

Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3rd ed >

## Chapter 14: Harm (Observational Studies)

Mitchell Levine; John P. A. Ioannidis; Alfred Theodore Haines; Gordon Guyatt

### Clinical Scenario

#### CLINICAL SCENARIO

##### Does Soy Milk (or Soy Formula) Increase the Risk of Developing Peanut Allergy in Children?

You are a general practitioner examining a 29-year-old patient who is 8 months pregnant with her second child. Her first child, who is now 3 years old, had an intolerance to cow's milk as an infant. He was switched to soy formula and then soy milk, which he subsequently tolerated well. At 2 years of age, cow's milk was reintroduced without any problems, and he has been receiving cow's milk since. The mother was planning to start feeding her next child soy formula at birth but heard from a neighbor that it can increase the risk of peanut allergy in her child—a potentially serious and lifelong problem. She asks for your advice on the topic. Because you are not familiar with this issue, you inform the patient that you will examine the *evidence* and discuss your findings with her when she returns for her next prenatal visit in 1 week.

### Finding the Evidence

You formulate the relevant question: In infants, is there an association between *exposure* to soy milk and the subsequent development of peanut allergy? You search a point-of-care clinician evidence synthesis tool with the term “peanut allergy.” Under the subtopic “Causes and Risk Factors,” you see that “consumption of soy milk or soy formula” is identified as a possible *risk factor* and a reference is provided. You click on the hypertext link to view the relevant article.<sup>1</sup>

The article describes a *case-control study* that used a geographically defined *cohort* of 13971 preschool children. The investigators identified children with a convincing history of peanut allergy who reacted to a peanut challenge in which they were blind to whether they were being exposed to peanut protein or a “placebo.” They collected detailed information from the children's parents and from 2 groups of control parents (a random sample from the geographically defined cohort and from a subgroup of children from the cohort who had eczema in the first 6 months of life and whose mothers had a history of eczema).

**Box 14-1** presents our usual 3-step approach to using an article about *harm* from the medical literature to guide your practice. You will find these criteria useful for a variety of issues that involve concerns of etiology or risk factors in which a potentially harmful exposure cannot be randomly assigned. These *observational studies* involve using cohort or case-control designs.

BOX 14-1

**Users' Guides for an Article About Harm**

**How serious is the risk of bias?**

- In a cohort study, aside from the exposure of interest, did the exposed and control groups start and finish with the same risk for the outcome?
- Were patients similar for prognostic factors that are known to be associated with the outcome (or did statistical adjustment address the imbalance)?
- Were the circumstances and methods for detecting the outcome similar?
- Was the follow-up sufficiently complete?
- In a case-control study, did the cases and control group have the same risk for being exposed in the past?
- Were cases and controls similar with respect to the indication or circumstances that would lead to exposure (or did statistical adjustment address the imbalance)?
- Were the circumstances and methods for determining exposure similar for cases and controls?

**What are the results?**

- How strong is the association between exposure and outcome?
- How precise was 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 follow-up sufficiently long?
- Is the exposure similar to what might occur in my patient?
- What is the magnitude of the risk?
- Are there any benefits that are known to be associated with exposure?

## How Serious Is the Risk of Bias?

Clinicians often encounter patients who face potentially harmful exposures to either medical interventions or environmental agents. These circumstances give rise to common questions: Do cell phones increase the risk of brain tumors? Do vasectomies increase the risk of prostate cancer? Do changes in health care policies (eg, activity-based funding) lead to harmful *health outcomes*? When examining these questions, clinicians and administrators must evaluate the *risk of bias*, the strength of the association between the assumed cause and the adverse outcome, and the relevance to patients in their practice or domain.

In answering any clinical question, our first goal should be to identify whether there is an existing *systematic review* of the topic that can provide a summary of the highest-quality available evidence (see the Summarizing the Evidence section). Interpreting such a *review* requires an understanding of the rules of evidence for individual or *primary studies*, *randomized clinical trials* (RCTs), and observational studies. The tests for judging the risk of bias associated with results of observational studies will help you decide whether exposed and *control groups* (or cases and controls) began and completed the study with sufficient similarities that we can obtain a minimally biased assessment of the influence of exposure on outcome (see [Chapter 6](#), Why Study Results Mislead: Bias and Random Error).

Randomized clinical 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 known and unknown determinants of the outcome (see [Chapter 7](#), Therapy [Randomized Trials]). Although investigators conduct RCTs to determine whether therapeutic agents are beneficial, they also should look for harmful effects and may sometimes make surprising discoveries about the adverse effects of the intervention on their primary outcomes (see [Chapter 11.2](#), Surprising Results of Randomized Trials).

There are 4 reasons why RCTs may not be helpful for determining whether a putative harmful agent truly has deleterious effects. First, we may consider it unethical to randomize patients to exposures that might result in harmful effects without benefit (eg, smoking).

Second, we are often concerned about rare and serious adverse effects that may become evident only after tens of thousands of patients have consumed a medication for a period of years. For instance, even a very large RCT failed to detect an association between clopidogrel and thrombotic thrombocytopenic purpura,<sup>2</sup> which appeared in a subsequent observational study.<sup>3</sup> Randomized clinical trials that address adverse effects may be feasible for adverse event rates as low as 1%,<sup>4,5</sup> but the RCTs needed to explore harmful events occurring in fewer than 1 in 100 exposed patients are logistically difficult and often prohibitively expensive because of the huge sample size and lengthy *follow-up* required. *Meta-analyses* may be helpful when the event rates are very low.<sup>6</sup> However, availability of large-scale evidence on specific harms in systematic reviews is not common. For example, in a report of nearly 2000 systematic reviews, only 25 had large-scale data on 4000 or more randomized participants regarding well-defined harms that might be associated with the interventions under study.<sup>7</sup>

Third, RCT duration of follow-up is limited, yet not infrequently we are interested in knowing effects years, or even decades, after the exposure (eg, long-term consequences of chemotherapy in childhood).<sup>8</sup>

Fourth, even when events are sufficiently frequent and occur during a time frame feasible for RCTs to address, study reports often fail to adequately provide information on harm.

Given that clinicians will not find RCTs to answer most questions about harm, they must understand the alternative strategies used to minimize *bias*. This requires a familiarity with observational study designs (Table 14-1).

TABLE 14-1

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

| Design                    | Starting Point       | Assessment           | Strengths  | Weaknesses                                   |
|---------------------------|----------------------|----------------------|--|--|
| Randomized clinical trial | Exposure status      | Outcome event status | Low susceptibility to bias   | Feasibility and generalizability constraints |
| Cohort                    | Exposure status      | Outcome event status | Feasible when randomization of exposure not possible, generalizability                 | Susceptible to bias                          |
| Case-control              | Outcome event status | Exposure status      | Overcomes temporal delays and the need for huge sample sizes to accumulate rare events | Susceptible to bias                          |

There are 2 main types of observational studies: cohort and case-control. In a cohort study, investigators identify exposed and nonexposed groups of patients, each a cohort, and then follow them forward in time, monitoring the occurrence of outcomes of interest in an attempt to identify whether there is an association between the exposure and the outcomes. The cohort design is similar to an RCT but without randomization; rather, the determination of whether a patient received the exposure of interest results from the patient's or investigator's preference or from happenstance.

Case-control studies also assess associations between exposures and outcomes. Rare outcomes or those that take a long time to develop can threaten the feasibility not only of RCTs but also of cohort studies. The case-control study provides an alternative design that relies on the initial identification of cases—that is, patients who have already developed the target outcome—and the selection of controls—persons who do not have the outcome of interest. Using case-control designs, investigators assess the relative frequency of previous exposure to the putative harmful agent in the cases and the controls.

For example, in addressing the impact of nonsteroidal anti-inflammatory drugs (NSAIDs) on clinically apparent upper gastrointestinal tract hemorrhage, investigators needed a cohort study to deal with the problem of infrequent events. Bleeding among those taking NSAIDs has been reported to occur approximately 1.5 times per 1000 person-years of exposure, in comparison with 1.0 per 1000 person-years in those not taking NSAIDs.<sup>9</sup> Because the event rate in unexposed patients is so low (0.1%), an RCT to study an increase in risk of 50% would require huge numbers of patients (sample size calculations suggested approximately 75000 patients per group) for adequate power to test the hypothesis that NSAIDs cause the additional bleeding.<sup>10</sup> Such an RCT would not be feasible, but a cohort study, in which the information comes from a large administrative database, would be possible.

## Cohort Studies

Cohort studies may be prospective or retrospective. In prospective cohort studies, the investigator enrolls patients or participants, starts the follow-up, and waits for the outcomes (events of interest) to occur. Such studies may take many years to complete, and thus they are difficult to conduct. An advantage, however, is that the investigators can plan how to monitor patients and collect data.

In retrospective cohort studies, the data regarding both exposures and outcomes have been previously collected; the investigator obtains the data and determines whether participants with and without the outcome of interest have been exposed to the putative causal agent or agents. These studies are easier to perform because they depend on the availability of data on exposures and outcomes that have already happened. On the other hand, the investigator has less control over the quality and relevance of the available data. In the end, clinicians need not pay too much attention to whether studies are prospective or retrospective but should instead focus on the risk of bias criteria in [Box 14-1](#).

### In a Cohort Study, Aside From the Exposure of Interest, Did the Exposed and Control Groups Start and Finish With the Same Risk for the Outcome?

#### Were Patients Similar for Prognostic Factors That Are Known to Be Associated With the Outcome (Or Did Statistical Adjustment Level Address This Imbalance)?

Cohort studies will yield biased results if the group exposed to the putative harmful agent and the unexposed group begin with additional differences in baseline characteristics that give them a different **prognosis** (ie, a different risk of the target outcome) and if the analysis fails to deal with this imbalance. For instance, in the association between NSAIDs and the increased risk of upper gastrointestinal tract bleeding, age may be associated with exposure to NSAIDs and gastrointestinal bleeding. In other words, because patients taking NSAIDs will be older and because older patients are more likely to bleed, this variable makes attribution of an increased risk of bleeding to NSAID exposure problematic. When a variable with prognostic power differs in frequency in the exposed and unexposed cohorts, we refer to the situation as *confounding*.

There is no reason that patients who self-select (or who are selected by their physicians) for exposure to a potentially harmful agent should be similar to the nonexposed patients with respect to important determinants of the harmful outcome. Indeed, there are many reasons to expect they will not be similar. Physicians are appropriately reluctant to prescribe medications they perceive will put their patients at risk.

In one study, 24.1% of patients who were given a then-new NSAID, ketoprofen, had received peptic ulcer therapy during the previous 2 years compared with 15.7% of the control population.<sup>11</sup> 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 comparison of ketoprofen to other agents would be subject to the risk of finding a spurious increase in bleeding with the new agent (compared with other therapies) because higher-risk patients would have been receiving the ketoprofen. This bias may be referred to as a selection bias or a bias due to confounding by indication.

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, sometimes referred to as the *channeling bias*.<sup>12</sup> Ray et al<sup>13</sup> found an association between long-acting benzodiazepines and risk of falls (*relative risk* [RR], 2.0; 95% *confidence interval* [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*) selectively received these benzodiazepines during the earlier period. Reports of associations between benzodiazepine use and falls led to greater caution, and the apparent association disappeared when physicians began to avoid using benzodiazepines in those at high risk of falling.

Therefore, investigators must document the characteristics of the exposed and nonexposed participants and either demonstrate their comparability (very unusual in cohort studies) or use statistical techniques to adjust for these differences. Effective adjusted analyses for prognostic factors require the accurate measurement of those prognostic factors. For prospective cohorts, the investigators may take particular care of the quality of this information. For retrospective databases, however, one has to make use of what is available. Large administrative databases, although providing a sample size that may allow ascertainment of rare events, often have limited quality of data concerning relevant patient characteristics, health care encounters, or diagnoses. For example, in a cross-sectional study designed to measure the accuracy of electronic reporting of care practices compared with manual review, electronic reporting significantly underestimated rates of appropriate asthma medication and pneumococcal vaccination and overestimated rates of cholesterol control in patients with diabetes.<sup>14</sup>

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. We call this *residual confounding*.

Returning to our earlier example, it may be that the illnesses that require NSAIDs, rather than the NSAIDs themselves, contribute to 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.

### Were the Circumstances and Methods for Detecting the Outcome Similar?

In cohort studies, ascertainment of outcome is the key issue. For example, investigators have reported a 3-fold increase in the risk of malignant melanoma in individuals who work with radioactive materials. One possible explanation for some of the increased risk might be that physicians, concerned about a possible risk, search more diligently and therefore detect disease that might otherwise go unnoticed (or they may detect disease at an earlier point). This could result in the exposed cohort having an apparent, but spurious, increase in risk—a situation known as *surveillance bias*.<sup>15</sup>

The choice of outcome may partially address this problem. In one cohort 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 study that assessed all of the men who had been members of the printers' unions in Oslo, Norway.<sup>16</sup> The investigators used job classification to categorize the fathers as being exposed to lead and organic solvents or not exposed to those substances. Investigators' awareness of whether the fathers had been exposed to the lead or solvents might bias their assessment of the infant's outcome for minor birth defects or defects that required special investigative procedures. On the other hand, an outcome such as preterm birth would be unlikely to increase simply as a result of *detection bias* (the tendency to look more carefully for an outcome in one of the comparison groups) because prior knowledge of exposure is unlikely to influence whether an infant is considered preterm or not. The study found that exposure was associated with an 8-fold increase in preterm births but no increase in birth defects, so detection bias was not an issue for the results that were obtained in this study.

### Was the Follow-up Sufficiently Complete?

As we pointed out in [Chapter 7](#), Therapy (Random-ized Trials), loss to follow-up can introduce bias because the patients who are lost may have different outcomes from those patients still available for assessment. This is particularly problematic if there are differences in follow-up between the exposed and nonexposed groups.

For example, in a well-executed study,<sup>17</sup> investigators determined the vital status of 1235 of 1261 white men (98%) employed in a chrysotile asbestos textile operation between 1940 and 1975. The RR 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. In this study, the 2% missing data were unlikely to affect the results, and the loss to follow-up did not threaten the strength of the inference that asbestos exposure caused lung cancer deaths.

## Case-Control Studies

Case-control studies are always retrospective in design. The outcomes (events of interest) have already happened and participants are designated to 1 of 2 groups: those with the outcomes (cases) and those where the outcome is absent (controls). Retrospectively, investigators ascertain prior exposure to putative causal agents. This design entails inherent risks of bias because exposure data require memory and recall or are based on a collection of data that were originally accumulated for purposes other than the intended study.

### In a Case-Control Study, Did the Cases and Control Group Have the Same Risk (Chance) for Being Exposed in the Past?

#### Were Cases and Controls Similar With Respect to the Indication or Circumstances That Would Lead to Exposure (or Did Matching or Statistical Adjustment Address the Imbalance)?

As with cohort studies, case-control studies are susceptible to unmeasured confounding. For instance, in looking at the association between use of  $\beta$ -agonists and mortality among patients with asthma, investigators need to consider—and match or adjust for—previous hospitalization and use of other medications to avoid confounding by disease severity. Patients who use more  $\beta$ -agonists may have more severe asthma, and this severity, rather than  $\beta$ -agonist use, may be responsible for increased mortality. As in cohort studies, however, matching and adjustment cannot eliminate the risk of bias, particularly when exposure varies over time. In other words, matching or adjustment for hospitalization or use of other medications may not adequately capture all of the variability in underlying disease severity in asthma. In addition, the adverse lifestyle behaviors of patients with asthma who use large amounts of  $\beta$ -agonists could be the real explanation for the association.

To further illustrate the concern about unmeasured confounding, consider the example of a case-control study that was designed to assess the association between diethylstilbestrol ingestion by pregnant women and the development of vaginal adenocarcinomas in their daughters many years later.<sup>18</sup> 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. Furthermore, given the infrequency of the disease, an RCT or a cohort study would have required hundreds of thousands of participants. By contrast, using the case-control strategy, the investigators delineated 2 relatively small groups of young women. Those who had 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, the investigators determined exposure rates to diethylstilbestrol for the 2 groups. They found a significant association between in utero diethylstilbestrol exposure and vaginal adenocarcinoma, and they found their answer without a delay of 20 years and by studying only 40 women.

An important consideration in this study would be whether the cases could have been exposed to diethylstilbestrol in any special circumstances that would not have affected women in the control group. In this situation, diethylstilbestrol had been prescribed to women at risk for miscarriages or premature births. Could either of these indications be a confounder? Before the introduction of diethylstilbestrol, vaginal adenocarcinoma in young women was uncommon, but miscarriages and premature birth were common. Thus, it would be unlikely that miscarriages and premature births were directly associated with vaginal adenocarcinoma, and in the absence of such an association, neither could be a confounder.

In another study, investigators used a case-control design relying on computer-record linkages between health insurance data and a drug insurance plan to investigate the possible association between use of  $\beta$ -adrenergic agonists and mortality rates in patients with asthma.<sup>19</sup> The database for the study included 95% of the population of the province of Saskatchewan, Canada. The investigators selected 129 patients who had experienced a fatal or near-fatal asthma attack to serve as cases and used a matching process to select another 655 patients who also had asthma but who had not had a fatal or near-fatal asthma attack to serve as controls.

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 before death (for the cases) or before the index date of entry into the study (for the 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 agonists through metered-dose inhalers and death from asthma (*odds ratio* [OR], 2.6 per canister of inhaler per month; 95% CI, 1.7–3.9), even after correcting for measures of disease severity.

### Were the Circumstances and Methods for Determining Exposure Similar for Cases and Controls?

In case-control studies, ascertainment of the exposure is a key issue. However, if case patients have a better memory for exposure than control patients, the result will be a spurious association.

For example, a case-control study found a 2-fold increase in risk of hip fracture associated with psychotropic drug use.<sup>20</sup> In this study, investigators established drug exposure by examining computerized claim files from the Michigan Medicaid program, a strategy that avoided selective memory of exposure—*recall bias*—and differential probing of cases and controls by an interviewer—*interviewer bias*.

Another example was a case-control study that evaluated whether the use of cell phones was associated with an increased risk of motor vehicle crash.<sup>21</sup> Suppose the investigators had tried to ask people who had a motor vehicle crash and control patients (who were in no crash at the same day and time) whether they were using their cell phone around the time of interest. People who were in a crash would have been more likely to recall such use because their memory might be heightened by the unfortunate circumstances. This would have led to a spurious association because of differential recall. Alternatively, they might specifically deny the use of a cell phone because of embarrassment or legal concerns, thus obscuring an association. Therefore, the investigators in this study used a computerized database of cell phone use instead of patient recall.<sup>21</sup> Moreover, the investigators used each person in a crash as his or her own control. The time of the crash was matched against corresponding times of the life of the same person when they were driving but when no crash occurred (eg, same time driving to work). This appropriate design established that use of cell phones was associated with an increased risk of having a motor vehicle crash.

Not all studies have access to unbiased information on exposure. For instance, in a case-control study of the association between coffee and pancreatic cancer, the patients with cancer may be more motivated to identify possible explanations for their problem and provide a greater recounting of coffee use.<sup>22</sup> Also, if the interviewers are not blinded to whether a patient is a case or a control patient, the interviewer may probe deeper for exposure information from cases. In this particular study, there were no objective sources of data regarding exposure. Recall or interviewer bias might have explained the apparent association.

As it happened, another bias provided an even more likely explanation for what turned out to be a spurious association. The investigators chose control patients from the practices of the physicians treating the patients with pancreatic cancer. 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, which explains the investigators' finding of an association between coffee (which the patients with pancreatic cancer consumed at general population levels) and pancreatic cancer. Subsequent investigations, using more appropriate controls, refuted this association.<sup>23</sup>

In addition to a biased assessment of exposure, *random error* in exposure ascertainment is also possible. In random misclassification, exposed and unexposed patients are misclassified, but the rates of misclassification are similar in cases and controls. Such nondifferential misclassification dilutes any association (ie, the true association will be larger than the observed association). Fortunately, unless the misclassification is extremely large, the reduction in the true association will not be important.

## What Is the Risk of Bias in Cross-sectional Studies?

Like the cohort and the case-control study, the *cross-sectional study* is also an observational study design. Like a cohort study, a cross-sectional study is based on an assembled population of exposed and unexposed participants. However, in the cross-sectional study, the exposure and the existing or prevalent outcome are measured at the same point in time. Accordingly, the direction of association may be difficult to determine. Another important limitation is that the outcome or the threat of experiencing an adverse outcome may have led patients assigned as cases to leave the study, so a measure of association may be biased against the association. However, cross-sectional studies are relatively inexpensive and quick to conduct and may be useful in generating and exploring hypotheses that will be subsequently investigated using other observational designs or RCTs.

## What Is the Risk of Bias in 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 groups, so it is impossible to determine whether the observed outcome would likely have occurred in the absence of the exposure. Although descriptive studies have been reputed to have significant findings mandating an immediate change in clinician behavior, this is rarely justified, and without availability of evidence from stronger study designs, there are potentially undesirable consequences when actions are taken in response to evidence warranting very low confidence. Recall the consequences of case reports of specific birth defects occurring in association with thalidomide exposure.<sup>24</sup>



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 consequence of case reports suggesting that it was teratogenic.<sup>25</sup> Later, although a number of comparative studies reported the drug's relative safety,<sup>26</sup> they could not eradicate the prevailing litigious atmosphere—which prevented the manufacturer from reintroducing Bendectin. Thus, many pregnant women who might have benefited from the drug's availability were denied the symptomatic relief it could have offered.

For some interventions, registries of adverse events may provide the best possible initial evidence. For example, there are vaccine registries that record adverse events among people who have received the vaccine. These registries may signal problems with a particular adverse event that would be very difficult to capture from prospective studies limited by sample sizes that were too small. Even retrospective studies might be too difficult to conduct if people who receive the vaccine differ substantially from those who do not and if adjustment or matching cannot deal with the differences. In this situation, investigators might conduct a *before-after study* using the general population before the introduction of the new vaccine occurred. Such comparisons using historical controls are, however, prone to bias because many other factors may have changed in the same period. If, however, changes in the incidence of an adverse event are very large, the signal may be real. An example is the clustering of *intussusception* cases among children receiving a particular type of rotavirus vaccine,<sup>27</sup> resulting in a decision to withdraw the vaccine. The association was subsequently supported by a case-control study.<sup>28</sup> Eventually, another type of rotavirus vaccine was developed that did not cause this adverse event.

In general, clinicians should not draw conclusions about relationships from case series, but rather, they should recognize that the results may generate questions, or even hypotheses, that clinical investigators can address with studies that have optimal safeguards against risk of bias. When the immediate risk of exposure outweighs the benefits (and outweighs the risk of stopping an exposure), the clinician may have to make a management decision with less than optimal data.

### How Serious Is the Risk of Bias: Summary

Just as it is true for the resolution of questions of therapeutic effectiveness, clinicians should first look to RCTs to resolve issues of harm. They will often be disappointed in the search and must make use of studies of weaker design. Regardless of the design, however, they should look for an appropriate control population. For cohort studies, the control group should have a similar baseline risk of outcome, or investigators should have used statistical techniques to adjust for differences. In case-control studies, the cases and the controls should have had a similar opportunity to have been exposed, so that if a difference in exposure is observed, one might legitimately conclude that the association could be due to a link between the exposure and the outcome and not due to a confounding factor. Nevertheless, investigators should routinely use statistical techniques to match cases and controls or adjust for differences.

Even when investigators have taken all of the appropriate steps to minimize bias, clinicians should bear in mind that residual differences between groups still may bias the results of observational studies. Because evidence, clinician preferences, and patient *values and preferences* determine the use of interventions in the real world, exposed and unexposed patients are likely to differ in prognostic factors.

## USING THE GUIDE

Returning to our earlier discussion, the study that we retrieved investigating the association between soy milk (or soy formula) and the development of peanut allergy used a case–control design.<sup>1</sup> Those with peanut allergy (cases) appeared to be similar to controls with respect to the indication or circumstances leading to soy exposure, but there were a few potentially important imbalances. In the peanut allergy group, a family history of peanut allergy and an older sibling with a history of milk intolerance were more common and could have biased the likelihood of a subsequent child's being exposed to soy. To avoid confounding, the investigators conducted an *adjusted analysis*.

The methods for determining exposure were similar for cases and controls because the data were collected with the interviewers and parents unaware of the hypothesis that related soy exposure to peanut allergy (thus avoiding interviewer bias and perhaps recall bias). With regard to access to soy, all of the children came from the same geographic region, although this does not ensure that cultural and economic factors that might determine soy access were similar in cases and controls. Overall, protection from risk of bias seemed adequate.

## What are the Results?

### How Strong Is the Association Between Exposure and Outcome?

We describe options that can be used for expressing an association between an exposure and an outcome—the *risk ratio* or RR and the OR—in other chapters of this book (see [Chapter 9](#), Does Treatment Lower Risk? Understanding the Results, and [Chapter 12.2](#), Understanding the Results: More About Odds Ratios).

For example, in a cohort study that assessed in-hospital mortality after noncardiac surgery in male veterans, 23 of 289 patients with a history of hypertension died compared with 3 of 185 patients without the condition.<sup>29</sup> The RR for mortality in hypertensive patients compared with normotensive patients (23/289 and 3/185, respectively) was 4.9 (95% CI, 1.5–16.1). The RR tells us that death after noncardiac surgery occurs almost 5 times more often in patients with hypertension than in normotensive patients.

The estimate of the RR 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 RR is 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, the RR, we use a ratio of odds, the OR, specifically the odds of a case patient being exposed divided by the odds of a control patient being exposed. Unless the risk of the outcome in the relevant population is high (20% or more), you can think of the OR as providing a good estimate of the much easier to conceptualize RR.

### How Precise Is the Estimate of the Risk?

Clinicians can evaluate the precision of the estimate of risk by examining the CI around that estimate (see [Chapter 10](#), Confidence Intervals: Was the Single Study or Meta-analysis Large Enough?). In a study in which investigators have found an association between an exposure and an adverse outcome, the lower limit of the estimate of RR associated with the adverse exposure provides an estimate of the lowest possible magnitude of the association. Alternatively, in a negative study (in which the results are not *statistically significant*), the upper boundary of the CI around the RR tells the clinician just how big an adverse effect still may be present, despite the failure to find a statistically significant association.

## USING THE GUIDE

The investigators calculated an OR of 2.6 (95% CI, 1.3–5.2) for the risk of peanut allergy in those exposed to soy vs those not exposed.<sup>1</sup> These results were adjusted for skin manifestations of allergy (ie, atopy). The consumption of soy by the infants was independently associated with peanut allergy and could not be explained as a dietary response to other atopic conditions. It nevertheless remains possible that the association with soy was confounded by other, unknown factors. Unfortunately, the investigators did not evaluate the possibility of a dose-response relationship for soy exposure and the development of peanut allergy.

## 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 should consider the extent to which the results might apply to the patient in your practice. Would your patient have met the eligibility criteria? Is your patient similar to those described in the study with respect to potentially important factors, such as patient characteristics or medical history? If not, is the biology of the harmful exposure likely to be different for the patient for whom you are providing care?

### Was Follow-up Sufficiently Long?

Studies can be pristine in terms of avoiding bias but of limited use if patients are not followed up for a sufficiently long period. That is, they may provide an unbiased estimate of the effect of an exposure during the short term, but the time frame in which we are interested is a substantially longer period. For example, most cancers take a decade or longer to develop from the original assault at the biologic level to the clinically detected malignant tumor. If the question is whether a specific exposure, say to an industrial chemical, is related to a subsequent cancer, one would not expect cancers detected in the first few years to reflect any of the effect of the exposure under question.

### Is the Exposure Similar to What Might Occur in My Patient?

Clinicians should ask whether there are important differences in the exposures of a study (eg, dose and duration) between their patient and the patients in the study. For example, it should be clear that the risk of thrombophlebitis associated with oral contraceptive use described in the 1970s may not be applicable to the patient in the 21st century because of the lower estrogen dose in oral contraceptives currently used. Another example of questionable applicability comes from a study that found that workers employed in a chrysotile asbestos textile operation between 1940 and 1975 had an increased risk of 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.<sup>17</sup> The study does not provide trustworthy information regarding what might be the risks associated with only brief or intermittent exposure to asbestos (eg, a person working for a few months in an office located in a building subsequently found to have abnormally high asbestos levels).

### What Is the Incremental Risk?

The RR and OR 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 vs the unexposed group. Even when we observe a large and statistically significant relative difference between the 2 groups, the results still may not be important if the adverse event is rare. Thus, we need a method for assessing the absolute impact of the exposure. In our discussion of therapy (see [Chapter 7](#), Therapy [Randomized Trials], and [Chapter 9](#), Does Treatment Lower Risk? Understanding the Results), we describe how to calculate the *risk difference* and the number of patients whom clinicians must treat to prevent an adverse event (*number needed to treat*). When the issue is harm, we can use data from a randomized trial or cohort study (but not a case-control study) in a similar way to calculate the number of patients who would have to be exposed to result in an additional harmful event. However, this calculation requires knowledge of the *absolute risk* in unexposed individuals in our population.

For example, during a mean of 10 months of follow-up, investigators conducting the Cardiac **Arrhythmia** Suppression Trial, an RCT of antiarrhythmic agents, 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 ( $100/4.7$ ) tells us that, on average, for every 21 patients treated with encainide or flecainide for approximately a year, there would be 1 excess death.

This contrasts with our example of the association between NSAIDs and upper gastrointestinal tract bleeding. Of 2000 unexposed patients in that study, 2 will have a bleeding episode each year. Of 2000 patients taking NSAIDs, 3 will have such an episode each year. Thus, if we treat 2000 patients with NSAIDs, we can expect a single additional bleeding event.<sup>9</sup>

## Are There Any Benefits That Offset the Risks Associated With Exposure?

Even after evaluating the evidence that an exposure is harmful and establishing that the results are potentially applicable to the patient in your practice, determining subsequent actions may not be simple. In addition to considering the magnitude of the risk, one must consider the adverse consequences of reducing or eliminating exposure to the harmful agent (ie, the magnitude of any potential benefit that patients will no longer receive).

Clinical decision making is simple when harmful consequences are unacceptable and benefit is absent. For example, because the evidence of increased mortality from encainide and flecainide came from an RCT with low risk of bias,<sup>30</sup> we can be at least moderately confident of a substantial increase in risk of death. Because treating only 21 people would 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. Even if the evidence warrants low confidence, the availability of an alternative substance can result in a clear decision.

## CLINICAL SCENARIO RESOLUTION

You determine that the patient's unborn child, once he or she reaches early childhood, would likely fulfill the eligibility criteria in the study. Also relevant to the clinical scenario, but perhaps unknown, is whether the soy products discussed in the study are similar to the ones that the patient is considering using. With regard to the magnitude of risk, the prevalence of peanut allergy is approximately 4 per 1000 children. An approximate calculation would suggest that, if the OR with exposure to soy is truly 2.6, 10 children per 1000 would be affected by peanut allergy, an additional 6 children in every 1000. In other words, the number of children needed to be exposed to soy to result in 1 additional child having peanut allergy is 167 (1000/6). Finally, there are no data regarding the negative consequences of withholding soy formula or soy milk products, and the use of these products would clearly be dependent on how severe and sustained an intolerance to cow's milk was in a particular child.

To decide on your course of action, you proceed through 3 steps of using the medical literature to guide your clinical practice. First, you consider the risk of bias in the study before you. Adjustments of known confounders did not diminish the association between soy exposure as a neonate and the development of peanut allergy. Also, the design of the study provides adequate safeguards against recall or interviewer bias. You conclude that, with the obvious limitations of the observational design (generally only warranting low confidence in estimates of effect), the study is at low risk of bias.

Turning to the results, you note a moderate association between soy exposure and the development of peanut allergy (moderate typically being considered ORs greater than 2 and less than 5) that is strong enough, despite the limitations of the observational design, that it leaves you moderately confident of an association between exposure to soy and peanut allergy. The lower boundary of the CI (1.3) and the uncertainty around the baseline risk estimate of 4 per 10000 children lead you to conclude that you have only low confidence in your estimate of the incremental harm of peanut allergy of 6 in 1000.

You proceed to the third step, and consider the implications of the study results for your patient. The study would appear to apply to a future child of your patient. Although the best estimate of the absolute increase in risk is only 6 in 1000, and you have only low confidence in this estimate, the consequences of peanut allergy can be a serious health threat to a patient and quite disruptive for a family because of the required precautions and food restrictions. You discuss the situation with the mother, who elects to start feedings with milk products. Together you agree that, given the limited confidence in estimates and the small absolute risk, should the child appear to have distressing milk allergy, she will probably switch to soy.

## References

1. Lack G, Fox D, Northstone K, Golding J; Avon Longitudinal Study of Parents and Children Study Team. Factors associated with the development of peanut allergy in childhood. *N Engl J Med*. 2003;348(11):977--985. [[PubMed: 12637607](#)]
2. CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet*. 1996;348(9038):1329--1339. [[PubMed: 8918275](#)]
3. Bennett CL, Connors JM, Carwile JM, et al. Thrombotic thrombocytopenic purpura associated with clopidogrel. *N Engl J Med*. 2000;342(24):1773--1777. [[PubMed: 10852999](#)]
4. 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(4):241--249. [[PubMed: 7611589](#)]
5. Bombardier C, Laine L, Reicin A, et al; VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med*. 2000;343(21):1520--1528. [[PubMed: 11087881](#)]
6. Langman MJ, Jensen DM, Watson DJ, et al. Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. *JAMA*. 1999;282(20):1929--1933. [[PubMed: 10580458](#)]

7. Papanikolaou PN, Ioannidis JP. Availability of large-scale evidence on specific harms from systematic reviews of randomized trials. *Am J Med.* 2004;117(8):582--589. [PubMed: 15465507]
8. Geenen MM, Cardous-Ubbink MC, Kremer LC, et al. Medical assessment of adverse health outcomes in long-term survivors of childhood cancer. *JAMA.* 2007;297(24):2705--2715. [PubMed: 17595271]
9. Carson JL, Strom BL, Soper KA, West SL, Morse ML. The association of nonsteroidal anti-inflammatory drugs with upper gastrointestinal tract bleeding. *Arch Intern Med.* 1987;147(1):85--88. [PubMed: 3492182]
10. 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(4):387--397. [PubMed: 848486]
11. 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(4):428--432. [PubMed: 1431721]
12. Joseph KS. The evolution of clinical practice and time trends in drug effects. *J Clin Epidemiol.* 1994;47(6):593--598. [PubMed: 7722572]
13. Ray WA, Griffin MR, Downey W. Benzodiazepines of long and short elimination half-life and the risk of hip fracture. *JAMA.* 1989;262(23):3303--3307. [PubMed: 2573741]
14. Kern LM, Malhotra S, Barrón Y, et al. Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study. *Ann Intern Med.* 2013;158(2):77--83. [PubMed: 23318309]
15. Hiatt RA, Fireman B. The possible effect of increased surveillance on the incidence of malignant melanoma. *Prev Med.* 1986;15(6):652--660. [PubMed: 3797396]
16. 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(2):134--144. [PubMed: 8452117]
17. Dement JM, Harris RL Sr, Symons MJ, Shy CM. Exposures and mortality among chrysotile asbestos workers. Part II: mortality. *Am J Ind Med.* 1983;4(3):421--433. [PubMed: 6846339]
18. 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(15):878--881. [PubMed: 5549830]
19. 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(8):501--506. [PubMed: 1346340]
20. Ray WA, Griffin MR, Schaffner W, Baugh DK, Melton LJ III. Psychotropic drug use and the risk of hip fracture. *N Engl J Med.* 1987;316(7):363--369. [PubMed: 2880292]
21. Redelmeier DA, Tibshirani RJ. Association between cellular-telephone calls and motor vehicle collisions. *N Engl J Med.* 1997;336(7):453--458. [PubMed: 9017937]
22. MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. Coffee and cancer of the pancreas. *N Engl J Med.* 1981;304(11):630--633. [PubMed: 7453739]
23. Baghurst PA, McMichael AJ, Slavotinek AH, Baghurst KI, Boyle P, Walker A'8M. A case-control study of diet and cancer of the pancreas. *Am J Epidemiol.* 1991;134(2):167--179. [PubMed: 1862800]
24. Lenz W. Epidemiology of congenital malformations. *Ann N Y Acad Sci.* 1965;123:228--236. [PubMed: 14329207]

- 
25. Soverchia G, Perri PF. [2 cases of malformations of a limb in infants of mothers treated with an antiemetic in a very early phase of pregnancy]. *Pediatr Med Chir*. 1981;3(1):97--99. [[PubMed: 7301625](#)]
- 
26. Holmes LB. Teratogen update: bendectin. *Teratology*. 1983;27(2):277--281. [[PubMed: 6346560](#)]
- 
27. Centers for Disease Control and Prevention (CDC). [Intussusception](#) among recipients of rotavirus vaccine-United States, 1998-1999. *MMWR Morb Mortal Wkly Rep*. 1999;48(27):577--581. [[PubMed: 10428095](#)]
- 
28. Murphy TV, Gargiullo PM, Massoudi MS, et al; Rotavirus [Intussusception](#) Investigation Team. [Intussusception](#) among infants given an oral rotavirus vaccine. *N Engl J Med*. 2001;344(8):564--572. [[PubMed: 11207352](#)]
- 
29. Browner WS, Li J, Mangano DT; The Study of Perioperative Ischemia Research Group. In-hospital and long-term mortality in male veterans following noncardiac surgery. *JAMA*. 1992;268(2):228--232. [[PubMed: 1608142](#)]
- 
30. 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(12):781--788. [[PubMed: 1900101](#)]
-