HARM (OBSERVATIONAL STUDIES)

Mitchell Levine, John Ioannidis, Ted Haines, and Gordon Guyatt

IN THIS CHAPTER:

Clinical Scenario

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

Finding the Evidence

Are the Results Valid?

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?

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

Cross-sectional Studies

Case Series and Case Reports

Design Issues: Summary

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 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 Offset the Risks Associated With Exposure?

Clinical Resolution

FINDING THE EVIDENCE

You formulate the relevant question: In infants, what is the association between exposure to soy milk and the subsequent development of peanut allergy? Searching Ovid (MEDLINE) with the terms “peanut” AND “soy” AND “allergy” AND “risk,” you identify 12 articles. One article appears to be particularly relevant to your target: factors associated with the development of peanut allergy in childhood. You print a copy of the abstract and then arrange to obtain a copy of the full-text article from your local hospital library.

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 blinded peanut challenge. 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).

Table 12-1 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 involving concerns of etiology or risk factors in which a potentially harmful exposure cannot be randomly assigned. These observational studies involve using either cohort or case-control designs.
Clinicians often encounter patients who face potentially harmful exposures either to medical interventions or environmental agents. These circumstances give rise to important questions. 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 changes in health care policies lead to harmful outcomes? When examining these questions, health care providers and administrators 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 domain.

In answering any clinical question, our first goal should be to identify any existing systematic review of the topic that can provide a summary of the highest-quality available evidence (see Chapter 19, Summarizing the Evidence). Interpreting such a review requires an understanding of the rules of evidence for individual or primary studies, randomized controlled trials (RCTs), and observational studies. The tests for judging the validity of observational study results will help you decide whether exposed and control groups (or cases and controls) began and finished the study with sufficient similarities that we obtain a minimally biased assessment of the influence of exposure on outcome (see Chapter 5, Why Study Results Mislead: Bias and Random Error).

### TABLE 12-1

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

Are the results valid?

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 the playing field)?
- 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 (chance) for being exposed in the past?
- Were cases and controls similar with respect to the indication or circumstances that would lead to exposure?
- 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 result 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?

**Are the Results Valid?**

Clinicians often encounter patients who face potentially harmful exposures either to medical interventions or environmental agents. These circumstances give rise to important questions. 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 changes in health care policies lead to harmful outcomes? When examining these questions, health care providers and administrators 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 domain.

In answering any clinical question, our first goal should be to identify any existing systematic review of the topic that can provide a summary of the highest-quality available evidence (see Chapter 19, Summarizing the Evidence). Interpreting such a review requires an understanding of the rules of evidence for individual or primary studies, randomized controlled trials (RCTs), and observational studies. The tests for judging the validity of observational study results will help you decide whether exposed and control groups (or cases and controls) began and finished the study with sufficient similarities that we obtain a minimally biased assessment of the influence of exposure on outcome (see Chapter 5, Why Study Results Mislead: Bias and Random Error).
RCTs 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 the outcome (see Chapter 6, Therapy). Although investigators conduct RCTs to determine whether therapeutic agents are beneficial, they should also look for harmful effects and may sometimes make surprising discoveries about the negative effects of the intervention on their primary outcomes (see Chapter 9.2, Surprising Results of Randomized Trials).

There are 3 reasons why RCTs may not be helpful for determining whether a putative harmful agent truly has deleterious effects. First, we would consider it unethical to randomize patients to exposures that we anticipate might result in harmful effects without benefit. 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. Even a very large RCT failed to detect an association between clopidogrel and thrombotic thrombocytopenic purpura, which appeared in a subsequent observational study. RCTs specifically addressing adverse effects may be feasible for adverse event rates as low as 1%. But the RCTs that we need to explore harmful events occurring in less than 1 in 100 exposed patients are logistically difficult and often prohibitively expensive because of huge sample size and lengthy follow-up. Meta-analyses may be very helpful when the event rates are very low. Across almost 2000 systematic reviews, however, only 25 reviews had large-scale data on 4000 or more randomized subjects regarding well-defined harms that might be associated with the assessed interventions. Third, RCTs often fail to adequately report 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, which we will now describe (Table 12-2).

| TABLE 12-2 |
| Directions of Inquiry and Key Methodologic Strengths and Weaknesses for Different Study Designs |
| Design | Starting Point | Assessment | Strengths | Weaknesses |
| Randomized controlled 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, 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. 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 or physician’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 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.

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 the Playing Field)?

In a cohort study, the investigator identifies exposed and nonexposed groups of patients, each a cohort, and then traces their outcomes forward in time. Cohort studies may be either prospective or retrospective. In prospective studies, the investigator starts the follow-up and waits for the outcome (events of interest) to occur. Such studies may take many years to complete and thus they are difficult to conduct. On the other hand, an advantage is that the investigator may have a better idea of how patients are to be monitored and data are to be collected. In retrospective studies, the outcomes (events of interest) have already happened at some point in the past; the investigator simply goes back even farther in the past and selects exposed and unexposed people; then the question is whether these differ in the development of these outcomes of interest. 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 for the research question being addressed.

Cohort studies of potentially harmful exposures will yield biased results if the group exposed to the putative harmful agent and the unexposed group begin with different baseline characteristics that give them a different prognosis (and the analysis fails to deal with this imbalance). Investigators rely on cohort designs when exposure has little or no possible benefit and possible harm (making randomization unethical) or when harmful outcomes occur infrequently.

In an example of the latter situation, clinically apparent upper gastrointestinal hemorrhage in nonsteroidal anti-inflammatory drug (NSAID) users 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.9 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 suggest about 75000 patients per group) for adequate power to test the hypothesis that NSAIDs cause the additional bleeding. Such an RCT would not be feasible, but a cohort study, in which the information comes from a large administrative database, would be possible.

One 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 with both 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 confounding variable makes attribution of an increased risk of bleeding to NSAID exposure problematic.

There is no reason that patients who self-select (or who are selected by their physician) for exposure to a potentially harmful agent should be similar to the nonexposed patients with respect to other important determinants of that outcome. 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 can 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 with 15.7% of the control population. 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.

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. Ray et al 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 or use statistical techniques to create a level playing field by adjusting for 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 allows ascertainment of rare events, sometimes have limited quality of data concerning relevant patient characteristics.

For example, Jollis et al.\(^\text{14}\) 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 a high degree of chance-corrected agreement between the fellow and the administrative database for the presence of diabetes: the \(\kappa\), a measure of chance-corrected agreement, was 0.83 (see Chapter 17.3, 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 (\(\kappa, 0.19\)).

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, for instance, it may be that the illnesses that require NSAIDs, rather than the NSAIDs themselves, can 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 RCTs and 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 working 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 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.\(^\text{15}\)

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 assessing all the men who had been members of the printers’ unions in Oslo.\(^\text{16}\) The investigators used job classification to categorize the fathers as either 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 baby’s outcome for minor birth defects or for defects that required
special investigative procedures. On the other hand, the outcome of preterm birth would be less susceptible to a detection bias. In the study, exposure was associated with an 8-fold increase in preterm births, but it was not linked with birth defects, so detection bias was unlikely.

Was the Follow-up Sufficiently Complete?
As we pointed out in Chapter 6, Therapy, 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.

In a well-executed study, 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, where exposure was on a continuum (ie, not dichotomous), the 2% missing data were unlikely to affect the results, and the loss to follow-up did not threaten the validity of the inference that asbestos exposure caused lung cancer deaths.

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?

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. 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 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, they determined exposure rates to DES for the 2 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 < .001) They found their answer without a delay of 20 years and by studying only 40 women.

A critical issue in that study would be whether the cases would have had any other special circumstances to be exposed to DES that controls would not. In this situation, DES had been prescribed to woman at risk for miscarriages or having premature births. It would be important in the assessment of this study
to be confident that those risk factors on their own could not account for the subsequent high rate of vaginal pathology in the female offspring.

In another study, 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 β-adrenergic agonists and mortality rates in patients with asthma. The database for the study included 95% of the population of the province of Saskatchewan in western Canada. The investigators used matching to choose 129 cases of fatal or near-fatal asthma attack with 655 controls that also had asthma but who had not had a fatal or near-fatal asthma attack.

The tendency of patients with more severe asthma to use more β-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 β-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 β-agonists could be the real explanation for the association.

**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. 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. In this study, investigators established drug exposure by examining computerized claims files of 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 is a study that evaluated whether the use of cellular phones increases the risk of motor vehicle crash. 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 cellular 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 relationship because of differential recall. Therefore, the investigators in this study instead
used a computerized database of cellular phone use. Moreover, they 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 cellular phones increases the risk of having a motor vehicle crash.\textsuperscript{21}

Not all studies have access to unbiased information on exposure. In a case-control study looking at 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.\textsuperscript{22} 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 may explain the apparent association.

As it turns out, another bias provides 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 looking after 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 pancreatic cancer patients consumed at general population levels) and pancreatic cancer. Subsequent investigations, using more appropriate controls, refuted the association.\textsuperscript{23}

The examples above relate to the biased assessment of exposure, but the inaccurate assessment of exposure may also be random. In other words, lots of exposed persons get classified as unexposed, and vice versa, but the rates of misclassification are similar in cases and controls. Such nondifferential misclassification tends to dilute the association (ie, the true association will be larger than the observed association). In the extreme case in which errors are very frequent, even associations that are very strong in reality may not be identified in the database.

**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 subjects. But in the cross-sectional study, the exposure and the existing or prevalent outcome are measured at the same time. Accordingly, the direction of association may be difficult to determine. Another important limitation is that the outcome, or the threat of getting it, may have led to a departure of cases, so that 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.
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, so it is impossible to determine whether the observed outcome would likely have occurred in the absence of the exposure. Although descriptive studies occasionally demonstrate dramatic findings mandating an immediate change in physician behavior as a precaution, before the availability of evidence from stronger study designs (eg, recall the consequences of case reports of specific birth defects occurring in association with thalidomide exposure), 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 consequence of case reports suggesting that it was teratogenic. Later, although a number of comparative studies demonstrated the drug’s relative safety, 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 evidence initially. 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 (too small sample size). Even retrospective studies might be too difficult to conduct if most people receive the vaccine or the people who do not receive the vaccine may be quite different from those who get it, and the differences cannot be accounted for adequately. In this case, a before/after comparison using the general population before the introduction of the new vaccine can be conducted. But such comparisons using historical controls are prone to bias because many other things may have changed in the same period. However, if 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 rotavirus vaccine, resulting in a decision to withdraw the vaccine. The association was subsequently strengthened by a case-control study.

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

Design Issues: 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 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. 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 causal link between the exposure and the outcome and not due to a confounding factor. 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 still bias the results of observational studies.29 Because evidence, provider 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. The extent of bias in observational studies vs randomized trials remains uncertain. An empirical evaluation of 15 harms in which both types of evidence were available showed that observational studies might give either smaller or larger risk estimates compared with RCTs, but it is more common for observational studies to underestimate rather than overestimate the absolute risk of harm.30 Therefore, evidence of harmful effects from well-designed observational studies should not be easily dismissed.

---

**USING THE GUIDE**

Returning to our earlier discussion, the study that we retrieved investigating the association between soy milk (or formula) and the development of peanut allergy used a case-control design.1 Those with peanut allergy (cases) appear to be similar to the 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 (cases), both a family history of peanut allergy and an older sibling with a history of milk intolerance were more common and could bias the likelihood of a subsequent child’s being exposed to soy. To avoid confounding, these factors were adjusted in the analysis to provide an independent assessment of the association between soy and peanut allergy.

The methods for determining exposure were similar for cases and controls because the data were collected prospectively and both the interviewers and parents were unaware of the hypothesis relating soy exposure to peanut allergy (thus avoiding interviewer and perhaps recall bias). With regard to access to soy, all the children came from the same geographic region, although this does not ensure that cultural and economic factors that might determine soy access were balanced between cases and controls. Thus, from the initial assessment, the validity of the study appears adequate with the appropriate adjustments being done.
What Are the Results?

How Strong Is the Association Between Exposure and Outcome?
We describe the alternatives for expressing the association between the exposure and the outcome—the RR and the OR—in other chapters of this book (see Chapter 6, Therapy; Chapter 7, Does Treatment Lower Risk? Understanding the Results; and Chapter 10.2, Understanding the Results: More About Odds Ratios).

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 3 of 185 patients without the condition. The RR for mortality in hypertensive patients was 4.9. 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 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 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 RR, we use OR, the odds of a case patient being exposed divided by the odds of a control patient being exposed (see Chapter 7, Does Treatment Lower Risk? Understanding the Results; and Chapter 10.2, Understanding the Results: More About Odds Ratios). In circumstances in which the outcome is rare in the population at large (<1%), the OR of a case-control study represents the risk ratio in the whole population from which the cases and controls have been sampled. Even when event rates are as high as 10%, the OR and RR may still be quite close.

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 if the study design is strong (such as in an RCT). A much greater increase in risk might be required of weaker designs (such as cohort or case-control studies) because subtle findings are more likely to be caused by the inevitably higher chance of bias. Very large values of RR or OR represent strong associations that are less likely to be the result of bias.

In addition to showing a large magnitude of RR or OR, there is a second finding that can strengthen an inference that an exposure is truly associated with harmful effect. If, when the quantity or the duration of exposure to the putative harmful agent increases, the risk for the adverse outcome also increases (ie, the data suggest a dose-response gradient), then 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.
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 6, Therapy; see also Chapter 8, 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 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 may still be present, despite the failure to show a statistically significant association (see Chapter 8, Confidence Intervals).

**USING THE GUIDE**

The investigators calculated the OR for the risk of peanut allergy in those exposed to soy vs those not exposed to be 2.6 (95% CI, 1.3-5.2). 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 address 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. Could 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 (see Chapter 11.1, Applying Results to Individual Patients)?

**Was Follow-up Sufficiently Long?**

Studies can be pristine in terms of validity 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 effect we are really interested in is during a 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 malignancy. For example, if the question is whether a specific exposure, say to an
industrial chemical, causes cancer to develop, 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?**

**Are There Important Differences in the Exposures, for Instance, Dose or Duration, Between Your Patients and the Patients in the Study?**

As an illustration, 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 comes from the study that showed that workers employed in chrysotile asbestos textile operation between 1940 and 1975 had 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. The study does not provide reliable information regarding what might be the risks associated with only brief or intermittent exposure to asbestos (e.g., 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 Magnitude of the 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 compared with the unexposed group. Thus, we need a method for assessing clinical importance. In our discussion of therapy (see Chapter 6, Therapy; and Chapter 7, Does Treatment Lower Risk? Understanding the Results), we described how to calculate 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 in a similar way, only this time to calculate the number of patients that would have to be exposed to result in 1 additional harmful event. We may even use data from case-control studies with OR, although the formula is a bit more complex, and we would need to know the event rate for the outcome in the unexposed population from which the cases and controls were drawn (see Chapter 10.2, Understanding the Results: More About Odds Ratios).

During an average of 10 months of follow-up, investigators conducting the Cardiac Arrhythmia Suppression Trial, an RCT of antiarrhythmic agents, found that the mortality rate at approximately 10 months 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 treated with encainide or flecainide for about a year, we would cause 1 excess death. This contrasts with our example of the association between NSAIDs and upper gastrointestinal bleeding. Of 2000 unexposed patients, 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.
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 what are the adverse consequences of reducing or eliminating exposure to the harmful agent; that is, 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. Because the evidence of increased mortality from encainide and flecainide came from an RCT, we can be confident of the causal connection. 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 is relatively weak, the availability of an alternative substance can result in a clear decision.

For instance, the early case-control studies demonstrating the association between aspirin use and Reye syndrome were relatively weak and left considerable doubt about the causal relationship. Although the strength of the inference was not great, the availability of a safe, inexpensive, and well-tolerated alternative, acetaminophen, justified the preference for using this alternative agent in lieu of aspirin in children at risk for Reye syndrome.

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 there may not be safer and equally efficacious anti-inflammatory alternatives available. We were therefore probably right in continuing to prescribe NSAIDs for the appropriate clinical conditions.

**USING THE GUIDE**

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, we are told that the prevalence of peanut allergy is approximately 4 per 1000 children. An approximate calculation would suggest that with exposure to soy, 10 children per 1000 would be affected by peanut allergy. In other words, the number of children needed to be exposed to soy that would result in 1 additional case of peanut allergy is 167. (This estimate is crude and relies...
on a number of unverified assumptions regarding the true incidence of peanut allergy. Finally, there are no data regarding the negative consequences of withholding soy formula or soy milk products, and this would clearly be dependent on how severe and sustained an intolerance to cow’s milk was in a particular child.

References


