Methodologic Challenges to Studying Patient Safety and Comparative Effectiveness

Brian L. Strom MD, MPH

Abstract: Studies of patient safety and comparative effectiveness entail unique methodologic challenges. These studies may be susceptible to systematic error, including selection bias, exposure misclassification, and outcome misclassification. They may also be vulnerable to random error, or confounding by a variable such as another drug, a disease, or the drug indication itself. Finally, special logistical issues can arise, including data access problems, difficulties in conveying the need for studies of certain interventions, and obstacles to gaining institutional review board approval. This article provides a conceptual overview of these methodologic issues.

Key Words: comparative effectiveness, bias, confounding

(*Med Care* 2007;45: S13–S15)

The study of patient safety and comparative effectiveness raises multiple methodologic challenges. Though researchers in this field have a broad spectrum of expertise and training, and address a wide variety of questions, they all deal with a common subject: the use and effects of pharmaceuticals.

Historically, those who investigate the effects of marketed pharmaceuticals have focused on the risk of drugs and, in particular, the detection and confirmation of the rare adverse reaction not discovered at the time of drug marketing.[1] In recent years, however, we have realized that this focus is misplaced, because most public health problems from iatrogenic drug-induced disease are not due to these rare, undiscovered adverse reactions, but rather to the common dose-related adverse reactions of drugs used suboptimally.[2] This has become a focus of patient safety researchers. In addition, it has become clear that the information available at the time of drug marketing is sufficient to prove drug efficacy (ie, the possibility of a beneficial effect in an ideal setting), but not drug effectiveness (ie, whether the drug works in reality).[3] Further, information on comparative drug effectiveness (ie, whether the drug works better than alternatives) is often inadequate at the time of marketing. Finally, in recent years, there has been an increased focus on risk management of drugs (ie, modification of prescription practices to improve the use of drugs, and by extension, patient safety).[4]

METHODOLOGIC ISSUES

Overview

The goal of clinical research, in general, is to differentiate between the absence of an association and the presence of a causal association. Problems arise, however, because of intermediate options such as artifactual associations or indirect associations. Artifactual associations, otherwise known as false associations, can be caused by random chance (unsystematic variation) or bias (systematic variation). Indirect associations occur because of confounding, which happens when the presence of a variable other than the exposure and outcome of primary interest is related to both of them in a way that it can create an artifactual association or mask a true one.

Thus, investigators need to worry about 3 types of study design errors: random error, systematic error (ie, bias), and confounding. Different types of bias are selection bias (sometimes caused by uncontrolled confounding), misclassification of exposure, and misclassification of outcome. Investigators should also be concerned with the study's generalizability or lack of generalizability. Another concern is the possibility of effect modification, which occurs when the association between the exposure and the outcome is modified by a third variable such that the magnitude of the association depends on this variable. Statistically, effect modification is observed as statistical interaction.

Unique Problems in the Study of Pharmaceuticals

Exposure misclassification is a unique challenge in the study of pharmaceuticals.[5] In epidemiologic studies, patient recall is often used to determine exposure. However, patients often do not know which drugs they take at the time of the interview, and know less about which drugs they took in the past, though this history would be critical to an investigator studying long-term drug effects. Although little information exists regarding the validity of patient recall of drugs taken,

data primarily from hormone studies suggest that patient recall varies with drug type, how long the drug has been taken, how recently the drug was taken, and demographics of the patient, among other things.[6]

Even in studies using claims databases, which track whether a drug was dispensed, misclassification of the amount of drug used can occur. This is especially true if a drug exists in liquid or inhaled preparations, or may be used on an "as needed" basis.[5,6] Further, issues can arise regarding adherence (does the patient take the drug, and is it taken as recommended?) and persistence (for how long does the patient take the drug?).[7]

In addition, epidemiologists and health services researchers are used to studying dichotomous exposures, but drug exposure is rarely dichotomous.[8] Indeed, most adverse drug effects having public health importance are dose-related.[2] For some drugs, a wide range exists between the highest and lowest dose, with an equally wide range of effects. Variation in risk can also depend on duration. Some drug effects (eg, carcinogenesis) occur only after long duration and cumulative exposure, whereas other effects (eg, anaphylaxis) emerge rapidly. Also, depletion of susceptibles can occur in a study of prevalent users. Patterns of risk related to last exposure can also vary. Risk may be present only with the first exposure, or it may exist after any exposure, or it may disappear after exposure ends.

Outcome misclassification is also a substantial threat to the validity of the study. As one example, many studies investigate not true clinical outcomes (eg, myocardial infarction) but rather intermediate outcomes (eg, blood pressure). Premarketing studies are especially likely to investigate surrogate outcomes, leaving future outcomes studies to be performed after marketing. Sometimes this could result in misclassification of outcome. As another example, many outcomes (eg, pain relief) are difficult to measure and quantify,[9] and other outcomes (eg, mild rashes) are not likely to appear in claims databases. Further, uncertain validity of diagnostic outcomes is a problem, especially when claims databases are used.[10] Inpatient hospital claims diagnoses are usually good representations of the corresponding discharge diagnoses, although correspondence is not perfect; outpatient clinical diagnoses are much less likely to be valid. Researchers need to obtain medical records routinely to validate these diagnoses.[10]

Confounding

Conventionally, confounding is controlled for in a study using randomization, design (via exclusion or matching), or analysis (via stratification or mathematical modeling).[11] Although many of these approaches are useful in studies of the effects of pharmaceuticals, all except randomization require measurement of the confounder. In other words, to exclude those with a confounding variable, or match on the confounder, or stratify the analysis by the confounder, or control for the confounder in logistic regression or any other mathematical approach, one must measure the confounder. However, measurement of confounders presents the same challenges as described above for measure-

ment of exposure and outcome. Indeed, the confounders of concern are usually other drugs or other diseases.

In addition, studies of pharmaceuticals, especially studies of the effectiveness of pharmaceuticals, present unique challenges involving confounding by indication.[12] People who receive a drug are different from people who do not, in that they have an indication for therapy. This indication is commonly associated with the outcome under study, so unless the indication can be measured in detail, uncontrolled confounding and selection bias may occur. This is a more subtle problem than it might at first seem. For example, although the pretreatment blood pressure of a patient is easy to measure, the choice of a hypertensive drug is dictated by additional factors (eg, patient tolerability of different side effects). Further, the selection of dose over time, and even the decision to continue or discontinue the drug, is based on a patient's blood pressure response and the presence or absence of different side effects. Complete measurement of confounding by indication in this situation, on a longitudinal basis, is extraordinarily difficult. Yet, few indications are as quantitative and precise as blood pressure.

SPECIAL LOGISTICAL ISSUES

Finally, unique logistical issues arise in the study of patient safety and comparative effectiveness. Data access is one such issue. For example, one obvious potential data source is Medicare Part D data, which, when linked to other Medicare claims data, will be a unique resource on a massive and stable population, valuable for the studies we seek to perform. Yet, these data are not accessible to the research community, and we do not know when or if they will be. Patient safety studies are another example. Although they often benefit from access to local health care system data, this access can be mired in issues of local control, data quality, peer-review protections, etc. A final example, as noted above, is the critical need in many claims studies for medical records to validate diagnoses. Yet, attempts to obtain records are inhibited by logistical issues such as cost, institutional review board (IRB) protections, and Health Insurance Portability and Accountability Act protections.

Another logistical issue is convincing people that studies on patient safety and comparative effectiveness need to be performed. This is especially true for evaluations of interventions such as information technology (IT) interventions. The appeal of using IT to improve patient safety grows as health systems computerize, but some of these interventions can cause harm,[13] and others are at best useless. Yet, important IT alerts will be less effective the more frequently we provide them, especially if we provide useless ones. These IT alerts must be evaluated, but it is often hard to convince people to conduct such evaluations.

IRB approval can also be logistically difficult to obtain. For example, it may be infeasible to obtain approval from geographically dispersed hospitals to access their medical records or enroll their patients. In other situations, IRBs are concerned about study design, and novel approaches need to be developed. In a randomized trial of an IT intervention in health systems, for example, an IRB may be concerned that

the study is unethical because it believes the intervention *must* work, though the control group would undergo usual care before existence of the intervention. Also, the study subjects in such situations are actually the physicians; yet, obtaining consent from physicians would not be practical or facilitate a valid study, because of a Hawthorne effect. In still other situations, IRBs may be nervous because of the need for a rigid boundary between research and peer review and the peer-review protection inherent in quality assurance activities.

CONCLUSIONS

The goal of our research community is to optimize the use of pharmaceuticals by improving patient safety and informing practice with better data regarding the comparative effectiveness of drugs. Studies of safety and drug effectiveness pose unique methodologic challenges, including selection bias, misclassification of exposure or outcome, confounding, and logistical problems. Addressing these issues will require novel research approaches such as new study designs, innovative risk-adjustment methods to control for confounding, active surveillance of adverse effects, and new ways to bias reduction in observational studies. Our success in meeting these challenges will facilitate improved risk management, better patient safety, and more complete knowledge about drug effectiveness and adverse effects.

The next section of this supplement will address new methods to evaluate drug safety and effectiveness (ie, to generate the evidence needed therapeutic decisions). This will include issues relating to databases and data analysis approaches applicable to Medicare Part D data. Subsequent sections will include exploration of new types of experimental studies, risk-adjustment methods to control confounding, and approaches to active surveillance of adverse effects. Other articles will then examine novel approaches to bias reduction in observational studies. Finally, we will discuss the use of comparative effectiveness and safety evidence for clinical and policy decisions.

These articles present and advance the state-of-the-art of new methods for this new field. In most articles, the authors show along with it, demonstrations of their new approach. In other articles, the methods have been carried forward far enough to have detailed simulations, wherein it becomes clear where and when the new approach is useful, and when it would be misleading. In yet other situations, these are still just concepts, and their utility is yet to be demonstrated. This is an evolving science, and we are delighted to present its state-of-the-art. In all cases, though, as readers seek to apply these approaches, we suggest it is imperative that the decision to use them be made a priori, rather than after results are already available, to avoid selecting the answers preferred by the researchers in that situation.

REFERENCES

1. Strom BL. What is pharmacoepidemiology? In: Strom BL, ed. *Pharmacoepidemiology*. 4th ed. Sussex, England: John Wiley; 2005:3–15.
2. Strom BL; The Centers for Education & Research on Therapeutics (CERTs) Risk Assessment Workshop Participants. Risk assessment of drugs, biologics, and therapeutic devices: current approaches and future directions. *Pharmacoepidemiol Drug Saf.* 2003;12:653–662.
3. Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? *Am J Med.* 1985;78:475–480.
4. US Food and Drug Administration. Managing the risks from medical product use: creating a risk management framework. http://www.fda.gov/oc/tfrm/1999report.html. Accessed January 16, 2007.
5. West SL, Savitz DA, Koch G, et al. Recall accuracy for prescription medications: self-report compared to data base information. *Am J Epidemiol.* 1995;142:1103–1112.
6. West SL, Savitz DA, Koch G, et al. Demographics, health behaviors, and past drug use as predictors of recall accuracy for previous prescription medication use. *J Clin Epidemiol.* 1997;50:975–980.
7. Urquhart J. The odds of the three nons when an aptly prescribed medicine isn't working: non-compliance, non-absorption, non-response. *Br J Clin Pharmacol.* 2002;54:212–220.
8. Guess HA. Exposure-time-varying hazard function ratios in case-control studies of drug effects. *Pharmacoepidemiol Drug Saf.* 2006;15:81–92.
9. Guyatt GH, Kirshner B, Jaeschke R. A methodologic framework for health status measures: clarity or oversimplification? *J Clin Epidemiol.* 1992;45:1353–1355.
10. Strom BL, Carson JL, Halpern AC, et al. Using a claims database to investigate drug-induced Stevens-Johnson syndrome. *Stat Med.* 1991; 10:565–576.
11. Rothman KJ, Greenland S. *Modern Epidemiology.* 2nd ed. Philadelphia, PA: Lippincott-Raven; 1998.
12. Strom BL, Miettinen OS, Melmon KL. Postmarketing studies of drug efficacy: when must they be randomized? *Clin Pharmacol Ther.* 1983; 34:1–7.
13. Koppel R, Metlay JP, Cohen A, et al. Role of computerized physician order entry systems in facilitating medication errors. *JAMA.* 2005;293: 1197–1203.