

HARM

Mitchell Levine, David Haslam, Stephen Walter,
Robert Cumming, Hui Lee, Ted Haines, Anne Holbrook,
Virginia Moyer, and Gordon Guyatt

The following EBM Working Group members also made substantive contributions to this section: Peter Pronovost and Sharon Straus

IN THIS SECTION

Finding the Evidence

Are the Results Valid?

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 Two Groups?

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 Patient in My Practice?

Was the Duration of Follow-up Adequate?

What Was the Magnitude of the Risk?

Should I Attempt to Stop the Exposure?

Clinical Resolution



CLINICAL SCENARIO

Do SSRIs Cause Gastrointestinal Bleeding?

You are a general practitioner considering the optimal choice of antidepressant medication. Your patient is a 55-year-old previously cheerful and well-adjusted individual who, during the past 2 months, has become sad and distressed for the first time in his life. He has developed difficulty concentrating and experiences early morning waking, but lacks thoughts of self-harm. The patient has attended your practice for the past 20 years and you know him well. You believe he is suffering from a major depressive episode and that he might benefit from antidepressant medication.

During recent years, you have been administering a selective serotonin reuptake inhibitor (SSRI), paroxetine, as your first-line antidepressant agent. However, recent reviews suggesting that the SSRIs are no more effective¹⁻³ and do not have lower discontinuation rates¹⁻⁴ than tricyclic antidepressants (TCAs) have led you to revert to your previous first choice, nortriptyline, in some patients. Patients in your practice usually consider the adverse effects in some depth before agreeing to any treatment decisions and many choose SSRIs on the basis of a preferable side-effect profile.

However, for the past 5 years the patient you are seeing today has been taking ketoprofen (a nonsteroidal anti-inflammatory drug, or NSAID), 50 mg three times per day, which has controlled the pain from his hip osteoarthritis. Your mind jumps to a review article suggesting that SSRIs may be associated with an increased risk of bleeding, and you become concerned about the risk of gastrointestinal bleeding when you consider that the patient is also receiving an NSAID. Unfortunately, an abstract from *Evidence Based Mental Health*,⁵ which you have used to obtain a summary of side effects of antidepressant medications, provides no information regarding this issue.

You remember the review article⁶ and locate a copy in your files, but at a glance you realize that it will not help answer your question for three reasons: It did not use explicit inclusion and exclusion criteria, it failed to conduct a systematic and comprehensive search, and it did not evaluate the methodologic quality of the original research it summarized (see Part 1E, "Summarizing the Evidence"). In addition, it did not cite any original studies specific to an association between SSRI treatment and gastrointestinal bleeding.

You consider that it is worth following up this issue before you make a final recommendation to the patient. You inform him that he will need antidepressant medication, but you explain your concern about the possible bleeding risk and your need to acquire more definitive information before making a final recommendation. You schedule a follow-up visit 2 days later and you commit to presenting a strategy at that time.

FINDING THE EVIDENCE

You formulate the following focused question:

Do adults suffering from depression and taking SSRI medications, compared to patients not taking antidepressants, suffer an increased risk of serious upper gastrointestinal bleeding?

Later that day, you begin your search using prefiltered evidence-based medicine resources—the journal *Evidence Based Mental Health*, Best Evidence⁴, Clinical Evidence, and the Cochrane Library. For each database, you enter the term “serotonin reuptake inhibitor.” Search of *Evidence Based Mental Health* yields eight reviews in volumes 1 (1998) and 2 (1999). Four of these deal with adverse effects associated with SSRI use, but none addresses gastrointestinal bleeding. Searching Best Evidence⁴ yields 17 equally unhelpful articles. A Clinical Evidence search identifies only a review on treatment of depressive disorders in adults. The Cochrane Library search locates four complete reviews and two abstracts of systematic reviews, but none addresses the issue of gastrointestinal bleeding in SSRI users.

You now turn to the PubMed version of MEDLINE and PreMEDLINE searching system (www.ncbi.nlm.nih.gov/entrez/query.fcgi). For optimum search efficiency, you click on “Clinical queries” under “PubMed Services” to access systematically tested search strategies, or you go to “Search hedges,” which will help you identify methodologically sound studies pertaining to your question on harm (see Part 1A1, “Finding the Evidence”). You enter the following: “selective serotonin reuptake inhibitor” AND “bleeding” for the subject search term; and you click on “Etiology” for study category and “Specificity” for emphasis. Your MEDLINE search (from 1966 through 2000) identifies one citation, an epidemiologic study assessing the association between SSRIs and upper gastrointestinal bleeding.⁷ This study describes a threefold increased risk of upper gastrointestinal bleeding associated with the use of SSRIs. Thinking that this article may answer your question, you download the full text free of charge from the *British Medical Journal (BMJ)* Web site (www.bmj.com) as a portable document format (PDF) file, an electronic version of a printed page or pages.

ARE THE RESULTS VALID?

Clinicians often encounter patients who are facing potentially harmful exposures, either to medical interventions or environmental agents, and important questions arise. Are pregnant women at increased risk of miscarriage if they work in front of video display terminals? Do vasectomies increase the risk of prostate cancer? Do hypertension management programs at work lead to increased absenteeism? When examining these questions, physicians must evaluate the validity of the data, the strength of the association between the assumed cause and the adverse outcome, and the relevance to patients in their practice.



As when answering any clinical question, our first goal should be to identify a systematic review of the topic that can provide an objective summary of all the available evidence (see Part 1E, “Summarizing the Evidence”). However, interpreting a systematic review requires an understanding of the rules of evidence for observational (nonrandomized) studies. The tests for judging the validity of observational study results, like the validity tests for randomized controlled trials, help you decide whether experimental and control groups began the study with a similar prognosis and whether similarity with respect to prognostic factors persisted after the study was started (see Table 1B-5).

TABLE 1B-5

Users' Guides for an Article About Harm

Are the results valid?

Did experimental and control groups begin the study with a similar 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 two 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 patient 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?

Did the Investigators Demonstrate Similarity in All Known Determinants of Outcome? Did They Adjust for Differences in the Analysis?

Studies of potentially harmful exposures will yield biased results if the group exposed to the putative harmful agent and the unexposed group begin with a different prognosis. Let us say we are interested in the impact of hospitalization on mortality rate. To investigate this question, we compare mortality in hospitalized individuals to that in people of similar age and sex in the community. Although an examination of the results would lead us to stay clear of hospitals, few would take these results seriously. The reason for skepticism is that people are admitted to hospitals because they are sick and, therefore, are at greater risk of dying. This higher risk results in a spurious (that is, noncausal) association between exposure

(hospitalization) and outcome (death). In general, people who seek health care or who take medicine are sicker than people who do not. If clinicians fail to take this into account, they are at high risk of making inaccurate inferences about causal relations between medications and adverse effects.

How can investigators ensure that their comparison groups start a study with a similar likelihood of suffering the target outcome? Randomized controlled trials provide less biased estimates of potentially harmful effects than other study designs because randomization is the best way to ensure that groups are balanced with respect to both known and unknown determinants of outcome (see Part 1B1, “Therapy”). Although investigators conduct RCTs to determine whether therapeutic agents are beneficial, RCTs can also demonstrate harm. The unexpected results of some randomized trials (for example, drugs that investigators expected to show benefit sometimes are associated with increased mortality) have demonstrated the potential of this study design for demonstrating harm (see Part 2B1, “Therapy and Validity, Surprising Results of Randomized Trials”).

There are two reasons that we cannot usually find RCTs to help us determine if a putative harmful agent truly has deleterious effects. First, we consider it unethical to randomize patients to exposures that may be harmful (not beneficial). Even if we did not hold these scruples, informed patients would not consent to such an experiment.

Second, we are often concerned about rare and serious adverse effects that occur over prolonged periods of time—ones that become evident only after tens of thousands of patients have consumed the medication. For instance, even a very large randomized trial⁸ failed to detect an association between clopidogrel and thrombotic thrombocytopenic purpura, which was detected by a subsequent observational study.⁹ Randomized trials specifically addressing side effects may be feasible for adverse event rates as low as 1%^{10,11} and meta-analyses may be very helpful when event rates are low.¹² The randomized trials we would need to explore harmful events that occur in less than one in 100 exposed patients—trials characterized by huge sample size and lengthy follow-up—are logistically difficult and prohibitively expensive.

Given that clinicians will not find RCTs to answer most questions about harm, they must understand alternative strategies for ensuring a balanced prognosis in the groups being compared. This understanding requires a familiarity with observational study designs, which we will now describe (Table 1B-6).



TABLE 1B-6

Directions of Inquiry and Key Methodologic Strengths and Weaknesses for Different Study Designs

Design	Starting Point	Assessment	Strengths	Weaknesses
Cohort	Exposure status	Outcome event status	Feasible when randomization of exposure not possible	Susceptible to bias, limited validity
Case-Control	Outcome event status	Exposure status	Overcomes temporal delays, may only require small sample size	Susceptible to bias, limited validity
RCT	Exposure status	Adverse event status	Low susceptibility to bias	Feasibility, generalizability

Cohort Studies

In a *cohort study*, the investigator identifies exposed and nonexposed groups of patients, each a cohort, and then follows them forward in time, monitoring the occurrence of the predicted outcome. In one such study, for example, investigators assessed perinatal outcomes among infants of men exposed to lead and organic solvents in the printing industry by means of a cohort of all males who had been members of printers' unions in Oslo.¹³ The investigators used job classification to categorize fathers as being either exposed to lead and solvents or not exposed to those substances. In this study, exposure was associated with an eightfold increase in preterm births, but it was not linked with birth defects.

Investigators may rely on cohort designs when harmful outcomes occur infrequently. For example, clinically apparent upper gastrointestinal hemorrhage in patients using NSAIDs occurs approximately 1.5 times per 1000 person-years of exposure, in comparison with 1.0 per 1000 person-years in those not taking NSAIDs.¹⁴ Because the event rate in unexposed patients is so low (0.1%), a randomized trial to study an increase in risk of 50% would require huge numbers of patients (sample size calculations suggest about 75,000 patients per group) for adequate power to test the hypothesis that NSAIDs cause the additional bleeding.¹⁵ Such a randomized trial would not be feasible, but a cohort study, in which the information comes from a large administrative database, would be possible.

The danger in using observational studies to assess a possible harmful exposure is that exposed and unexposed patients may begin with a different risk of the target outcome. For instance, in the association between NSAIDs and the increased risk of upper gastrointestinal bleeding, age may be associated both with exposure to NSAIDs and with gastrointestinal bleeding. In other words, since patients taking NSAIDs will be older and older patients are more likely to bleed,

this *confounding variable* makes attribution of an increased risk of bleeding to NSAID exposure problematic.

There is no reason patients who self-select (or who are selected by their physician) for exposure to a potentially harmful agent should be similar, with respect to other important determinants of outcome, to the nonexposed patients. Indeed, there are many reasons to expect they will not be similar. Physicians are reluctant to prescribe medications they perceive will put their patients at risk and will selectively prescribe low-risk medications. In one study, for instance, 24.1% of patients who were given a then-new NSAID, ketoprofen, had received peptic ulcer therapy during the previous 2 years in comparison to 15.7% of the control population.¹⁶ The likely reason is that the ketoprofen manufacturer succeeded in persuading clinicians that ketoprofen was less likely to cause gastrointestinal bleeding than other agents. A subsequent comparison of ketoprofen to other agents would be subject to the risk of finding a spurious increase in bleeding with the new agent because higher-risk patients would have been receiving the drug.

The prescription of benzodiazepines to elderly patients provides another example of the way that selective physician prescribing practices can lead to a different distribution of risk in patients receiving particular medications. This is referred to as the *channeling effect*.^{17,18} Ray and colleagues¹⁹ found an association between long-acting benzodiazepines and risk of falls (relative risk [RR], 2.0; 95% CI, 1.6-2.5) in data from 1977 to 1979, but not in data from 1984 to 1985 (RR, 1.3; 95% CI, 0.9-1.8). The most plausible explanation for the change is that patients at high risk for falls (those with dementia and anxiety or agitation) selectively received benzodiazepines during the earlier time period. Reports of associations between benzodiazepine use and falls led to greater caution, and the apparent association disappeared when physicians began to avoid benzodiazepine use in those at high risk of falling.

Therefore, investigators must document the characteristics of the exposed and nonexposed participants and either demonstrate their comparability or use statistical techniques to adjust for differences. Since investigators cannot recruit groups that are age-balanced, they must use statistical techniques that correct or adjust for the imbalances.

Effective adjustment for prognostic factors requires the accurate measurement of those prognostic factors. Large administrative databases, while providing a sample size that allows ascertainment of rare events, sometimes have limited quality of data concerning relevant patient characteristics. For example, Jollis and colleagues²⁰ wondered about the accuracy of information about patient characteristics in an insurance claims database. To investigate this issue, they compared the insurance claims data with prospective data collection by a cardiology fellow. They found that a high degree of chance corrected agreement between the fellow and the administrative database for the presence of diabetes: kappa, a measure of chance-corrected agreement, was 0.83 (see Part 2C, "Diagnosis, Measuring Agreement Beyond Chance"). They also found a high degree of agreement for myocardial infarction (kappa, 0.76), and moderate agreement for hypertension (kappa, 0.56). However, agreement was poor for heart failure (kappa, 0.39) and very poor for tobacco use (kappa, 0.19). We expand on the limitations of



administrative databases in another section of this book (see Part 2B, “Therapy and Harm, Outcomes of Health Services”).

Even if investigators document the comparability of potentially confounding variables in exposed and nonexposed cohorts and even if they use statistical techniques to adjust for differences, important prognostic factors that the investigators do not know about or have not measured may be unbalanced between the groups and, thus, may be responsible for differences in outcome. Returning to our earlier example, for instance, it may be that the illnesses that require NSAIDs, rather than the NSAIDs themselves, are responsible for the increased risk of bleeding. Thus, the strength of inference from a cohort study will always be less than that of a rigorously conducted RCT.

Case-Control Studies

Rare outcomes, or those that take a long time to develop, threaten cohort studies' feasibility. An alternative design relies on the initial identification of *cases*—that is, patients who have already developed the target outcome. The investigators then choose *controls*—persons who, as a group, are reasonably similar to the cases with respect to important determinants of outcome such as age, sex, and concurrent medical conditions, but who have not suffered the target outcome. Using this *case-control* design, investigators then assess the relative frequency of exposure to the putative harmful agent in the cases and controls, adjusting for differences in the known and measured prognostic variables. This design permits the simultaneous exploration of multiple exposures that have a possible association with the target outcome.

For example, investigators used a case-control design to demonstrate the association between diethylstilbestrol (DES) ingestion by pregnant women and the development of vaginal adenocarcinomas in their daughters many years later.²¹ An RCT or prospective cohort study designed to test this cause-and-effect relationship would have required at least 20 years from the time when the association was first suspected until the completion of the study. Further, given the infrequency of the disease, either an RCT or a cohort study would have required hundreds of thousands of participants. By contrast, using the case-control strategy, the investigators delineated two groups of young women. Those who had suffered the outcome of interest (vaginal adenocarcinoma) were designated as the cases ($n = 8$) and those who did not experience the outcome were designated as the controls ($n = 32$). Then, working backward in time, they determined exposure rates to DES for the two groups. The investigators found a strong association between in utero DES exposure and vaginal adenocarcinoma, which was extremely unlikely to be attributable to the play of chance ($P < .00001$). They found their answer without a delay of 20 years and by studying outcomes in only 40 women.

In another example, investigators used a case-control design relying on computer record linkages between health insurance data and a drug plan to investigate the possible relationship between use of beta-adrenergic agonists and mortality rates in patients with asthma.²² The database for the study included 95% of the population of the province of Saskatchewan in western Canada. The investigators

matched 129 cases of fatal or near-fatal asthma with 655 controls who also suffered from asthma but who had not had a fatal or near-fatal asthma attack.

The tendency of patients with more severe asthma to use more beta-adrenergic medications could create a spurious association between drug use and mortality rate. The investigators attempted to control for the confounding effect of disease severity by measuring the number of hospitalizations in the 24 months prior to death (cases) or the index date of entry in to the study (control group) and by using an index of the aggregate use of medications. They found an association between the routine use of large doses of beta-adrenergic agonist metered-dose inhalers and death from asthma (odds ratio [OR], 2.6 per canister per month; 95% CI, 1.7-3.9), even after correcting for their measures of disease severity.

As with cohort studies, case-control studies are susceptible to unmeasured confounding variables, particularly when exposure varies over time. For instance, previous hospitalization and medication use may not adequately capture all the variability in underlying disease severity in asthma. In addition, adverse lifestyle behaviors of asthmatic patients who use large amounts of beta agonists could contribute to the association. Furthermore, choice of controls may inadvertently create spurious associations. For instance, in a study that examined the association between coffee and pancreatic cancer, the investigators chose control patients from the practices of the physicians looking after the patients with pancreatic cancer.²³ These control patients had a variety of gastrointestinal problems, some of which were exacerbated by coffee ingestion. The control patients had learned to avoid coffee; as a result, the investigators found an association between coffee (which the pancreatic cancer patients consumed at general population levels) and cancer. Subsequent investigations, using more appropriate controls, refuted the association.^{24, 25} These problems illustrate why clinicians can draw inferences of only limited strength from the results of observational studies, even after adjustment for known determinants of outcome.

Case Series and Case Reports

Case series (descriptions of a series of patients) and *case reports* (descriptions of individual patients) do not provide any comparison group, and are therefore unable to satisfy the requirement that treatment and control groups share a similar prognosis. Although descriptive studies occasionally demonstrate dramatic findings mandating an immediate change in physician behavior (eg, recall the consequences when a link was associated between thalidomide and birth defects²⁶), there are potentially undesirable consequences when actions are taken in response to weak evidence. Consider the case of the drug Bendectin (a combination of doxylamine, pyridoxine, and dicyclomine used as an antiemetic in pregnancy), whose manufacturer withdrew it from the market as a result of case reports suggesting it was teratogenic.²⁷ Later, even though a number of comparative studies demonstrated the drug's relative safety,²⁸ they could not eradicate the prevailing litigious atmosphere—which prevented the manufacturer from reintroducing Bendectin. Thus, many pregnant women who potentially could have benefited from the drug's availability were denied the symptomatic relief it could have offered.



In general, clinicians should not draw conclusions about cause-and-effect relationships from case series but, rather, should recognize that the results may generate questions for regulatory agencies and clinical investigators to address.

Design Issues—Summary

Just as is true for the resolution of questions of therapeutic effectiveness, clinicians should look first for randomized trials to resolve issues of harm. They will often be disappointed in this search and must make use of studies of weaker design.

Regardless of the design, however, they should look for an appropriate control population before making a strong inference about a putative harmful agent. For RCTs and cohort studies, the control group should have a similar baseline risk of outcome, or investigators should use statistical techniques to adjust or correct for differences. Similarly, in case-control studies the derived exposed and nonexposed groups should be similar with respect to determinants of outcome other than the exposure under study. Alternatively, investigators should use statistical techniques to adjust for differences. Even when investigators have taken all the appropriate steps to minimize bias, clinicians should bear in mind that residual differences between groups may always bias the results of observational studies.²⁹ Since prescribing in the real world is carried out on the basis of evidence, clinician values, and patient values, exposure opportunities in nonrandomized medication studies are likely to differ among patients (channeling bias or effect).

Were Exposed Patients Equally Likely to Be Identified in the Two Groups?

In case-control studies, ascertainment of the exposure is a key issue. For example, when patients with leukemia are asked about prior exposure to solvents, they may be more likely to recall exposure than would control group members, either because of increased patient motivation (*recall bias*) or because of greater probing by an interviewer (*interviewer bias*). Clinicians should note whether investigators used bias-minimizing strategies such as blinding participants and interviewers to the hypothesis of the study. For example, a case-control study found a twofold increase in risk of hip fracture associated with psychotropic drug use. In this study, investigators established drug exposure by examining computerized claims files of the Michigan Medicaid program, a strategy that avoided both recall and interviewer bias.³⁰ The study of beta-adrenergic agonist use in patients with asthma suggesting an association with mortality also relied on an administrative database to ascertain exposure.⁹ In both cases, the assurance of unbiased exposure status increases our confidence in the studies' findings.

Were the Outcomes Measured in the Same Way in the Groups Being Compared?

In RCTs and cohort studies, ascertainment of outcome is a key issue. For example, investigators have reported a threefold increase in the risk of malignant melanoma in individuals working with radioactive materials. One possible explanation for

some of the increased risk might be that physicians, aware of a possible risk, search more diligently and, therefore, detect disease that might otherwise go unnoticed (or they may detect disease at an earlier point in time). This could result in the exposed cohort having an apparent, but spurious, increase in risk—a situation we refer to as *surveillance bias*.³¹

Was Follow-up Sufficiently Complete?

As we pointed out in Part 1B1, “Therapy,” loss to follow-up can introduce bias because the patients who are lost may have very different outcomes from those still available for assessment. The longer the required follow-up period, the greater the possibility that the follow-up will be incomplete.

For example, in a well-executed study, investigators determined the vital status of 1235 of 1261 white males (98%) employed in a chrysotile asbestos textile operation between 1940 and 1975.³² The relative risk for lung cancer death over time increased from 1.4 to 18.2 in direct proportion to the cumulative exposure among asbestos workers with at least 15 years since first exposure. Because the 2% missing data were unlikely to affect the results, the loss to follow-up does not threaten the validity of the inference that asbestos exposure causes lung cancer deaths.

USING THE GUIDE

Returning to our earlier discussion, the study that we retrieved investigating the association between SSRIs and risk of upper gastrointestinal bleeding used a case-control design.⁶ Data came from a general practitioner electronic medical record database in the United Kingdom, which included data from more than 3 million people, most of whom had been entered prospectively during a 5-year period.³³⁻³⁵ The investigators identified cases of upper gastrointestinal bleeding ($n=1651$) and ulcer perforation ($n=248$) among patients aged 40 to 79 years between 1993 and 1997. They then randomly selected 10,000 controls from the at-risk source population that gave rise to cases, choosing their sample so that age, sex, and the year patients were identified were similar among the cases and control groups.

The analysis controlled for a number of possible prognostic factors: previous dyspepsia, gastritis, peptic ulcer and upper gastrointestinal bleeding or perforation, smoking status, and current use of NSAIDs, anticoagulants, corticosteroids, and aspirin. The database included prescription drugs only. The investigators examined the relative frequency of SSRI prescription use in the 30 days before the *index date* (that is, the date of the reported bleeding or perforation) in patients with and without bleeding and perforation after controlling for the prognostic variables. Control patients received a random date as their index date.



Although the investigators controlled for a number of prognostic factors, there are other potential important determinants of bleeding for which they did not control. For example, more patients being treated for depression or anxiety suffer from painful medical conditions than those without depression and anxiety. Patients may have been using over-the-counter NSAIDs for these problems. The database the investigators used does not capture the use of self-medication with over-the-counter analgesics.

Alcohol use is another potential confounder. Although the investigators excluded patients with known alcoholism, many persons afflicted with alcoholism remain unrevealed to their primary care physician, and alcoholism is associated with an increased prevalence of depression and anxiety that could lead to the prescription of SSRIs. Since alcoholism is associated with increased bleeding risk, this prognostic variable fulfills all the criteria for a confounding variable that could bias the results of the study. Finally, it is possible that patients returning for prescription of SSRIs would be more likely to have their bleeding diagnosed in comparison to patients under less intense surveillance (a state of affairs known as *detection bias*).

These biases should apply to all three classes of antidepressants (ie, SSRIs, nonselective serotonin reuptake inhibitors, and a miscellaneous group of other drugs) that the investigators considered. The results of the study, which we will discuss later in this section, showed an association only between gastrointestinal bleeding and SSRIs, rather than between gastrointestinal bleeding and other antidepressant medications. One would expect all these biases to influence the association between any antidepressant agent and bleeding. Thus, the fact that the investigators found the association only with SSRIs decreases our concern about the threats to validity from possible differences in prognostic factors in those receiving—and not receiving—SSRIs.

At the same time, most physicians make decisions regarding the prescription of SSRIs or tricyclic antidepressant agents based on particular patient characteristics. Thus, it remains possible that these characteristics include some that are associated with the incidence of gastrointestinal bleeding. This would be true, for instance, if clinicians differentially used SSRI rather than other antidepressant medications in patients in whom they suspected alcohol abuse.

The major strength of the use of a large database for this study is that it eliminates the possibility of biased assessment of exposure (or recall bias) to SSRIs in the patients who suffered the outcomes as well as in those who did not. The outcomes and exposures were probably measured in the same way in both groups, as most clinicians are unaware that UGI bleeding may be associated with SSRI use. We have no idea, however, about the number of patients lost to follow-up. Although the investigators included

only those patients who stayed in the practices of the participating primary care physicians from the beginning to the end of the study, we do not know, for instance, how many people in the database began to receive SSRIs but subsequently left those practices.

In summary, the study suffers from the limitation inherent in any observational study: that exposed and unexposed patients may differ in prognosis at baseline. In this case, at least two unmeasured variables, over-the-counter NSAID use and alcohol consumption, might create a spurious association between SSRIs and gastrointestinal bleeding. The other major limitation of the study is the lack of information regarding completeness of follow-up. That said, although these limitations weaken any inferences we might make, we are likely to conclude that the study is strong enough to warrant a review of the results.

WHAT ARE THE RESULTS?

How Strong Is the Association Between Exposure and Outcome?

We have described the alternatives for expressing the association between the exposure and the outcome, the relative risk and the odds ratio, in other sections of this book (see Part 1B1, “Therapy”; see also Part 2B2, “Therapy and Understanding the Results, Measures of Association”). In a cohort study assessing in-hospital mortality after noncardiac surgery in male veterans, 23 of 289 patients with a history of hypertension died, compared with three of 185 patients without the condition. The relative risk for hypertension and mortality, $(23/289)/(3/185)$, was 4.9.³⁶ The relative risk tells us that death after noncardiac surgery occurs almost five times more often in patients with hypertension than in normotensive patients.

The estimate of relative risk depends on the availability of samples of exposed and unexposed patients, where the proportion of the patients with the outcome of interest can be determined. The relative risk is therefore not applicable to case-control studies in which the number of cases and controls—and, therefore, the proportion of individuals with the outcome—is chosen by the investigator. For case-control studies, instead of using a ratio of risks (*relative risk*), we use a ratio of odds (*odds ratio*): the odds of a case-patient being exposed, divided by the odds of a control patient being exposed (see Part 2B2, “Therapy, Measures of Association”).

When considering both study design and strength of association, we may be ready to interpret a small increase in risk as representing a true harmful effect when the study design is strong (such as in a RCT). A much higher increase in risk might be required of weaker designs (such as cohort or case-control studies), as subtle findings are more likely to be caused by the inevitably higher risk of bias.



Very large values of relative risk or odds ratio represent strong associations that are less likely to be caused by confounding variables or bias.

In addition to showing a large magnitude of relative risk or odds ratio, a second finding will strengthen an inference that we are dealing with a true harmful effect. If, as the quantity or the duration of exposure to the putative harmful agent increases, the risk of the adverse outcome also increases (that is, the data suggest a dose-response gradient), we are more likely to be dealing with a causal relationship between exposure and outcome. The fact that the risk of dying from lung cancer in male physician smokers increases by 50%, 132%, and 220% for 1 to 14, 15 to 24, and 25 or more cigarettes smoked per day, respectively, strengthens our inference that cigarette smoking causes lung cancer.³⁷

How Precise Is the Estimate of the Risk?

Clinicians can evaluate the precision of the estimate of risk by examining the confidence interval around that estimate (see Part 1B1, “Therapy”; see also Part 2B2, “Therapy and Understanding the Results, Confidence Intervals”). In a study in which investigators have shown an association between an exposure and an adverse outcome, the lower limit of the estimate of relative risk associated with the adverse exposure provides a minimal estimate of the strength of the association. By contrast, in a study in which investigators fail to demonstrate an association (a negative study), the upper boundary of the confidence interval around the relative risk tells the clinician just how big an adverse effect may still be present, despite the failure to show a statistically significant association (see Part 2B2, “Therapy and Understanding the Results, Confidence Intervals”).

USING THE GUIDE

Returning to our earlier discussion, the investigators calculated odds ratios (ORs) of the risk of bleeding in those exposed to SSRIs vs those not exposed, but they reported the results as relative risks (RR). Unfortunately, this practice is not unusual. Fortunately, when event rates are low, relative risks and odds ratios closely approximate one another (see Part 2B2, “Therapy, Measures of Association”). The investigators found an association between current use of SSRIs and upper gastrointestinal bleeding (adjusted OR, 3.0; 95% CI, 2.1-4.4). They noted a weak association with nonselective serotonin reuptake inhibitors (adjusted OR, 1.4; 95% CI, 1.1-1.9), but found no association with antidepressant medications that had no action on the serotonin reuptake mechanism. The investigators found that the association between NSAID use and bleeding (adjusted OR, 3.7; 95% CI, 3.2-4.4) was of similar magnitude to the association between bleeding and SSRIs. The current use of SSRIs with prescription NSAID drugs further increased the risk of upper gastrointestinal bleeding (adjusted OR, 15.6; 95% CI, 6.6-36.6). The dose and duration of SSRI use had little influence on the risk of this adverse outcome.

HOW CAN I APPLY THE RESULTS TO PATIENT CARE?

Were the Study Patients Similar to the Patient in My Practice?

If possible biases in a study are not sufficient to dismiss the study out of hand, you must consider the extent to which they might apply to the patient in your office. Is the patient before you similar to those described in the study with respect to morbidity, age, sex, race, or other potentially important factors? If not, is the biology of the harmful exposure likely to differ in the patient you are attending (see Part 2B3, “Therapy and Applying the Results, Applying Results to Individual Patients”)? Are there important differences in the treatments or exposures between the patients you see and the patients studied? For example, the risk of thrombophlebitis associated with oral contraceptive use described in the 1970s may not be applicable to the patient of the 1990s because of the lower estrogen dose in oral contraceptives used in the 1990s. Similarly, increases in uterine cancer secondary to postmenopausal estrogen replacement do not apply to women who are also taking concomitant progestins tailored to produce monthly withdrawal bleeding with chronic, noncyclic use.

Was the Duration of Follow-up Adequate?

Let us return for a moment to the study that showed that workers employed in chrysotile asbestos textile operation between 1940 and 1975 showed an increased risk for lung cancer death, a risk that increased from 1.4 to 18.2 in direct relation to cumulative exposure among asbestos workers with at least 15 years since first exposure.³² The fact that the follow-up was sufficiently long to capture a large proportion of the lung cancers destined to occur enhances our confidence in application of the results to patients in our practice. By contrast, excessively short follow-up may fail to detect harmful effects that emerge with longer observation.

What Was the Magnitude of the Risk?

The relative risk and the odds ratio do not tell us how frequently the problem occurs; they tell us only that the observed effect occurs more or less often in the exposed group compared to the unexposed group. Thus, we need a method for assessing clinical importance. In our discussion of therapy (see Part 1B1, “Therapy”; see also Part 2B2, “Therapy and Understanding the Results, Measures of Association”), we described the way to calculate the number of patients who must be treated to prevent an adverse event. When the issue is harm, we can use data from a randomized trial or cohort study, but not a case-control study, to make an analogous calculation to determine how many people must be exposed to the harmful agent to cause an adverse outcome.

For example, over an average of 10 months of follow-up, investigators conducting the Cardiac Arrhythmia Suppression Trial (CAST), a RCT of antiarrhythmic agents,^{38, 39} found that the mortality rate was 3.0% for placebo-treated patients and



7.7% for those treated with either encainide or flecainide. The *absolute risk* increase was 4.7%, the reciprocal of which tells us that, on average, for every 21 patients we treat with encainide or flecainide for about a year, we will cause one excess death. This contrasts with our example of the association between NSAIDs and upper gastrointestinal bleeding. Of 2000 unexposed patients, two will suffer a bleeding episode each year. Of 2000 patients taking NSAIDs, three will suffer such an episode each year. Thus, if we treat 2000 patients with NSAIDs, we can expect a single additional bleeding event.⁷

Should I Attempt to Stop the Exposure?

After evaluating the evidence that an exposure is harmful and after establishing that the results are potentially applicable to the patient in your practice, determining subsequent actions may not be simple. There are at least three aspects to consider in making a clinical decision.

First is the strength of inference: how strong was the study or studies that demonstrated harm in the first place? Second, what is the magnitude of the risk to patients if exposure to the harmful agent continues? Third, what are the adverse consequences of reducing or eliminating exposure to the harmful agent—that is, the magnitude of the benefit that patients will no longer receive?

Clinical decision making is simple when both the likelihood of harm and its magnitude are great. Because the evidence of increased mortality from encainide and flecainide came from a randomized trial,³⁸ we can be confident of the causal connection. Since treating only 21 people will result in an excess death, it is no wonder that clinicians quickly curtailed their use of these antiarrhythmic agents when the study results became available.

The clinical decision is also made easier when an acceptable alternative for avoiding the risk is available. For example, beta blockers prescribed for the treatment of hypertension can result in symptomatic increase in airway resistance in patients with asthma or chronic airflow limitation. This risk mandates the use of an alternative drug, such as a thiazide diuretic, in susceptible patients.⁴⁰

Even if the evidence is relatively weak, the availability of an alternative can result in a clear decision. The early case-control studies demonstrating the association between aspirin use and Reye syndrome, for example, were relatively weak and left considerable doubt about the causal relationship. Although the strength of inference was not great, the availability of a safe, inexpensive, and well-tolerated alternative, acetaminophen, justified use of this alternative agent in children at risk of Reye syndrome.⁴¹

In contrast to the early studies regarding aspirin and Reye syndrome, multiple well-designed cohort and case-control studies have consistently demonstrated an association between NSAIDs and upper gastrointestinal bleeding; therefore, our inference about harm has been relatively strong. However, the risk of an upper gastrointestinal bleeding episode is quite low, and until recently we have not had safer and equally efficacious anti-inflammatory alternatives available. We were therefore probably right in continuing to prescribe NSAIDs for the appropriate

clinical conditions. Depending on both their safety profile after longer experience and cost-effectiveness considerations, COX 2-inhibiting NSAIDs may prove to be an appropriate alternative class of agents.

CLINICAL RESOLUTION

To decide on your course of action, you proceed through the three steps of using the medical literature to guide your clinical practice. First, you consider the validity of the study before you. The antidepressant and upper gastrointestinal bleeding study addressed multiple classes of antidepressant agents and the risk of upper gastrointestinal bleeding or ulcer perforation. You decide that the limitations of the case-control design, along with the lack of information about loss to follow-up, leave you uncertain about a causal relationship between SSRIs and gastrointestinal bleeding. Furthermore, this is a single study and, as we have previously mentioned, in other areas of medicine subsequent investigations^{11, 12, 42-45} have failed to confirm many apparent harmful associations.^{10, 46, 47}

Turning to the results, you note the very strong association between the combined use of SSRIs and NSAIDs. Despite the methodologic limitations of this single study, you believe the association is too strong to ignore. You therefore proceed to the third step and consider the implications of the results for the patient before you.

The primary care database from which the investigators drew their sample suggests that the results are readily applicable to the patient before you. You consider the magnitude of the risk to which you would be exposing this patient if you prescribed an SSRI and it actually did cause bleeding. Using the baseline risk reported by Carson et al in a similar population,¹⁴ you calculate that you would need to treat about 625 patients with SSRIs for a year to cause a single bleeding episode in patients not using NSAIDs, and about 55 patients a year taking NSAIDs along with an SSRI for a year to cause a single bleeding episode.

From previous experience with the patient before you, you know that he is risk averse. When he returns to your office, you note the equal effectiveness of the SSRIs and tricyclic antidepressants that you can offer him, and you describe the side-effect profile of the alternative agents. You note, among the other considerations, the possible increased risk of gastrointestinal bleeding with the SSRIs. The patient decides that, on balance, he would prefer a tricyclic antidepressant and leaves your office with a prescription for nortriptyline.



References

1. Geddes JR, Freemantle N, Mason J, Eccles MP, Boynton J. SSRIs versus other antidepressants for depressive disorder. *Cochrane Database of Systematic Reviews*. 2000:1-26.
2. Trindale E, Menon D. *Selective Serotonin Reuptake Inhibitors (SSRIs) for Major Depression*. Ottawa: Canadian Coordinating Office for Health Technology Assessment; 1997.
3. Mulrow CD, Williams JW Jr, Trivedi M. *Treatment of Depression: Newer Pharmacotherapies*. San Antonio, Texas: San Antonio Evidence-based Practice Centre; 1999.
4. Hotopf M, Hardy R, Lewis G. Discontinuation rates of SSRIs and tricyclic antidepressants: a meta-analysis and investigation of heterogeneity. *Br J Psychiatry*. 1997;170:120-127.
5. Trindale E, Menon D. Review: selective serotonin reuptake inhibitors differ from tricyclic antidepressants in adverse events. *Evidence Based Mental Health*. 1998;1:50.
6. Goldberg RJ. Selective serotonin reuptake inhibitors: infrequent medical adverse effects. *Arch Fam Med*. 1998;7:78-84.
7. de Abajo FJ, Rodriguez LA, Montero D. Association between selective serotonin reuptake inhibitors and upper gastrointestinal bleeding: population based case-control study. *BMJ*. 1999;319:1106-1109.
8. CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet*. 1996;348:1329-1339.
9. Bennett CL, Connors JM, Carwile JM, et al. Thrombotic thrombocytopenic purpura associated with clopidogrel. *N Engl J Med*. 2000;342:1773-1777.
10. Silverstein FE, Graham DY, Senior JR, et al. Misoprostol reduces serious gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal anti-inflammatory drugs: a randomized, double-blind, placebo-controlled trial. *Ann Intern Med*. 1995;123:241-249.
11. Merck and Co. VIGOR Study Summary. Paper presented at: Digestive Disease Week Congress; May 24, 2000; San Diego.
12. Langman MJ, Jensen DM, Watson DJ, et al. Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. *JAMA*. 1999;282:1929-1933.
13. Kristensen P, Irgens LM, Daltveit AK, Andersen A. Perinatal outcome among children of men exposed to lead and organic solvents in the printing industry. *Am J Epidemiol*. 1993;137:134-143.
14. Carson JL, Strom BL, Soper KA, et al. The association of nonsteroidal anti-inflammatory drugs with upper gastrointestinal tract bleeding. *Arch Intern Med*. 1987;147:85-88.

15. Walter SD. Determination of significant relative risks and optimal sampling procedures in prospective and retrospective comparative studies of various sizes. *Am J Epidemiol*. 1977;105:387-397.
16. Leufkens HG, Urquhart J, Stricker BH, Bakker A, Petri H. Channelling of controlled release formulation of ketoprofen (Oscorel) in patients with history of gastrointestinal problems. *J Epidemiol Community Health*. 1992;46:428-432.
17. Joseph KS. The evolution of clinical practice and time trends in drug effects. *J Clin Epidemiol*. 1994;47:593-598.
18. Leufkens HG, Urquhart J. Variability in patterns of drug usage. *J Pharm Pharmacol*. 1994;46(suppl 1):433-437.
19. Ray WA, Griffin MR, Downey W. Benzodiazepines of long and short elimination half-life and risk of hip fracture. *JAMA*. 1989;262:3303-3307.
20. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical information systems: implications for outcomes research. *Ann Intern Med*. 1993;119:844-850.
21. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med*. 1971;284:878-881.
22. Spitzer WO, Suissa S, Ernst P, et al. The use of beta-agonists and the risk of death and near death from asthma. *N Engl J Med*. 1992;326:501-506.
23. MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. Coffee and cancer of the pancreas. *N Engl J Med*. 1981;304:630-633.
24. Baghurst PA, McMichael AJ, Slavotineck AH, Baghurst KI, Boyle P, Walker AM. A case-control study of diet and cancer of the pancreas. *Am J Epidemiol*. 1991;134:167-179.
25. Zheng W, McLaughlin JK, Gridley G, et al. A cohort study of smoking, alcohol consumption and dietary factors for pancreatic cancer. *Cancer Causes Control*. 1993;4:477-482.
26. Lenz W. Epidemiology of congenital malformations. *Ann NY Acad Sci*. 1965;123:228-236.
27. Soverchia G, Perri PF. Two cases of malformation of a limb in infants of mothers treated with an antiemetic in a very early phase of pregnancy. *Pediatr Med Chir*. 1981;3:97-99.
28. Holmes LB. Teratogen update: Bendectin. *Teratology*. 1983;27:277-281.
29. Kellermann AL, Rivara FP, Rushforth NB, Banton JG, et al. Gun ownership as a risk factor for homicide in the home. *N Engl J Med*. 1993;329:1084-1091.
30. Ray WA, Griffin MR, Schaffner W, et al. Psychotropic drug use and the risk of hip fracture. *N Engl J Med*. 1987;316:363-369.
31. Hiatt RA, Fireman B. The possible effect of increased surveillance on the incidence of malignant melanoma. *Prev Med*. 1986;15:652-660.



32. Dement JM, Harris RL Jr, Symons MJ, Shy CM. Exposures and mortality among chrysotile asbestos workers. Part II: mortality. *Am J Ind Med.* 1983;4:421-433.
33. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised resource in the United Kingdom. *BMJ.* 1991;302:766-768.
34. Garcia Rodriguez LA, Perez Gutthann S. Use of the UK general practice research database for pharmacoepidemiology. *Br J Clin Pharmacol.* 1998;45:419-425.
35. Jick H, Terris B, Derby LE, Jick SS. Further validation of information recorded on a general practitioner database resource in the United Kingdom. *Pharmacoepidemiol Drug Saf.* 1992;1:347-349.
36. Browner WS, Li J, Mangano DT. In-hospital and long-term mortality in male veteran following noncardiac surgery. *JAMA.* 1992;268:228-232.
37. Doll R, Hill AB. Mortality in relation to smoking: ten years' observation of British doctors. *BMJ.* 1964;1:1399-1410, 1460-1467.
38. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: the Cardiac Arrhythmia Suppression Trial. *N Engl J Med.* 1991;324:781-788.
39. The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med.* 1992;327:227-233.
40. Ogilvie RI, Burgess ED, Cusson JR, Feldman RD, Leiter LA, Myers MG. Report of the Canadian Hypertension Society Consensus Conference, 3: pharmacologic treatment of essential hypertension. *Can Med Assoc J.* 1993;149:575-584.
41. Soumerai SB, Ross-Degnan D, Kahn JS. Effects of professional and media warnings about the association between aspirin use in children and Reye's syndrome. *Milbank Q.* 1992;70:155-182.
42. Danesh J, Appleby P. Coronary heart disease and iron status: meta-analyses of prospective studies. *Circulation.* 1999;99:852-854.
43. Klebanoff MA, Read JS, Mills JL, Shiono PH. The risk of childhood cancer after neonatal exposure to vitamin K. *N Engl J Med.* 1993;329:905-908.
44. Passmore SJ, Draper G, Brownbill P, Kroll M. Case-control studies of relation between childhood cancer and neonatal vitamin K administration: retrospective case-control study. *BMJ.* 1998;316:178-184.
45. Parker L, Cole M, Craft AW, Hey EN. Neonatal vitamin K administration and childhood cancer in the north of England. *BMJ.* 1998;316:189-193.
46. Salonen JT, Nyssonen K, Korpela H, Tuomilehto J, Seppanen R, Salonen R. High stored iron levels are associated with excess risk of myocardial infarction in eastern Finnish men. *Circulation.* 1992;86:803-811.
47. Golding J, Greenwood R, Birmingham K, Mott M. Childhood cancer, intramuscular vitamin K, and pethidine given during labour. *BMJ.* 1992;305:341-346.