

Bias: Considerations for research practice

TOBIAS GERHARD

Pharmacy practice research is multifaceted and includes a variety of settings, study questions, and study designs. This report serves as an introduction to the concept of bias for pharmacy practice researchers to help facilitate the conduct of methodologically sound research projects. Regardless of the specific characteristics of a research project, investigators must recognize and understand potential biases that threaten the validity of their study. Depending on the nature of the bias and the specifics of the study, investigators can use this information to avoid bias through changes in study design, subject selection, or measurement; adjust for the effects of bias during the analysis if bias cannot be feasibly avoided; or, as a last resort, acknowledge the presence of bias and examine and discuss the magnitude of its effect on the study's results. This report provides practical information and tools for novice researchers and thus sometimes oversimplifies complex issues. References are provided for more in-depth discussion of specific topics. Lastly, this report is not comprehensive but rather highlights select concepts and biases most relevant for applied pharmacy practice research.

Purpose. The concept of bias and its effects on study results are reviewed.

Summary. Understanding bias is paramount for the conduct of sound research studies. In the context of research methodology, bias refers to the presence of systematic error in a study. This report serves as an introduction to the concept of bias and describes three major areas through which bias is introduced into research studies of health care interventions: (1) factors that relate to the exposure of patients to treatments in the population, (2) factors that influence inclusion of patients in the study, and (3) factors related to assessment and measurement. Methods to address bias in both design and analysis stages of a study are presented, including restriction, matching, multivariate analysis, instrumental variable estimation, and the prevention of misclassification of information. While some biases are specific to certain study

designs and approaches, the general concepts of how bias is introduced into a study apply equally to randomized controlled trials, prospective observational or quasi-experimental studies, and retrospective studies. An important difference between the designs is how much opportunity the investigator has to avoid bias and influence accuracy of measurements, ranging from ideally full control in randomized controlled trials to almost no control in retrospective studies of automated databases.

Conclusion. Investigators should aim to avoid bias in the design of a study, adjust for bias in the study analysis if bias cannot feasibly be avoided, and quantify and discuss the effects of residual bias on study results.

Index terms: Clinical studies; Errors; Methodology; Quality assurance; Research
Am J Health-Syst Pharm. 2008;65:2159-68

What is bias?

Unlike its colloquial use as “prejudice” or “unfairness,” *bias*, in the context of research methodology, refers to the presence of systematic error in a study. In practice, this means that an estimate of association obtained from a biased study differs systematically from the true association in the source population of the

study.¹ For example, for many years, numerous observational studies suggested a protective effect of hormone replacement therapy (HRT) on heart disease in postmenopausal women; however, in 2002, a large randomized trial (Women's Health Initiative) found that HRT increased the risk for heart disease, drastically altering the use of one of the most commonly

TOBIAS GERHARD, PH.D., is Associate Director, Center for Education and Research on Mental Health Therapeutics, Institute for Health, Health Care Policy, and Aging Research, and Assistant Professor, Ernest Mario School of Pharmacy, Rutgers University, 30 College Avenue, New Brunswick, NJ 08901 (tgerhard@rci.rutgers.edu).

Supported in part through Agency for Healthcare Research and Quality grant HS016097.

Copyright © 2008, American Society of Health-System Pharmacists, Inc. All rights reserved. 1079-2082/08/1102-2159\$06.00.
DOI 10.2146/ajhp070369

The Research Fundamentals section comprises a series of articles on important topics in pharmacy research. These include valid research design, appropriate data collection and analysis, application of research findings in practice, and publication of research results. Articles in this series have been solicited and reviewed by guest editors Lee Vermeulen, M.S., and Almut Winterstein, Ph.D.

prescribed drug classes.^{2,3} While there is complex and controversial discussion regarding the sources of discrepancies between the observational studies and the clinical trial, the failure to adequately account for differences in lifestyle and socioeconomic status between women receiving and not receiving HRT in many of the observational studies has been suggested to explain how the observational studies could have produced biased (i.e., systematically false) results.^{4,5}

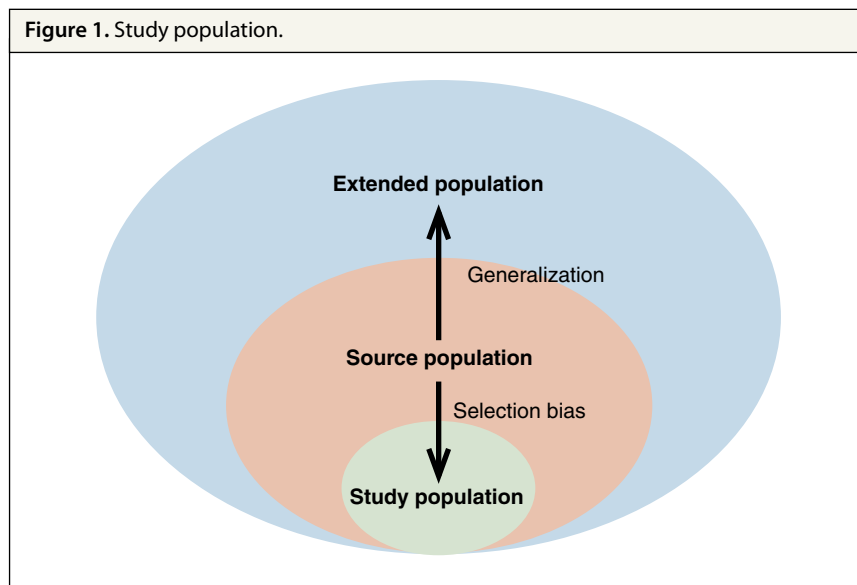
Bias is distinguished from random error, which denotes error due to sampling variability (i.e., chance) and is addressed by statistical hypothesis testing. Statistical hypothesis testing and bias resulting from inappropriate use of statistical methods are not the topics of this review and will be addressed in another article of this series on research fundamentals. In contrast to random error, which diminishes as sample size increases, systematic error is independent of the size of the study population. Importantly, statistical significance does not reflect the presence or absence of bias. In other words, a low *p* value does not guarantee or even suggest the absence of bias. Another common misconception is that a study is either biased or unbiased; this is an oversimplification. Because bias is a quantitative problem, the pertinent questions to ask are often how much bias exists in a study and to what extent existing bias could influence the conclusions drawn from a study, not whether the study is biased.^{1,6,7}

Numerous specific biases have been described. In 1979, Sackett⁸ proposed 35 specific biases; more recently, Delgado-Rodriguez and Llorca⁹ identified over 70 types of bias. While several classification systems have been proposed,^{6,9-11} the most common distinguishes among confounding, selection, and information bias.¹¹ However, the attribution of specific biases to these three categories is not always clear-cut, and the definitions of these categories vary by discipline and among authors. To avoid confusion, this report will rarely emphasize this nomenclature but rather present bias based on how it is introduced into the study. Conceptually, bias can be introduced by factors that determine who is exposed to the drug or intervention of interest in the population (confounding bias), factors related to who is included in the study (selection bias), and errors of assessment and measurement (information bias). Bias can also be classified by its effects on the association being studied. Bias may produce estimates of association that are smaller in magnitude than the true association (bias toward the null hypothesis) or larger in magnitude than the true association (bias away

from the null hypothesis). In extreme cases, bias can reverse an association and, as in the HRT example, make a harmful drug appear beneficial.

Study population

Bias can only be adequately discussed in the context of a study's population (Figure 1). The study population comprises the subjects included in the study's final analysis, excluding subjects who drop out of a prospective study if they are not included in the analysis by an intent-to-treat approach. The source population is the population from which the study population originated. For example, in a hospital-based observational study, all patients admitted to the hospital in the time frame of the study are considered the source population. In a randomized controlled trial (RCT), the randomized subjects are considered the source population; in the ideal case of an RCT without loss to follow-up, the source population is identical to the study population. The terms *internal validity* and *external validity* relate to these defined populations. Internal validity is defined as the absence of bias in relation to the source population of the study; external validity



refers to bias relating to an extended population. In the literature, the terms *study population* and *source population* are sometimes treated interchangeably, but this may lead to problems if the study population is not representative of the source population.

Consider the example of a study that aims to assess the effect of a quality-improvement program on diabetes outcomes in an outpatient clinic with a large percentage of minority patients. Assume that the researchers decide to include only patients with complete records for glycosylated hemoglobin and cholesterol in the analysis. In this case, the source population of the study (all patients with diabetes in the clinic) differs from the study population (the clinic's diabetic patients with complete laboratory records). Such sampling can introduce bias, as patients with complete laboratory records may systematically differ from patients with incomplete records. In contrast, *generalization* refers to the application of the study results to a different population (e.g., other clinics with differing patient populations).

Confounding variables

A variable is defined as a confounder if it is a risk factor (or surrogate risk factor^a) for the outcome of interest and associated with the exposure but not part of the causal pathway through which the exposure affects the outcome (i.e., not a mediating variable).¹² As a consequence of the association of the confounder with the exposure, the estimate of the association between exposure and outcome will represent a mix of effects—the effect of the exposure on the outcome and the effect of the confounder on the outcome. As a result, the estimate of association between the exposure and the outcome may be artificially increased, reduced, or even reversed. A fictitious example of a study investigating the associa-

tion between carrying matches and lung cancer is presented in Table 1. In the example, the true relative risk (RR) (proportion of people with the outcome [lung cancer] among those exposed [carried matches] divided by the proportion of people with the outcome among those not exposed [did not carry matches]) between carrying matches and lung cancer is 1.0 (i.e., there is no association); however, due to the confounding variable “smoking,” the observed association between carrying matches and lung cancer is 6.0. This association occurs because smokers are more than 10 times more likely to carry matches than nonsmokers ($[13,000/20,000]/[5,000/80,000]$) and are 25 times more likely to be diagnosed with lung cancer than nonsmokers ($[100/20,000]/[16/80,000]$).

Effect modification

Although it oftentimes occurs with confounding variables, the concept of effect modification has to be clearly distinguished from confounding variables. A pure effect modifier is not a risk factor for the outcome but rather changes the effect of the exposure on the outcome. Con-

sider the fictitious example in Table 2 where we observe a beneficial effect of treatment A on mortality for the entire study cohort (RR = 0.75). As in the example in Table 1, the RR for the full study cohort differs from the RRs observed in the stratified subgroups. However, in the case of an effect modifier, the RR also varies between the two strata, because the magnitude of the association between the exposure and the outcome depends on the level of the modifying variable. Unlike confounding, effect modification is a meaningful finding for a study because it may potentially provide guidance for treatment. Effect modification is also the reason for potential problems in applying the results of a study to an extended population (i.e., generalizability). If a study is conducted in only young men, most clinicians would be hesitant to apply results from this study to elderly women. This reluctance results from the assumption that effect modification exists and is associated with sex or age. Limited generalizability is often confused with bias, resulting in the misconception that study results are flawed because the study population was not representa-

Table 1.
Association Between Carrying Matches and Lung Cancer Is Confounded by Smoking Status

Study Group	Do Subjects Carry Matches?	No. Subjects Diagnosed with Lung Cancer		Event Rate	Relative Risk
		No	Yes		
Smokers (n = 20,000)	Yes (n = 13,000)	12,935	65	3.7%	1.0
	No (n = 7,000)	6,965	35	0.61%	
Nonsmokers (n = 80,000)	Yes (n = 5,000)	4,999	1	5%	1.0
	No (n = 75,000)	74,985	15	5%	
All (n = 100,000)	Yes (n = 18,000)	17,934	66	0.2%	6.0
	No (n = 82,000)	81,950	50	0.2%	

Table 2.
An Example of Effect Modification by Disease Severity

Study Group	Treatment	Death		Death Rate	Relative Risk
		Yes	No		
All (n = 1000)	A (n = 200)	15	185	7.5%	0.75
	B (n = 800)	80	720	10%	
Mild disease (n = 200)	A (n = 100)	5	95	5%	0.5
	B (n = 100)	10	90	10%	
Severe disease (n = 800)	A (n = 100)	10	90	10%	1.0
	B (n = 700)	70	630	10%	

tive of the general population (e.g., studies conducted in Medicaid claims databases) or restricted (e.g., a study restricted to mostly men because it was conducted in a Veterans Affairs setting). However, bias is not inherent in studies with unrepresentative or restricted populations but often arises only if results are generalized inappropriately.

In practice, effect modifiers are often confounding variables. For example, consider the role of race and ethnicity in heart failure. The prognosis for black patients with heart failure is worse than for their white counterparts (i.e., race and ethnicity are risk factors for mortality in heart failure patients). In addition, response to drug therapy varies by patient race and ethnicity (i.e., race and ethnicity are effect modifiers), which ultimately resulted in the selective approval of isosorbide dinitrate-hydralazine exclusively for black patients.^{13,14}

RCTs: Bastions against bias

The core elements of a standard RCT—random assignment of treatment, blinding of patients and treating physicians, and intent-to-treat analysis—have been explicitly developed to produce unbiased estimates of treatment effects by com-

paring outcomes between groups of patients who do not differ in any characteristic except the exposure to the treatment under study.¹⁵ This does not imply that all RCTs are free of bias—regrettably, many poorly conducted RCTs show severe bias; however, well-designed and properly conducted RCTs are the gold standard to obtain unbiased results. Specifically, in studies with a sufficiently large population, randomized assignment of treatment ensures the absence of significant confounding at the beginning of follow-up by preventing associations between known and unknown risk factors for the outcome and the exposure, and blinding ensures that treatments other than the intervention being studied are implemented identically in both groups and that knowledge of treatment assignment does not influence patients' perception of their health status. Intent-to-treat analysis ensures that the comparability of groups established at the beginning of the trial through randomization is not compromised by selection processes occurring during the study's follow-up period (e.g., discontinuation of the study drug due to adverse effects in a nonrandom subset of the treatment group).¹² This ideal situation is generally unattainable

in observational studies of general practice where (1) drug use is not randomly assigned, (2) only a sample of treated and untreated patients is studied, and (3) neither patients nor physicians are blinded. Prospective observational or quasiexperimental studies may implement some of these techniques (e.g., blinding of researchers who conduct measurements or intent-to-treat analysis) but by definition will not be able to randomly assign treatment and are therefore inherently vulnerable to confounding.

Sources of bias in nonrandomized studies

Bias can be introduced by having a comparison group that differs from the treatment group beyond the exposure to the drug and through measures of either drug or disease that differ between groups. In observational research, differences between patients taking the drug and patients not taking the drug can be introduced at two levels: determinants of drug use in the source population and determinants of inclusion in the study that lead to a different relation between exposure and outcome in the study population and the source population.

Patient and treatment. In real life, and thus in observational research, drugs and other therapeutic regimens are not used randomly but rather for specific indications and in specific populations. Access to care may differ across geographic and socioeconomic strata, and cultural customs and beliefs may affect the acceptance and use of specific treatments. In addition, risk factors for adverse outcomes are commonly considered when making treatment decisions. Adherence and survival are additional factors that may affect the comparability of subjects exposed to a drug versus those who are not exposed. For example, if the initiation of a drug is not associated with age but the drug has an adverse effect that

is more common in older patients, older patients are more likely to discontinue the drug in response to the adverse effect. On average, prevalent users of the drug (i.e., those who are taking the drug at a given point in time) will be younger than nonusers, even though age was not a factor that influenced drug initiation.

Because in medical practice neither drug initiation nor drug adherence is a random occurrence, users of a specific drug will likely differ from nonusers of a drug in numerous characteristics other than the use of the drug. Any of these characteristics may then introduce bias if they are (direct or indirect) risk factors for the outcome (i.e., if they are confounders). Several specific terms deserve mention in this context.

Confounding by indication describes the various factors related to an individual's condition that determine the initiation and choice of a specific drug (e.g., statins for patients with high cholesterol, antibiotics for patients with an infection), and thus comparisons of drug users to nonusers usually have to be made for patients sharing similar indications, as the condition itself is commonly one of the most important risk factors for the outcome of interest. In addition, the term refers to the more subtle considerations affecting the choice of a specific drug among patients with similar indications (sometimes referred to as *channeling*), relating to disease severity (*confounding by severity*), comorbid conditions, contraindications, patient demographics, the drug's adverse-effect profile, and subjective impressions of the prescribing physician. *Confounding by indication* has to be distinguished from *protopathic bias*, a specific bias that occurs when the initiation of a drug under study occurs in response to a symptom of the (at this point undiagnosed) outcome (e.g., the use of analgesics in response to pain caused by an undiagnosed tumor). Protopathic bias thus reflects

a reversal of cause and effect. The distinction between confounding and protopathic bias is important because the latter cannot be addressed through the methods used to control for confounding.

Prevalent user bias incorporates a number of problems that arise when prevalent drug users are included in a study (i.e., when patients who already took the drug under study at the beginning of follow-up are included in the analysis). Since drug use occurs over time, not only the initiation of a drug can give rise to confounding but also factors that are associated with the continuation of a drug. As a result, confounding may emerge among prevalent users, though it was not present at drug initiation. In addition to the potential introduction of confounding, the inclusion of prevalent drug users in observational studies has other pitfalls. For many adverse events, the risk of experiencing the adverse event changes over time. Gastric bleeding and ulcers occur commonly early after initiation of therapy with nonsteroidal anti-inflammatory drugs (NSAIDs) and are comparatively rare in long-term users. Since patients who experience gastric bleeding or ulcers will often discontinue their medication while patients who tolerate their medication will continue its use, an assessment of the NSAID-induced risk for gastric bleeding among prevalent users will systematically underestimate the true risk associated with the use of these medications.

Patient and study. Bias can also be introduced into a study whenever the study population is not representative of the source population. For example, a risk factor for the outcome that is not associated with the exposure in the source population (and thus does not act as a confounder) is associated with the exposure in the study population. This can occur whenever the exposure and a risk factor for the outcome are both associated with subjects' in-

clusion into the study.^{1,16} Imagine, for example, a clinician-researcher who has collected data on patients hospitalized for arthritis and treated with cyclooxygenase-2 (COX-2)-selective NSAIDs at his institution. He plans a study to assess the risk of adverse effects associated with the use of these drugs. The researcher decides that the control group will consist of an equal number of patients with a diagnosis of arthritis who have been treated with ibuprofen (ascertained from a claims database). Such a design will almost certainly result in a distorted association of the exposure (COX-2-selective NSAID versus ibuprofen) and several confounders (e.g., age, general health status) in the study because all patients in the COX-2-selective NSAID group were identified in the hospital and are therefore likely older and of poorer health than the patients treated with ibuprofen who were identified from a claims database. Practically, it does not matter whether a confounder is present in the source population or whether it is introduced during the selection of the study population. However, the distinction is useful, because it explains why a variable may act as a confounder in one study but not in another, even when exposure and outcome are defined identically.¹⁶

Bias can also be introduced if selection procedures change the distribution of an effect modifier from the source population to the study population. For example, if a drug works better in women than in men and the study population contains more women than the source population, then the overall estimate of the drug effect for the source population will be biased if the results are not stratified by sex. In summary, the selection of subjects from the source population into the study population may alter the association of a risk factor or effect modifier for the outcome with the exposure and, thus, the estimate of association between exposure and outcome

obtained from the study population does not reflect the true association in the source population.

Nonrandom selection of patients into a study can occur for a number of reasons. *Self-selection bias* is the common term used to denote selection of patients who volunteer, or self-select, for study participation. The associations among exposure, outcome, and confounders within these patients often differ from the associations in the source population. For example, patients may be more likely to seek treatment and consequently participate in a cohort study of a drug to treat their condition if their disease is severe. *Referral bias* describes a situation where exposure to the drug of interest is associated with the likelihood of referral into the institution where the study takes place (often a hospital). This is frequently the case when there is already a known or suspected association of the drug with the outcome of interest, so referral to an institution where the diagnosis is made is more likely for patients taking this drug (e.g., NSAIDs and gastric ulcers).

In practice, the availability of a certain measure (e.g., a preintervention serum creatinine value) is commonly an inclusion criterion for the study. This may then change the distribution of an effect modifier or confounder between the study population and the source population if the availability of this measure is associated with the study outcome (e.g., if a measurement is not routinely made for emergency procedures but always occurs for scheduled procedures whose outcomes are more favorable). In prospective studies, patient dropout often has similar effects, resulting in a study population at the time of analysis that may not be representative of the population that was enrolled into the study. In RCTs, bias arising from the loss of participants after randomization is generally referred to as *attrition bias*.

Descriptive studies

Most comments in this review pertain to bias as it relates to measures of association (i.e., effects of treatments). It is important first to understand how bias relates to participants in descriptive studies. As discussed above, selection processes create bias in the estimation of treatment effects only when the study population is not representative of the source population in relation to confounders or effect modifiers. In contrast, descriptive results may be biased even when that is not the case. Consider a hospital-based study of a drug effect. If the hospitalized patient population is representative of the source population regarding confounders and effect modifiers, a properly conducted inferential study will not produce biased results, even if the disease of interest is much more common in the hospital setting, which may be the very reason that the study was conducted in this setting. However, if the goal is to estimate the prevalence of a certain disease, the hospital-based study will systematically overestimate the proportion of patients with the disease in the source population. Careful consideration of the relation between study and source populations, as well as sampling characteristics, is therefore central for descriptive studies if the results are intended to be valid for the source population. However, if the goal of the study is to describe disease prevalence within the study population (e.g., a specific hospital), these considerations do not apply.

Misclassification of information

All information used in the analysis of patient data is subject to error in measurement, including exposure to drugs, the outcome of interest, and confounding factors. For categorical variables such as sex or the occurrence of an adverse event, this is usually referred to as misclassification. For example, a patient who took a nonprescription NSAID could be

(mis)classified as an NSAID non-user if the automated data source of the study included only patients who took prescription medications. When errors in measurement are related to exposure or outcome, they are called “differential,” while errors unrelated to exposure and outcome are termed “nondifferential.” For example, if the outcome is more likely to be observed among patients who have the exposure under study (e.g., an adverse effect has been previously reported in association with the drug being studied), there is differential misclassification, while nondifferential misclassification would exist if the adverse effect remained undiagnosed in some of the patients who experienced it, but the likelihood of diagnosis was not affected by exposure. This distinction is of practical importance because the consequences of the most common form of nondifferential misclassification, namely the misclassification of a binary variable, are predictable in that they always lead to a bias toward the null hypothesis (i.e., an estimate of association smaller in magnitude than the true estimate). This means that, in studies of effectiveness, nondifferential misclassification of a binary exposure or outcome will always underestimate the benefits of the treatment and thus provide a conservative estimate. Nondifferential misclassification of multicategory variables and differential misclassification, however, may result in bias toward the null hypothesis, away from the null hypothesis, or even reversal of an association, depending on the specific situation.

Examples of differential misclassification include recall and detection biases. Recall bias may occur whenever information relies on patients’ memory (i.e., information is collected retrospectively through surveys or interviews). Recollection of important study variables is often affected by factors related to the outcome. For example, a patient di-

agnosed with lung cancer may more accurately recall his smoking habits over the past decades compared with a healthy subject.

While recall bias commonly affects the accuracy of the data collected regarding exposure and confounding factors, *detection bias* can affect the accuracy of both exposure and outcome when the assessment of the exposure varies depending on the outcome (in case-control studies) and vice versa (in cohort studies). For example, patients who take a drug that requires frequent monitoring are likely to see their physician more frequently than nonusers of such medications and are consequently more likely to be diagnosed with other conditions. Detection bias is especially common when a concern about a drug's safety has been previously reported in the literature, as patients exposed to this drug are more likely to be assessed for the previously reported adverse outcome than are unexposed patients.

Misclassification of confounding variables. In crude analysis, the misclassification of confounders will not introduce bias but will affect the ability to adjust for confounding factors as discussed below. Nondifferential misclassification of a binary confounder will reduce the investigator's ability to control for the confounder's effect, while differential misclassification may result in a situation in which the adjustment made for a confounder leads to an increase in bias over the unadjusted result.

Immortal time bias. This type of bias is a special case of exposure misclassification. It occurs when the definition of exposure creates periods of "immortal time" (i.e., time periods during which subjects cannot experience the outcome of interest). For example, consider an observational study aiming to assess outcomes after hospital discharge conditional on drug use (e.g., the risk for a second myocardial infarction among patients discharged with or without a

prescription for a β -blocker). In this example, drug exposure is defined as filling the β -blocker prescription sometime within 60 days after hospital discharge. Such a definition of drug exposure introduces periods of immortal time into the study. A patient who fills the first prescription after 50 days of discharge is assigned to the exposure cohort. Had the same patient died after 40 days, she would have been assigned to the unexposed cohort because at that time she had not filled the prescription. In contrast, patients who do not receive a prescription in the window of time used to define exposure cannot have periods of immortal time. In other words, a definition of exposure as described above leads to selective misclassification of unexposed "person time" as exposed person time (in the example, the 50 days after discharge but prior to the dispensing of the prescription). This bias can be avoided by adequately defining person-time exposure.^{17,18}

Dealing with bias

Because exposure to pharmacologic and other treatments in populations does not occur at random and is not under the control of the investigator, the presence of confounding factors cannot be avoided in observational studies of medical interventions and is arguably one of the biggest challenges in pharmacoepidemiologic research. However, if confounding factors are known and measured with sufficient accuracy, their effect can be controlled for either through study design or analysis.

Control of measured confounding. The simplest approach to control a categorical confounder is *restriction*. A variable has to be able to vary to be associated with an exposure. If an analysis is restricted to one particular value of the variable of interest (e.g., if the analysis is conducted only in men), then the variable can, by definition, not act as a confounder

(in a study of a male population, sex cannot be a confounder). The logical extension of this approach is stratification, where analysis is conducted within each stratum of a categorical variable (Table 1). As with restriction, stratification controls for confounding, because it does not allow the confounder to vary within a stratum. A further advantage of stratified analysis is that the presence of effect modification is readily demonstrated when the measure of association differs significantly between strata (Table 2). However, feasibility of stratification is limited when there are multiple confounders, because a stratum has to be constructed for each possible confounder combination (e.g., a fully stratified analysis of four three-category confounders would require the calculation of 81 [3⁴] measures of association).

Matching attempts to produce a control group that is similar to the study group with regard to the distribution of selected confounding factors.¹ It is most commonly applied in case-control studies but can also be used in cohort studies. Several matching techniques exist, but specifics are beyond the scope of this report. Of note, matched designs generally require inclusion of the matching variables in the model. As the number of confounding factors increases, matching becomes increasingly difficult and time-consuming. Matching on exposure propensity scores greatly simplifies matching on multiple variables. Propensity-score matching is a two-step process that calculates the probability of exposure (the propensity score) based on all observed confounding factors and then matches subjects with nearly identical scores.^{19,20} Propensity scores are also commonly used for stratification or as covariates in regression analysis. They are an efficient tool, especially when large numbers of confounding variables are included in a study, but do not control for unmeasured confounding.

Multivariate analysis applies mathematical models to assess the association of multiple predictors (i.e., the exposure variable of interest and multiple potential confounding variables) with an outcome. The estimate of association for each variable (including the exposure variable of interest) reflects the individual association of a variable with the outcome, controlling for the contribution of other variables in the model. Effect modification is not identifiable through simple inclusion of the modifier in a regression model but requires the inclusion of an interaction term between the exposure and the effect modifier in the model. Regression models will produce unbiased results for the variable of interest if all confounders are included and adequately measured and the model is correctly specified.

Time-dependent confounders. Like exposure to treatment, confounders can change their value over time and, as such, be included as time-dependent covariates in appropriate models. However, recall that timing plays an important role in the definition of a confounder. For example, blood pressure measured before the initiation of antihypertensive therapy meets the definition of a confounder for the risk of stroke (and can be adjusted for using the methods described above), but blood pressure measured after the initiation of antihypertensive medication would not meet this definition because its association with treatment is at least partially a result of the blood-pressure-lowering effects of antihypertensive therapy (i.e., it is an intermediate in the causal pathway between the exposure and the outcome). In such a situation where a time-dependent confounder is itself affected by previous exposure, the aforementioned methods of confounder control will not provide unbiased estimates. Instead, more complex methods (e.g., marginal structural models) that account for

the intermediate nature of some confounders have to be used.²¹

Unmeasured confounding. When a confounder is not measured, it cannot be adjusted for by any of the previously discussed approaches. Confounders may not be measured if they are not known or are difficult or expensive to measure. Under these circumstances, the appropriate definition of an active control group can offer a chance to minimize the effect of unmeasured confounders. For example, research has shown that medications that largely treat asymptomatic, chronic conditions (e.g., statins) are underutilized in patients with unrelated comorbidities, which may lead to a healthier population of users of a chronic medication.²² Investigators have addressed this problem by using a comparison group of users of another chronic medication, instead of a group of nonusers of the study drug. For example, while Ray et al.²³ found a 38% (95% confidence interval (CI), 15–55%) reduction in the risk of hip fractures among new users of statins compared with nonusers, this beneficial effect disappeared when the comparison was made between new statin users and new users of other lipid-lowering drugs (RR, 1.42; 95% CI, 0.83–2.43). Similarly, in a study investigating a hypothesized beneficial effect of statins on the risk of lung, breast, and colorectal cancer, Setoguchi et al.²⁴ used a comparison group of new users of glaucoma medication instead of statin nonusers. In certain situations, a similar level of control can be achieved by comparing users of the drug of interest with former users of the drug. This group should be similar to current users in difficult-to-measure factors associated with receiving the drug.²⁵

Instrumental variable estimation, a method currently not widely used in the medical literature, allows adjustment for measured and unmeasured confounding.²⁶ The method requires the identification of a valid instru-

ment (a variable that is highly associated with the exposure of interest and not independently associated with the outcome). A recent study investigating the effect of cardiac catheterization in more than 120,000 patients with acute myocardial infarction (AMI) compared several analysis approaches, including an instrumental variable analysis.²⁷ Based on the fact that strong regional differences in cardiac catheterization rates exist despite similar AMI severity, the authors used regional catheterization rates as their instrument. The authors then presented the association of cardiac catheterization with mortality without adjustments, using multivariate and several propensity-score-based methods (i.e., approaches that adjust exclusively for measured confounders) and an instrumental variable analysis. Patients receiving cardiac catheterization were younger and had lower AMI severity than those who did not, and accordingly the unadjusted analysis would be expected to overestimate the benefit of the procedure. The authors reported an unadjusted RR of 0.36 (95% CI, 0.36–0.37), with RRs ranging from 0.51 (95% CI, 0.50–0.52) to 0.54 (95% CI, 0.53–0.55) for the multivariate and propensity-score-adjusted models and a RR of 0.84 (95% CI, 0.79–0.90) for the instrumental variable analysis, suggesting that the standard adjustment approaches remove some but not all of the confounding present in the unadjusted estimate.

Patient selection

Unlike the exposure to treatment in a population, the selection of subjects into a study and the retention of subjects in prospective studies may be within the control of the investigators, albeit often with limitations. Prospective studies with control of subject selection should consider the following strategies to minimize bias arising from the selection of patients: random sampling from

the source population, recruitment of a consecutive series of patients, minimizing loss to follow-up, and procedures to attempt outcome measurement in dropouts. However, such measures are often infeasible, costly, and generally imperfect for studies whose patient accrual is limited to a specific setting (e.g., a specific hospital). As pointed out before, factors associated with hospitalization may become problematic if exposure to the treatment of interest is also a risk factor for the hospitalization. (This applies to studies that only evaluate an exposure that was initiated before hospitalization.)

As in clinical trials, prospective nonrandomized studies may use an intent-to-treat analysis to prevent additional selection bias after the initiation of the study. This involves the inclusion of all patients who were assessed at baseline into the final analysis in the study group to which they were assigned, regardless of whether they completed the study or discontinued or switched treatment.

Very little control over patient selection is commonly available in retrospective studies with automated data collection. However, it is paramount that equal inclusion criteria and ascertainment procedures for cases and controls (for case-control studies) and exposed and unexposed patients (for cohort studies) are applied and that all patients ideally stem from the same source population. In addition, multiple biases arising from the inclusion of prevalent users can be avoided through a new user design, which limits the study to incident (i.e., new) users of the study medication.²⁸

Preventing misclassification

Several approaches can be used to minimize misclassification of information in prospective observational studies and clinical trials: blinding; use of standardized and validated measures; and definitions for exposure, outcome, and covariates. In

many circumstances, especially in the analysis of prescription claims data sets over longer time periods, it is advantageous to use a person-time approach to classify exposure. For example, classify a patient as a user of the drug of interest only for time periods when he or she is actually taking the drug to reduce misclassification. Investigators have limited control over measurement, especially in retrospective studies, where the measurement has already occurred, and in studies of automated databases. It is therefore crucial for the investigator to learn as much as possible about how measurements were made and how data were coded. In smaller settings (e.g., a single hospital), the investigators may be able to locally assess the specifics and the quality of data generation; in other settings, it may be necessary to formally validate data entries against medical records.

Other considerations

Bias has to be viewed and treated as a quantitative problem, and it is important to explore and report the magnitude of bias present in a study. Sensitivity analysis can be used to provide confidence limits around estimates of association based on informed assumptions about unmeasured confounders or variable misclassification. As such, sensitivity analysis provides information about the extent to which biases that remained uncontrolled in the design, data collection, and analysis of a study may change the reported results.⁷

Lastly, in order to assess the potential of uncontrolled confounding and other systematic bias, it is often useful to assess the specificity of the study results to the outcome studied. Risks or benefits that are similar in magnitude across a wide range of outcomes are a common sign of bias. For example, if the drug of interest shows a hypothesized protective effect against hip fractures but

an equally protective effect for all-cause mortality, myocardial infarction, and cancer, this may suggest that the observed protective effect for hip fractures was the result of a generally healthier treatment group (e.g., due to a healthy-user effect) and not the result of drug-specific effects. Generally, it is advisable to use an endpoint that is as specific to the hypothesized biological pathway as possible (e.g., ischemic stroke versus all-cause mortality).²⁹

While some biases are specific to certain study designs and approaches, the general concepts of how bias is introduced into a study apply equally to RCTs, prospective observational or quasiexperimental studies, and retrospective (including case-control) studies. An important difference between the designs is how much opportunity the investigator has to avoid bias and influence accuracy of measurements, ranging from full control in RCTs to almost no control in retrospective studies of automated databases.

Conclusion

Investigators should aim to avoid bias in the design of a study, adjust for bias in the study analysis if bias cannot feasibly be avoided, and quantify and discuss the effects of residual bias on study results.

²⁹Consider race; depending on the outcome of interest, race may be a surrogate risk factor for socioeconomic status, dietary habits, or genetic makeup rather than a direct risk factor.

References

1. Rothman KJ, Greenland S. *Modern epidemiology*. 2nd ed. Philadelphia: Lippincott-Raven; 1998.
2. Nelson HD, Humphrey LL, Nygren P et al. Postmenopausal hormone replacement therapy: scientific review. *JAMA*. 2002; 288:872-81.
3. Rossouw JE, Anderson GL, Prentice RL et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002; 288:321-33.
4. Garbe E, Suissa S. Hormone replacement therapy and acute coronary outcomes:

- methodological issues between randomized and observational studies. *Hum Reprod.* 2004; 19:8-13.
5. Prentice RL, Langer R, Stefanick ML et al. Combined postmenopausal hormone therapy and cardiovascular disease: toward resolving the discrepancy between observational studies and the Women's Health Initiative clinical trial. *Am J Epidemiol.* 2005; 162:404-14.
 6. Maclure M, Schneeweiss S. Causation of bias: the episcopo. *Epidemiology.* 2001; 12:114-22.
 7. Schneeweiss S. Confounding. In: Hartzema A, Tilson H, Chan K, eds. *Pharmacoepidemiology and Therapeutic Risk Management*. 4th ed. Cincinnati: Harvey Whitney; 2008:273-300.
 8. Sackett DL. Bias in analytic research. *J Chronic Dis.* 1979; 32:51-63.
 9. Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health.* 2004; 58:635-41.
 10. Gail MH, Benichou J. *Encyclopedia of epidemiologic methods*. New York: Wiley; 2000.
 11. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods*. Belmont, CA: Lifetime Learning; 1982.
 12. Last JM, ed. *A dictionary of epidemiology*. 4th ed. New York: Oxford Univ. Press; 2001.
 13. Taylor AL, Wright JT Jr. Should ethnicity serve as the basis for clinical trial design? Importance of race/ethnicity in clinical trials: lessons from the African-American Heart Failure Trial (A-HeFT), the African-American Study of Kidney Disease and Hypertension (AASK), and the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *Circulation.* 2005; 112:3654-60,3666.
 14. BiDil (isosorbide dinitrate and hydralazine hydrochloride) package insert. Lexington, MA: NitroMed; 2005.
 15. Machin D, Day S, Green SB. *Textbook of clinical trials*. 2nd ed. Hoboken, NJ: Wiley; 2006.
 16. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004; 15:615-25.
 17. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf.* 2007; 16:241-9.
 18. Suissa S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: immortal time bias in observational studies. *Am J Respir Crit Care Med.* 2003; 168:49-53.
 19. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol.* 2006; 98:253-9.
 20. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med.* 1997; 127(8, pt. 2):757-63.
 21. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000; 11:550-60.
 22. Redelmeier DA, Tan SH, Booth GL. The treatment of unrelated disorders in patients with chronic medical diseases. *N Engl J Med.* 1998; 338:1516-20.
 23. Ray WA, Daugherty JR, Griffin MR. Lipid-lowering agents and the risk of hip fracture in a Medicaid population. *Inj Prev.* 2002; 8:276-9.
 24. Setoguchi S, Glynn RJ, Avorn J et al. Statins and the risk of lung, breast, and colorectal cancer in the elderly. *Circulation.* 2007; 115:27-33.
 25. Ray WA, Meredith S, Thapa PB et al. Cyclic antidepressants and the risk of sudden cardiac death. *Clin Pharmacol Ther.* 2004; 75:234-41.
 26. Martens EP, Pestman WR, de Boer A et al. Instrumental variables: application and limitations. *Epidemiology.* 2006; 17:260-7.
 27. Stukel TA, Fisher ES, Wennberg DE et al. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA.* 2007; 297:278-85.
 28. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol.* 2003; 158:915-20.
 29. Yusuf S, Negassa A. Choice of clinical outcomes in randomized trials of heart failure therapies: disease-specific or overall outcomes? *Am Heart J.* 2002; 143:22-8.