ELSEVIER
SAUNDERS

Prim Care Clin Office Pract
33 (2006) 839–862

PRIMARY CARE:
CLINICS IN
OFFICE PRACTICE

Finding Truth from the Medical Literature: How to Critically Evaluate an Article

William F. Miser, MD, MA

Department of Family Medicine, The Ohio State University College of Medicine, 2231 North High Street, Room 203, Columbus, OH 43201, USA

With Internet access available to all, patients are increasingly gaining access to medical information, and then looking to their primary care physician for its interpretation. Gone are the days when what the physician says goes unchallenged by a patient. Our society 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." Patients are searching for answers even before they come to the office, and are bringing with them articles they have downloaded from the Internet for interpretation.

Primary care physicians also encounter an "information jungle" when it comes to the medical literature [1,2]. The amount of information available can be overwhelming [3]. There were 682,121 articles recorded in Pub MED in 2005. If clinicians, trying to keep up with the medical literature, were to read two articles per day, in just 1 year they would be over nine centuries behind in their reading!

Despite the volume of medical literature, fewer than 15% of all articles published on a particular topic are useful for clinical practice [4]. Most articles are not peer-reviewed, are sponsored by those with commercial interests, or arrive free in the mail (the so-called "throwaways"). Even articles published in the most prestigious journals are far from perfect. Analyses of clinical trials published in a wide variety of journals have identified large deficiencies in design, analysis, and reporting; although improving over time, the average quality score of clinical trials over the past 2 decades is less than 50% [5–7]. This has resulted in diagnostic tests and therapies becoming established as a routine part of practice before being rigorously

E-mail address: miser.6@osu.edu

doi:10.1016/j.pop.2006.09.012

evaluated; which has led to the widespread use of tests with uncertain efficacy, and treatments that are either ineffective or that may do more harm than good [8]. A good recent example is the widespread use of hormonal replacement therapy to prevent cardiovascular disease, dementia, and other chronic diseases; the Women's Health Initiative studies showed that this practice did more harm than good [9].

Although several excellent services are available to physicians that sift through and critically assess the medical literature, they are not helpful when a patient brings in the latest article that is "hot off the presses." Thus, physicians must have basic skills in judging the validity and clinical importance of these articles. The two major types of articles (Fig. 1) found in the medical literature are those that (1) 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). This article provides an overview of a systematic, efficient, and effective approach to the critical review of original research. This information is pertinent to physicians no matter the clinical setting. Because of space limitations, this article cannot cover everything in exhaustive detail, and the reader is encouraged to refer to the suggested readings in Appendix 1 for further assistance.

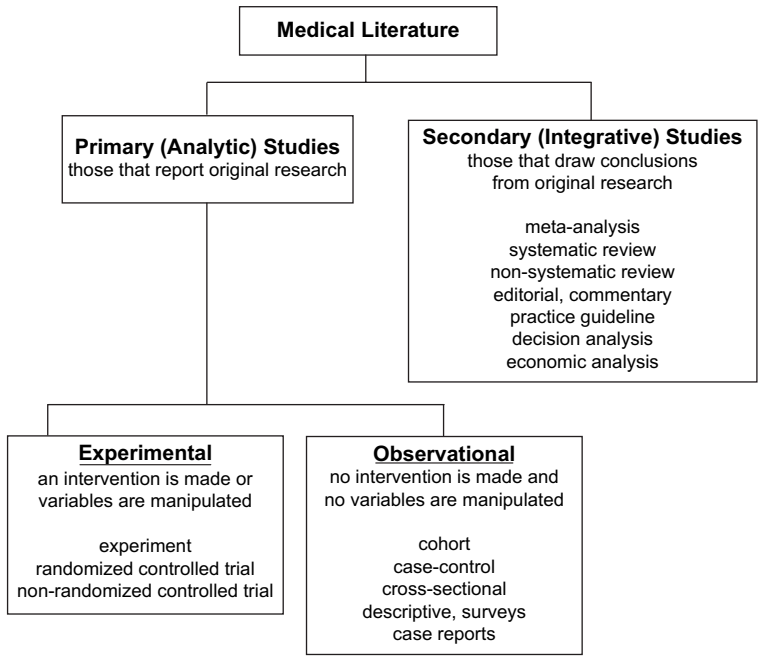

Fig. 1. The major types of studies found in the medical literature.

Critical assessment of an original research article

It is important for clinicians to master the ability to critically assess an original research article if they are to apply "evidence-based medicine" to the daily clinical problems they encounter. Most busy clinicians, however, do not have the hours required to fully critique an article; they need a brief and efficient screening method that allows them to know if the information is valid and applicable to their practice. By applying the techniques offered here, one can approach the literature confidently and base clinical decisions on "evidence rather than hope" [10].

This approach is modified and adapted from several excellent sources. The Department of Clinical Epidemiology and Biostatistics at McMaster University in Hamilton, Ontario, Canada in 1981 published a series of useful guides to help the busy clinician critically read clinical articles about diagnosis, prognosis, etiology, and therapy [11–15]. 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 (see Appendix 1) [10]. The recommendations from these users' guides form the foundation upon which techniques developed by Slawson and colleagues are modified and added [1,2]. With an article in hand, the process involves three steps: (1) conduct an initial validity and relevance screen, (2) determine the intent of the article, and (3) evaluate the validity of the article based on its intent.

Step one: conduct an initial validity and relevance screen

The first step when looking at an article is to ask, "Is this article worth taking the time to review in depth?" This can be answered within a few seconds by asking six simple questions (Appendix 2). A "stop" or "pause" answer to any of these questions should prompt one to seriously consider whether time should be spent to critically assess the article.

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, journals sent to clinicians unsolicited and free of charge are known as "throwaway" journals. These journals, although attractive in appearance, are not peer-reviewed, but instead are often geared toward generating income from advertising, and consist of "expert opinions" [3,10].

Articles published in the major peer-reviewed journals have already undergone an extensive process to sift 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 first establishes whether the manuscript is suitable for that journal, and then, if

acceptable, sends it to several reviewers for assessment. Peer reviewers are not part of the editorial staff, but usually are volunteers who have expertise in both the subject matter and research design. This peer review process acts as a sieve by detecting those studies that are flawed by poor design, are trivial, or are uninterpretable. This process, along with subsequent revisions and editing, improves the quality of the paper and its statistical analyses [16–19]. The *Annals of Internal Medicine*, for example, receives more than 1200 original research manuscript submissions each year. The editorial staff reject half after an internal review, and the remaining half are sent to at least two peers for review. Of the original 1200 submissions, only 15% are subsequently published [20].

Because of these strengths, peer review has become the accepted method for improving the quality of the science reported in the medical literature [21]; however, this mechanism is far from perfect, and it does not guarantee that the published article is without flaw or bias [4]. 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 [22–25]. Also, the speed of publication depends on the direction and strength of the trial results; trials with negative results may take twice as long to be published as do positive trials [26]. Finally, no matter how good the peer review system, fraudulent research, although rare, is extremely hard to identify [27].

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 one is in a rural general practice and the study was performed in a university subspecialty clinic, one may want to pause and consider the potential biases that may be present. This is a ''soft'' area, and rarely will one want to reject an article outright at this juncture; however, large differences in types of populations should raise caution in accepting the final results.

Is the study sponsored by an organization that may influence the study design or results?

This question considers the potential bias that may occur from outside funding. In most journals, investigators are required to identify 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 [28]. Also, randomized clinical trials

(RCTs) published in journal supplements are generally of inferior quality compared with articles published in the parent journal [29]. 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, however, a pharmaceutical company or other commercial organization funded the study, look for assurances from investigators that this association did not influence the design and results.

The answers to the next three questions deal with clinical relevance to one's practice, and can be obtained by reading the conclusion and selected portions of the abstract. Clinical relevance is important to not only physicians, but to patients. Rarely is it worthwhile to read an article about an uncommon condition one never encounters in practice, or about a treatment or diagnostic test that is not, and never will be, available because of cost or patient preference. Reading these types of articles may satisfy one's intellectual curiosity, but will not impact significantly on the practice. Slawson and colleagues [1,30] 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). 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 undergo this screening enjoy an improved quality and length of life (a POEM), one would probably want to choose the latter.

Will this information, if true, have a direct impact on the health of my patients, and is it something they will care about?

Typically the abstract will contain this information. Outcomes such as quality of life, overall mortality, and cost are ones that physicians and patients often consider important.

Is the problem addressed one that is common to my practice, and is the intervention or test feasible and available to me?

Problems addressed should be something commonly encountered in practice, tests should be feasible, and therapy should be easily available.

Will this information, if true, require me to change my current practice?

If one's practice already includes this diagnostic test or therapeutic intervention, this article reinforces what is being done; if not, however, then time should be spent on determining whether or not the results are valid before making any changes.

In only a few seconds, one can quickly answer six pertinent questions that allow one to decide if more time is needed to critically assess the article. This "weeding" tool allows one to discard those articles that are not relevant to practice, thus allowing more time to examine the validity of those few articles that may have a direct impact on the care of one's patients.

Step two: determine the intent of the article

If the physician decides to continue with the article after completing step one, the next task is to determine why the study was performed, and what clinical questions the investigators were addressing [31]. 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 1). The answer to this step can 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.

Step three: evaluate the validity of the article based on its intent

After an article has successfully passed the first two steps, it is now time to critically assess its validity and applicability to one's practice setting. Each of the four clinical categories found in Table 1 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 Appendices 3–6.

To get started on this step, read the entire abstract, survey the boldface headings, review the tables, graphs, and illustrations, and then skim-read the first sentence of each paragraph to quickly grasp the organization of

Table 1
Major clinical categories of primary research and preferred study designs

Clinical category	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 (see Fig. 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 "gold standard") (Fig. 3)
Causation—Determines whether an agent is related to the development of an illness	Cohort or case-control study, depending on how the rarity of disease; case reports may also provide crucial information (Figs. 4, 5)
Prognosis—Determines what is likely to happen to someone whose disease is detected at an early stage.	Longitudinal cohort study (see Fig. 4)

Adapted from Greenhalgh T. How to read a paper—getting your bearings (deciding what the paper is about). BMJ 1997;315:243–6; with permission.

the article. One then needs to focus on the methods section, answering a specific list of questions based on the intent of the article.

Is the study a randomized controlled trial?

Randomized controlled trials (RCTs) (Fig. 2) are considered the "gold 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 diabetic medication) or another (eg, an established diabetic medication or placebo). Both groups are then followed for a specified period, and defined outcomes (eg, glucose control, quality of life, death) are measured and analyzed at the conclusion.

Randomization diminishes the potential for investigators selecting individuals in a way that would unfairly bias one treatment group over another (selection bias). It is important to determine how the investigators actually

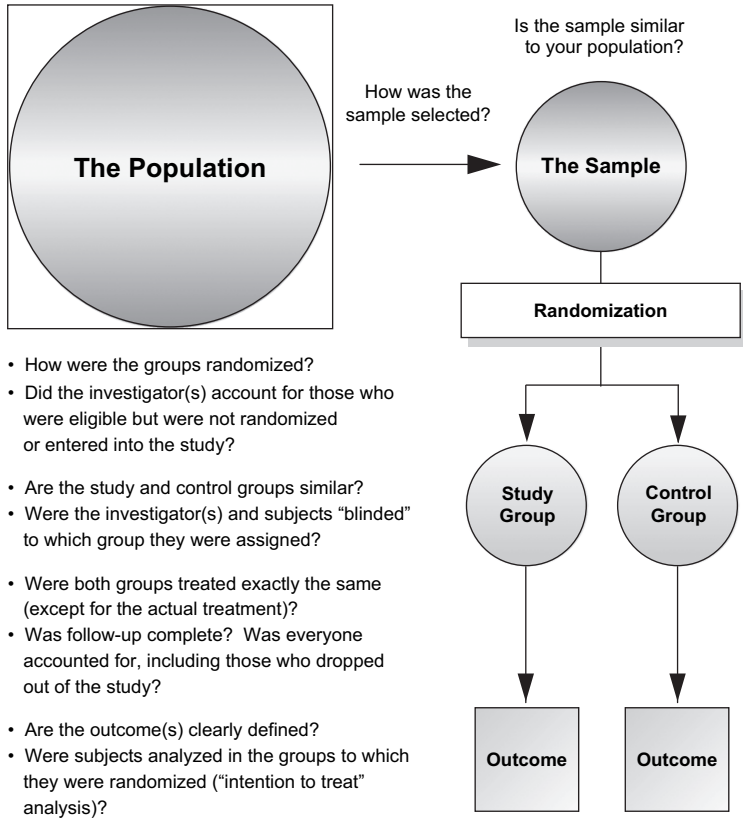

Fig. 2. The randomized controlled trial, considered the "gold standard" for studies dealing with treatment or other interventions.

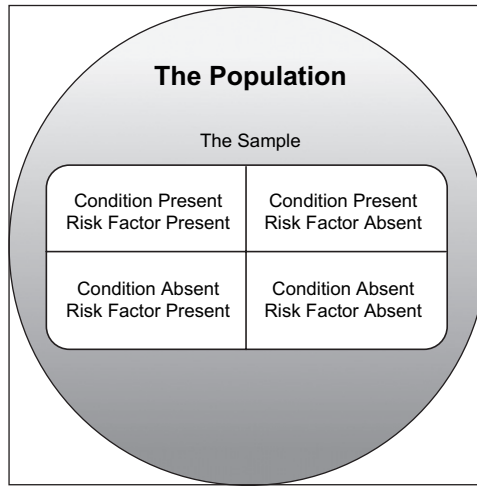

Fig. 3. The cross-sectional (prevalence) study. This design is most often used in studies on diagnostic or screening tests.

performed the randomization. Although infrequently reported in the past, most journals now require a standard format that provides this information [6]. Various techniques can be used for randomization [32]. Investigators may 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 may affect the study. To avoid these problems, investigators may use blocked randomization (groups are equal in size) or stratified randomization (subjects are randomized within groups based on potential confounding factors such as age or gender).

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 subjects are aware of the allocation, they may unwittingly (or otherwise) assign those who have a better prognosis to the treatment group and those who have a worse prognosis to the control group. RCTs that have inadequate allocation concealment will yield an inflated treatment effect that is up to 30% better than those trials with proper concealment [33,34].

Are the subjects in the study similar to mine?

To be generalizable (external validity), the subjects in the study should be similar to the patients in one's practice. A common problem encountered by

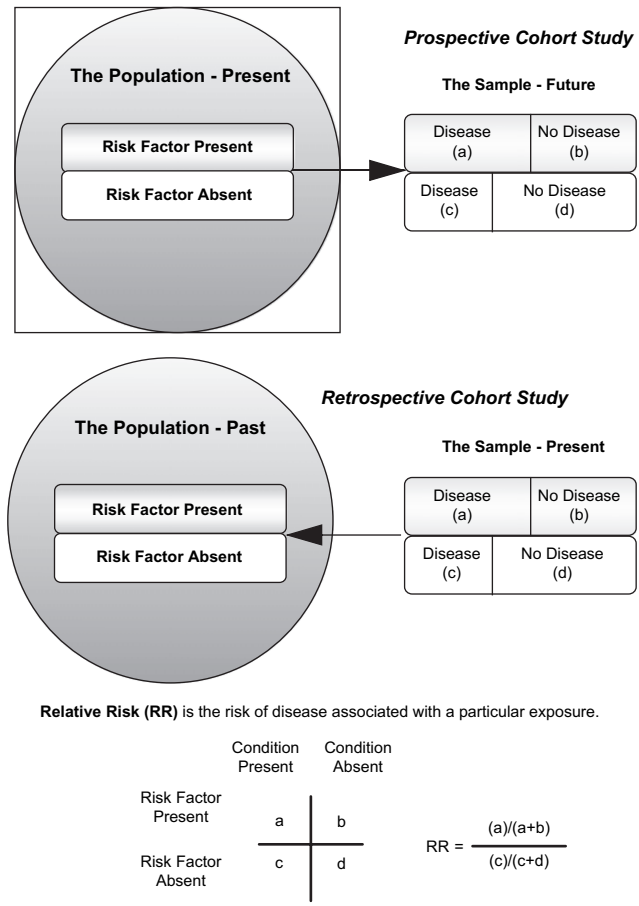

Fig. 4. Prospective and retrospective cohort study. These types of studies are often used for determining causation or prognosis. Data are typically analyzed using relative risk.

primary care physicians is interpreting the results of studies done on patients in subspecialty care clinics. For example, the group of men participating in a study on early detection of prostate cancer at a university urology practice may be different from the group of men seen in a typical primary care office. It is important to determine who was included and who was excluded from the study.

Are all participants who entered the trial properly accounted for at its conclusion?

Another strength of RCTs is that participants are followed prospectively; however, it is important that these participants be accounted for at the end of the trial to avoid a "loss-of-subjects bias," which can occur through the

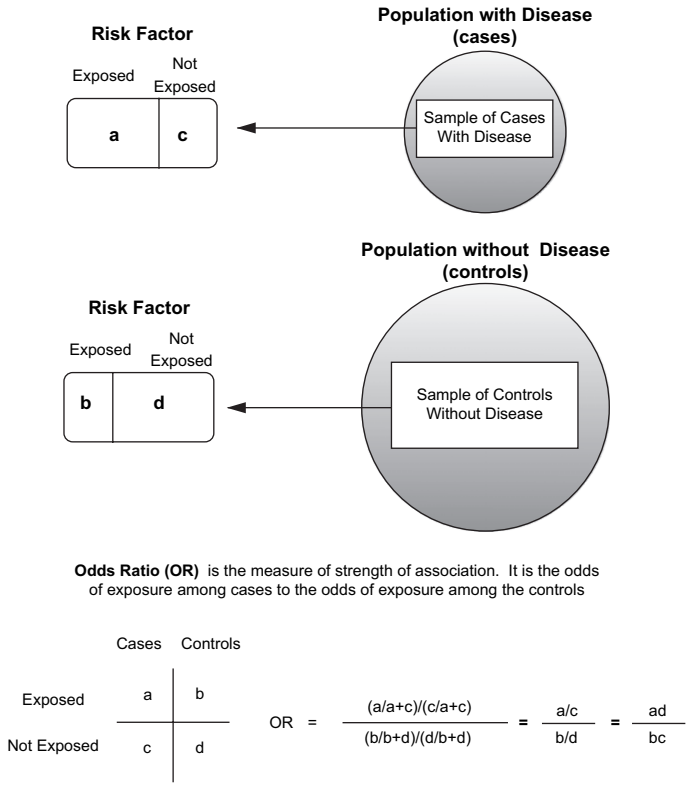

Fig. 5. The case-control study, a retrospective study in which the investigator selects a group with disease (cases) and one without disease (controls) and looks back in time at exposure to potential risk factors to determine causation. Data are typically analyzed using the odds ratio.

course of a prospective study as subjects drop out of the investigation for various reasons. Subjects may lose interest, move out of the area, develop intolerable side effects, or die. The subjects who are lost to follow-up may be different from those who remain in the study to the end, and the groups studied may have different rates of dropouts. An attrition rate of greater than 10% for short-term trials and 15% for long-term trials may invalidate the results of the study.

At the conclusion of the study, subjects should be analyzed in the group in which they were originally randomized, even if they were noncompliant or switched groups (intention-to-treat analysis). For example, a study wishes 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, individuals randomized to endarterectomy were found to be

poor surgical candidates, they may instead be treated medically; however, at the conclusion of the study, their outcomes (stroke, death) should be included in the surgical group, even if they didn't have surgery—to do otherwise would unfairly inflate the benefit of the surgical approach. Most journals now require a specific format for reporting RCTs, which includes a chart that allows you to easily follow the flow of subjects through the study [6].

Was everyone involved in the study (subjects and investigators) "blind" to treatment?

Investigator bias may occur when those making the observations may unintentionally "shade" the results to confirm the hypothesis or to influence the subjects. The process of masking, in which neither the investigators nor the subjects are aware of group assignment (ie, double-blinding), prevents this bias. For example, in a study comparing a new diabetic medication to a placebo, neither the investigators nor the subjects should be aware of what the subjects 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 taste or other side effect, and the comparison medication does not, subjects may be able to guess what medicine they are on, which may then influence how they perceive their improvement.

Were the intervention and control groups similar at the start of the trial?

Through the process of randomization, one would anticipate the groups to be similar at the beginning of a trial. Because this may not always be the case, investigators should provide a group comparison. This information is usually found in the first table of the article.

Typically, comparisons will be made for demographic factors, other known risk factors, and disease severity. If differences exist between groups, one must use clinical experience and judgment to determine if small differences are likely to influence outcomes.

Were the groups treated equally (aside from the experimental intervention)?

To ensure both proper blinding and that other unknown determinants are not a factor, groups should be treated equally except for the therapeutic intervention. Everyone should be seen with the same frequency, and interventions should be similar. One should look for assurances that the groups were treated equally except for the experimental intervention.

Are the results clinically as well as statistically significant?

Statistics are mathematical techniques of gathering, organizing, describing, analyzing, and interpreting numerical data [35]. 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 if the results reflect truth and did not occur by chance alone. External validity considers whether the results are generalizable to patients outside of 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 individual or group is "independent" (eg, measuring serum cholesterol in two groups of subjects), whereas making more than one observation on an individual is "paired" (eg, measuring serum cholesterol in an individual before and after treatment). Based on this information, one can then select an appropriate statistical test (Table 2). 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 may be attempting to prove something statistically significant that truly has no significance [36].

There are two types of errors that can potentially occur when comparing the results of a study to "reality." 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% risk that any differences 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 antihypertensives in over 1000 subjects may find a "statistically significant" difference in mean blood pressures of only 3 mmHg, which in the clinical realm is trivial. A *P* value of less than 0.0001 is no more clinically significant than a value of less than 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 (CI) 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 [37]. Also, *P* values provide a dichotomous answer—either the results are "significant" or "not significant." In contrast,

Table 2
A practical guide to commonly used statistical tests

Types of data	Categorical, 2 samples	Categorical, ≥3 samples	Ordinal	Continuous
Tests for association between two independent variables				
Categorical, 2 samples	• Chi-square • Fisher's exact	-	-	-
Categorical, ≥3 samples	Chi-square (r × r)	Chi-square (r × r)	-	-
Ordinal	• Mann-Whitney U • Wilcoxon rank-sum	Kruskal-Wallis one-way analysis of variance (ANOVA)	• Spearman'sr • Kendall's Tau	-
Continuous	Student's t	ANOVA	• Kendall's Tau • Spearman'sr • ANOVA	• Pearson correlation • Linear regression • Multiple regression
Tests for association between paired observations				
	McNemar's	Cochran Q	• Wilcoxon signed rank • Friedman two-way ANOVA	Paired t

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, gender, 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).

the CIl provides a range that will, with high probability, contain the true value, and provides more information than *P* values alone [38–40]. The larger the sample size, the narrower and more precise is the CI. A standard method used is the 95% CI, which provides the boundaries in which we can be 95% certain that the true value falls within that range. For example, a randomized clinical trial demonstrates that 50% of patients treated with drug A are cured, compared with 45% of those treated with drug B. Statistical analysis of this 5% difference shows a *P* value of less than 0.001 and a 95% CI of 0% to 10%. The investigators conclude this is a statistically significant improvement based on the *P* value; however, as a reader, you decide that a potential range of 0% to 10% is not clinically significant based on the 95% CI.

If a negative trial, was a power analysis done?

A negative trial is one in which no differences were found using the intervention between the groups. A Type II error occurs when the study finds no difference between groups when, in reality, there is a difference [41]. This type of error is similar to a jury finding a criminal innocent of a crime. The odds of reaching a false-negative conclusion (known as "beta") is typically set at 0.20 (20% chance). The power of a test (1-beta) is the ability to find a difference when in reality one exists, and depends on: (1) the number of subjects in the study (the more subjects, the greater the power), and (2) the 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%–25%), medium (26%–50%), and large (>50%) [42]. When looking at the effect size chosen by the investigators, ask whether you consider this difference to be clinically meaningful.

Before the start of a study, the investigators should do a "power analysis" to determine how many subjects should be included in the study. Unfortunately, this was often not done in the past. Only 32% 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, which led to insufficient statistical power to detect a 25% or 50% difference [43]. Other studies have shown similar deficiencies in other journals and disciplines [5,19,44,45]. Whenever one reads 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 see if the study had the required number of subjects. If a power analysis was not done, view the conclusions with skepticism—it may be that the sample size was not large enough to detect a difference.

Were there other factors that might have affected the outcome?

At times, an outcome may be caused by factors other than the intervention. For example, the simple act of observation can affect an outcome

(Hawthorne effect). This effect occurs when subjects change their normal behavior because they are aware of being observed. To minimize this effect, 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 one reads through an article, think about potential influences that could impact one group more than another, and thus affect the outcome.

Are the treatment benefits worth the potential harms and costs?

This final question forces one to consider the cost benefit and potential harm of the therapy. The number needed to treat (NNT) takes into consideration the likelihood of an outcome or side effect [46]. Generally, the less common a potential outcome (eg, death), the greater the number of patients that would require treatment to prevent one outcome. 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.

If, based upon a critical review of an article, one decides to implement a new test or therapy, one must also make a commitment to monitor its benefits and risks to patients, and to scan the literature for future articles that may offer additional findings. Consistency of the results in one's 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 [47]. 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 presented in this section, however, you can more confidently recognize the various shades of gray, and reject those articles that are seriously flawed.

Appendix 1

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 Phys 1997; 55:2155–60. [Provides useful techniques on reading a review article.]

Items 4–8 are from "How to read clinical journals," original McMaster series from *The Canadian Medical Association Journal*. [Despite being published in 1981, this series still has some great information!]

4. Why to read them and how to start reading them critically. Can Med Assoc J 1981;124:555–58.

5. To learn about a diagnostic test. Can Med Assoc J 1981;124:703–10.

6. To learn the clinical course and prognosis of disease. Can Med Assoc J 1981;124:869–72.

7. To determine etiology or causation. Can Med Assoc J 1981;124: 985–90,

8. To distinguish useful from useless or even harmful therapy. Can Med Assoc J 1981;124:1156–62.

Items 9–14 are from "How to keep up with the medical literature," in *Annals of Internal Medicine*. [A good series on the approach to keeping up with the medical literature.]

9. Haynes RB, McKibbon KA, Fitzgerald D, et al. Why try to keep up and how to get started. Ann Intern Med 1986;105:149–53.

10. Haynes RB, McKibbon KA, Fitzgerald D, et al. Deciding which journals to read regularly. Ann Intern Med 1986;105:309–12.

11. Haynes RB, McKibbon KA, Fitzgerald D, et al. Expanding the number of journals you read regularly. Ann Intern Med 1986;105:474–8.

12. Haynes RB, McKibbon KA, Fitzgerald D, et al. Using the literature to solve clinical problems. Ann Intern Med 1986;105:636–40.

13. Haynes RB, McKibbon KA, Fitzgerald D, et al. Access by personal computer to the medical literature. Ann Intern Med 1986;105:810–6.

14. Haynes RB, McKibbon KA, Fitzgerald D, et al. How to store and retrieve articles worth keeping. Ann Intern Med 1986;105:978–84.

Items 15–45 are from The McMaster's series—"User's guide to the medical literature" in *JAMA: The Journal of the American Medical Association*. This material can now be found in an interactive format at http://pubs.ama-assn.org/misc/usersguides.dtl. [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.]

15. Oxman AD, Sackett DL, Guyatt GH. How to get started. JAMA 1993;270:2093–8.

16. Guyatt GH, Sackett DL, Cook DJ. How to use an article about therapy or prevention. A. Are the results of the study valid? JAMA 1993;270:2598–601

17. Guyatt GH, Sackett DL, Cook DJ. 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 1994;271:59–63.

18. Jaeschke R, Guyatt GH, Sackett DL. How to use an article about a diagnostic test. A. Are the results of the study valid? JAMA 1994;271:389–91.

19. Jaeschke R, Guyatt GH, Sackett DL. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? JAMA 1994;271:703–07.

20. Levine M, Walter S, Lee H, et al. How to use an article about harm. JAMA 1994;271:1615–9.

21. Levine M, Walter S, Lee H, et al. How to use an article about prognosis. JAMA 1994;272:234–37.

22. Oxman AD, Cook DJ, Guyatt GH. How to use an overview. JAMA 1994; 272:1367–71.

23. Richardson WS, Detsky AS. How to use a clinical decision analysis. A. Are the results of the study valid? JAMA 1995;273:1292–5.

24. Richardson WS, Detsky AS. How to use a clinical decision analysis. B. What are the results and will they help me in caring for my patients? JAMA 1995;273:1610–23.

25. Hayward RS, Wilson MC, Tunis SR, et al. How to use clinical practice guidelines. A. Are the recommendations valid? 1995;274:570–4.

26. Wilson MC, Hayward RS, Tunis SR, et al. How to use clinical practice guidelines. B. What are the recommendations and will they help you in caring for your patients? JAMA 1995;274:1630–62.

27. Guyatt GH, Sackett DL, Sinclair JC. A method for grading health care recommendations. JAMA 1995;274:1800–4.

28. Naylor CD, Guyatt GH. How to use an article reporting variations in the outcomes of health services. JAMA 1996;275:554–8.

29. Naylor CD, Guyatt GH. How to use an article about a clinical utilization review. JAMA 1996;275:1435–9.

30. Guyatt GH, Naylor CD, Juniper E, et al. How to use articles about health-related quality of life. JAMA 1997;277:1232–7.

31. Drummond MF, Richardson WS, O'Brien BJ, et al. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? JAMA 1997;277:1552–7.

32. O'Brien BJ, Heyland D, Richardson WS, et al. 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? JAMA 1997;277:1802–06.

33. Dans AL, Dans LF, Guyatt GH, et al. How to decide on the applicability of clinical trial results to your patients. JAMA 1998;279:545–9.

34. Richardson WS, WIlson MC, Guyatt GH, et al. How to use an article about disease probability for differential diagnosis. JAMA 1999;281:1214–9.

35. Guyatt GH, Sinclair J, Cook DJ, et al. How to use a treatment recommendation. JAMA 1999;281:1836–43.

36. Randolph AG, Haynes RB, Wyatt JC. How to use an article evaluating the clinical impact of a computer-based clinical decision support system. JAMA 1999;282:67–74.

37. Bucher HC, Guyatt GH, Cook DJ. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. JAMA 1999;282:771–8.

38. McAlister FA, Laupacis A, Wells GA, et al. Applying clinical trial results. B. Guidelines for determining whether a drug is exerting (more than) a class effect. JAMA 1999;282(9):1371–7.

39. Hunt DL, Jaeschke R, McKibbon KA. Using electronic health information resources in evidence-based practice. JAMA 2000;283:1875–9.

40. McAlister FA, Strauss SE, Guyatt GH, et al. Integrating research evidence with the care of the individual patient. JAMA 2000;282:2829–36.

41. McGinn TG, Guyatt GH, Wyer PC, et al. How to use articles about clinical decision rules. JAMA 2000;284:79–84.

42. Giacomini MK, Cook DJ. Qualitative research in health care. A. Are the results of the study valid? JAMA 2000;284:357–62.

43. Giacomini MK, Cook DJ. Qualitative research in health care. What are the results and will they help me in caring for my patients? JAMA 2000;284:478–82.

44. Richardson WS, WIlson MC, Williams JW, et al. How to use an article on the clinical manifestation of disease. JAMA 2000;284:869–75.

45. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Evidence-based medicine: principles for applying the Users' Guides to patient care. JAMA 2000;284:1290–6.

Items 46–55 are from "How to read a paper" in the *British Medical Journal*. [A great series that compliments the User's guide.]

46. Greenhalgh T. The MEDLINE database. Br Med J 1997;315(7101): 180–3.

47. Greenhalgh T. Getting your bearings (deciding what the paper is about). Br Med J 1997;315(7102):24–6.

48. Greenhalgh T. Assessing the methodological quality of published papers. Br Med J 1997;315(7103):305–8.

49. Greenhalgh T. Statistics for the non-statistician. Br Med J 1997;315(7104):364–6.

50. Greenhalgh T. Statistics for the non-statistician. II: "Significant" relations and their pitfalls. Br Med J 1997;315(7105):422–5.

51. Greenhalgh T. Papers that report drug trials. Br Med J 1997;315(7106):480–3.

52. Greenhalgh T. Papers that report diagnostic or screening tests. Br Med J 1997;315(7107):540–3.

53. Greenhalgh T. Papers that tell you what things cost (economic analyses). Br Med J 1997;315(7108):596–9.

54. Greenhalgh T. Papers that summarize other papers (systemic reviews and meta-analyses). Br Med J 1997;315(7109):672–5.

55. Greenhalgh T. Papers that go beyond numbers (qualitative research). Br Med J 1997;315(7110):740–3.

56. Hulley SB, Cummings SR. Browner WS, Get al. Designing clinical research—an epidemiologic approach. Baltimore (MD): Lippincott, Williams & Wilkins; 2000. [An excellent textbook on understanding research methods and statistics.]

57. Fletcher RH, Fletcher SW. Clinical epidemiology: the essentials. 4th edition. Baltimore (MD): Lippincott, Williams and Wilkins; 2005. [A basic textbook written for clinicians and organized by clinical questions: diagnosis, treatment, and so on.]

58. Haynes RB, Sackett DL, Guyatt GH, et al. Clinical epidemiology: how to do clinical practice research. 4th edition. Baltimore (MD): Lippincott, Williams and Wilkins; 2006. [A lively introduction to clinical epidemiology, with special emphasis on diagnosis and treatment, by leading proponents of "evidence-based medicine."]

59. Riegelman RK. Studying a study and testing a test: how to read the medical evidence. 4th edition. Baltimore (MD): Lippincott, Williams & Wilkins; 2000. [A clear description of an approach to studies of diagnosis and treatment.]

60. Gelbach SH. Interpreting the medical literature. 4th edition. New York: McGraw-Hill; 2002. [A basic introduction.]

Appendix 2

Step one in critically assessing an original research article

Initial validity and relevance screen: is this article worth taking the time to review in depth? A "stop" or "pause" answer to any of the following should prompt one to seriously question whether one should spend the time to critically review the article.

1. Is the article from a peer-reviewed journal? Articles published in a peer-reviewed journal have already gone through an extensive review and editing process.	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 (stop)
3. Is the study sponsored by an organization that may influence the study design or results?	Yes (pause)	No (stop)
Read the conclusion of the abstract to determine relevance.		
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)

Questions 4–6 *adapted from* Slawson D, Shaughnessy A, Ebell M, et al. Mastering medical information and the role of POEMs—Patient-Oriented Evidence that Matters. J Fam Pract 1997;45:195–6.

Appendix 3

Determining validity of an article about therapy

If the article passes the initial screen in Appendix 2, proceed with the following critical assessment by reading the Methods section. A ''stop'' answer to any of the following should prompt one to seriously question whether the results of the study are valid and whether one should use this therapeutic intervention.

1. Is the study a randomized controlled trial? a. How were patients selected for the trial? b. Were they properly randomized into groups using concealed assignment?	Yes (go on)	No (stop)
2. Are the subjects 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? 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)?	Yes (go on)	No (stop)
4. Was everyone involved in the study (subjects and investigators) ''blind'' to treatment?	Yes (go on)	No (stop)
5. Were the intervention and control groups similar at the start of the trial? (Check Appendix 1)	Yes (go on)	No (stop)
6. Were the groups treated equally (aside from the experimental intervention)?	Yes (go on)	No (stop)
7. Are the results clinically as well as statistically significant? Were the outcomes measured clinically important?	Yes (go on)	No (stop)
8. If a negative trial, was a power analysis done?	Yes (go on)	No (stop)
9. Were there other factors that might have affected the outcome?	Yes (go on)	No (stop)
10. Are the treatment benefits worth the potential harms and costs?	Yes (go on)	No (stop)

Adapted from Slawson D, Shaughnessy A, Bennett J. Becoming a medical information master: feeling good about not knowing everything. J Fam Pract 1994;38:505–13, and Guyatt G, Sackett D, Cook D. 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? The Evidence-Based Medicine Working Group. JAMA 1993;270:2598–601.

Appendix 4

Determining validity of an article about a diagnostic test

If the article passes the initial screen in Appendix 2, proceed with the following critical assessment by reading the Methods section. A "stop" answer to any of the following should prompt one to seriously question whether the results of the study are valid and whether one 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 "gold 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 subjects similar to my patients? Varying prevalences will affect the predictive value of the test in my practice. Yes (go on) No (stop)
6. Will my patients be better off as a result of this test? Yes (go on) No (stop)
7. What are the sensitivity, specificity, and predictive values of the test?

Sensitivity = (a)/(a + c) = _____

Specificity = (d)/(b + d) =_____

Positive predictive value = (a)/(a + b) = _____

Negative predictive value = (c)/(c + d) = _____

Test result	"Gold standard" result: Positive	Negative
Positive	a	b
Negative	c	d

Adapted from Slawson D, Shaughnessy A, Bennett J. Becoming a medical information master: feeling good about not knowing everything. J Fam Pract 1994;38:505–13, and Jaeschke R, Guyatt G, Sackett D. User's guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? The Evidence-Based Medicine Working Group. JAMA 1994;271:389–91.

Appendix 5

Determining validity of an article about causation

If the article passes the initial screen in Appendix 2, proceed with the following critical assessment by reading the Methods section. A "stop" answer to any of the following should prompt one to seriously question 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 relationship correct? Does the exposure to the agent precede the outcome?	Yes (go on)	No (stop)
6. Is there a dose-response gradient? As the quantity or the duration of exposure to the agent increases, does the risk of outcome likewise increase?	Yes (go on)	No (stop)
7. How strong is the association between exposure and outcome? Is the relative risk (RR) or odds ratio (OR) large?	Yes (go on)	No (stop)

Adapted from Levine M, Walter S, Lee H, et al. User's guides to the medical literature. IV. How to use an article about harm. The Evidence-Based Medicine Working Group. JAMA 1994;271:1615–9.

Appendix 6

Determining validity of an article about prognosis

If the article passes the initial screen in Appendix 2, proceed with the following critical assessment by reading the Methods section. A "stop" answer to any of the following should prompt one to seriously question whether the results of the study are valid.

1. Was an "inception cohort" assembled? Did the investigators identify a specific group of people initially free of the outcome of interest, and follow them forward in time?	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 subjects 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 subjects come from? (was the referral pattern specified?)	Yes (go on)	No (stop)
6. Were outcomes assessed objectively and blindly?	Yes (go on)	No (stop)

Adapted from Slawson D, Shaughnessy A, Bennett J. Becoming a medical information master: feeling good about not knowing everything. J Fam Pract 1994;38:505–13, and Laupacis A, Wells G, Richardson W, et al. User's guides to the medical literature. V. How to use an article about prognosis. The Evidence-Based Medicine Working Group. JAMA 1994;272:234–37.

References

[1] Slawson D, Shaughnessy A, Bennett J. Becoming a medical information master: feeling good about not knowing everything. J Fam Pract 1994;38:505–13.

[2] Shaughnessy A, Slawson D, Bennett J. Becoming an information master: a guidebook to the medical information jungle. J Fam Pract 1994;39:489–99.

[3] Fletcher R, Fletcher S. Keeping clinically up-to-date. Evidence-based approach to the medical literature. J Gen Intern Med 1997;12:S5–14.

[4] Lock S. Does editorial peer review work? [editorial]. Ann Intern Med 1994;121:60–1.

[5] Sonis J, Jones J. The quality of clinical trials published in *The Journal of Family Practice*, 1974–1991. J Fam Pract 1994;39:225–35.

[6] Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637–9.

[7] Altman D. The scandal of poor medical research: we need less research, better research, and research done for the right reasons. BMJ 1994;308:283–4.

[8] Reid M, Lachs M, Feinstein A. Use of methodological standards in diagnostic test research. Getting better but still not good. JAMA 1995;274:645–51.

[9] Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in health postmenopausal women: principal results from the Women's Health Initiative randomized controlled Trial. JAMA 2002;288(3):321–33.

[10] Guyatt G, Rennie D. Users' guides to the medical literature [editorial]. JAMA 1993;270: 2096–7.

[11] Department of Clinical Epidemiology and Biostatistics. McMaster University. How to read clinical journals: I. Why to read them and how to start reading them critically. Can Med Assoc J 1981;124(5):555–8.

[12] Department of Clinical Epidemiology and Biostatistics. McMaster University. How to read clinical journals: II. To learn about a diagnostic test. Can Med Assoc J 1981;124: 703–10.

[13] Department of Clinical Epidemiology and Biostatistics. McMaster University. How to read clinical journals: III. To learn the clinical course and prognosis of disease. Can Med Assoc J 1981;124:869–72.

[14] Department of Clinical Epidemiology and Biostatistics. McMaster University. How to read clinical journals: IV. To determine etiology or causation. Can Med Assoc J 1981;124:985–90.

[15] Department of Clinical Epidemiology and Biostatistics. McMaster University. How to read clinical journals: V. To distinguish useful from useless or even harmful therapy. Can Med Assoc J 1981;124:1156–62.

[16] Kassirer J, Campion E. Peer review—crude and understudied, but indispensable. JAMA 1994;272:96–7.

[17] Abby M, Massey M, Galandiuk S, Polk H. Peer review is an effective screening process to evaluate medical manuscripts. JAMA 1994;272:105–7.

[18] Goodman S, Berlin J, Fletcher S, et al. Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. Ann Intern Med 1994;121:11–21.

[19] Gardner M, Bond J. An exploratory study of statistical assessment of papers published in the *British Medical Journal*. JAMA 1990;263:1355–7.

[20] Justice A, Berlin J, Fletcher S, et al. Do readers and peer reviewers agree on manuscript quality? JAMA 1994;272:117–9.

[21] Colaianni L. Peer review in journals indexed in *Index Medicus*. JAMA 1994;272:156–8.

[22] Dickersin K, Min Y, Meinert C. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. JAMA 1992; 267(3):374–8.

[23] Jadad A, Rennie D. The randomized controlled trial gets a middle-aged checkup [editorial]. JAMA 1998;279:319–20.

[24] Rennie D, Flanagin A. Publication bias—the triumph of hope over experience. JAMA 1992; 267:411–2.
[25] Scherer R, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts—a meta-analysis. JAMA 1994;272:158–62.
[26] Ioannidis J. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279:281–6.
[27] Whitely W, Rennie D, Hafner A. The scientific community's response to evidence of fraudulent publication. The Robert Slutsky case. JAMA 1994;272:170–3.
[28] Bero L, Galbraith A, Rennie D. The publication of sponsored symposiums in medical journals. N Engl J Med 1992;327:1135–40.
[29] Rochon P, Gurwitz J, Cheung M, et al. 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.
[30] Slawson D, Shaughnessy A, Ebell M, et al. Mastering medical information and the role of POEMs—Patient-Oriented Evidence that Matters. J Fam Pract 1997;45:195–6.
[31] Greenhalgh T. How to read a paper—getting your bearings (deciding what the paper is about). BMJ 1997;315:243–6.
[32] Franks P. Clinical trials. Fam Med 1988;20:443–8.
[33] Schulz K, Chalmers I, Grimes D, et al. Assessing the quality of randomization from reports of controlled trials published in oObstetrics and gynecology journals. JAMA 1994;272:125–8.
[34] Schulz K, Chalmers I, Hayes R, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408–12.
[35] O'Brien P, Shampo M. Statistics for clinicians—1. Descriptive statistics. Mayo Clin Proc 1981;56:47–9.
[36] Greenhalgh T. How to read a paper—statistics for the non-statistician. BMJ 1997;315:364–6.
[37] Grimes D. The case for confidence intervals [editorial]. Obstet Gynecol 1992;80:865–6.
[38] Simon R. Confidence intervals for reporting results of clinical trials. Ann Intern Med 1986; 105:429–35.
[39] Braitman L. Confidence intervals assess both clinical significance and statistical significance. Ann Intern Med 1991;114:515–7.
[40] Gehlbach S. Interpreting the medical literature. 3rd edition. New York: McGraw-Hill; 1993.
[41] Detsky A, Sackett D. 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.
[42] Raju R, Langenberg P, Sen A, et al. How much "better" is good enough? The magnitude of treatment effect in clinical trials. Am J Dis Child 1992;146:407–11.
[43] Moher D, Dulberg C, Wells G. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994;272:122–4.
[44] Freiman J, Chalmers T, Smith H, et al. 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.
[45] Mengel M, Davis A. The statistical power of family practice research. Fam Pract Res J 1993; 13:105–11.
[46] Guyatt G, Sackett D, Cook D. 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? The Evidence-Based Medicine Working Group. JAMA 1994;271:59–63.
[47] Oxman A, Sackett D, Guyatt G. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. JAMA 1993;270:2093–5.