Evidence-Based Medicine, Part 4. An Introduction to Critical Appraisal of Articles on Harm

Roberto Cardarelli, DO, MPH
Margaret M. Seater, OMS IV

This article provides an introductory step-by-step process to appraise an article on harm. The authors introduce these principles using a systematic approach and case-based format. The process of assessing the validity of an article on harm, determining its importance, and applying it to an individual patient is reviewed. The concepts of study population homogeneity, equal treatment, sufficient follow-up periods, and completeness are discussed to help physicians determine an article's validity. Instruction on calculating odds ratios, relative risk, absolute risk increase, and the number needed to harm is provided and applied to the clinical scenario. Finally, information that is learned from the previous two steps is applied to patient care. Study generalizability and the role of patient values, patient expectations, and patient concerns are also addressed. The skills learned from appraising an article on harm in the manner outlined provides a solid basis for lifelong learning and improved patient care.

As noted, in this article, we introduce a strategy for busy physicians, physician residents, and medical students to critically assess the medical literature on harm. In-depth details of research methods are beyond the scope of this introductory series on EBM. Readers are encouraged to seek further training on these topics with supplemental learning opportunities and continuing medical education. Finally, the clinical scenario described has been simplified to provide readers with an illustrative example for the general concepts introduced.

Searching the Evidence
As with other types of clinical questions, the most authoritative evidence for treatment decisions comes from systematic reviews.1 This standard is especially true for clinical questions that include harm because individual randomized controlled trials are seldom sufficiently powered to allow researchers to conduct thorough assessments for potential harm to patients. Unfortunately, well-designed systematic reviews are uncommon and individualized studies frequently must be used in EBM. For this reason, the present article will focus on how to use such studies in lieu of preferred systematic reviews.

Validity of Articles on Harm
A valid study aims to discover objective truth while also attempting to exclude competing explanations.1,2 Only a study that is deemed valid should be further reviewed. The following set of questions will help a clinician elucidate the validity of a study on harm:

- Are the defined groups of patients similar—other than through their exposure status to the treatment under study? When assessing the effectiveness or harm of an exposure or intervention, factors that may influence the outcome of interest must be balanced or accounted for in all study groups to avoid confounding results and conclusions. Treatment and comparison groups should be clearly defined, and crossover between allocated groups should be minimized.

Both groups should share similar baseline characteristics, especially characteristics that may confound the outcome of interest. In a study examining a potentially harmful treatment, healthier subjects should not be overrepresented in the treatment group because such a group allocation may confound the study’s outcome (ie, healthier people may be able to tolerate the treatment better). Information on baseline charac-
teristics can usually be ascertained quickly from a study population characteristics table. In most research articles, subject characteristics are placed in the first table.

In addition, patients assigned to receive active treatment should not also receive the control drug.

- Were the measures of interest ascertained in the same way for both groups?
  To properly compare study results among study groups, outcome assessments should be objective. Such clarity can be achieved only through the use of clearly stated case definitions.

  For example, in a study investigating cardiovascular outcome, the investigators decide that a myocardial infarction (MI) must meet specific criteria such as EKG changes or positive troponins. The definition of an MI cannot be left open for interpretation.

- Were participants and researchers blinded to the measures of interest?
  Assessment should also be blinded to prevent observer bias. Observer bias occurs when the researcher unconsciously (or consciously) looks harder for outcomes in the treatment group than in the control group. Blinding means that the researcher is unaware of the participants’ group assignment. Double-blinding means that study participants are also masked to group allocation.

- Was the follow-up period sufficiently long and complete?
  Follow-up in a valid study should be sufficiently long and complete. Short follow-up periods may allow too little time for the disease under investigation to manifest. The appropriate length of the study is dependent on the study question, the intervention used, the outcomes of interest, and special circumstances (eg, funding).

  A clinical investigation is considered complete when all study participants are accounted for at the study’s completion. Patients who dropped out of the study early because of death or adverse effects should be included and analyzed with their original group assignment. This is called an intention-to-treat analysis.

  A failure to perform an intention-to-treat analysis may direct researchers to providing readers with misleading results. The “5-and-20 rule” can be used by a critical reader to evaluate a study’s completeness. If less than 5% of the study population is lost to follow-up, one can be assured that the loss minimally impacted the results. If, however, more than 20% of the study population is lost to follow-up, caution is advised when making clinical decisions based on study findings. An attrition rate of 5% to 20%—and its impact on the researchers’ results—must be determined by the reader based on other specifics of the study.

- Do the results of the study satisfy some of the criteria for causation?
  While most studies provide some insight as to whether an exposure is associated with an outcome, that contact may not determine causality (ie, the exposure causes the outcome). The following five questions will help a cautious reviewer determine if the criteria are met to presume causality:

  - Does the exposure precede the outcome?
    To establish causation, the exposure must take place before the outcome.

  - Is a dose-response phenomenon present?
    The more one is exposed (eg, dosage), the greater the effect.

  - Is there evidence in the change of the outcome when the exposure is removed and reintroduced?
    Causation is suggested when the health status or clinical outcome improves (or deteriorates) after exposure is discontinued. In addition, the outcome recurs if the exposure is reintroduced.

  - Do other studies find consistent results?
    The assumption of causation is strengthened when more than one group of researchers reports similar findings.

  - Does the association make biological sense?
    It is important for the relationship between the exposure and the study outcome to be a logical one from the perspective of basic biology.

Study Results

Having ascertained a study’s validity, it is necessary to determine if the study’s findings are important. In the medical literature, importance refers to the study’s results and is measured in terms of the magnitude and precision of the association.

- What is the magnitude of the association between the exposure and the outcome?
  The magnitude of association quantifies the benefit or risk of an intervention when the results seen in the treatment group are compared with that of the control group. In retrospective studies, this data is reported as an odds ratio (hazards ratio). An odds ratio of less than 1 indicates that the intervention confers a protective effect. An odds ratio greater than 1 indicates that the intervention confers a detrimental effect. An odds ratio equal to 1 indicates that there is no difference in the outcome between the treatment and control groups. In the everyday practice of EBM, correctly interpreting the odds ratio is more important than knowing how to calculate it (Figure 1).

  Relative risk, which is used in prospective studies (ie, clinical trials and cohort studies), is another way to report the magnitude of an association. Relative risk is defined as “the ratio of the risk in the treated group (experimental event rate) to the risk in the control group (control event rate).” Interpretation of the relative risk is the same as the odds ratio (Figure 1).

  For patients and physicians alike, the odds ratio and relative risk are difficult to conceptualize. These measures are more digestible when reported as the number needed to harm...
SPECIAL COMMUNICATION

The NNH is defined as “the number of patients who need to be exposed to the causative agent to produce one additional harmful event.”1,2 The following methods can only be used for clinical trials and cohort studies. The NNH can also be calculated from case-control studies. Calculations are not shown because they are complicated and unrealistic to perform in real practice. However, free EBM calculators are widely available online (http://www.cebm.utoronto.ca/). Such calculators can easily convert an odds ratio into a NNH.

What is the precision of the relationship between the exposure and the outcome?

Measures of the magnitude of this association (eg, odds ratio, relative risk, and number needed to harm) are estimates of some unknown “true” value. If one were to repeat the experiment on different samples, they would yield similar, but not identical, results.1 The results of repeat experiments are dis-

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Adverse Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
</tr>
</tbody>
</table>

Formulas Used in Calculations of Patient Risk

- Odds Ratio: \( \frac{(a \times d)}{(b \times c)} \)
- Relative Risk: \( \frac{a/(a+b)}{c/(c+d)} \)
- Experimental Event Rate: \( a/(a+b) \)
- Control Event Rate: \( c/(c+d) \)
- Absolute Risk Increase*: \( \frac{a/(a+b)-c/(c+d)}{c/(c+d)} \)
- Number Needed to Harm†: \( \frac{1}{a/(a+b)-c/(c+d)} \)

Figure 1. Simple calculations used to analyze the results of clinical trials and cohort studies for risk to patients (ie, patient harm). *The absolute risk increase is the experimental event rate minus the control event rate. †The number needed to harm is 1 divided by the absolute risk increase.

Clinical Scenario

A 56-year-old retired engineer presents to your family medicine clinic after undergoing a colon polypectomy. He has no history of cardiovascular disease or diabetes and is not currently taking any medications. He neither smokes nor drinks. The patient jogs 3 miles every morning on a home treadmill. His blood pressure is 137 mm Hg systolic and 84 mm Hg diastolic, and his body weight is within the healthy target range for men of his height.

During his physical examination, the patient mentions that he has been following a story about the cyclooxygenase-2 (COX-2) isoenzyme inhibitor celecoxib on television news. Online, he has read that though COX-2 inhibitors may cause heart attacks and strokes, they might also prevent colon cancer.

The patient reports that his father had a colon resection—and that he is now worried about having his colon removed if it is determined in the future that he also has colon cancer.

The patient wants to learn more about COX-2 inhibitors and the potential for cardiovascular risk. Specifically, he wants to know if it would be safe for him to take a COX-2 inhibitor for cancer prevention.

Based on this information, and as a physician practicing evidence-based medicine, you might begin formulating a four-part question for later research:

What is the risk of death to a 56-year-old man with minimal cardiovascular risk who begins taking COX-2 inhibitors for cancer prevention?

In the meantime, you let the patient know that you would like to do some research before making a recommendation to him. You promise that you will have an answer for him within the next week and that you will e-mail him if he leaves his e-mail address with the receptionist on his way out.

When you return to your office, you type the following text into PubMed:

COX-2 inhibitors and cancer prevention and cardiovascular risk

You restrict the search to studies in English that are clinical trials. Your search yields multiple items. In the first group of 10, you locate one study that seems to be most appropriate to your clinical question. In this particular study, patients at risk for colon cancer were randomly assigned to one of three study groups: (1) 200 mg celecoxib twice daily, (2) 400 mg celecoxib twice daily, or (2) placebo twice daily. Subjects were monitored for multiple endpoints: myocardial infarction, stroke, death, and other cardiovascular outcomes.

(continued)
Clinical Scenario (continued)

In reviewing the article, you note that there were three clearly defined study groups to which participants were randomly assigned in a double-blind manner. Furthermore, as indicated in the article’s first table, subjects’ baseline characteristics (ie, demographic data, medical histories, comorbidities, concurrent medication use) were similar among the three study groups. Subject assessments were objective and blinded. A cardiovascular safety committee developed definitions and guidelines for evaluating cardiovascular endpoints, and clear definitions existed for all other outcomes assessed.

In addition, follow-up was sufficiently long and complete. With the exceptions of a common close-out date or subject death (both of which were documented), the authors reported that patients were observed for 2.8 years. Of 2035 study enrollees, 77% completed the study. Participants who dropped out of the study early or died were observed for a minimum of 2.8 years. Data was analyzed on an intention-to-treat basis.

The authors satisfied several criteria for causation. This was a prospective study, and therefore cause and effect were distinguished from effect-cause. By comparing two dosing levels of celecoxib (200 mg versus 400 mg) with a placebo control, a dose-response effect was also measured. The authors compared their findings with those of similar studies in which the cardiovascular risk of cyclooxygenase-2 (COX-2) isoenzyme inhibitors were assessed, thus demonstrating consistency of association.

In addition, the authors described their results in biologically plausible terms, proposing the following mechanism of action:

Celecoxib selectively inhibits vasodilatory prostacyclins but not platelet-aggregating thromboxane, leading to a prothrombotic state and predisposing patients to cardiovascular mortality.

These criteria help establish causality between celecoxib and cardiovascular events, eliminating the possibility that the association is a merely coincidental.

After this critical evaluation of the study design and methodology, you conclude that the study and the authors’ conclusions are valid.

(continued)

Figure 3. Clinical scenario (continued).

...voked by reviewing the article, you note that there were three clearly defined study groups to which participants were randomly assigned in a double-blind manner. Furthermore, as indicated in the article’s first table, subjects’ baseline characteristics (ie, demographic data, medical histories, comorbidities, concurrent medication use) were similar among the three study groups. Subject assessments were objective and blinded. A cardiovascular safety committee developed definitions and guidelines for evaluating cardiovascular endpoints, and clear definitions existed for all other outcomes assessed.

In addition, follow-up was sufficiently long and complete. With the exceptions of a common close-out date or subject death (both of which were documented), the authors reported that patients were observed for 2.8 years. Of 2035 study enrollees, 77% completed the study. Participants who dropped out of the study early or died were observed for a minimum of 2.8 years. Data was analyzed on an intention-to-treat basis.

The authors satisfied several criteria for causation. This was a prospective study, and therefore cause and effect were distinguished from effect-cause. By comparing two dosing levels of celecoxib (200 mg versus 400 mg) with a placebo control, a dose-response effect was also measured. The authors compared their findings with those of similar studies in which the cardiovascular risk of cyclooxygenase-2 (COX-2) isoenzyme inhibitors were assessed, thus demonstrating consistency of association.

In addition, the authors described their results in biologically plausible terms, proposing the following mechanism of action:

Celecoxib selectively inhibits vasodilatory prostacyclins but not platelet-aggregating thromboxane, leading to a prothrombotic state and predisposing patients to cardiovascular mortality.

These criteria help establish causality between celecoxib and cardiovascular events, eliminating the possibility that the association is a merely coincidental.

After this critical evaluation of the study design and methodology, you conclude that the study and the authors’ conclusions are valid.

Figure 3. Clinical scenario (continued).

...voked by reviewing the article, you note that there were three clearly defined study groups to which participants were randomly assigned in a double-blind manner. Furthermore, as indicated in the article’s first table, subjects’ baseline characteristics (ie, demographic data, medical histories, comorbidities, concurrent medication use) were similar among the three study groups. Subject assessments were objective and blinded. A cardiovascular safety committee developed definitions and guidelines for evaluating cardiovascular endpoints, and clear definitions existed for all other outcomes assessed.

In addition, follow-up was sufficiently long and complete. With the exceptions of a common close-out date or subject death (both of which were documented), the authors reported that patients were observed for 2.8 years. Of 2035 study enrollees, 77% completed the study. Participants who dropped out of the study early or died were observed for a minimum of 2.8 years. Data was analyzed on an intention-to-treat basis.

The authors satisfied several criteria for causation. This was a prospective study, and therefore cause and effect were distinguished from effect-cause. By comparing two dosing levels of celecoxib (200 mg versus 400 mg) with a placebo control, a dose-response effect was also measured. The authors compared their findings with those of similar studies in which the cardiovascular risk of cyclooxygenase-2 (COX-2) isoenzyme inhibitors were assessed, thus demonstrating consistency of association.

In addition, the authors described their results in biologically plausible terms, proposing the following mechanism of action:

Celecoxib selectively inhibits vasodilatory prostacyclins but not platelet-aggregating thromboxane, leading to a prothrombotic state and predisposing patients to cardiovascular mortality.

These criteria help establish causality between celecoxib and cardiovascular events, eliminating the possibility that the association is a merely coincidental.

After this critical evaluation of the study design and methodology, you conclude that the study and the authors’ conclusions are valid.

(continued)
Clinical Scenario (continued)

Now that you have determined that the study design, methodology, and the authors’ conclusions are valid and reasonable, you begin to determine the study’s applicability to your particular patient.

You find that the authors’ demonstrated a dose-related risk of mortality due myocardial infarction, stroke, heart failure, or other cardiovascular causes in patients taking two daily doses of celecoxib for prevention of colorectal adenoma. The relative risk in the group assigned to receive 200 mg of celecoxib twice daily was 2.3 (95% confidence interval [CI]=0.9-5.5). The risk for subjects in the higher dose group (ie, 400 mg twice daily) was 3.4 (95% CI=1.4-7.8). Thus, though low-dose exposure resulted in a trend toward cardiovascular harm, it is not considered significant because the CI includes 1.0.

At the higher dose, however, the increased risk should be considered significant because the CI does not include 1.0. This high-dose exposure translates to a number needed to harm of 42 people. In other words, when 42 individuals with the same baseline characteristics described by these researchers are given 400 mg of celecoxib twice daily for 2.8 years, at least one additional cardiovascular event should be expected.

Researchers found no independent association between celecoxib exposure and subjects’ demographic data, medical histories, comorbidities, or concurrent medication use (P>.05 for each comparison). You deem these results to be clinically important to your particular patient.

Rather than e-mail the patient, you decide to call him to discuss the findings of this clinical trial. Although the patient’s initial (fear-based) benefit-versus-risk calculation seemed acceptable to him, you now conclude together that larger studies designed to define a clear risk-benefit ratio are required before you can make a confident decision in favor of treatment.

Figure 4. Clinical scenario (continued).

participants (eg, comorbidities), reducing his or her overall risk of cardiovascular events by half (Figure 3). This personalized rate of risk is called the \( F \) statistic, and it allows clinicians to personalize a study’s NNH for individual patients easily, by dividing the NNH by the \( F \) statistic: \( NNH/F \).

If a study’s NNH is 42 (Figure 4), for example, a personalized NNH for the patient could be determined by dividing 42 by 5 for an answer of 8.4. Therefore, according to this theoretical study (\( N=42 \)), 8 people would need to take celecoxib for 2.8 years to cause one adverse event or complication.

What are the patient’s concerns and expectations from the treatment?

When a study is declared a landmark by journalists, its results are irrelevant if incorporating them in practice would violate a patient’s preferences, concerns, or expectations. For example, any “wonder drug” would probably be unacceptable to a patient who is completely opposed to the idea of taking medication.

Are other alternatives available?

The physician and the patient, together, should explore the relative safety of alternative treatment options (Figure 4). Alternative choices may include lifestyle modifications and other nonpharmacologic treatment modalities.

Conclusion

Although most clinicians are already incorporating EBM principles in their practices, often instinctively, some physicians may require a more organized approach to integrating this relatively new model of self-education. Improved comfort levels and true expertise in the practice of EBM are the result of additional education, repetition, and self-assessment. The principles of EBM allow physicians to stay informed while also improving the quality of the information communicated to patients during patient encounters. The systematic approach that is used to appraise an article on harm is but one step in practicing EBM. Remember, the goal is always to provide the best care possible to patients—using one’s clinical expertise to address patient values and expectations for treatment.

References