APPLYING EVIDENCE TO THE USE OF CELECOXIB IN A PATIENT WITH OSTEOARTHRITIS: HOW TO ADDRESS CLINICAL QUESTIONS OF HARM

Case and Discussion: *David A. Feldstein, MD, and Shobhina G. Chheda, MD, MPH*

Quality medical practice requires that we strive to provide care that benefits our patients and avoid doing harm. It is important, therefore, to understand and explain to our patients the potential harms that therapeutic and other exposures may pose to them. For example, we know that aspirin may prevent a heart attack, but can it increase the risk of hemorrhagic stroke? The media may highlight environmental harms such as breast cancer from living near high-voltage power lines or lung cancer from secondhand smoke, but when our patients turn to us for advice about these purported health risks, how should we respond? Fortunately, evidence-based medicine (EBM) provides the tools to help our patients understand what are true risks to their health, so that they can make appropriate decisions about taking medications or modifying exposures.

This article uses a case of a patient with osteoarthritis who is currently taking celecoxib to illustrate the EBM process for addressing questions of harm. The aim of harm questions is to determine the effects of potentially harmful exposures (including therapies) on patient function, morbidity, and mortality. As the case unfolds, we examine search strategies for questions of harm, review the different study types that provide evidence for questions of harm as well as the critical appraisal criteria for harm studies, and highlight types of results often presented in these studies. We also examine the importance of understanding baseline risk in order to provide a patient with an estimate of the overall risk of an exposure. Finally, we demonstrate how patient preferences regarding trade-offs between the potential benefits and risks of specific therapies affect treatment decisions.

David A. Feldstein, MD, and Shobhina G. Chheda, MD, MPH; both from the Division of General Internal Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI.

Assessing the Clinical Problem

You are a third-year medical resident on a primary care block rotation seeing a new patient. Ms. Johnson recently changed insurance plans and is here to establish care with you as her new primary care provider.

Ms. Johnson is a 63-year-old woman with a history of osteoarthritis of her knees. She has been taking celecoxib for several months and is concerned that she only has enough pills for 3 more weeks. She was previously taking rofecoxib until it was removed from the market. She had problems with ulcers and gastrointestinal (GI) bleeding when she was in her early 50s and was instructed to avoid traditional nonsteroidal anti-inflammatory drugs (NSAIDs). She tried acetaminophen, but it did not control her pain.

Ms. Johnson is healthy except for her arthritis. She does not use any other medications and reports that her cholesterol and blood sugar were tested a year ago and were "excellent." She exercises regularly and mentions that good control of her knee pain is important so that she can continue golfing, swimming, and playing with her grandchildren.

You complete your physical examination, which reveals a blood pressure of 120/70 mm Hg and no abnormalities. Records from Ms. Johnson's previous physician confirm that fasting blood tests performed a year prior revealed lipid and blood glucose levels all within normal range. When asked whether she has any other concerns besides keeping her arthritis under control, Ms. Johnson cheerfully quips, "No, should I?"

You smile and tell Ms. Johnson that you will be back shortly with your preceptor, Dr. Finney, to discuss her arthritis treatment.

While waiting for your preceptor, you encounter Sara, an intern who is also waiting to speak with Dr. Finney. The 2 of you discuss the recent removal of rofecoxib for causing myocardial infarction (MI). Sara wonders if this risk would apply to celecoxib, since it is a cyclooxygenase-2 (COX-2) inhibitor as well. You agree that you would not want to prescribe a medication that could harm your patient.

You and Sara recall that the first step of the EBM process is assessing the patient and the problem. Ms. Johnson is in good health and does not seem to have any risk factors for heart disease, but you do not know her actual risk. Sara suggests using the National Cholesterol Education Program (NCEP) risk calculator, which uses information from the Framingham Heart Study to predict a person's risk of having an MI or coronary death within the next 10 years. You log on to the site (http://hp2010.nhlbihin.net/atpiii/calculator.asp?usertype=prof) and enter Ms. Johnson's information. The calculator determines that your patient's 10-year risk of an MI or coronary death is 2%.

Recognizing Issues of Potential Harm

It is important to recognize clinical questions that arise during patient encounters. These questions may be based on our own curiosity, or patients may specifically ask them. In this case, Ms. Johnson has not voiced any urgent concern during her initial visit and appears to have come in solely to request a refill of an arthritis medication prescribed by her previous doctor. She clearly believes that the medication is relieving her osteoarthritis pain, and she has given no indication that she is aware of any potential adverse effects of the drug.

When assuming the care of a new patient, it is important to review the appropriateness of any current medications and to make sure that the patient understands the benefits as well as potential harms of the drugs. If we are unsure about potential adverse effects of the drugs a patient is taking, we need to explore these areas before we can proceed with confidence that a treatment plan is not only appropriate but also safe. Questions about potential adverse effects of medications are only 1 type of question about harm. Anytime we ask whether an exposure increases a patient's risk of an adverse outcome we are asking a question about harm. This includes naturally occurring exposures such as radon gas or high-voltage power lines as well as physician-initiated exposures such as vaccinations or medications.

Assessing Baseline Risk

Before we can answer a question about harm, we need to determine a patient's baseline risk, which is the risk of developing the target outcome if the patient were not exposed to the risk factor. There are several ways to assess baseline risk. For example, we could estimate our patient's risk using our own clinical expertise and knowledge of the adverse target outcome. Unfortunately, this estimate may not be very accurate, especially if we have limited clinical experience with that adverse outcome. We can also use data from clinical studies that enrolled patients similar to ours and followed them over time to determine their rate of adverse events (prognosis studies). We can adjust the rate if we think that our patient has a different risk than the average patient in the study. For instance, if we think our patient is 50% less likely to have the adverse event than the average patient in the study, we can divide the rate by 2 to approximate our patient's baseline risk. Finally, the most accurate way to determine baseline risk is by using risk calculators or clinical prediction rules. These calculators are developed from clinical studies that assess the impact of multiple factors on a patient's risk for an adverse event. We must be cautious when using these calculators, because they are only as good as the studies on which they were based and they need to be validated before they can be broadly used. The NCEP's risk calculator is one example, but others include a breast cancer risk calculator from the National Cancer Institute (http://bcra.nci.nih.gov/brc/q1.htm) and a lung cancer risk calculator for smokers from the Memorial Sloan-Kettering Cancer Center (www.mskcc.org/mskcc/html/12463.cfm).

Once we have determined our patient's baseline risk, we need to find evidence to determine how much this risk will increase if the patient has the exposure. If a patient's baseline risk is very low, even a large relative increase could translate into a low overall risk. Thus, in the case of Ms. Johnson, we want to know how much, if at all, celecoxib increases her risk of a heart attack or coronary death so that we can help her make the best decision about remaining on the medication.

Asking a Focused Clinical Question

Dr. Finney is still busy with another patient, so you and Sara take a few minutes to build a focused question about Ms. Johnson using the PICO format. Soon after, Dr. Finney emerges from the exam room. You present Ms. Johnson's case to her and relate your question about whether to prescribe celecoxib given its potential risks. You show Dr. Finney the harm question you've developed, and she suggests that you search for evidence to answer the question and bring it tomorrow to discuss with her and Sara.

You and Dr. Finney go back in to see Ms. Johnson. You discuss your concerns about celecoxib and let her know that you want to evaluate the available evidence so you can help her make an informed decision regarding continuing with this medication.

Ms. Johnson appreciates your concern and agrees to return next week to discuss her arthritis treatment.

Using the PICO format (patient/population, intervention, comparison, outcome) allows us to develop focused questions to answer our specific need. In developing a PICO question related to harm, it is important to identify the appropriate *patient population*, so that we may increase the likelihood that the evidence found in our search will apply to our patient. Also, it is essential that we are specific about our *intervention* (ie, exposure). The *comparison* will often be lack of exposure in harm questions, but occasionally we will want to know if one exposure is more harmful than another exposure. For example, we might use a comparison if we are evaluating the risk of 2 different medications. Lastly, we need to identify an *outcome* that is clinically important to our patient. We should choose the outcome that we are most interested in, but when we find evidence to answer our question we may need to extrapolate from the outcomes that the study measured.

The PICO components for the case of Ms. Johnson may be defined as:

P = women with osteoarthritis, otherwise healthy
I = use of celecoxib
C = use of no medication
O = increase in MI or coronary death

Using this framework, our harm question regarding Ms. Johnson would be:

In an otherwise **healthy woman with osteoarthritis**, does the **use of celecoxib** compared with **use of no medication** lead to an **increase in MI or coronary death?**

Acquiring the Current Best Evidence

You remember seeing an article not long ago that reported that celecoxib did not increase the risk of nonfatal MI and was safer than rofecoxib. When you go home that night, you search through your piles of medical journals and find the article in a February 2005 issue of the *Annals of Internal Medicine,* by Kimmel et al [1].

The next day, when you show Dr. Finney the article, she tells you there are 3 types of studies available for evaluating harm. The study you found was a case-control study. She suggests you search for a randomized controlled trial (RCT) or cohort study, since these study types provide stronger evidence. She reminds you that it is often worthwhile to search databases of prefiltered evidence as a first step in seeking answers to clinical questions.

You start with the Cochrane Database of Systematic Reviews, but no systematic reviews are available. You move on to *ACP Journal Club,* enter a search for "celecoxib," and find a summary [2] of an RCT that was published in 2005 [3]. Although the title of the summary ("Celecoxib increased cardiovascular events in patients with colorectal adenomas") makes you question whether the study directly applies to your patient, you decide to print and read the summary and commentary. A statement in the commentary makes you believe this RCT may in fact be relevant to your patient ("it seems sensible to restrict the use of COX-2 inhibitors to a small minority of patients without overt vascular disease who require an anti-inflammatory but are at high risk for GI hemorrhage or intolerant of other NSAIDs" [4]). You access the full text of the RCT online and print it.

You decide to check for other RCTs, so you turn to PubMed to search using the Clinical Queries option. Clicking the "etiology" category and "narrow, specific search" emphasis, you enter the terms "celecoxib AND (cardiovascular OR myocardial infarction)." This search retrieves 45 articles. There are no RCTs, but you do find a cohort study looking at the risk of MI in patients taking COX-2 inhibitors [5]. You read the abstract for this study and decide it may also be suitable to your patient, so you access the full text of the article and print it.

You show Dr. Finney the 2 new articles you found, and she recommends appraising all 3 studies to show the differences between them. She suggests dividing the work among the clinic group and sharing the findings at the next preclinic conference. She offers to look at the RCT and you choose the cohort study, leaving Sara with the case-control trial.

Study Types for Answering Harm Questions

To effectively search for evidence to answer clinical questions about harm, it is important to consider the different types of studies available. Three main study types can help us answer harm questions (**Figure 1**). Each study type has its strengths and weaknesses (**Table 1**), an understanding of which helps in choosing the appropriate study type to address the specific question at hand.

RCTs are experimental studies in which investigators randomly assign participants to be exposed to either an experimental or a control intervention. The investigators predetermine the outcomes of interest, and then the 2 groups are followed over time and the

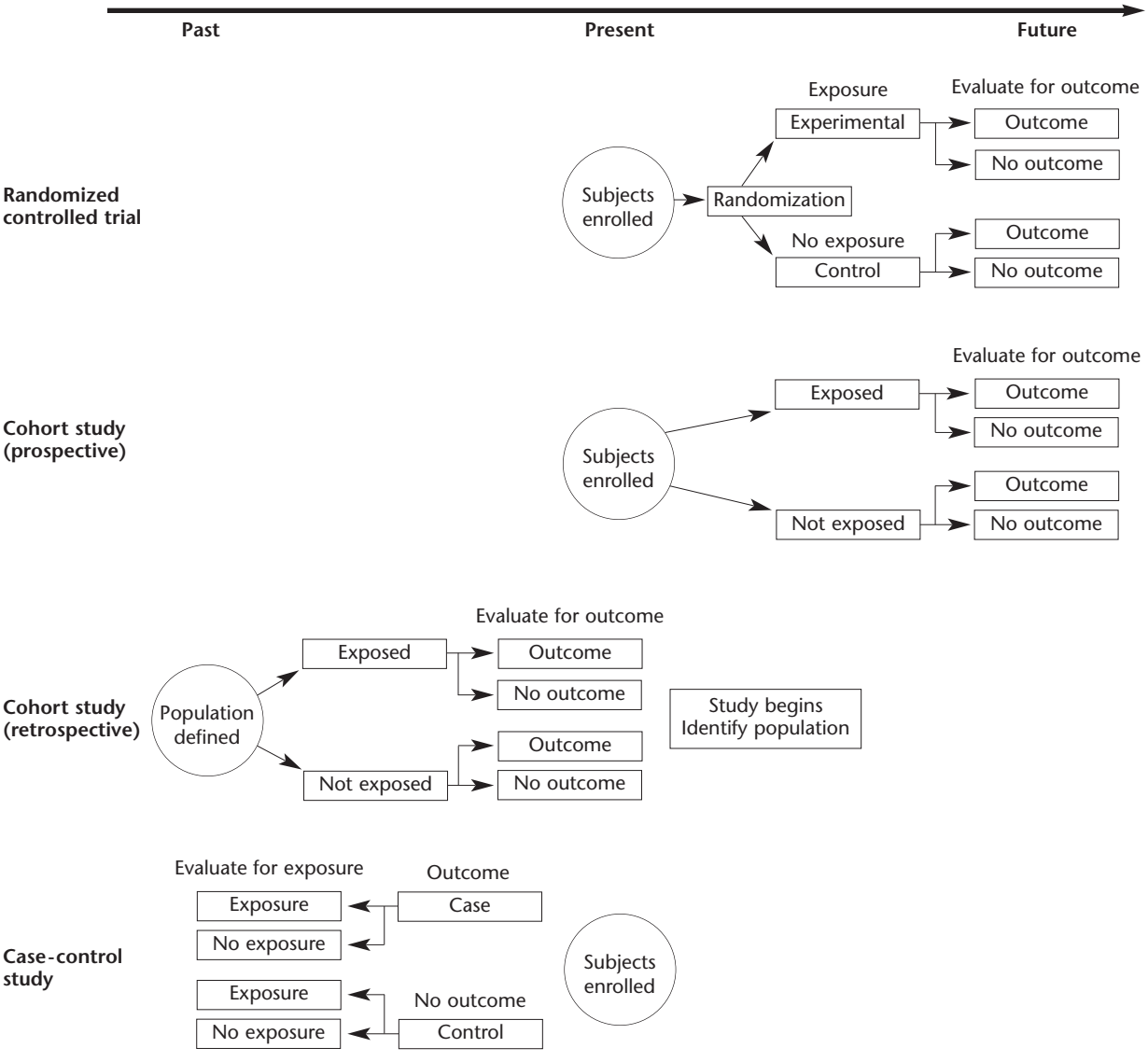


Figure 1. Design and timeline for randomized controlled trials, cohort studies, and case-control studies.

number of outcomes in each group is compared. RCTs are the least biased studies to evaluate harm, since randomization minimizes differences between study groups. However, RCTs are expensive and take a long time to perform and would require prohibitively large numbers of patients to evaluate rare outcomes. Ethical concerns can also limit the use of RCTs, since it would be unethical to randomize patients to an exposure that could be harmful, with no possible benefit. For these reasons, RCTs are rarely performed to investigate questions of harm. However, occasionally an RCT that is undertaken to evaluate a therapeutic intervention yields a finding of unexpected adverse outcomes in the study participants. In these cases, the safety committee for the therapy trial will evaluate the adverse outcomes and can stop the trial early if the intervention is causing harm.

Cohort studies. Prospective cohort studies enroll participants who have not yet had the outcome being studied. Study participants are followed over time; some are exposed to the risk factor of interest and others are not. Investigators do not control whether or not participants are exposed. The number of outcomes in exposed study participants is then compared with the number of outcomes in unexposed participants. Cohort studies are more open to bias, because participants

Table 1. Strengths and Weaknesses of Different Study Types for Evaluating Harm

Type of Study	Strengths	Weaknesses
RCT	Randomization decreases bias due to confounding variables Exposure clearly precedes outcome Able to generate incidence rates Allows accurate measurement of risk factors	Expensive to perform Volunteer bias Ethical considerations Not powered to evaluate rare outcomes Long time to perform
Cohort study (prospective)	Less expensive than an RCT Able to evaluate less common outcomes Relieves ethical concerns Exposure clearly precedes outcome Able to generate incidence rates Allows accurate measurement of risk factors	Increased bias due to unknown confounders Not powered to evaluate very rare outcomes Long time to perform
Case-control study	Relatively inexpensive Able to evaluate outcomes that take a long time to occur Quicker results—do not need to wait for outcomes to occur Able to evaluate very rare outcomes Relieves ethical concerns	Relies on participant recall of exposure (recall bias) Increased bias due to unknown confounders Unclear temporal relationship between exposure and outcome Choice of control group is difficult Cannot generate incidence rates

RCT = randomized controlled trial.

who are exposed to the risk factor may differ in their baseline risk for the target outcome from participants who are not exposed. These differences may affect the outcome and are called *confounders* (also, *confounding factors* or *variables)*. Investigators can try to adjust for the known confounders (eg, tobacco use contributing to heart disease), but unknown confounders can still cause bias. Cohort studies can more efficiently follow large numbers of patients for many years, so they are able to evaluate less common outcomes that may have a long lag time from exposure to occurrence. However, if an outcome is extremely rare, it may not be practical to enroll enough participants in a cohort study to demonstrate it. Cohort studies also avoid ethical concerns, as patients are not chosen by investigators to receive the exposure.

Retrospective cohort studies use previously collected data to define study groups and then determine exposures and outcomes in each group. The investigators compare the outcomes in the exposed versus unexposed groups. The data were not collected with the study in mind, so the databases may be inaccurate or incomplete.

Case-control studies enroll participants who have already had the outcome of interest (cases). The investigators then assemble another group of participants who are as similar as possible to the case patients, except that they have not yet developed the outcome of interest (controls). Investigators usually use interviews to determine retrospectively which participants in each group had the exposure of interest. They compare the rates of exposure between the 2 groups. Case-control studies are very susceptible to bias. They have the same risk of unknown confounders as cohort studies, but the retrospective evaluation of exposure causes other possible biases. Case patients who developed the outcome are more likely to remember even the smallest exposure, which can cause a recall bias. Case-control studies, however, do have benefits. They are relatively inexpensive to perform, and because they are retrospective, investigators do not have to wait a long time for outcomes to occur. Most importantly, as case patients already have the outcome of interest, even extremely rare outcomes can be studied. A case-control study was used to show that phenylpropanolamine, a common component in appetite suppressants and cold remedies, increased the risk of hemorrhagic stroke in young women aged 18 to 49 years [6]. The risk of hemorrhagic stroke is so low in this patient population that even an extremely large cohort study may have missed the association. This study led to the U.S. Food and Drug Administration's decision to remove phenylpropanolamine from the U.S. market.

In summary, an RCT is best if it is ethical to randomize study participants to the exposure and the

target outcome is relatively common. If the exposure causes a rarer outcome or carries ethical concerns, a cohort study would be best because it is less biased than a case-control study. If there is a very long lag time from exposure to outcome or if the outcome is very rare, a case-control study is best, since its retrospective nature prevents long delays in determining harm.

Search Strategies

Prefiltered evidence. Systematic reviews of RCTs would be an ideal source of evidence to answer a harm question, but it is rare to find one. Nevertheless, it is worthwhile to check the Cochrane Database of Systematic Reviews (www.cochrane.org), since this resource will provide the highest quality evidence, if available. The *ACP Journal Club* (www.acpjc.org) reviews articles about harm. It is an excellent evidence resource that provides a structured abstract and a brief commentary that can provide an answer quickly. Both of these resources require a subscription.

MEDLINE/PubMed. If no evidence is found by searching prefiltered evidence resources, we must search unfiltered databases of medical information, such as MEDLINE. PubMed (www.ncbi.nlm.nih.gov/entrez/query.fcgi) is a free search engine for MEDLINE. Literature searches on PubMed using the Clinical Queries service (on left side menu) applies filters to limit the search to articles relevant to the type of question of interest (ie, etiology, diagnosis, therapy, prognosis, or clinical prediction rules). Selecting the "etiology" limit is appropriate when searching for articles about harm.

Appraising Evidence for Validity

At the preclinic conference the following week, Dr. Finney begins by summarizing the RCT by Solomon et al [3]. The study investigated the occurrence of all potentially serious cardiovascular events in a randomized, double-blind, controlled trial performed initially to assess the impact of celecoxib on prevention of colorectal adenomas. In the initial study, 2035 patients (aged 32–88 years) with a history of colorectal neoplasia were randomized to 200 mg celecoxib twice daily, 400 mg celecoxib twice daily, or placebo; study participants did not necessarily have a clinical reason for celecoxib use. The data and safety monitoring board recommended early discontinuation of the study, resulting in a 77% completion rate; remaining study patients were followed for a minimum of 2.8 years. Two independent assessors blinded to the patients' initial treatment assignments evaluated cardiovascular end points based on predefined criteria. The study showed that celecoxib use was associated with an increase in the composite cardiovascular end point of death from cardiovascular causes, MI, stroke, or heart failure.

You are excited about presenting the cohort study by Mamdani et al [5] because the results were very different! This study looked at the risk for acute MI in elderly patients (aged > 66 years) who were new users of celecoxib. The study was very large, with 15,271 participants in the celecoxib arm and 100,000 in the control arm. It was a retrospective cohort study done in Ontario, Canada, where computerized pharmacy and hospitalization records and physician billing information are readily available. Patients in the celecoxib arm were assumed to have had osteoarthritis or rheumatoid arthritis and to have failed other NSAIDs or to have had a history of GI bleeding or an ulcer, since these are the criteria that would allow prescription of celecoxib to be paid for in Ontario. The control group had not been prescribed any NSAIDs for at least 1 year prior to inclusion. The maximum follow-up period was 1 year. The study used statistical analysis to account for differences in the groups related to hospitalizations, procedures, other medication use, and demographics (eg, age, sex, long-term care, and low-income status). The study results indicated no significant differences in short-term risk of acute MI among those who used celecoxib compared with those who had not used NSAIDs.

Sara then presents a summary of the case-control trial by Kimmel et al [1]. This study was performed at 36 hospitals in a 5-county area and included 1718 case patients (aged 40–75 years) with a first nonfatal MI and 6800 control patients with no previous history of MI; controls were selected by random-digit calling from a similar geographic region as case patients. The participation rate among eligible case patients and controls was 55% and 50%, respectively. Data regarding medication use, clinical history, and demographics were collected from both groups through structured telephone interviews. Multiple clinical and demographic factors were assessed and adjusted for statistically. The study results indicated that celecoxib users had a lower risk of MI than patients who did not use NSAIDs.

You all agree that the next step should be to assess the validity of the 3 studies using the *Users' Guides* criteria for evaluating harm studies.

Criteria for Appraising Validity

It is important to have a structured method for evaluating a study's validity. Fortunately, the Evidence-Based Medicine Working Group—an international

group of EBM scholars—has developed appraisal criteria to support this essential step in the EBM process. Published first as 25-part series of articles in *Journal of the American Medical Association*, the *Users' Guides to the Medical Literature* were compiled and expanded into a book form [7]. The specific questions outlined in the *Users' Guides* for evaluating studies of harm are shown in **Table 2**. As we review these questions, we will consider how each of the 3 studies found in Ms. Johnson's case measures up.

- **Did investigators demonstrate similarity in all known determinants of outcome? Did they adjust for differences in the analysis?**

In assessing the validity of a study of harm, it is important to first determine if the experimental and control groups began the study with a similar prognosis. If one of the groups is at higher risk of having the target outcome from the outset (eg, one group is older, overweight, or more sedentary and the outcome of interest is a cardiovascular event), the study may yield biased results that show an increased risk of the target outcome that is not attributable to the exposure. Randomization is the best way to ensure that both groups are similar at the beginning of a study. The study by Solomon et al [3] was an RCT. Patients were randomly assigned to intervention groups by a computer-generated randomization schedule, and a detailed baseline assessment revealed similar cardiovascular disease status and risk factors in both groups. Therefore, it is likely that both groups had a similar prognosis at the beginning of the study. Because RCTs are rarely available to answer clinical questions about harm, we need to be familiar with how other study types try to ensure that experimental and control groups begin the study with a similar prognosis.

In a cohort study, there may be inherent differences between participants who tend to be exposed and participants who tend not be exposed to a potentially harmful agent. In fact, it is more likely than not that there are differences in those who are prescribed versus those who are not prescribed a medication. For example, patients who are prescribed celecoxib are likely to be older, have arthritis, be more sedentary, have known coronary artery disease, and have a prior history of GI bleeding or ulcers than patients not taking celecoxib [1]. It is possible that the increased cardiovascular risk seen could be attributable to some of these confounding variables instead of actual exposure to celecoxib. Therefore, it is important for investigators to measure these other variables so that they can determine whether they are similar in both groups; if they are not, the results can be statistically adjusted for the differences. The large cohort study by Mamdani et al [5] assessed multiple important variables and appropriately compared them between groups using statistical analysis to adjust for them when necessary. The investigators found that more celecoxib users had cardiovascular hospitalizations before cohort entry and were prescribed more cardiovascular medications than the control group. By combining multiple administrative databases, the investigators were able to extract many of the potential confounding variables. However, some important information, such as smoking status and obesity, was not available. Even after adjusting for all known confounders, unknown confounders can still bias the results.

In case-control studies, the investigators try to minimize the differences between the case patients and controls by choosing controls who are reasonably similar to the case patients except that they have not had the outcome of interest, in this case MI or coronary

Table 2. *Users' Guides* Criteria for Appraising Articles About Harm

Are the results valid?
Did experimental and control groups begin the study with similar a prognosis?
Did the investigators demonstrate similarity in all known determinants of outcome; did they adjust for differences in the analysis?
Were exposed patients equally likely to be identified in the 2 groups?
Did experimental and control groups retain a similar prognosis after the study started?
Were the outcomes measured in the same way in the groups being compared?
Was follow-up sufficiently complete?
What are the results?
How strong is the association between exposure and outcome?
How precise is the estimate of the risk?
How can I apply the results to patient care?
Were the study patients similar to the patients under consideration in my practice?
Was the duration of follow-up adequate?
What was the magnitude of the risk?
Should I attempt to stop the exposure?

death. Then, they would explore for exposure to celecoxib as well as other potential prognostic variables and statistically adjust for the known confounders. In the study by Kimmel et al [1], most important variables were accounted for. However, as with cohort studies, it is still not certain that both experimental and control groups were similar with respect to determinants of outcome.

- **Were exposed patients equally likely to be identified in the 2 groups?**

A major concern in case-control studies that does not apply to the other study types is that exposures may not be equally identified in both groups. Case-control studies are open to both *recall bias* and *interviewer bias*. In the study by Kimmel et al [1], case patients and controls were asked about drug exposure in a structured telephone interview. To attempt to minimize recall bias, the investigators used techniques such as prompting participants with indication-specific questions and having participants examine photographs of medications to help them accurately recall their medication usage. They also asked participants in both groups to have all medication bottles available at the time of the interview. In order to negate the potential for interviewer bias, the trained telephone interviewers were not informed of the study hypothesis. Had interviewers been so informed, they may have pursued questioning differently in case patients and controls.

- **Were the outcomes measured in the same way in the groups being compared?**

In RCTs and cohort studies, it is essential that outcomes are measured in a similar fashion in the experimental and control groups. If investigators measuring the outcomes are aware of the patients' exposure, they may look more diligently for the outcome in the exposed group. This would lead to *surveillance bias*, as the outcome will be diagnosed more often if it is measured more frequently. In the RCT by Solomon et al [3], a cardiovascular safety committee developed end point definitions for adjudicating cardiovascular end points. Throughout the adjudication and safety evaluation process, the entire cardiovascular safety committee was unaware to which intervention group the patients had been assigned. In the retrospective cohort study by Mamdani et al [3], acute MI was identified through a database for hospital admissions that included diagnostic and procedural information. This same database was used for both treatment and control cohorts. It is unclear in the study who exactly extracted the data and whether or not they were familiar with exposure data while they were extracting outcome data.

- **Was follow-up sufficiently complete?**

Finally, we must consider whether follow-up was sufficiently complete. Patients lost to follow-up tend to have a different prognosis than patients who remain in the study. If a large number of patients are lost, the results may be biased. In the study by Solomon et al [3], 77% of the study population had completed the study at the time of its early termination and all remaining surviving participants completed at least 2.8 years of follow-up. Since significant differences in harm were seen in this timeframe, it seems that this follow-up period was long enough. If no differences were seen, one may be concerned that the 2.8-year follow-up for 23% of the study population was not sufficient. Kimmel et al [1] had complete data on only 55% of eligible case participants and 50% of eligible control patients; there was no information on nonparticipants presented, so it is unclear whether they differed systematically from participants. In the study by Mamdani et al [5], complete follow-up was reported.

After you, Sara, and Dr. Finney review the *Users' Guides* criteria for validity of a harm study, you all agree that the RCT (Solomon) represents the strongest design to answer the question relevant to Ms. Johnson. The cohort study (Mamdani) and the case-control study (Kimmel) are limited, as groups in observational studies may have differences other than the variable of interest that could have accounted for the differences in outcomes. Additionally, the case-control study (Kimmel) is limited by potential recall bias that could lead to failure in identifying exposure to celecoxib. You decide as a group to review the study results in more depth to understand what Ms. Johnson's risk would be if she took celecoxib. You wonder if design and validity factors themselves account for what initially seems like differing results among the 3 studies.

The case-control study (Kimmel) found that celecoxib use was actually associated with a lower risk of MI when compared with no NSAID use, with an adjusted odds ratio of 0.43 (95% confidence interval [CI], 0.23–0.79). The cohort study (Mamdani) found an MI rate of 8.2/1000 person-years in the no NSAID group versus an MI rate of 10.7/1000 person-years in the celecoxib group; this gave a rate ratio (dividing one rate by the other rate) of 1.3. When this was adjusted for possible confounding factors,

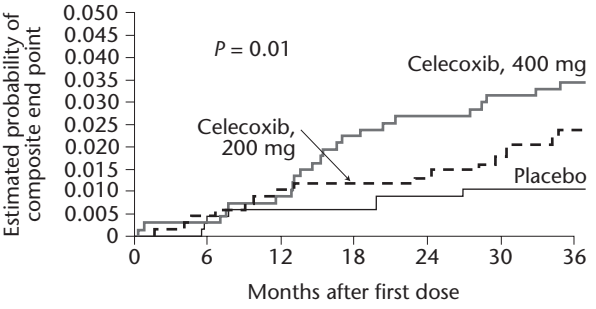
the adjusted rate ratio was 0.9 (95% CI, 0.6–1.7), which was not significant for a difference between celecoxib users and no NSAID users. The findings from the RCT (Solomon) show that for the combined end point of death from cardiovascular causes or nonfatal MI, the combined celecoxib groups (200 mg twice daily and 400 mg twice daily) had a higher risk than the placebo group, with a hazard ratio of 3.4 (95% CI, 1.2–9.7).

Criteria for Appraising Results

- **How strong is the association between exposure and outcome? How precise is the estimate of risk?**

The results from studies of harm differ from those usually seen in therapy studies. Trials of harm seek to find a relationship between the exposure and the outcome of interest. There are several possible methods to determine whether such a relationship exists. Which methods are used depend in part on the type of study performed.

Case-control studies have a predetermined outcome rate because the investigators determine the number of cases and the number of controls. Therefore, the results of these studies are typically expressed as an *odds ratio* (OR), since they compare the odds of patients being exposed in the case group to the odds of patients being exposed in the control group. The OR is often misinterpreted as the increased risk of developing the outcome in patients who are exposed, which is actually the *relative risk* (RR). When the incidence of disease is low (< 10%), the OR is a reasonable approximation of the RR. However, as the incidence increases, the OR overestimates the RR. For example, the Women's Health Initiative trial compared the effects of hormone replacement with estrogen plus progestin versus placebo in healthy postmenopausal women and found an increase in deep venous thrombosis (DVT) in the treatment group, although the overall rates of DVT were very low (1.3% in the estrogen/progestin group; 0.6% in the placebo group) [8]. Because the outcome was so rare, the OR and RR were very similar (OR = 2.12; RR = 2.10). Using OR and RR interchangeably in this case would not cause a problem. However, if the incidence of DVT had been 20 times greater (26% in the estrogen/progestin group; 12% in the placebo group), the OR for DVT would increase to 2.52, whereas the RR would remain at 2.10. In that case, it would be misleading to use the OR like an RR, since it overestimates the risk. Cohort studies and RCTs can generate incidence rates for the targeted outcome in each group or, in the case of the 2 studies found in this example, the number of outcomes per person-year.



No. at risk							
Celecoxib, 400 mg	671	669	665	655	651	648	576
Celecoxib, 200 mg	685	681	676	675	673	670	595
Placebo	679	677	675	672	668	667	585

Figure 2. Kaplan-Meier curve showing estimates of the risk of the composite end point of death from cardiovascular causes, myocardial infarction, stroke, or heart failure among patients who received celecoxib (200 mg twice daily or 400 mg twice daily) or placebo. (Adapted with permission from Solomon SD, McMurray JJ, Pfeffer MA, et al. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. Adenoma Prevention with Celecoxib [APC] Study Investigators. N Engl J Med 2005;352:1077. Copyright © 2005 Massachusetts Medical Society. All rights reserved.)

When evaluating harm, it is important to know not only whether the exposed group is more likely to develop the outcome but also whether the events occur sooner in one group versus the other. In a study evaluating mortality as the target outcome, if the follow-up was long enough, 100% of the patients would experience the outcome. Therefore, it is important to know whether the exposed group died sooner than the control group. A graphic way to represent the timing of events in each group is with a Kaplan-Meier curve (**Figure 2**), which shows time on the x-axis and the probability of developing the outcome on the y-axis. A Kaplan-Meier curve allows one to see whether events occurred early or late and when the 2 groups separate. The extent of follow-up is shown by listing the number of patients available for each time period. The statistical certainty of the plotted lines decreases as time progresses, since less patients are available for analysis. The graph should have error bars for each group at fixed time points so one can evaluate the level of uncertainty.

The *hazard ratio* is the statistical method for comparing the 2 groups at multiple points in time over

the course of the study. The hazard ratio is weighted for the number of patients available for analysis at any time, since patients who developed the outcome already and those who are lost before study completion are not available to have the outcome. The hazard ratio tells us how much more likely the exposed group is to develop the outcome at any point in time.

How to Interpret Lower Quality Evidence

The results of the 3 studies evaluated in the case of Ms. Johnson vary significantly. The differences are likely due to the different study designs and the biases inherent in them. Because they are the least biased of the 3 study designs, RCTs should be used whenever they are available. However, when we cannot find an RCT we may have no choice but to use lesser quality evidence. It is crucial to incorporate the quality of the evidence into our overall clinical decision making. The greater the potential bias in the study, the less certain we can be that the results of the study are true and that we will see similar effects in our patient. As cohort and case-control studies are more susceptible to bias, we have to look at the results with greater scrutiny. If the risk of developing the outcome after the exposure is very high, with an OR greater than 5, even though the study may have some bias it probably does not account for this large increased risk in the adverse outcome. On the other hand, if the OR is relatively small (< 1.2), the bias in the study may be causing the increased risk and we should be cautious interpreting the risk of exposure. Ultimately, we can only use the evidence that is available to us, but if the quality is very poor, we may decide that it does not help us in treating our patient.

Applying Evidence to Patient Care

After further discussing the results with Dr. Finney, you decide that the quality of the RCT is superior to that of the other 2 studies. You would like to use these results to make your decision, but first you must decide if the results of the study can be extrapolated to your patient.

The patients in the study were slightly younger than Ms. Johnson and at higher baseline risk, with almost 50% having a history of previous cardiovascular events. Even with these differences, there is no physiologic reason why Ms. Johnson would have a different response to celecoxib, so you feel comfortable using the results of the study to calculate her risk. You already calculated her risk of MI or coronary death over the next 10 years to be 2%. As the rates of cardiovascular death or MI were low in the study (0.6%–2%), you assume the hazard ratio will closely approximate the RR. Therefore, by multiplying Ms. Johnson's baseline risk (2%) by the hazard ratio (3.4), you estimate that her risk of MI or coronary death over 10 years is approximately 7%. This risk seems a little bit high but still uncommon.

You realize that to make an informed decision about whether or not she should take celecoxib, Ms. Johnson needs to understand the information you have learned, and you need to determine how she views the benefits and risks of staying on the medication.

Once we determine that a study is valid and has important results, we must decide if and how we should apply the evidence to our patient. The first step in this process is evaluating whether the patients in the study group are similar to our patient and whether the exposure is similar to what our patient might receive. It is important to remember that our patient does not necessarily have to meet the criteria for inclusion in the study. Instead, we should consider whether our patient is so different from those in the study that the results do not apply and decide whether there is a physiologic reason why our patient would respond differently. Interventions tend to have similar relative effects across a range of baseline risk for disease [9,10]. Therefore, although the patients in the RCT were at higher baseline risk for MI or cardiovascular death than Ms. Johnson, the hazard ratios should still be similar in our patient.

The next step in applying the evidence is determining all of the potential benefits and harms for the intervention. In this case, the potential benefit of celecoxib use in Ms. Johnson is effective control of arthritis pain. Based on the findings from the study by Solomon et al [3], the potential harm of celecoxib use is an increase in adverse cardiovascular events. When considering a treatment that has a potential harm, we must also consider other possible treatments. In this case, discontinuing all treatment is not a reasonable option because uncontrolled arthritis pain would decrease Ms. Johnson's quality of life. Another possibility would be to prescribe a standard NSAID with a protective agent such as a proton pump inhibitor.

The following week, when Ms. Johnson returns to the clinic, you discuss the information you have found about potential risks of taking celecoxib. You tell her that a well-performed study showed an association between use of celecoxib and an increase

in several serious adverse cardiovascular effects, including heart attack, stroke, and death. You add that the patients in the study were probably at a higher risk than she is for these problems, but that you estimate her risk of MI or death from heart disease to be about 7% over the next 10 years. You also acknowledge that celecoxib seems to have provided her with good pain control for her osteoarthritis and that other pain medicines have not worked as well.

Ms. Johnson says that the most important thing for her is to be able to keep active and play with her grandchildren. While she might consider switching to a standard NSAID and a proton pump inhibitor, she would rather continue with celecoxib, which has worked for her pain. She obviously does not want to have an MI and certainly does not want to die, but she feels that the 7% risk over 10 years is low.

You discuss Ms. Johnson's preferences in more detail to make sure that she truly understands the increased risk of cardiovascular events if she takes celecoxib. She is still certain that she will take the increased risk for the benefits of improved pain control and increased activity, so you prescribe celecoxib and tell her to call anytime if she needs to discuss the decision further.

Patient Preferences

An often overlooked part of practicing EBM is the incorporation of the patient's beliefs and preferences into clinical decisions. It is impossible for us to know a patient's concerns without asking about them. It is not appropriate for us to make decisions based solely on our value systems and what we determine is beneficial or harmful. Discussing the potential benefits and harms with our patients allows them to be active participants in their medical care, which increases the likelihood of compliance with treatment recommendations and can improve outcomes [11].

Conclusion

A summary of the EBM process used to evaluate the decision regarding Ms. Johnson's use of celecoxib is shown in **Table 3**. Our patients often have important questions about harm, and we need to find evidence to help answer these questions. Understanding which types of studies provide the best evidence for a particular harm question is important, as it can change with the type of exposure we are evaluating and the frequency of the outcome. Using a structured system like the *Users' Guides* for critically appraising the evidence allows us to systematically evaluate a study for bias. It is critical to be able to interpret the results of studies of harm, including ORs, hazard ratios, and Kaplan-Meier curves. If misinterpreted, these results can cause inaccurate assessments of patients' risks. This case example illustrates the importance of explaining the risks to a patient and taking the time to truly understand his or her preferences. While Ms. Johnson chose to except the increased risk and to continue to take celecoxib, not all patients would have made the same choice. Understanding the fundamentals of harm will allow us to help all our patients make better-informed decisions.

Table 3. Summary of the Evidence-Based Medicine Approach to Evaluate the Decision Regarding Ms. Johnson's Use of Celecoxib

Assess	Ms. Johnson has osteoarthritis and is taking celecoxib for pain relief. She is at low risk for MI or coronary death, but how celecoxib might affect her risk is unclear.
Ask	In an otherwise healthy woman with osteoarthritis, does the use of celecoxib compared with use of no medication lead to an increase in MI or coronary death?
Acquire	Cochrane Library—No systematic reviews relevant to our question *ACP Journal Club*—1 RCT evaluating cardiovascular risk in patients taking celecoxib to prevent colorectal adenomas PubMed Clinical Queries (MEDLINE)—No RCTs relevant to our question but 1 cohort study and several case-control studies
Appraise	An RCT, cohort study, and case-control study are all appraised. The RCT is determined to be less biased than the other studies. The study was randomized, and patients were similar for all known determinants of outcomes. Outcomes were measured the same way in both groups, and everyone completed at least 2.8 years of follow-up. The combined celecoxib group (400 mg/day or 800 mg/day) had an increased risk of death from cardiovascular causes or nonfatal MI, with a hazard ratio of 3.4.
Apply	The patients in the study were at higher risk of cardiovascular events than Ms. Johnson, but the results should still apply. Celecoxib increases Ms. Johnson's risk of cardiovascular events, but after an informed discussion, the patient is willing to accept this increased risk in order to continue taking celecoxib, which has controlled her pain.

MI = myocardial infarction; RCT = randomized controlled trial.

Address correspondence to: David Feldstein, MD, University of Wisconsin Hospital and Clinics, J5/210c, MC 2454, 600 Highland Ave., Madison, WI 53792 (email: df2@medicine.wisc.edu).

References

1. Kimmel SE, Berlin JA, Reilly M, et al. Patients exposed to rofecoxib and celecoxib have different odds of nonfatal myocardial infarction. Ann Intern Med 2005; 142:157–64.
2. Celecoxib increased cardiovascular events in patients with colorectal adenomas [abstract]. ACP J Club 2005 Jul/Aug;143:3. Abstract of Solomon SD, McMurray JJ, Pfeffer MA, et al. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. Adenoma Prevention with Celecoxib (APC) Study Investigators. N Engl J Med 2005;352:1071–80.
3. Solomon SD, McMurray JJ, Pfeffer MA, et al. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. Adenoma Prevention with Celecoxib (APC) Study Investigators. N Engl J Med 2005;352:1071–80.
4. Juurlink DN. Commentary on "Celecoxib increased cardiovascular events in patients with colorectal adenomas." ACP J Club 2005 Jul/Aug;143:3. Comment on: Solomon SD, McMurray JJ, Pfeffer MA, et al. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. Adenoma Prevention with Celecoxib (APC) Study Investigators. N Engl J Med 2005;352:1071–80.
5. Mamdani M, Rochon P, Juurlink DN, et al. Effect of selective cyclooxygenase 2 inhibitors and naproxen on short-term risk of acute myocardial infarction in the elderly. Arch Intern Med 2003;163:481–6.
6. Kernan WN, Viscoli CM, Brass LM, et al. Phenylpropanolamine and the risk of hemorrhagic stroke. N Engl J Med 2000;343:1826–32.
7. Guyatt G, Rennie D, editors. Users' guides to the medical literature: a manual for evidence-based clinical practice. The Evidence-Based Medicine Working Group, American Medical Association. Chicago (IL): AMA Press; 2002.
8. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 2002;288:321–3.
9. Yusuf S, Sleight P, Pogue J, et al. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients [published errata appear in N Engl J Med 2000;342:748 and 2000; 342:1376]. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 2000;342:145–53.
10. Simes J, Furberg CD, Braunwald E, et al. Effects of pravastatin on mortality in patients with and without coronary heart disease across a broad range of cholesterol levels. The Prospective Pravastatin Pooling project. Eur Heart J 2002;23:207–15.
11. Stewart MA. Effective physician-patient communication and health outcomes: a review. CMAJ 1995;152: 1423–33.

How to cite this article:
Feldstein DA, Chheda SG. Applying evidence to the use of celecoxib in a patient with osteoarthritis: how to address clinical questions of harm. Semin Med Pract 2006;9:9–20. Available at www.turner-white.com.