0.01
Urgent revascularization	30	28	>0.05
Bypass surgery	17	14	>0.05

28. The simvastatin ARR and NNT values for the composite endpoint in the above trial are:

- a. 9%, 19
- b. 5.12%, 20
- c. 50%, 20
- d. 23.5%, 11

29. The beneficial effect of simvastatin reflected in the composite endpoint is most likely due to

- a. simvastatin's beneficial effect on mortality
- b. simvastatin's ability to reduce recurrent ischemic events
- c. simvastatin reducing the need for urgent revascularization
- d. simvastatin's effects on the need for bypass surgery

30. Which of the following statements is most true, based on the trial data presented:

- a. Simvastatin may help prevent the need for bypass surgery and urgent revascularization in patients with a previous heart attack.
- b. Simvastatin reduces cardiovascular mortality during a 1-year period by 7%, but each patient treated has only a 1 in 1000 chance of avoiding death by taking simvastatin over that 1-year period.
- c. Simvastatin reduces the likelihood of recurrent ischemia during a 1-year period by 51%, and each patient treated has a 1 in 20 chance of experiencing a benefit when taking simvastatin over that 1-year period.