

## Critical Appraisal of the Literature

William F. Miser, MD, MA

---

### Case 1

*A 47-year-old perimenopausal woman, in your office for a well-woman examination, has a newspaper clipping given to her by a friend. The clipping reviews a recent article published in a well-known national medical journal that warns against the use of hormonal replacement therapy (HRT) because of an increased risk of breast cancer.<sup>1</sup> Although she is at low risk for this cancer, and findings of her breast examination are normal, she resists your recommendation to begin HRT. When you discuss with her the results of an article showing that postmenopausal use of estrogen reduces the risk of severe coronary heart disease,<sup>2</sup> she counters with another article from the same issue that concludes that cardiovascular mortality is increased in estrogen users.<sup>3</sup> As you review these studies, you fail to recognize that all have serious flaws. Also, you do not have available articles that are more methodologically sound that show the overwhelming benefit of HRT<sup>4-6</sup> with no increased risk in breast cancer<sup>7-9</sup> She leaves triumphantly from your office without a prescription, and you feel confused about the overall benefit of HRT.*

*After you make a mental note to read more about HRT, you see your next patient, a 28-year-old man with allergic rhinitis. He hands you a study he obtained from the Internet, which concludes that the latest anti-histamine is far superior in efficacy to all of the other antihistamines currently available on the market. He asks you for this new prescription, realizing that his health insurance company will not approve it unless you justify to them why he should take this particular anti-histamine. You promise to review the article and call him later in the week with his prescription.*

*The mother of your next patient, a 12-year-old boy, requests a test that you have never heard of. She hands you yet another article, which suggests that physicians who do not offer this test are guilty of negligence. As you*

*review this study, you wish that you remembered more about how to assess an article critically, and you hope that the rest of the day goes better.*

The above scenarios are occurring more frequently as patients are increasingly gaining access to medical information and then looking to their physicians for its interpretation. Gone are the days when what the physician says goes unchallenged by a naive patient. The public is inundated with medical advice and contrary views from the newspaper, radio, television, popular lay journals, and the Internet, and physicians are faced with the task of damage control.

Physicians also encounter constantly changing recommendations for clinical practice and an information jungle.<sup>10,11</sup> With 6 million medical articles published each year, the amount of information available is overwhelming.<sup>12</sup> If clinicians, trying to keep up with all of the literature, were to read two articles per day, in just 1 year, they would fall 82 centuries behind in their reading!

Despite this gargantuan volume of medical literature, less than 15 percent of all articles published on a particular topic are useful.<sup>13</sup> Most articles are not peer-reviewed, are sponsored by those with commercial interests, or arrive free in the mail. Even articles published in the most prestigious journals are far from perfect. Analyses of clinical trials published in a wide variety of journals have described large deficiencies in the design, analysis, and reporting; although improving with time, the average quality score of clinical trials during the past two decades is less than 50 percent.<sup>14-16</sup> As a result, many diagnostic tests and therapies are not rigorously evaluated before becoming established as a routine part of practice, which leads to the widespread use of tests with uncertain efficacy and treatments that are either ineffective or that may do more harm than good.<sup>17</sup> Readers must thus take personal responsibility for judging the validity and clinical importance of the medical literature.

---

Submitted, revised, 16 December 1998.

From the Department of Family Medicine, University of Ohio Hospitals Clinic, Columbus, Ohio. Address reprint requests to William F. Miser, MD, MA, Department of Family Medicine, The Ohio State University, 456 West 10th Ave, Columbus, OH 43210.

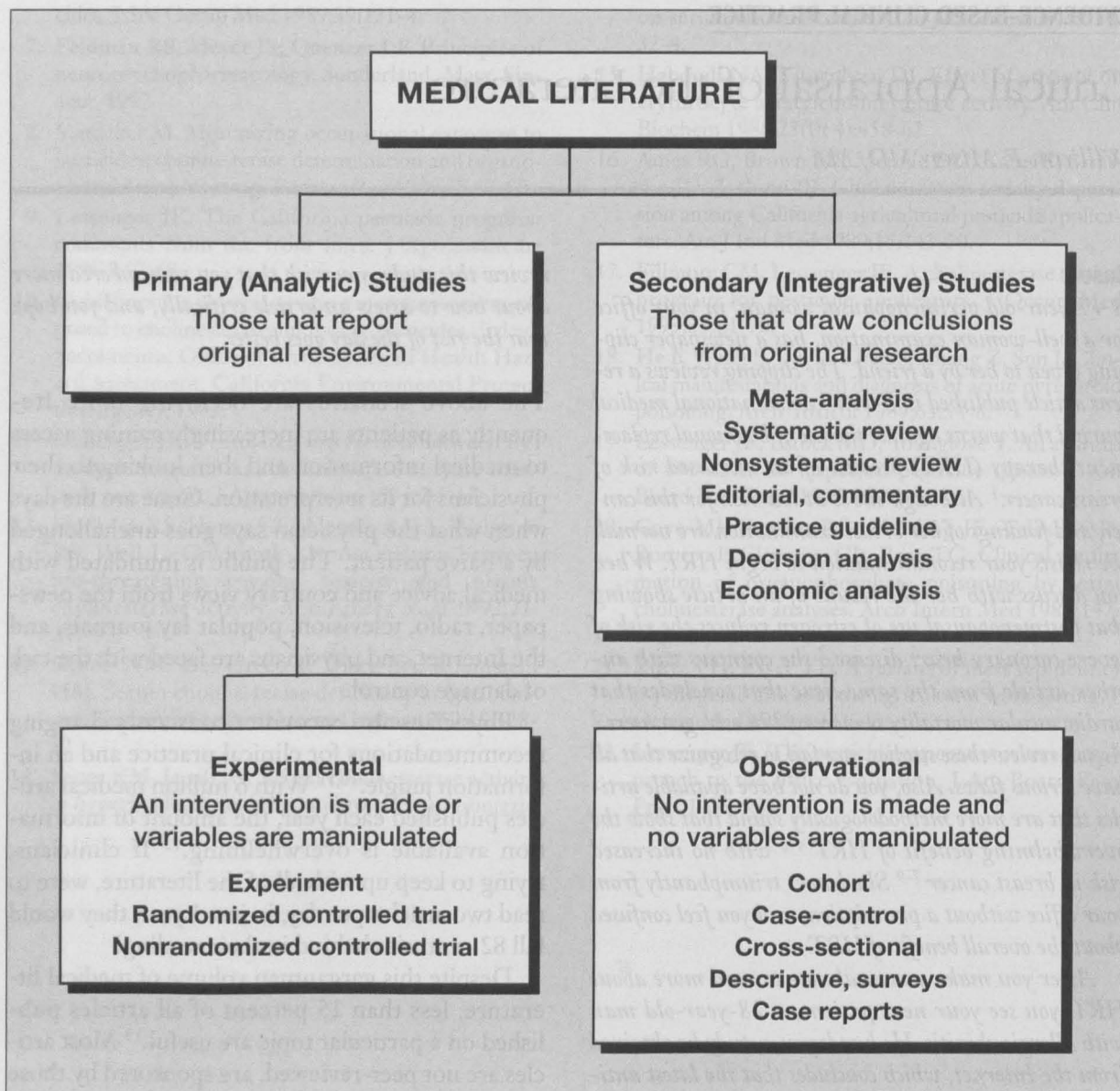


Figure 1. Major types of studies found in the medical literature.

The challenge to physicians is to provide up-to-date medical care incorporating valid new information. Our ultimate goal as clinicians should be to help patients live long, functional, satisfying, pain- and symptom-free lives. To do so requires us to balance compassion with competence. One of the essential skills needed to maintain competence, to provide patients with the best possible care, and to do more good than harm is the ability to critically appraise the literature. We must be able to find potentially relevant information, filter out the best from the much larger volume of less credible information, and then judge whether to believe the information that remains.<sup>12</sup>

The two major types of studies (Figure 1) reported in the medical literature are (1) those that report original research (analytic, primary studies), and (2) those that summarize or draw conclusions from original research (integrative, secondary studies). Primary studies can be either experimental (an intervention is made) or observational (no intervention is made). The purpose of this article is to provide an overview of a systematic, efficient, and effective approach to the critical review of original research. This information is pertinent to physicians no matter what their setting, be it an academic medical center or a rural solo practice. Because of space limitations, this article cannot address everything in exhaustive detail,

and the reader is encouraged to refer to the suggested readings at the end for further assistance.

### Critical Appraisal of an Article

It is important that clinicians master the skills of critical appraisal of the literature if they are to apply evidence-based medicine to the daily clinical problems they encounter. Most busy clinicians do not have hours to spend critiquing an article, however; they need a brief and efficient screening method that allows them to know whether the information is valid and applicable to their practice. By applying the techniques offered here, it is possible to approach the literature confidently and base clinical decisions on "evidence rather than hope."<sup>18</sup>

This approach is modified and adapted from several excellent sources. The Department of Clinical Epidemiology and Biostatistics at McMaster University in 1981 published a series of useful guides to help the busy clinician critically read clinical articles about diagnosis, prognosis, etiology, and therapy.<sup>19-23</sup> These guides have subsequently been updated and expanded to focus more on the practical issues of first finding pertinent articles and then validating (believing) and applying the information to patient care.<sup>18,24-43</sup> The recommendations from these users' guides form the foundation upon which techniques developed by Slawson, Shaughnessy, and Bennett<sup>10,11</sup> are modified and added.

With an article in hand, the process involves three steps: (1) screen for initial validity and relevance, (2) determine the intent of the article, and (3) evaluate the validity of the article based on its intent. This paper focuses on the type of study most germane to clinical practice: a therapeutic intervention. To make the most of this exercise, it would be helpful for the reader to obtain a copy of the article mentioned in case 2, and to follow the steps outlined below. The users' guides and other resources listed at the end of this paper are helpful in learning how to appraise other types of articles.

### Case 2

*Croup season is approaching, and you have a rather large pediatric population in your practice. Since you finished your residency, you have been treating croup with mist therapy but have been dissatisfied with its results. As you talk to a colleague about this problem, she hands you the following article recently published in 1998 in the Journal of the American Medical Association,*

**Table 1. Step 1 in Critically Assessing an Article: Screen for Initial Validity and Relevance.**

Is this article worth taking the time to review in depth? <i>A "stop" or "pause" answer to any of the following should prompt you to question seriously whether you should spend the time to review the article critically</i>		
1. Is the article from a peer-reviewed journal? <i>Articles published in a peer-reviewed journal have already gone through an extensive review and editing process.</i>	Yes (go on)	No (stop)
2. Is the location of the study similar to mine so the results, if valid, would apply to my practice?	Yes (go on)	No (pause)
3. Is the study sponsored by an organization that might influence the study design or results? <i>Read the conclusion of the abstract to determine relevance.</i>	Yes (pause)	No (go on)
4. Will this information, if true, have a direct impact on the health of my patients, and is it something they will care about?	Yes (go on)	No (stop)
5. Is the problem addressed one that is common to my practice, and is the intervention or test feasible and available to me?	Yes (go on)	No (stop)
6. Will this information, if true, require me to change my current practice?	Yes (go on)	No (stop)

Note: Questions 4 through 6 were adapted from Slawson and his Information Mastery Working Group.<sup>45</sup>

*ciation, "Nebulized Budesonide and Oral Dexamethasone for Treatment of Croup—A Randomized Controlled Trial."<sup>44</sup> You were taught that the use of corticosteroids for croup is controversial and should be reserved for those in the hospital. You have a few minutes before seeing your next patient but are unsure whether you have the time to read this article.*

### Step 1 - Screen for Initial Validity and Relevance

The first step when looking at an article is to ask whether the article is worth taking the time to review in depth. This question can be answered within a few seconds by asking six simple questions (Table 1). A stop or pause answer to any of these questions should prompt you to seriously consider whether you should spend the time to critically review the study. The article mentioned in case 2 will be used to illustrate these points.

1. *Is the article from a peer-reviewed journal?* Most national and specialty journals published in the United States are peer-reviewed; if in doubt, this answer can be found in the journal's instructions for authors section. Typically, those journals sent to clinicians unsolicited and free of charge are throwaway journals, so called because that is exactly what you should do with them. These journals, although attractive in appearance, are not

peer-reviewed but instead are geared toward generating income from advertising.<sup>12,18</sup>

Articles published in the major peer-reviewed journals have already undergone an extensive process to weed out flawed studies and to improve the quality of the ones subsequently accepted for publication. When an investigator submits a manuscript to a peer-reviewed journal, the editor typically will first establish whether the manuscript is suitable for that journal, and then, if acceptable, send it to several reviewers for analysis. Peer reviewers are not part of the editorial staff but usually are volunteers who have expertise in both the subject matter and research design. The purpose of the peer review is to act as a sieve by detecting those studies that are flawed by poor design, are trivial, or are uninterpretable. This process, along with the subsequent revisions and editing, improves the quality of the paper and its statistical analyses.<sup>46-49</sup> The *Annals of Internal Medicine*, for example, receives more than 1200 original research manuscript submissions each year. The editorial staff reject one half after an internal review, and the remaining half are sent to at least 2 peers for review. Of the original 1200 submissions, only 15 percent are subsequently published.<sup>50</sup>

Because of these strengths, peer review has become the accepted method for improving the quality of the science reported in the medical literature.<sup>51</sup> This mechanism, however, is far from perfect, and it does not guarantee that the published article is without flaw or bias.<sup>13</sup> Other types of publication biases are inherent in the process despite an adequate peer-review process. Studies showing statistically significant (positive) results and having larger sample sizes are more likely to be written and submitted by authors, and subsequently accepted and published than are nonsignificant (negative) studies.<sup>52-55</sup> Also, the speed of publication depends on the direction and strength of the trial results; trials with negative results take twice as long to be published as positive trials.<sup>56</sup> Finally, no matter how good the peer-review system, fraudulent research, although rare, is extremely hard to recognize.<sup>57</sup>

The article you are assessing is published in the *Journal of the American Medical Association (JAMA)*. You are almost certain that this journal is peer-reviewed, which is confirmed in their Instructions for Authors (“*JAMA* is an international, peer-reviewed, general medical journal...”). You answer “yes” to this question.

2. *Is the location of the study similar to mine so that the results, if valid, would apply to my practice?* This question can be answered by reviewing information about the authors on the first page of an article (typically at the bottom of the page). If you have a rural general practice and you are assessing a study performed in a university subspecialty clinic, you might want to pause to consider the potential biases that might be present. This is a soft area, and rarely will you want to reject an article outright at this juncture; however, large differences in location should raise caution in your mind.

In the article you are assessing, you notice at the bottom of the first page that the study was performed in two university hospitals in Canada. There is no reason to believe children with croup for whom you provide care are different from those seen in Canada, but you begin to wonder whether the study done in a tertiary care center is applicable to your practice. You decide to continue critiquing this article, but make a mental note to consider this issue later.

3. *Is the study sponsored by an organization that might influence the study design or results?* This question considers the potential bias that could occur from outside funding. In most journals investigators are required to state sources of funding for their study. Clinicians need to be wary of published symposiums sponsored by pharmaceutical companies. Although found in peer-reviewed journals, they tend to be promotional in nature, to have misleading titles, to use brand names, and are less likely to be peer-reviewed in the same manner as other articles in the parent journal.<sup>58</sup> Also, randomized clinical trials (RCTs) published in journal supplements are generally of inferior quality compared with articles published in the parent journal.<sup>59</sup> This is not to say that all studies sponsored by commercial interests are biased; on the contrary, numerous well-designed studies published in the literature are sponsored by the pharmaceutical industry. If a pharmaceutical company or other commercial organization funded the study, however, look for assurances from investigators that the design and results were not influenced by this association.

You again review the information about the authors, and look at the end of the article for this information. You find that funding support was from several foundations, but none from a com-

**Table 2. Major Clinical Categories of Primary Research and Their Preferred Study Designs.**

Clinical Category	Description	Preferred Study Design
Therapy	Tests the effectiveness of a treatment, such as a drug, surgical procedure, or other intervention	Randomized, double-blinded, placebo- controlled trial (Figure 2)
Diagnosis and screening	Measures the validity (is it dependable?) and reliability (will the same results be obtained every time?) of a diagnostic test, or evaluates the effectiveness of a test in detecting disease at a presymptomatic stage when applied to a large population	Cross-sectional survey (comparing the new test with a reference standard) (Figure 3)
Causation	Assesses whether a substance is related to the development of an illness or condition	Cohort or case-control (Figures 4 and 5)
Prognosis	Determines the outcome of a disease	Longitudinal cohort study (Figure 4)

Adapted from Greenhalgh.<sup>60</sup>

pany that has commercial interests in the drugs used in the study.

The answers to the next three questions dealing with clinical relevance to your practice can be obtained by reading the conclusion and selected portions of the abstract. Clinical relevance is important not only to physicians but also to their patients. Rarely is it worthwhile to read an article about an uncommon condition you have never encountered in your practice, or about a treatment or diagnostic test that is not and never will be available to you. Reading these types of articles might satisfy your intellectual curiosity but will not impact your practice. Slawson and his colleagues have emphasized that for a busy clinician, articles concerned with patient-oriented evidence that matters (POEMs) are far more useful than those articles that report disease-oriented-evidence (DOE).<sup>10,45</sup> So, given a choice between reading an article that describes the sensitivity and specificity of a screening test in detecting cancer (a DOE) and one that shows that those who undergo this screening enjoy an improved quality and length of life (a POEM), you would probably want to choose the latter.

4. *Will this information, if true, have a direct impact on the health of my patients, and is it something they will care about?* You read this conclusion of the abstract, “Based on the similar outcomes in the 3 groups, oral dexamethasone is the preferred intervention because of its ease of administration, lower cost, and more widespread availability.” You scan the rest of the abstract and find that the outcomes were a croup score, hospital admission rates, time spent in the emergency department,

return visits, and ongoing symptoms at 1 week. Because these are outcomes that you and your patients care about, you answer this question “yes.”

5. *Is the problem addressed one that is common to my practice, and is the intervention or test feasible and available to me?* If you were in a practice that sees very few children or rarely sees croup, you might decide the answer is “no” and go on to read other articles. You decide the answer to this is “yes,” however, because croup is a common problem seen in your practice, and oral dexamethasone is something you could easily stock in your office.

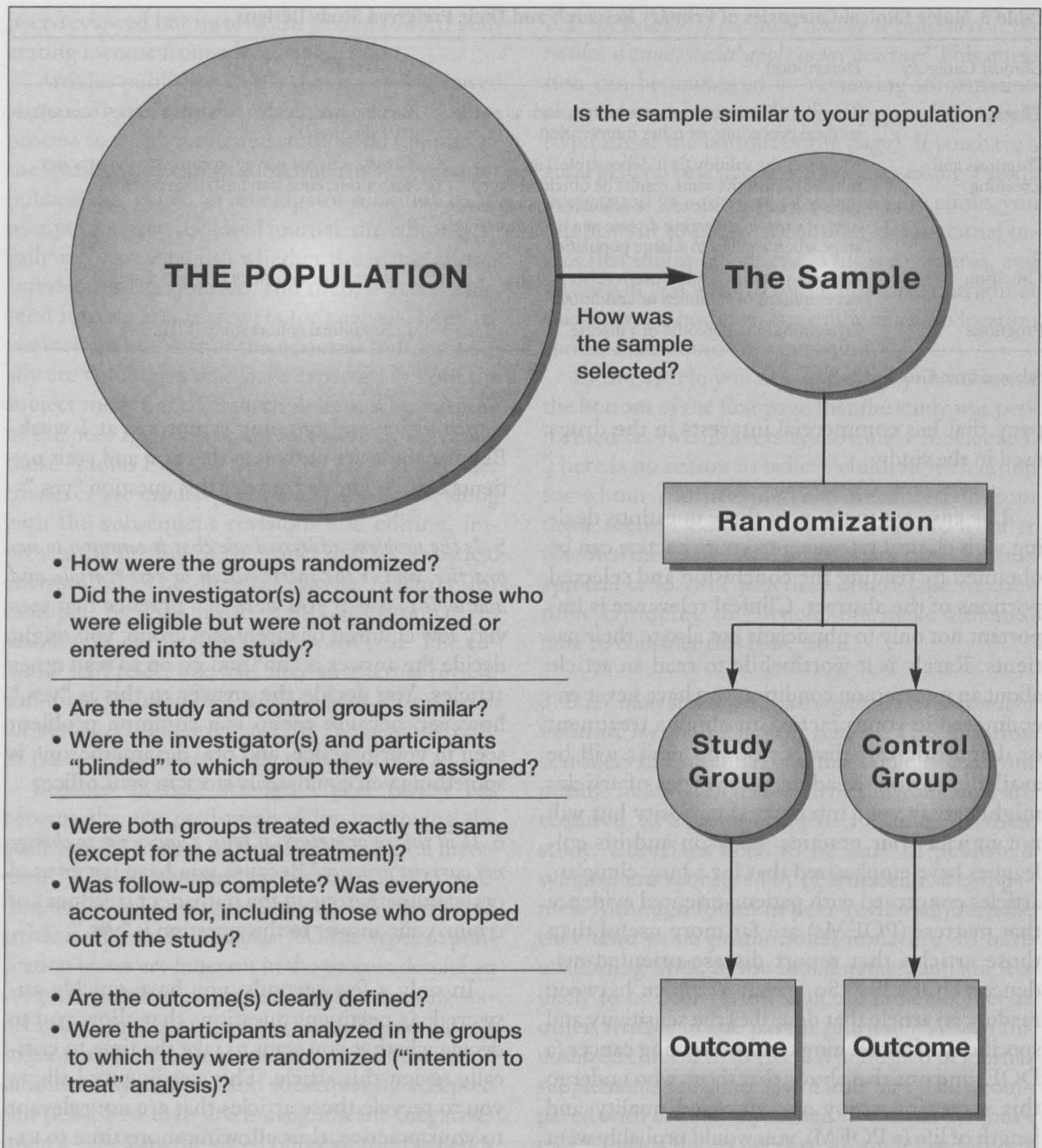
6. *Will this information, if true, require me to change my current practice?* Because you have never used oral dexamethasone in the outpatient treatment of croup, your answer to this question is “yes.”

In only a few seconds, you have quickly answered six pertinent questions that allow you to decide whether you want to take the time to critically review this article. This weeding tool allows you to recycle those articles that are not relevant to your practice, thus allowing more time to examine the validity of those few articles that might have an impact on the care of your patients.

### **Step 2 - Determine the Intent of the Article**

If you decide to continue with the article after completing step 1, your next task is to determine why the study was performed and what clinical question(s) the investigators were addressing.<sup>60</sup> The four major clinical categories found in articles of primary (original) research are (1) therapy, (2) diagnosis and screening, (3) causation, and (4) prognosis (Table 2). The intent of the article can





**Figure 2. Randomized controlled trial, considered the reference standard for studies dealing with treatment or other interventions.**

usually be found by reading the abstract and, if needed, by skimming the introduction (usually found in the last paragraph) to determine the purpose of the study.

For the article mentioned in case 2, the investigators address a therapeutic intervention (the use of oral dexamethasone in treating mild-to-moderate croup). Because you are seriously considering including this therapeutic intervention in your

practice, you decide you need to spend the time to validate critically the conclusions of the study.

**Step 3 - Evaluate the Validity of the Article Based on Its Intent**

After an article has successfully passed the first two steps, it is time to assess critically its validity and applicability to your practice setting. Each of the four clinical categories found in Table 2 (and

illustrated in Figures 2 through 5) has a preferred study design and critical items to ensure its validity. The *Users' Guides* published by the Department of Clinical Epidemiology and Biostatistics at McMaster University provide a useful list of questions to help you with this assessment. Modifications of these lists of questions are found in Tables 3 through 6.

To get started on this step, read the entire abstract; survey the boldface headings; review the tables, graphs, and illustrations; and then skim the first sentence of each paragraph to grasp quickly the organization of the article. You then need to focus on the methods section, answering a specific list of questions based on the intent of the article. Because the article from case 2 deals with a therapeutic intervention, you begin reading the methods section of the article and address the questions listed in Table 3.

1. *Is the study a randomized controlled trial?* RCTs (Figure 2) are considered the reference standard design to determine the effectiveness of treatment. The power of RCTs lies in their use of randomization. At the start of a trial, participants are randomly allocated by a process equivalent to the flip of a coin to either one intervention (eg, a new antihypertensive medication) or another (eg, an established antihypertensive medication or placebo). Both groups are then observed for a specified period, and defined outcomes (eg, blood pressure, myocardial infarction, death) are measured and analyzed at the conclusion.

Randomization diminishes the potential for investigators selecting participants in a way that would unfairly bias one treatment group over another (selection bias). It is important to determine how the investigators actually performed the randomization. Although infrequently reported in the past, most journals now require a standard format that provides this information.<sup>15</sup> Various techniques can be used for randomization.<sup>61</sup> Investigators can use simple randomization; each participant has an equal chance of being assigned to one group or another without regard to previous assignments of other participants. Sometimes this type of randomization will result in one treatment group being larger than another, or by chance, one group having important baseline differences that might affect the study. To avoid these problems, investigators can use blocked random-

**Table 3. Determining Validity of an Article About Therapy.**

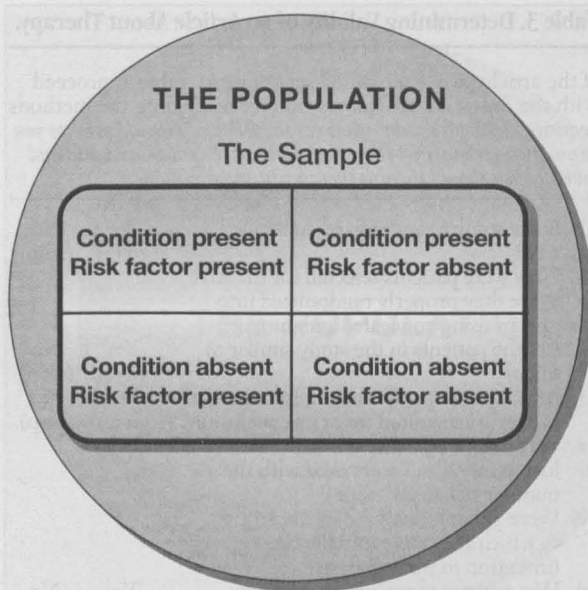
If the article passes the initial screening in Table 1, proceed with the following critical assessment by reading the methods section. A "stop" answer to any of the following should prompt you to question seriously whether the results of the study are valid and whether you should use this therapeutic intervention

1. Is the study a randomized controlled trial?	Yes (go on)	No (stop)
a. How were patients selected for the trial?		
b. Were they properly randomized into groups using concealed assignment?		
2. Are the patients in the study similar to mine?	Yes (go on)	No (stop)
3. Are all participants who entered the trial properly accounted for at its conclusion?	Yes (go on)	No (stop)
a. Was follow-up complete and were few lost to follow-up compared with the number of bad outcomes?		
b. Were patients analyzed in the groups to which they were initially randomized (intention to treat analysis)?		
4. Was everyone involved in the study (participants and investigators) "blind" to treatment?	Yes	No
5. Were the intervention and control groups similar at the start of the trial? (check Table 1)	Yes	No
6. Were the groups treated equally (aside from the experimental intervention)?	Yes	No
7. Are the results clinically as well as statistically significant?	Yes	No
a. Were the outcomes measured clinically important?		
8. If a negative trial, was a power analysis done?	Yes	No
9. Were there other factors that might have affected the outcome?	Yes	No
10. Are the treatment benefits worth the potential harms and costs?	Yes	No

Adapted from material developed by The Department of Clinical Epidemiology and Biostatistics at McMaster University<sup>25</sup> and by the Information Mastery Working Group.<sup>10</sup>

ization (groups are equal in size) or stratified randomization (subjects are randomized within groups based on potential confounding factors such as age or sex).

To determine the assignment of participants, investigators should use a table of random numbers or a computer that produces a random sequence. The final allocation of participants to the study should be concealed from both investigators and participants. If investigators responsible for assigning participants are aware of the allocation, they might unwittingly (or otherwise) assign those with a better prognosis to the treatment group and those with a worse prognosis to the control group. RCTs that have inadequate allocation concealment will yield an inflated treatment effect that is up to 30 percent better than those trials with proper concealment.<sup>62,63</sup>



**Figure 3. Cross-sectional (prevalence) study. This design is most often used in studies on diagnostic or screening tests.**

In the article you are assessing, you find in the second paragraph of the methods section that the study design was an RCT, and that participants were randomized to one of three groups: nebulized budesonide and oral placebo, placebo nebulizer and oral dexamethasone, and nebulized budesonide and oral dexamethasone. A central pharmacy randomized the patients into these groups using computer-generated random numbers in random blocks of 6 or 9 to help ensure equal distribution among the groups, and then stratified them by study site. The randomization list was kept in the central pharmacy to ensure allocation concealment. You answer “yes” to this question and proceed with your assessment.

2. *Are the participants in the study similar to my patients?* To be generalizable (external validity), the study participants should be similar to the patients you care for in your practice. A common problem encountered by primary care physicians is interpreting the results of studies done on patients in subspecialty care clinics. The group of men in a university urology clinic participating in a study on early detection of prostate cancer might be different from the group of men seen in a typical primary care clinic. It is important to determine who was included and who was excluded from the study. You find that the study participants were children, aged 3 months to 5 years, who had mild-

to-moderate croup. Because you provide care for children in this age group, and after noting the exclusion criteria, you answer this question “yes.”

3. *Are all participants who entered the trial properly accounted for at its conclusion?* Another strength of RCTs is that participants are observed prospectively. It is important, however, that these participants be accounted for at the end of the trial to avoid a loss-of-subjects bias, which can occur through the course of a prospective study as participants drop out of the investigation for various reasons. They might have lost interest, moved out of the area, developed intolerable side effects, or

**Table 4. Determining Validity of an Article About a Diagnostic Test.**

If the article passes the initial screening in Table 1, proceed with the following critical assessment by reading the methods section. A “stop” answer to any of the following should prompt you to question seriously whether the results of the study are valid and whether you should use this diagnostic test

1. What is the disease being addressed and what is the diagnostic test? \_\_\_\_\_
2. Was the new test compared with an acceptable reference standard test, and were both tests applied in a uniformly blind manner? Yes (go on)    No (stop)
3. Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice? Yes (go on)    No (stop)
4. Is the new test reasonable? What are its limitations? Explain: \_\_\_\_\_
5. In terms of prevalence of disease, are the study participants similar to my patients? Yes    No  
*Varying prevalences will affect the predictive value of the test in my practice.*
6. Will my patients be better off as a result of this test? Yes    No
7. What are the sensitivity, specificity, and predictive values of the test?

Sensitivity =  $a/(a + c) =$  \_\_\_\_\_

Specificity =  $d/(b + d) =$  \_\_\_\_\_

		Reference Standard Result	
		Positive	Negative
Test Result	Positive	a	b
	Negative	c	d

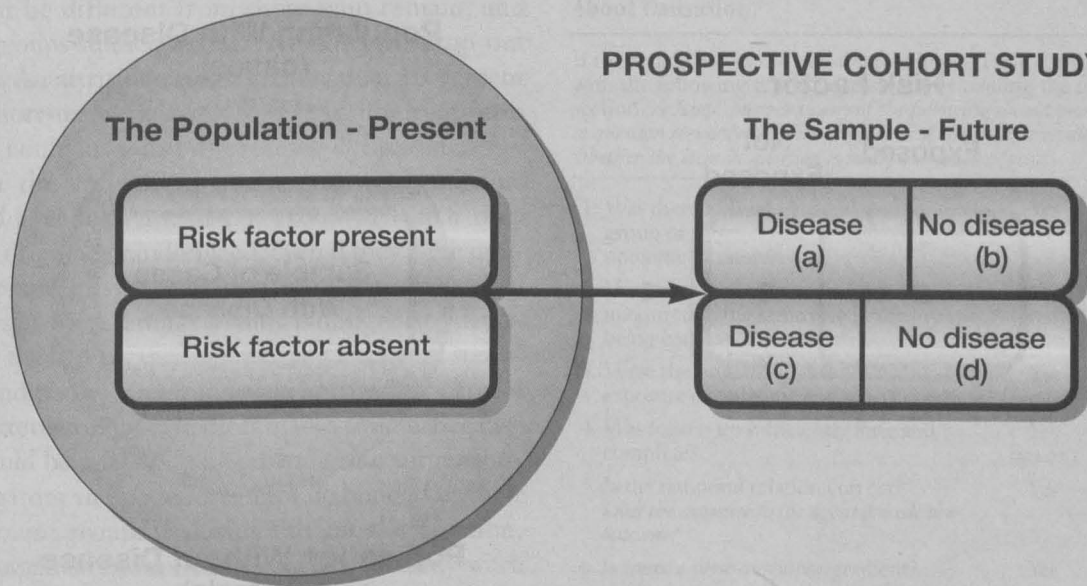
Positive predictive value =  $a/(a + b) =$  \_\_\_\_\_

Negative predictive value =  $c/(c + d) =$  \_\_\_\_\_

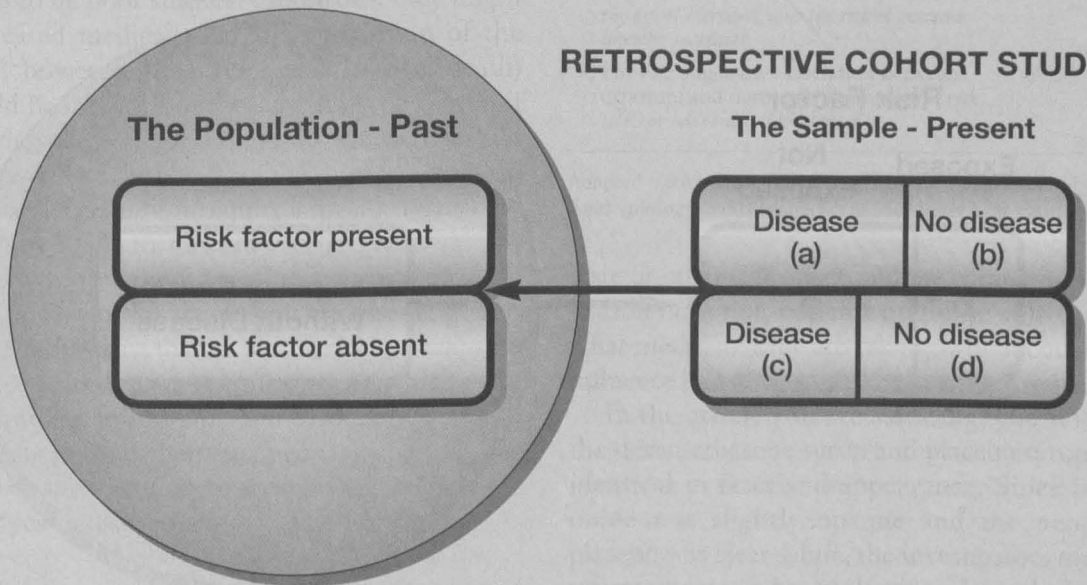
Adapted from material developed by the Department of Clinical Epidemiology and Biostatistics at McMaster University<sup>27</sup> and by the Information Mastery Working Group<sup>10</sup>



**PROSPECTIVE COHORT STUDY**



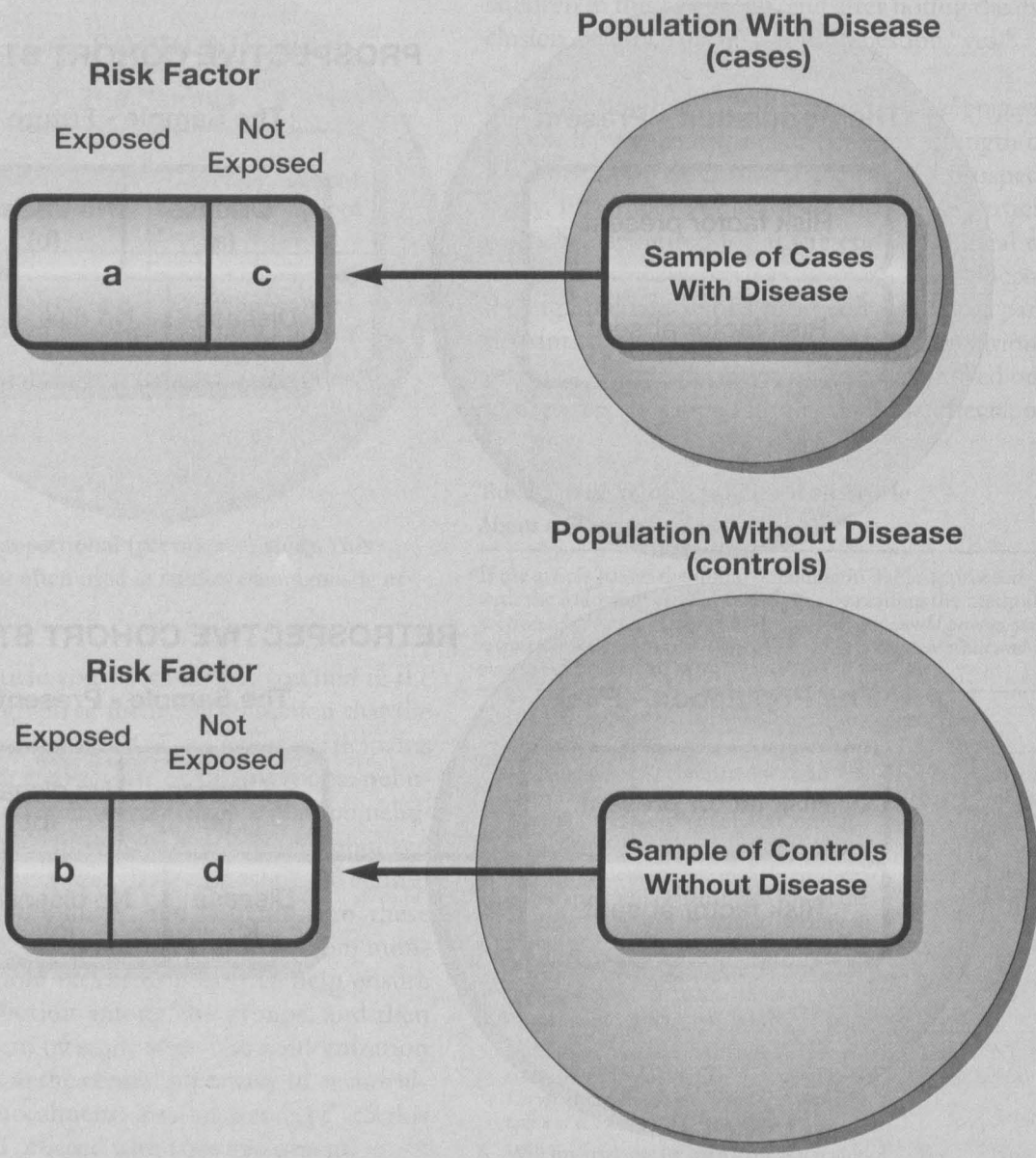
**RETROSPECTIVE COHORT STUDY**



**Relative Risk (RR)** is the risk of disease associated with a particular exposure.

	Condition present	Condition absent	
Risk factor present	a	b	RR = $\frac{(a)/(a+b)}{(c)/(c+d)}$
Risk factor absent	c	d	

**Figure 4.** Prospective and retrospective cohort study. These types of studies are often used for determining causation or prognosis. Data are typically analyzed using relative risk.



**Odds Ratio (OR)** is the measure of strength of association. It is the odds of exposure among cases to the odds of exposure among the controls.

	Cases	Controls	
Exposed	a	b	$OR = \frac{(a/a+c)/(c/a+c)}{(b/b+d)/(d/b+d)} = \frac{a/c}{b/d} = \frac{ad}{bc}$
Not Exposed	c	d	

Figure 5. Case-control study, a retrospective study in which the investigator selects a group with disease (cases) and one without disease (controls) and looks back in time at exposure to potential risk factors to determine causation. Data are typically analyzed using the odds ratio.

died. The participants who are lost to follow-up might be different from those who remain, and the groups studied might have different drop-out rates. An attrition rate of greater than 10 percent for short-term trials and 15 percent for long-term trials could invalidate the results of the study.

At the conclusion of the study, participants should be analyzed in the group in which they were originally randomized, even if they were non-compliant or switched groups (intention-to-treat analysis). For example, a study is designed to determine the best treatment approach to carotid stenosis, and patients are randomized to either carotid endarterectomy or medical management. Because it would be unethical to perform sham surgery, investigators and patients cannot be blinded to their treatment group. If, during the initial evaluation, participants randomized to endarterectomy were found to be poor surgical candidates, they might be treated medically. At the conclusion of the study, however, their outcomes (stroke, death) should be included in the surgical group, even if they did not have surgery; to do otherwise would unfairly inflate the benefit of the surgical approach.

Most journals now require a specific format for reporting RCTs that includes a chart allowing you to easily follow the flow of participants through the study.<sup>15</sup> In the article you are assessing, you notice in the chart that all but 1 of 198 participants were observed to study completion, which is an outstanding follow-up. You also notice in the methods section that the “primary analysis was based on the intention-to-treat principle.” You answer “yes” to this question.

4. *Was everyone involved in the study (participants and investigators) “blind” to treatment?* Investigator bias can occur when those making the observations might unintentionally shade the results to confirm the hypothesis or to influence the participants. This bias can be prevented by the process of blinding in which neither the investigators nor the participants are aware of group assignment (double-blinded). For example, in a study comparing a new antihypertensive drug with a placebo, neither the investigators nor the participants should be aware of what the participants are taking. The study medication should be indistinguishable from the comparison medication or placebo; it should have the same look and taste and be taken at the same frequency. If the study medication has a certain bitter

**Table 5. Determining Validity of an Article About Causation.**

If the article passes the initial screening in Table 1, proceed with the following critical assessment by reading the methods section. A “stop” answer to any of the following should prompt you to question seriously whether the results of the study are valid and whether the item in question is really a causative factor.

1. Was there a clearly defined comparison group or those at risk for or having the outcome of interest?	Yes (go on)	No (stop)
2. Were the outcomes and exposures measured in the same way in the groups being compared?	Yes (go on)	No (stop)
3. Were the observers blinded to the exposure of outcome and to the outcome?	Yes (go on)	No (stop)
4. Was follow-up sufficiently long and complete?	Yes (go on)	No (stop)
5. Is the temporal relation correct? <i>Does the exposure to the agent precede the outcome?</i>	Yes	No
6. Is there a dose-response gradient? <i>As the quantity or the duration of exposure to the agent increases, does the risk of outcome likewise increase?</i>	Yes	No
7. How strong is the association between exposure and outcome? <i>Is the relative risk (RR) or odds ratio (OR) large?</i>	Yes	No

Adapted from material developed by The Department of Clinical Epidemiology and Biostatistics at McMaster University.<sup>29</sup>

taste or other side effect, and the comparison medication does not, patients might be able to guess what medicine they are taking, which could then influence how they perceive their improvement.

In the article you are assessing, you find that the dexamethasone syrup and placebo syrup were identical in taste and appearance. Since budesonide was slightly opaque and the nebulized placebo was clear saline, the investigators took extra precautions by packaging the solutions in brown syringes. The investigators went further by asking the research assistants and participants to guess which intervention the patients received; their responses were no greater than chance alone, indicating the blinding was successful. Assured that this study was properly conducted and double-blinded, you answer “yes” to this question.

5. *Were the intervention and control groups similar at the start of the trial?* Through the process of randomization, you would anticipate the groups to be similar at the beginning of a trial. Since this might not always be the case, investigators should provide a group comparison. This information is usually found in Table 1 of the article.

**Table 6. Determining Validity of an Article About Prognosis.**

If the article passes the initial screening in Table 1, proceed with the following critical assessment by reading the methods section. A "stop" answer to any of the following should prompt you to question seriously whether the results of the study are valid.

1. Was an inception cohort assembled? <i>Did the investigators select a specific group of people initially free of the outcome of interest, and observe them over time?</i>	Yes (go on)	No (stop)
2. Were the criteria for entry into the study objective, reasonable, and unbiased?	Yes (go on)	No (stop)
3. Was follow-up of participants adequate? (at least 70% - 80%)	Yes (go on)	No (stop)
4. Were the patients similar to mine, in terms of age, sex, race, severity of disease, and other factors that might influence the course of the disease?	Yes (go on)	No (stop)
5. Where did the participants come from? <i>Was the referral pattern specified?</i>	Yes	No
6. Were outcomes assessed objectively and blindly?	Yes	No

Adapted from material developed by the Department of Clinical Epidemiology and Biostatistics at McMaster University<sup>30</sup> and by the Information Mastery Working Group.<sup>10</sup>

In the article you are assessing, you find the groups to be similar, but not exact, in sex, age, history, croup score, and vital signs. Those in the dexamethasone-treated group had a slightly higher percentage of preceding upper respiratory tract infections than did those in the budesonide-treated group (67 percent vs 54 percent). The investigators do not include an analysis on whether this difference is statistically significant, but it is unlikely that this small difference would be clinically significant. It is in areas such as these that you must use your clinical experience and judgment to determine whether small differences are likely to influence outcomes. You are satisfied that the groups are similar enough, and answer "yes" to this question.

6. *Were the groups treated equally (aside from the experimental intervention)?* To ensure that both proper blinding and that other unknown determinants are not a factor, the groups should be treated equally except for the therapeutic intervention. In the study you are assessing, you find that every participant was treated in the same manner—everyone received an oral syrup (dexamethasone or placebo) and a nebulized solution (budesonide or placebo), and each was assessed and observed equally. Had the investigators not given the participants randomized to the oral dexamethasone a

nebulized solution, both the investigators and participants would know which therapeutic group they were in (which would introduce a bias). Also, one could not exclude the possibility that the actual treatment benefit was due to the process of nebulization itself and not to the budesonide. Because the investigators took these precautions, you answer "yes" to this question.

7. *Are the results clinically as well as statistically significant?* Statistics are mathematical techniques of gathering, organizing, describing, analyzing, and interpreting numerical data.<sup>64</sup> By their use, investigators try to convince readers that the results of their study are valid. Internal validity addresses how well the study was done and whether the results reflect truth and did not occur by chance alone. External validity considers whether the results are generalizable to patients outside the study. Both types of validity are important.

The choice of statistical test depends on the study design, the types of data analyzed, and whether the groups are independent or paired. The three main types of data are categorical (nominal), ordinal, and continuous (interval). An observation made on more than one participant or group is independent (eg, measuring serum cholesterol in two groups of participants), whereas making more than one observation on a single participant is paired (eg, measuring serum cholesterol in a participant before and after treatment). Based on this information, one can then select an appropriate statistical test (Table 7). Be suspicious of a study that has a standard set of data collected in a standard way but is analyzed by a test that has an unpronounceable name and is not listed in a standard statistical textbook; the investigators might be attempting to prove something statistically significant that truly has no significance.<sup>65</sup>

There are two types of errors that can potentially occur when comparing the results of a study with reality (Figure 6). A type I error occurs when the study finds a difference between groups when, in reality, there is no difference. This type of error is similar to a jury finding an innocent person guilty of a crime. The investigators usually indicate the maximum acceptable risk (the  $\alpha$  level) they are willing to tolerate in reaching this false-positive conclusion. Usually, the  $\alpha$  level is arbitrarily set at 0.05 (or lower), which means the investigators are willing to take a 5 percent risk that any differences

**Table 7. A Practical Guide to Commonly Used Tests for Association Between Two Independent Variables or Paired Observations.\***

Types of Data	Categorical 2 Samples	Categorical ≥ 3 Samples	Ordinal	Continuous
Independent variables				
Categorical, 2 samples	Chi-square Fisher exact	—	—	—
Categorical, ≥ 3 samples	Chi-square ( $r \times r$ )	Chi-square ( $r \times r$ )	—	—
Ordinal	Mann-Whitney <i>U</i> Wilcoxon rank sum	Kruskal-Wallis one- way analysis of variance (ANOVA)	Spearman <i>r</i> Kendall tau	—
Continuous	Student <i>t</i>	ANOVA	Kendall tau Spearman <i>r</i> ANOVA	Pearson correlation Linear regression Multiple regression
Paired observations	McNemar	Cochran <i>Q</i>	Wilcoxon signed rank Friedman two-way ANOVA	Paired <i>t</i>

\*The test chosen depends on study design, types of variables analyzed, and whether observations are independent or paired. Categorical (nominal) data can be grouped, but not ordered (eg, eye color, sex, race, religion, etc). Ordinal data can be grouped and ordered (eg, sense of well-being: excellent, very good, fair, poor). Continuous data have order and magnitude (eg, age, blood pressure, cholesterol, weight, etc).

found were due to chance. At the completion of the study, the investigators then calculate the probability (known as the *P* value) that a type I error has occurred. When the *P* value is less than the  $\alpha$  value (eg, < 0.05), the investigators conclude that the results are statistically significant.

Statistical significance does not always correlate with clinical significance. In a large study, very small differences can be statistically significant. For example, a study comparing two antihypertensive drugs in more than 1000 participants might find a statistically significant difference in mean blood pressures of only 3 mmHg, which in the clinical realm is trivial. A *P* value of < 0.0001 is no more clinically significant than a *P* value of < 0.05. The smaller *P* value only means there is less risk of drawing a false-positive conclusion (less than 1 in 1000). When analyzing an article, beware of being seduced by statistical significance in lieu of clinical significance; both must be considered.

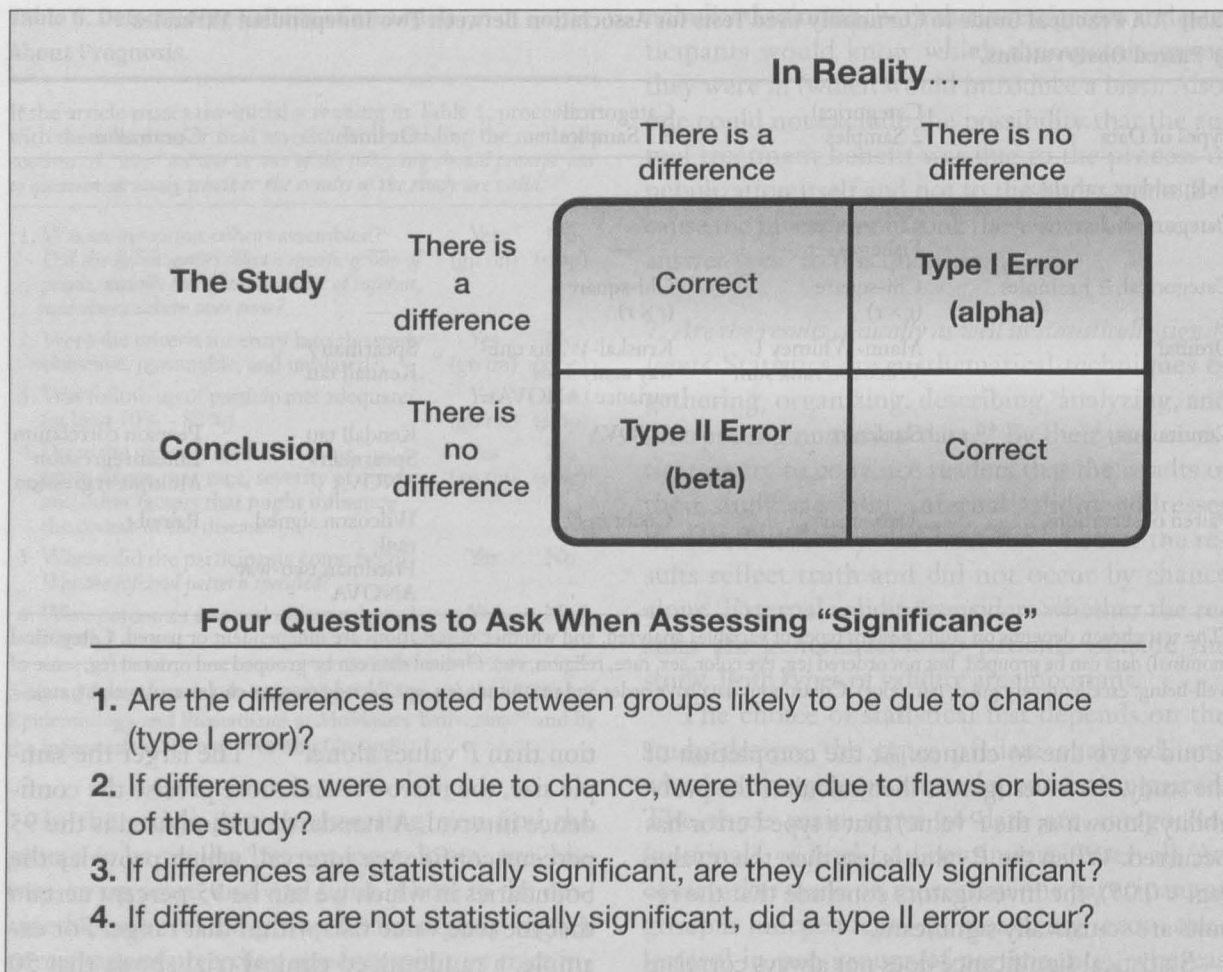
Instead of using *P* values, investigators are increasingly using confidence intervals (CIs) to determine the significance of a difference. The problem with *P* values are they convey no information about the size of differences or associations found in the study.<sup>66</sup> Also, *P* values provide a dichotomous answer—either the results are significant or not significant. In contrast, the confidence interval provides a range that will, with high probability, contain the true value and provide more informa-

tion than *P* values alone.<sup>67-69</sup> The larger the sample size, the narrower and more precise the confidence interval. A standard method used is the 95 percent confidence interval, which provides the boundaries in which we can be 95 percent certain that the true value falls within that range. For example, a randomized clinical trial shows that 50 percent of patients treated with drug A are cured compared with 45 percent of those treated with drug B. Statistical analysis of this 5 percent difference shows a *P* value of < 0.001 and a 95 percent confidence interval of 0 percent to 10 percent. The investigators conclude this improvement is statistically significant based on the *P* value. As a reader, however, you decide that a potential range of 0 percent to 10 percent is not clinically significant based on the 95 percent confidence interval.

In the article you are assessing, there was no statistical difference found among the groups in the change in croup score from baseline to final study assessment, time in the emergency department, hospitalization, and use of supplemental glucocorticoids. This trial is considered negative (no differences found). As such, you go on to the next question, which addresses these types of studies.

8. *If a negative trial, was a power analysis done? A type II error (Figure 6) occurs if the study finds no difference between groups when, in reality, there is a difference.*<sup>70</sup> This type of error is similar to a





**Figure 6. Potential outcomes of statistical decisions. If the study concludes there is a difference when in reality there is not, a type I error has occurred; a type II error happens if the study concludes there is no difference when in reality there is one.**

jury finding a criminal innocent of a crime. The odds of reaching a false-negative conclusion (known as  $\beta$ ) are typically set at 0.20 (20 percent chance). The power of a test ( $1-\beta$ ) is the ability to find a difference when in reality one exists, and depends on the (1) number of participants in the study (the more participants, the greater the power), and (2) size of the difference (known as effect size) between groups (the larger the difference, the greater the power). Typically, the effect size investigators choose depends on ethical, economic, and pragmatic issues, and can be categorized into small (10 to 25 percent), medium (26 to 50 percent), and large (greater than 50 percent).<sup>71</sup> When looking at the effect size chosen by the investigators, ask whether you consider this difference to be clinically meaningful.

Prior to the start of a study, the investigators should do a power analysis to determine how

many participants should be included in the study. Unfortunately, this step is often not done. Only 32 percent of the RCTs with negative results published between 1975 and 1990 in *JAMA*, *Lancet*, and *New England Journal of Medicine* reported sample size calculations; on review, the vast majority of these trials had too few patients that led to insufficient statistical power to detect a 25 percent or 50 percent difference.<sup>72</sup> Other studies have shown similar deficiencies in other journals and disciplines.<sup>14,48,73,74</sup> Whenever you read an article reporting a negative result, ask whether the sample size was large enough to permit investigators to draw such a conclusion. If a power analysis was done, check to find out whether the study had the required number of participants. If a power analysis was not done, view the conclusions with skepticism—it might be that the sample size was not large enough to detect a difference.

In the article you are assessing, you find that the investigators did perform a power analysis, which, using the criteria established above, required a minimum sample size of 62 participants per group. You notice that in the final analysis, each group had more than this number. You are assured that this study had adequate power to detect a type II error, and answer “yes” to this question.

9. *Were there other factors that might have affected the outcome?* At times an outcome might be due to factors other than the intervention. For example, the simple act of observation can affect an outcome (Hawthorne effect). This effect occurs when participants change their normal behavior because they are aware of being observed. To minimize this effect, the study groups should be observed equally. Also, randomization and sufficiently large sample size assure that both known and unknown determinants of an outcome are evenly distributed between groups. As you read through an article, think about potential influences that could impact one group more than another and thus affect the outcome.

In the article you are assessing, the investigators treated each of the groups equally (except for the intervention drugs). They also looked at such factors as earlier upper respiratory tract infections and episodes of croup, which could have had a potential impact. Since you can think of none, you answer “no” to this question.

10. *Are the treatment benefits worth the potential harms and costs?* This final question forces us to weigh the cost of the treatment versus the potential benefit and to consider the potential harm of the therapy. The common method used to weigh the benefits of treatment is the number needed to treat (NNT). The NNT takes into consideration the likelihood of an outcome or side effect.<sup>26</sup> Generally, the less common a potential outcome (eg, death), the more patients needing treatment to prevent that outcome it would require. For example, it might require 30 patients with severe stenosis to receive treatment with an anticoagulant to prevent one stroke. If sudden death is a potential risk of a medication used to treat a benign condition, one must question the actual benefit of that drug.

The investigators addressed this issue in the results section. Because the therapeutic interven-

tions were equal, oral dexamethasone was recommended as the preferred therapy because it is less expensive and easier to administer.

### **Conclusion of Case 2**

After a thorough assessment of this article, you conclude it is well designed with valid results. You feel confident that oral dexamethasone should be stocked in your office during croup season and that you will institute this treatment as a standard within your practice. As you apply this therapy, you also make a commitment to monitor its benefits and risks to your patients and to scan the literature for future articles that might offer additional information about croup therapy. Consistency of the results in your practice, as well as across multiple published studies, is one characteristic of the scientific process that leads to acceptance and implementation.

### **A Final Word**

With some practice and the use of the worksheets, one can quickly (within a few minutes) perform a critical assessment of an article. While performing this appraisal, it is important to keep in mind that few articles will be perfect. A critical assessment is rarely black-and-white, but often comes in shades of gray.<sup>24</sup> Only you can answer for yourself the exact shade of gray that you are willing to accept when deciding to apply the results of the study to your practice. By applying the knowledge, principles, and techniques described in this paper, however, you can more confidently recognize the various shades of gray and reject those articles that are seriously flawed.

### **Suggested Readings on Critical Reading Skills**

1. Slawson DC, Shaughnessy AF, Bennett JH. Becoming a medical information master: feeling good about not knowing everything. *J Fam Pract* 1994; 38:505-13.  
*A superb article that addresses the concepts of POEMs and DOEs.*
2. Shaughnessy AF, Slawson DC, Bennett JH. Becoming an information master: a guidebook to the medical information jungle. *J Fam Pract* 1994;39:489-99.  
*An excellent article that reviews how to manage one's way through the medical information jungle without getting lost or eaten alive.*
3. Shaughnessy AF, Slawson DC. Getting the most from review articles: a guide for readers and writers. *Am Fam Physician* 1997;55:2155-60.

- Provides useful techniques on reading a review article.*
4. How to read clinical journals: I-V. (Can Med Assoc J). *Original McMaster Series—despite being published in 1981, this series still has some great information!*
    - a. I. Why to read them and how to start reading them critically. 1981;124:555-8.
    - b. II. To learn about a diagnostic test. 1981;124:703-10.
    - c. III. To learn the clinical course and prognosis of disease. 1981;124:869-72.
    - d. IV. To determine etiology or causation. 1981;124:985-90.
    - e. V. To distinguish useful from useless or even harmful therapy. 1981;124:1156-62.
  5. Haynes RB, McKibbon KA, Fitzgerald D, Guyatt GH, Walker CJ, Sackett DL. How to keep up with the medical literature: I - VI (Ann Intern Med). *A good series on the approach to keeping up with the medical literature.*
    - a. I. Why try to keep up and how to get started. 1986;105:149-53.
    - b. II. Deciding which journals to read regularly. 1986;105:309-12.
    - c. III. Expanding the number of journals you read regularly. 1986;105:474-8.
    - d. IV. Using the literature to solve clinical problems. 1986;105:636-40.
    - e. V. Access by personal computer to the medical literature. 1986;105:810-6.
    - f. VI. How to store and retrieve articles worth keeping. 1986;105:978-84.
  6. The McMaster's series by the Evidence-Based Medicine Working Group - User's guide to the medical literature: I - XV. (JAMA). *The ultimate series written from the perspective of a busy clinician who wants to provide effective medical care but is sharply restricted in time for reading .*
    - a. Oxman AD, Sackett DL, Guyatt GH. I. How to get started. 1993;270:2093-5.
    - b. Guyatt GH, Sackett DL, Cook DJ. II. How to use an article about therapy or prevention. A. Are the results of the study valid? 1993;270:2598-601; B. What were the results and will they help me in caring for my patients? 1994;271:59-63.
    - c. Jaeschke R, Guyatt GH, Sackett DL. III. How to use an article about a diagnostic test. A. Are the results of the study valid? 1994;271:389-91; B. What are the results and will they help me in caring for my patients? 1994;271:703-7.
    - d. Levine M, Walter S, Lee H, Haines T, Holbrook A, Mayer V. IV. How to use an article about harm. 1994;271:1615-9.
    - e. Laupacis A, Wells G, Richardson WS, Tugwell P. V. How to use an article about prognosis. 1994;272:234-7.
    - f. Oxman AD, Cook DJ, Guyatt GH. VI. How to use an overview. 1994;272:1367-71.
    - g. Richardson WS, Detsky AS. VII. How to use a clinical decision analysis. A. Are the results of the study valid? 1995;273:1292-5; B. What are the results and will they help me in caring for my patients? 1995;273:1610-3.
    - h. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G. VII. How to use clinical practice guidelines. A. Are the recommendations valid? 1995;274:570-4; B. What are the recommendations and will they help you in caring for your patients? 1995;274:1630-2.
    - i. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. IX. A method for grading health care recommendations. 1995;274:1800-4.
    - j. Naylor CD, Guyatt GH. X. How to use an article reporting variations in the outcomes of health services. 1996;275:554-8.
    - k. Naylor DC, Guyatt GH. XI. How to use an article about a clinical utilization review. 1996;275:1435-9.
    - l. Guyatt GH, Naylor DC, Juniper E, Heyland DK, Jaeschke R, Cook DJ. XII. How to use articles about health-related quality of life. 1997;277:1232-7.
    - m. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? 1997;277:1552-7; B. What are the results and will they help me in caring for my patients? 1997;277:1802-6.
    - n. Dans AL, Dans LF, Guyatt GH, Richardson S. XIV. How to decide on the applicability of clinical trial results to your patient . 1998;279:545-9.
    - o. Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J. XV. How to use an article about disease probability for differential diagnosis. 1999;281:1214-9.
  7. How to read a paper (BMJ). *A great series that complements the User's Guide.*
    - a. Greenhalgh T. The MEDLINE database. 1997; 315:180-3.
    - b. Greenhalgh T. How to read a paper. Getting your bearings (deciding what the paper is about). 1997;315:243-6.
    - c. Greenhalgh T. Assessing the methodological quality of published papers. 1997;315:305-8.
    - d. Greenhalgh T. How to read a paper. Statistics for the non-statistician. I: different types of data need different statistical tests. 1997;315:364-6.
    - e. Greenhalgh T. Statistics for the non-statistician. II: "Significant" relations and their pitfalls. 1997: 315:422-5.
    - f. Greenhalgh T. Papers that report drug trials. 1997;315:480-3.
    - g. Greenhalgh T. Papers that report diagnostic or screening tests. 1997;315:540-3.
    - h. Greenhalgh T. Papers that tell you what things cost (economic analyses). 1997;315:596-9.
    - i. Greenhalgh T. Papers that summarize other pa-

pers (systematic reviews and meta-analyses). 1997;315: 672-5.

- j. Greenhalgh T, Taylor R. Papers that go beyond numbers (qualitative research). 1997;315:740-3.
8. Hulley ST, Cummings SR. Designing clinical research—an epidemiologic approach. Baltimore: Williams & Wilkins, 1988.  
*An excellent textbook on understanding research methods and statistics.*
9. Fletcher RH, Fletcher SW, Wagner EH. Clinical epidemiology: the essentials. 3rd ed. Baltimore: Williams & Wilkins, 1996.  
*A basic textbook written for clinicians and organized by clinical questions: diagnosis, treatment, etc.*
10. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston: Little, Brown, 1991.  
*A lively introduction to clinical epidemiology with special emphasis on diagnosis and treatment, by leading proponents of evidence-based medicine.*
11. Riegelman RK, Hirsch RP. Studying a study and testing a test: how to read the medical literature. 2nd ed. Boston: Little, Brown, 1989.  
*A clear description of an approach to studies of diagnosis and treatment.*
12. Gelbach SH. Interpreting the medical literature. 3rd ed. New York: McGraw-Hill, 1992.  
*A basic introduction.*

## References

1. Colditz GA, Hankinson SE, Hunter DJ, Willett WC, Manson JE, Stampfer MJ, et al. The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. *N Engl J Med* 1995;332: 1589-93.
2. Stampfer MJ, Willett WC, Colditz GA, Rosner B, Speizer FE, Hennekens CH. A prospective study of postmenopausal estrogen therapy and coronary heart disease. *N Engl J Med* 1985;313:1044-9.
3. Wilson P, Garrison R, Castelli W. Postmenopausal estrogen use, cigarette smoking, and cardiovascular morbidity in women over 50. The Framingham Study. *N Engl J Med* 1985;313:1038-43.
4. Stampfer MJ, Colditz GA, Willett WC, Manson JE, Rosner B, Speizer FE, et al. Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the Nurses' Health Study. *N Engl J Med* 1991;325:756-62.
5. Guidelines for counseling postmenopausal women about preventive hormone therapy. American College of Physicians. *Ann Intern Med* 1992;117:1038-41.
6. Gorsky RD, Koplan JP, Peterson HB, Thacker SB. Relative risks and benefits of long-term estrogen replacement therapy: a decision analysis. *Obstet Gynecol* 1994;83:161-6.
7. Dupont WD, Page DL. Menopausal estrogen replacement therapy and breast cancer. *Arch Intern Med* 1991;15:67-72.
8. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women: The Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. The Writing Group for the PEPI Trial. *JAMA* 1995;273:199-208.
9. Stanford JL, Weiss NS, Voigt LF, Daling JR, Habel LA, Rossing MA. Combined estrogen and progestin hormone replacement therapy in relation to risk of breast cancer in middle-aged women. *JAMA* 1995; 274:137-42.
10. Slawson DC, Shaughnessy AF, Bennett JH. Becoming a medical information master: feeling good about not knowing everything. *J Fam Pract* 1994;38: 505-13.
11. Shaughnessy AF, Slawson DC, Bennett JH. Becoming an information master: a guidebook to the medical information jungle. *J Fam Pract* 1994;39:489-99.
12. Fletcher R, Fletcher S. Keeping clinically up-to-date. Evidence-based approach to the medical literature. *J Gen Intern Med* 1997;12(Suppl):S5-S14.
13. Lock S. Does editorial peer review work? *Ann Intern Med* 1994;121:60-1.
14. Sonis J, Joines J. The quality of clinical trials published in *The Journal of Family Practice*, 1974-1991. *J Fam Pract* 1994;39:225-35.
15. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-9.
16. Altman DG. The scandal of poor medical research: we need less research, better research, and research done for the right reasons. *BMJ* 1994;308:283-4.
17. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;274:645-51.
18. Guyatt GH, Rennie D. Users' guides to the medical literature. *JAMA* 1993;270:2096-7.
19. How to read clinical journals: I. Why to read them and how to start reading them critically. *Can Med Assoc J* 1981;124:555-8.
20. How to read clinical journals: II. To learn about a diagnostic test. *Can Med Assoc J* 1981;124:703-10.
21. How to read clinical journals: III. To learn the clinical course and prognosis of disease. *Can Med Assoc J* 1981;124:869-72.
22. How to read clinical journals: IV. To determine etiology or causation. *Can Med Assoc J* 1981;124:985-90.
23. How to read clinical journals: V. To distinguish useful from useless or even harmful therapy. *Can Med Assoc J* 1981;124:1156-62.
24. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *JAMA* 1993;270:2093-5.
25. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention? A. Are the results of the study valid? *The Evidence-Based Medicine*

- Working Group. JAMA 1993;270:2598-601.
26. Guyatt GH, Sackett DL, Cook DJ. Users' 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? Evidence-Based Medicine Working Group. JAMA 1994;271:59-63.
  27. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1994;271:389-91.
  28. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. JAMA 1994;271:703-7.
  29. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. JAMA 1994;271:1615-9.
  30. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. JAMA 1994;272:234-7.
  31. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 1994;272:1367-71.
  32. Richardson WS, Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1995;273:1292-5.
  33. Richardson WS, Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. JAMA 1995;273:1610-3.
  34. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt GH. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? Evidence-Based Medicine Working Group. JAMA 1995;274:570-4.
  35. Wilson MC, Hayward RS, Tunis SR, Bass EB, Guyatt GH. Users' 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? Evidence-Based Medicine Working Group. JAMA 1995;274:1630-2.
  36. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. JAMA 1995;274:1800-4.
  37. Naylor CD, Guyatt GH. Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services. Evidence-Based Medicine Working Group. JAMA 1996;275:554-8.
  38. Naylor CD, Guyatt GH. Users' guides to the medical literature. XI. How to use an article about a clinical utilization review. Evidence-Based Medicine Working Group. JAMA 1996;275:1435-9.
  39. Guyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook DJ. Users' guides to the medical literature. XII. How to use articles about health-related quality of life. Evidence-Based Medicine Working Group. JAMA 1997;277:1232-7.
  40. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1997;277:1552-7.
  41. O'Brien BJ, Heyland D, Richardson WS, Levine M, Drummond MF. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. JAMA 1997;277:1802-6.
  42. Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature. XIV. How to decide on the applicability of clinical trial results to your patients. Evidence-Based Medicine Working Group. JAMA 1998;279:545-9.
  43. Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J. Users' guide to the medical literature. XV. How to use an article about disease probability for differential diagnosis. The Evidence-Based Medicine Working Group. JAMA 1999;281:1214-19.
  44. Klassen TP, Craig WP, Moher D, Osmond MH, Pasterkamp H, Sutcliffe T, et al. Nebulized budesonide and oral dexamethasone for treatment of croup: a randomized controlled trial. JAMA 1998;279:1629-32.
  45. Slawson DC, Shaughnessy AF, Ebell MW, Barry HC. Mastering medical information and the role of POEMs - Patient-Oriented Evidence that Matters. J Fam Pract 1997;45:195-6.
  46. Kassirer JP, Champion EW. Peer review. Crude and understudied, but indispensable. JAMA 1994;272:96-7.
  47. Abby M, Massey MD, Galandiuk S, Polk HC Jr. Peer review is an effective screening process to evaluate medical manuscripts. JAMA 1994;272:105-7.
  48. Goodman SN, Berlin J, Fletcher SW, Fletcher RH. Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. Ann Intern Med 1994;121:11-21.
  49. Gardner MJ, Bond J. An exploratory study of statistical assessment of papers published in the *British Medical Journal*. JAMA 1990;263:1355-7.
  50. Justice AC, Berlin JA, Fletcher SW, Fletcher RH, Goodman SN. Do readers and peer reviewers agree on manuscript quality? JAMA 1994;272:117-9.



51. Colaianni LA. Peer review in journals indexed in *Index Medicus*. JAMA 1994;272:156-8.
52. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. JAMA 1992;267: 374-8.
53. Jadad AR, Rennie D. The randomized controlled trial gets a middle-aged checkup. JAMA 1998;279: 319-20.
54. Rennie D, Flanagin A. Publication bias. The triumph of hope over experience. JAMA 1992;267: 411-2.
55. Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. JAMA 1994;272:158-62.
56. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279:281-6.
57. Whitely WP, Rennie D, Hafner AW. The scientific community's response to evidence of fraudulent publication. The Robert Slutsky case. JAMA 1994; 272:170-3.
58. Bero LA, Galbraith A, Rennie D. The publication of sponsored symposiums in medical journals. N Engl J Med 1992;327:1135-40.
59. Rochon PA, Gurwitz JH, Cheung CM, Hayes JA, Chalmers TC. Evaluating the quality of articles published in journal supplements compared with the quality of those published in the parent journal. JAMA 1994;272:108-13.
60. Greenhalgh T. How to read a paper - Getting your bearings (deciding what the paper is about). BMJ 1997;315:243-6.
61. Franks P. Clinical trials. Fam Med 1988;20:443-8.
62. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. JAMA 1994;272:125-8.
63. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12.
64. O'Brien PC, Shampo MA. Statistics for clinicians. 1. Descriptive statistics. Mayo Clin Proc 1981;56:47-9.
65. Greenhalgh T. How to read a paper. Statistics for the non-statistician. I: different types of data need different statistical tests. BMJ 1997;315:364-6.
66. Grimes DA. The case for confidence intervals. Obstet Gynecol 1992;80:865-6.
67. Simon R. Confidence intervals for reporting results of clinical trials. Ann Intern Med 1986;105:429-35.
68. Braitman LE. Confidence intervals assess both clinical significance and statistical significance. Ann Intern Med 1991;114:515-7.
69. Gehlbach SH. Interpreting the medical literature. 3rd ed. New York: McGraw-Hill, 1992.
70. Detsky AS, Sackett DL. When was a "negative" clinical trial big enough? How many patients you needed depends on what you found. Arch Intern Med 1985;145:709-12.
71. Raju TN, Langenberg P, Sen A, Aldana O. How much "better" is good enough? The magnitude of treatment effect in clinical trials. Am J Dis Child 1992;146:407-11.
72. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994;272:122-4.
73. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. N Engl J Med 1978;299:690-4.
74. Mengel MB, Davis AB. The statistical power of family practice research. Fam Pract Res J 1993;13:105-11.