43

Advanced Approaches to Controlling Confounding in Pharmacoepidemiologic Studies

Sebastian Schneeweiss[1] *and Samy Suissa*[2]

[1] *Harvard Medical School; Harvard TH Chan School of Public Health; Department of Medicine, Brigham & Women's Hospital, Boston, MA, USA*
[2] *Department of Epidemiology, Biostatistics and Medicine, McGill University; Centre for Clinical Epidemiology, Lady Davis Research Institute – Jewish General Hospital, Montreal, Quebec, Canada*

The past two decades have witnessed an explosion of methodological advances in the design and analysis of epidemiologic studies. Some of these contributions have been fundamental to the field of epidemiology in general while others have arisen specifically from questions posed by pharmacoepidemiologic applications. Several of these advances have already played an important role in the conduct of research on drug effects, and will take an even greater place in future applications. In this chapter, we introduce some of these approaches with a focus on confounding control, one of the major methodologic challenges in drug safety and effectiveness research with noninterventional studies.

We start out by describing a robust study design that will exemplify several aspects of confounding control and other biases and point out critical decision points in the choice of study designs. Second, we describe efficient sampling strategies within cohort studies (case–control, case–cohort, and two-stage sampling) and self-controlled designs (case–crossover and case–time–control designs) and how they will help reduce confounding bias. Third, we introduce several analytic methods that have gained wider use in pharmacoepidemiologic studies and others that only recently have made inroads into pharmacoepidemiology.

Clinical Problems to Be Addressed by Pharmacoepidemiologic Research

Pharmacoepidemiologic analyses are in principle no different from analyses in any other subject area within epidemiology. They are concerned with valid estimation of associations between an exposure and outcome, and methods to minimize systematic and random error that may cloud causal conclusions. Some issues specific to pharmacoepidemiology stem from the constraints of the frequently used secondary data sources, in particular large electronic longitudinal healthcare databases from insurance health plans, electronic medical records systems, or registries (see Chapters 11–14). Another difference is the often unusually direct interdependency of treatment choice with health status, severity of disease, and prognosis that may lead to strong, sometimes intractable confounding by indication (see Chapter 3) [1].

Pharmacoepidemiology, Sixth Edition. Edited by Brian L. Strom, Stephen E. Kimmel and Sean Hennessy.

Pharmacoepidemiologists try to reduce biases by appropriate choices of study design and analytic strategies. Challenges arise if not all confounder information is captured in the available data. This chapter provides an overview of selected options that fit typical pharmacoepidemiologic data sources and study questions.

Methodologic Problems to Be Addressed by Pharmacoepidemiologic Research

The ready and relatively cheap availability of large longitudinal patient-level healthcare databases make the new-user cohort design a natural design choice as a starting point that mimics the classic parallel group controlled trial, except of course for the randomized treatment assignment (Figure 43.1) [2]. Efficient sampling designs within such cohorts, including case–control, case–cohort, and two-stage sampling designs, are important extensions to assess additional covariate or outcome information in a subset of patients. Such sampling usually provides no advantage if secondary data are the only source for exposure, covariate, or outcome assessment because there is no additional cost or time to analyze the entire dataset rather than a subsample [3].

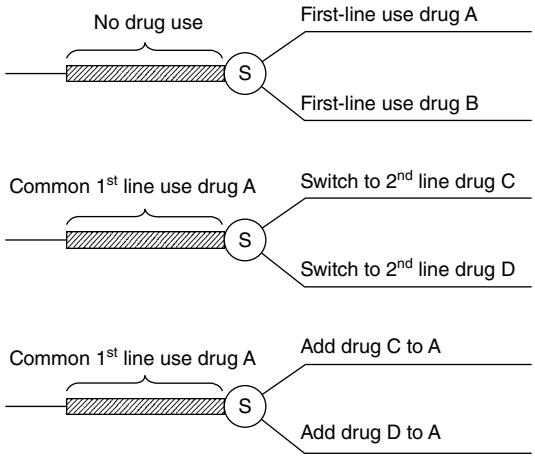

Figure 43.1 Principle of the new user design and its variations when studying second-line therapies. *Source:* Reproduced from Schneeweiss [3] with permission from John Wiley & Sons.

Bias can be reduced by appropriate design choices. Considerations about the sources for exposure variation will lead to fundamental decisions on the appropriate study design. In a causal experiment, one would theoretically expose a patient to an agent and observe the agent's effect on his or her health, then rewind time, leave the patient unexposed, and keep all other factors constant to establish a counterfactual experience. Since this experiment is impossible, the next logical expansion of the experiment is to randomly introduce or observe exposure variation within the same patient but over time (Figure 43.2). If we observe sporadic drug use resulting in fluctuations of exposure status within a patient over time, if that drug has a short washout period, and if the adverse event of interest has a rapid onset, then we may consider a case-crossover design or related approaches (see later). For most pharmacoepidemiologic studies, we will utilize variation in exposure between individual patients, and we will therefore apply a cohort study design. Any exposure variation among higher-level entities (provider, region, etc.) can be exploited using instrumental variable analyses (described later in the chapter) if unrelated to patient characteristics either directly or indirectly [4].

In a cohort design, there are several advantages to identifying patients who start a new drug and begin follow-up after initiation, similar to a parallel group randomized controlled trial that establishes inception cohorts. As patients in both the study group and the comparison group have been newly started on medications, they have been evaluated by physicians who concluded that these patients might benefit from the newly prescribed drug. This makes treatment groups more similar in characteristics that might not be observable in the study database if medication use is not new.

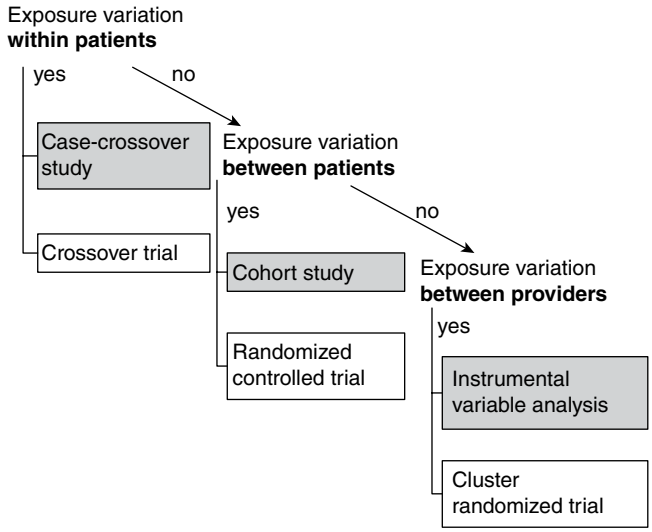

Figure 43.2 Study design choice by source of exposure variation. Shaded boxes indicate noninterventional study designs while clear boxes are the randomized design versions. *Source:* Reproduced from Schneeweiss [3] with permission from John Wiley & Sons.

The clear temporal sequence of confounder measurement before treatment initiation in an incident user design also avoids mistakenly adjusting for consequences of treatment (intermediates) rather than predictors for treatment, a possible reason for "overadjustment" [5]. Comparing two active treatment groups further reduces the chances of immortal time bias, a mistake that most frequently emerges if future information is used to define earlier exposure status in healthcare databases, particularly when defining a "nonuser" comparison group [6]. A common example of immortal time bias is to define nonusers as patients who have not used the study medication during the first six months of follow-up. By definition, these nonuser patients cannot die during the first six months of follow-up, and therefore their inclusion can bias the findings. Because of the well-defined starting point of inception cohorts, it is possible to assess whether and in what form hazards vary over time by stratifying on duration of treatment. Studying new users is also useful when studying newly marketed medications; the incident user design avoids comparing populations predominantly composed of first-time users of a newly marketed drug with a population predominantly composed of prevalent users of the old drug. Such a comparison would be prone to bias because patients who stay on treatment for longer may be less susceptible to the event of interest [7].

A common criticism of the incident user design is that excluding prevalent users will restrict and thus reduce the study size, in some cases substantially [8]. While this is true, researchers should be aware that by including ongoing (prevalent) users, they might gain precision at the cost of validity [9]. Screening and identifying incident users in secondary databases is not costly except for a bit more computing time. In some situations, particularly studies of second-line treatments in chronic conditions, we can only study patients who switch from one drug to another, as very few patients will be treatment naive. Such switching is often not random, but rather is determined by progressing disease and treatment failure or by side effects that may be related to the study outcome. A fairer treatment comparison may be achieved by comparing new switchers to the study drug with new switchers to a comparison drug (see Figure 43.1). Consequently, prevalent new-user cohort designs are being developed to minimize bias when one needs to include as many new users of the study drug as possible.

Even with appropriate designs, however, all observational pharmacoepidemiologic studies still must consider carefully how to approach the problems of potential confounding, in order to prevent bias. Approaches to addressing these methodologic challenges, and their limitations, will be the primary focus of this chapter.

Currently Available Solutions

The solutions available to minimize confounding in pharmacoepidemiologic database studies can be broadly categorized into (1) approaches that collect more information on potential confounders and apply efficient sampling designs to reduce the time and resources it takes to complete the study, and (2) analytic approaches that try to make better use of the existing data with the goal of improved control of confounding.

Efficient Sampling Designs Within a Cohort Study

In any cohort study, the cost, time, and resources necessary to collect data on all cohort members can be prohibitive. Even with cohorts formed from computerized databases, there may be a need to supplement and validate data with information from hospital records, medical records, and physician or patient interview questionnaires, with the goal of improved confounding control. When the cohort size is considerable, such additional data gathering can become a formidable task. Moreover, even if no additional data are needed, the data analysis of a cohort with multiple and time-dependent drug exposures can become technically unfeasible, particularly if the cohort size and number of outcome events are large. For example, a study of the long-term effect of antihypertensive drugs and the risk of cancer involved a cohort of over 1.1 million patients where 41 059 developed cancer during 14 years of follow-up, a size that necessitated sampling within the cohort [10]. Finally, there are situations with multiple confounding factors that may require accurate matching rather than simply modeling adjustment.

To counter these constraints, designs based on sampling subjects within a cohort have been proposed and applied successfully in pharmacoepidemiology. These designs are based on the selection of all cases with the outcome event from the cohort, but differ in the selection of a small subset of "noncases." Generally, they permit the precise estimation of relative risk measures with negligible losses in precision. Below, we discuss structural aspects of cohorts and present three sampling designs within a cohort the nested case–control, the multi-time case–control, and case–cohort designs.

Structures of Cohorts

Figure 43.3 illustrates graphically a cohort of 21 newly diagnosed diabetics over the period 1995 to 2010. This cohort is plotted in terms of calendar time, with subjects ranked according to their date of entry into the cohort, which can correspond to the date of disease diagnosis or treatment initiation. Such *calendar-time cohorts* depict the natural chronological nature of the cohort accrual. An alternative depiction of this same cohort could be based on duration of disease (i.e., follow-up time from diagnosis or first

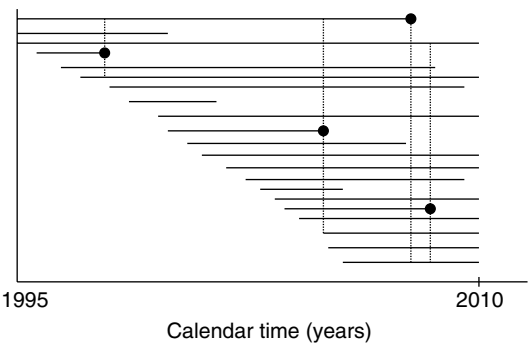

Figure 43.3 Illustration of a *calendar-time* cohort of 21 subjects followed from 1978 to 1990 with four cases (●) occurring and related risk-sets (---).

exposure to a drug), which may be more relevant to the drug effect under study. In this instance, the illustration given in Figure 43.4 for the same cohort, using follow-up time as the new time axis, is significantly different from the previous one. In these *follow-up-time cohorts*, the same subjects are ranked according to the length of follow-up time in the study with zero-time being the time of diagnosis or treatment start.

The question of which of the two forms one should use for the purposes of data analysis rests on one's judgment of the more relevant of the two time axes, essentially the one for which the risk varies most over time, called the primary time axis, with respect to risk and drug exposure. This decision is important, since it affects the demarcation of "risk-sets," which are fundamental to the analysis of data from cohorts and consequently the sampling designs within cohorts. A risk-set is formed by the members of the cohort who are at risk of the outcome event at a given point in time; namely they are free of the outcome event and are members of the cohort at that point in time, called the index date. Drug exposure measures are then anchored at this index date. It is clear that Figures 43.3 and 43.4 produce distinct risk-sets for the same cases in the same cohort, as illustrated by the different sets of subjects crossed by the vertical broken line for the same case under the two forms of the cohort. In Figure 43.3, for example, the first chronological case to occur has in its risk-set only the first six subjects to enter the cohort, while in Figure 43.4, all 21 cohort members belong to its risk-set at the time that the first case arises. While the second form based on disease duration is often used, because in pharmacoepidemiologic drug exposure can vary substantially over calendar time, the first form may be as relevant for the formation of risk-sets and data analysis as the second form. Regardless, an advantage of having data on the entire cohort is that the primary time axis can be changed according to the study question, using calendar time for one analysis, duration of disease or drug exposure for another, with respective adjustment in the analysis for the effect of the other time axis.

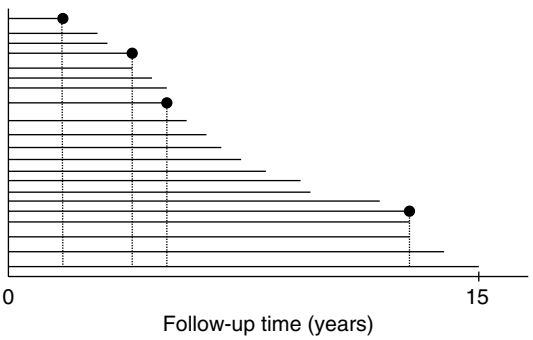

Figure 43.4 Illustration of *follow-up-time* cohort representation after rearranging the cohort in Figure 43.3, with the new risk-sets (---) for the four cases.

The Nested Case–Control Design

The notion of a nested case–control design within a cohort was first introduced by Mantel [11], who proposed an unmatched selection of controls and called it a synthetic retrospective study. It was developed further and formalized by Liddell *et al.* [12] in the context of a cohort study of asbestos exposure and the risks of lung cancer and mortality. The modern nested case–control design involves four steps:

1) defining the cohort time axis, as above
2) selecting all cases in the cohort, i.e., all subjects with an outcome event of interest
3) forming a risk-set for each case and
4) *randomly* selecting one or more controls from each risk-set.

Figure 43.5 illustrates the selection of a nested case–control sample from a cohort, with one control per case (1:1 matching). It is clear from the definition of risk-sets that a future case is eligible to be a control for a prior case, as illustrated in the figure for the fourth case (the circle occurring last in time), and that a subject may be selected as a control more than once. A bias

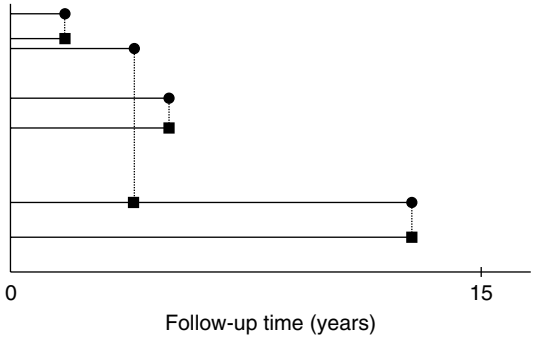

Figure 43.5 Nested case–control sample of one control (■) per case (●) from cohort in Figure 43.4.

is introduced in the estimation of the relative risk if controls are forced to be selected only from the noncases and subjects are not permitted to be used more than once in the nested case–control sample, since the control exposure prevalence will be slanted to that of longer term subjects who do not become cases during the study follow-up [13]. The magnitude of the bias depends on the frequency of the outcome event in the cohort; the more frequent the event, the larger the potential for bias.

This property leading to subjects possibly being selected more than once in the sample may be challenging when the exposure and covariate factors are time dependent, particularly when the data are obtained by questionnaire where the respondent would have to answer questions regarding multiple time points in their history. This issue arose in a study of the risks of severe adverse events in asthma associated with the use of inhaled beta-agonists [14]. A cohort of 12 301 asthmatics spanning the period 1978–87 was identified from the Saskatchewan Health computerized databases, of whom 129 were cases (death or near-death from asthma). A nested case–control approach was needed to permit the collection of additional data from hospital charts and questionnaires sent to all physicians who saw these patients. These additional data were time dependent, focusing on the two-year period prior to the index (risk-set) date. A standard nested case–control sample of six controls per case, as described above, would have likely produced some case and control subjects who contributed multiple times as controls in the sample. This would have added complexity to the questioned physicians who, for example, would have had to respond to questions about the same patient's asthma severity in different two-year periods, a potentially confusing data collection scheme. In part to circumvent this difficulty, the cohort was stratified according to various potential confounding factors, namely age, area of residence, social assistance, prior asthma hospitalization and calendar date of entry into the cohort. This fine stratification resulted in 129 mutually exclusive subcohorts, one leading to each case, and between two and eight controls per case (some risk-sets contained only two eligible controls). Since each subcohort contained a single risk-set (only one case) and the subcohorts were mutually exclusive, a selected subject was guaranteed to appear only once in the nested case–control sample.

The analysis of data from a nested case–control study must preserve the matched nature of the selection of cases and controls, particularly if the risk of the event changes with disease duration and drug exposure varies in calendar time. The method of analysis is identical to that of a conventional matched case–control study, not nested within a cohort. The conditional logistic regression method for this design is appropriate, as it uses the risk-set as the fundamental unit of analysis, in agreement with the proportional hazards model of the full cohort [15]. Simple formulae exist to estimate the relative risk for 1:1 matching [16].

The question of the required number of controls per case is important (see also Chapter 4). Although selecting one control per case will greatly simplify the data analysis, a large number of cases will be required to attain an acceptable level of power. Since the number of cases in the cohort is fixed and cannot be increased to

satisfy this requirement, the only remaining alternative is to increase the control-to-case ratio. Tables for determining the power for given numbers of controls are given in Breslow and Day [17], and Appendix A in this book. It can be readily seen from these sample size tables that the gain in power is significant for every additional control up to four controls per case, but becomes negligible beyond this ratio. For example, if we consider an exposure prevalence in the controls to be 30% and we target detecting a relative risk of 2 with 5% significance and 80% power, the required numbers of cases are 122, 90, 74, 65, and 62, respectively, for 1:1, 2:1, 4:1, 10:1, and 20:1 control-to-case ratios. These translate to total study sizes (cases and controls combined) of 244, 270, 370, 715, and 1302, with clear cost implications and related optimality decisions. Of course, the number of cases in a cohort is frequently fixed *a priori* by the study constraints, thus eliminating this option to increase the number of cases.

However, although this general rule of an optimal 4:1 control-to-case ratio is appropriate in the majority of instances, one should be prudent when exposure to the drug under study is infrequent, or when several factors or other drugs are being assessed simultaneously. In these situations, the ratio could easily be required to increase to 10 or more controls per case. This was the case in two recent nested case-control studies, within a cohort of over 40 000 patients with rheumatoid arthritis, where 100 controls per case were used to obtain sufficiently stable estimates of the rate ratios of serious hepatic events (n=25 cases) and interstitial lung disease (n=74 cases) associated with the use of disease-modifying antirheumatic drugs (DMARD) [18,19].

The appropriate method to perform external comparisons using data from a nested case–control design has been described [20]. It uses knowledge about the sampling structure to yield an unbiased estimate of the outcome event rate in the full cohort, thus permitting the estimation of the necessary standardized relative measure with respect to the selected external population.

Finally, the "nested case–control" label has led to some misunderstandings, including the usual presentation of data as a comparison between "cases" and "controls" rather than by exposure, as well as the convoluted way that forward-looking associations from exposure to outcome extracted from backward-looking data. Moreover, the nested case–control approach provides estimates of the odds ratio, not a rate difference. However, the fact that it is nested within a clearly defined cohort with known sampling fraction allows estimation of risks and rates [21]; the quasi-cohort approach utilizes this property to address these concerns [22]. A quasi-cohort approach was used to assess the risk of pneumonia associated with inhaled corticosteroids in patients with asthma [23].

The Multi-time Case–Control Design

The multi-time case–control design has been introduced recently as an alternative strategy to improve the precision of the odds ratio in a case–control study with transient time-varying exposures, in a setting where increasing the number of control subjects is too costly. This approach is based on increasing the number of observations per control subject, by measuring drug exposure at many different points in time. Indeed, several case–control studies will collect extensive data on time-dependent exposures but use only a portion of these data in estimating the rate ratio.

For example, the International Agranulocytosis and Aplastic Anemia Study (IAAAS) assessed the risk of agranulocytosis associated with the use of analgesics using a case–control study of 221 cases of agranulocytosis and 1425 controls [24]. While the study collected data on exposure for four weeks prior to the index date, only one week's worth of data was used in the analysis. The multi-time case–control approach allows the use of all available exposure data during the four weeks (i.e., four control person-moments)

rather than only one week (i.e., one control person-moment) to improve the precision of the odds ratio estimate, which must however be corrected for within-subject correlation.

This design increases the number of control observations per case, thus potentially also increasing the power of the study without adding additional subjects [25]. For example, in a nested case–control study within a cohort of 12090 patients with chronic obstructive pulmonary disease (COPD), there were 245 incident cases of acute myocardial infarction (AMI) that occurred during follow-up, for whom one and 10 controls per case were identified [25]. The rate ratio of AMI associated with use of antibiotics in the month prior to the index date was 2.00 (95% confidence interval [CI] 1.16–3.44) with one control per case. The precision (as reflected in the confidence intervals) was improved by increasing to 10 controls per case with a rate ratio of 2.13 (95% CI 1.48–3.05). Alternatively, keeping only one control patient per case but increasing the number of control time windows per subject from one to 10 (taken as 10 control exposure measures, one for each of the 10 months prior to the index date) also improved the precision with a rate ratio of 1.99 (95% CI 1.36–2.90).

The Case–Cohort Design

The first recognized application of a sampling design we currently call case–cohort was made by Hutchison [26], in performing external comparisons of leukemia rates in patients treated by radiation for cervical cancer. It was ultimately developed and formalized by Prentice [27], who coined the name "case–cohort." Although recent, this design has already been used effectively in some drug risk studies [28–31]. The case–cohort design involves two steps:

1) selecting all cases in the cohort, i.e., all subjects with an adverse event; and
2) randomly selecting a sample of predetermined size of subjects from the cohort, irrespective of case or control status.

Figure 43.6 depicts the selection of a case–cohort sample of six subjects from the illustrative cohort. Note that it is possible that some cases selected in step 1 are also selected in the step 2 sample, as illustrated in the figure for the third case.

The case–cohort design resembles a reduced version of the cohort, with all cases from the full cohort included. It can also be perceived as an unmatched version of the nested case–control design, with all cases compared with a random sample of the cohort used as controls though not at a specific person-moment. Although these aspects suggest a possible resemblance of the data analysis approach with either the established cohort or case–control methods, the techniques are in fact distinct, each requiring specific software. The analysis of case–cohort sampled data is complex, as it must take into account the overlap of cohort members between successive risk-sets induced by this sampling strategy [32].

The first advantage of the case–cohort design is its capacity to use the same sample to study several different types of events. Indeed, the cases can be split into several subcategories and each can be analyzed with the same "control" subcohort [33]. In contrast, the nested case–control design requires different control groups for each type of event because the selection

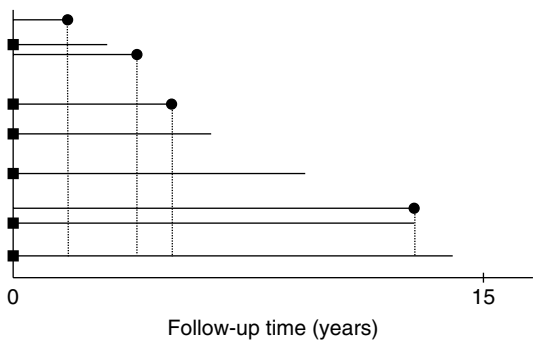

Figure 43.6 Case–cohort sample with six controls (■) from cohort in Figure 43.4.

depends on event times. For example, the beta-agonist risks nested case–control study had two distinct control groups, one of size 233 for the 44 asthma deaths, the other of size 422 for the 85 asthma near-deaths [14]. Another useful advantage is that the case–cohort design permits one to change the primary time axis of analysis from calendar to disease time and vice versa, depending on either the assumed model or the targeted outcome. This is not possible with the nested case–control study, where the primary time axis must be set *a priori* to permit the risk-set construction. This is less of a problem in pharmacoepidemiology, however, where the cohort can be divided into subcohorts of successive calendar time, as was discussed earlier. Yet another example is its simplicity in sampling, which has advantages in both comprehensibility and computer programming. Finally, external comparisons are simple to perform with the case–cohort approach [34].

The nested case–control design does have some advantages over the case-cohort design. The first is the simplicity of power calculation, or equivalently sample size determination. The nested case–control design is independent of the size of the cohort, while for the case–cohort design knowledge about overlap in risk-sets is essential, thus greatly complicating these calculations. Second, data on time-dependent exposure and covariates need only be collected up to the time of the risk-set for the nested case–control study, while the collection must be exhaustive for the case–cohort. Finally, despite the accessibility of software for data analysis of case–cohort data, these can quickly become surpassed and even infeasible with some of the huge sample sizes in some databases and multiple time-dependent exposures. In this situation, the nested case–control design, with its single risk-set per case, is not only advantageous but also the only solution. A study of benzodiazepine use and motor vehicle crashes, initially designed as a case–cohort study, had to be analyzed as a nested case–control study because of technical limitations of the case–cohort analysis software and hardware [35].

An obvious practical consideration is that the case–cohort sampling design can be used to study multiple endpoints in a single analysis (in contrast to case–control sampling) while the case–control study can easily consider many exposures. Depending on the clinical context, one might have strong preferences. As pointed out earlier, in database analyses the main use of cohort sampling designs is when additional information is needed that is time consuming or expensive to collect. If, for example, one engages in outcome validation via expensive chart review, a case–control analysis is often embedded in a cohort study [36]. On the other hand, if baseline biomarkers need to be obtained to determine patient subgroups or improve confounding control the case–cohort design is more suitable [37].

Prevalent New-User Designs

A common situation in pharmacoepidemiology involves the study of the effect of a new drug entering the market, with the best comparator being an older drug. Most often, patients prescribed the new drug will have been switched from the older comparator drug. An incident new-user cohort design based on incident new users of the study and comparator drugs, including only patients who were treatment naive to both drugs, would be optimal. However, it would exclude the possibly large number of subjects who switched from the older to the new drug, a clinically relevant subset. The prevalent new-user design provides an approach to include these switchers [38].

A prevalent new-user cohort is formed from the base cohort of all users of the comparator drug and of the drug under study, which inherently includes the subjects who switched from the comparator to the study drug those who initiated the study drug *de novo*. These latter subjects can directly be matched to contemporaneous initiators on covariates or propensity scores (see

below). For the subjects who switched from the comparator to the study drug, comparators can be selected from the base cohort by matching conditional on exposure sets. Time-based exposure sets can be defined, within the base cohort, by the time from the first prescription of the comparator drug up to the point of switching, while prescription-based exposure set are defined by the number of prescriptions of the comparator drug received up to the point of switching. Because of the granularity of the time scale, time-based exposure sets must be defined with a time interval (such as ±1 month) where all patients with a comparator prescription in the time interval belong to the exposure set and the set is defined by each patient's prescription date. Thus, with either type of exposure set, each switcher to the study drug will belong to an exposure set that includes subjects of similar duration or prescription history with a dispensing of the comparator drug. The importance of the exposure sets is that a visit occurred where the physician decided to either continue the comparator treatment or switch to the new study drug. The exposure set provides equivalent time points in the disease course at which confounding patient characteristics can be measured and controlled for.

To identify, within the exposure sets, the comparator drug users most similar to the patients who switched to the study drug, time-conditional propensity scores (TCPS) can be used [38]. The time-dependent Cox proportional hazards model can be used to compute the "propensity" of switching to the study drug, versus continuing on the comparator drug, as a function of the time-varying patient characteristics measured at the point of the exposure set, thus conserving the matching induced by the exposure set and avoiding adjusting for causal intermediates. The model is used to compute the time-conditional propensity scores within each exposure set, thus identifying their matched comparator as the one with the closest value to that of the switcher. For the purposes of the positivity assumption, the time-conditional propensity score of the switcher should lie within the range of the time-conditional propensity scores of the members of the corresponding exposure set. To emulate the randomized trial, the selection process can be initiated with the first chronological index study drug subject and repeated sequentially. Additionally, once a patient has been selected into the comparator group, they are not considered any longer in subsequent exposure sets as potential comparators. Thus, each subject who initiated the study drug will have a comparator user, matched on the time-conditional propensity score. Cohort entry is taken as the date of the first prescription of the study drug and the corresponding date for the matched comparator. If the exposure sets are too large to compute the time-conditional propensity scores by the time-varying Cox model, an alternative is to select random samples of prescriptions from each exposure set using conditional logistic regression, matching on the exposure set, with the relative odds estimating the relative hazards. The computed propensity odds score for the index switcher is used to identify the corresponding matched patient as the subject with the closest value from all members of the exposure set, not only the sampled ones.

This approach is useful for studies having a "nonuse" comparator, by using a physician visit or prescription for any drug other than the study drug as the comparator. Several questions remain regarding this design [38]. In particular, potential bias from using the prevalent users as comparators should be investigated by stratification on the incident/prevalent new-user status.

Within-Subject Designs

When dealing with the study of transient drug effects on the risk of acute adverse events, Maclure asserts that the best representatives of the source population that produced the cases would be the case subjects themselves; this is

the premise of the case–crossover design [39]. This is a design where comparisons between exposures are made within subjects, thus significantly attenuating the problem of confounding. An extension to the case–crossover design, the case–time–control design, has been proposed and is also presented here.

Case-Crossover Design

The case–crossover study is simply a crossover study *in the cases only*. The subjects alternate at varying frequencies between exposure and nonexposure to the drug of interest, until the adverse event occurs, which happens for all subjects in the study sample, since all are cases by definition. With respect to the timing of the adverse event, each case is investigated to determine whether exposure was present within the predetermined effect period, namely within the previous four hours in our example. This occurrence is then classified as having arisen either under drug exposure or nonexposure on the basis of the effect period. Thus, for each case, we have either an exposed or unexposed status, which represents for data analysis the first column of a 2 × 2 table, one for each case. Since each case will be matched to itself for comparison, the analysis is matched and thus we must create separate 2 × 2 tables for each case.

With respect to control information, the data on the average drug use pattern are necessary to determine the typical probability of exposure during the time window of effect. This is done by obtaining data for a sufficiently stable period of time. In our example, we may find out the average number of times a day each case has been using beta-agonists (two inhalations of 100 μg each) in the past year. Note that there are six four-hour periods (the duration of the effect period) in a day. Such data will determine the proportion of time that each asthmatic is usually spending time in the effect period and thus potentially "at risk" of ventricular tachycardia. This proportion is then used to obtain the number of cases expected on the basis of time spent in these "at-risk" periods, for comparison with the number of cases observed during such periods. This is done by forming a 2 × 2 table for each case, with the corresponding control data as defined above, and combining the tables using the Mantel–Haenszel technique as described in detail by Maclure [39].

To carry out a case–crossover study, three critical points must be considered. First, the study must necessarily be dealing with an acute adverse event that is alleged to be the result of a transient drug effect. Thus, drugs with chronic or regular patterns of use which vary only minimally between and within individuals are not easily amenable to this design. Nor are latent adverse events, which only occur long after exposure. Second, since a transient effect is under study, the effect period (or time window of effect) must be precisely determined. For example, in a study of the possible acute cardiotoxicity of inhaled beta-agonists in asthmatics, this effect period can be determined to be four hours after having taken the usual dose of two inhalations of 100 μg of the product. An incorrect specification of this time window can have important repercussions on the risk estimate, as we will show in the example below. Third, one must be able to obtain reliable data on the usual pattern of drug exposure for each case, over a sufficiently long period of time (as discussed further below). For our example, we could seek the frequency of use of beta-agonists during the year preceding the adverse event.

We generated data for a hypothetical case–crossover study of 10 asthmatic patients who experienced ventricular tachycardia. These were all queried (also hypothetically) regarding their use of two puffs of inhaled beta-agonist in the last four hours and on average over the past year. The data are displayed in Table 43.1. The fact of drug use within the effect period for the event classification is straightforward. The usual frequency of drug use per year is converted to a ratio of the number of "at-risk" periods to the number of "no-risk" periods, the total number of

four-hour periods being 2190 in one year. Thus, for example, the content of the 2 × 2 table for the first case, who is not found to have been exposed in the prior four-hour period, is (0,1,365,1825), while for the second case, who is exposed, it is (1,0,6,2184). Using the Mantel–Haenszel technique to combine the 10 2 × 2 tables, the estimate of relative risk is 3.0 (95% CI 1.2–7.6).

This method is sensitive to the specification of the time window of effect. For example, if this effect period is in fact only two hours, then the data of Table 43.1 would be affected in two ways: some cases may not be considered exposed any more, and the exposure probabilities will change. By considering as unexposed cases 2 and 4, for instance, who may have been exposed three hours before ventricular tachycardia, and recalculating the appropriate exposure probabilities, the relative risk becomes 2.0 (95% CI 0.3–12.0). On the other hand, if this effect period is in fact six hours long, then the data of Table 43.1 would be affected in two ways: some new cases may now be considered exposed, and the exposure probabilities will change. By considering as exposed cases 3 and 5, for instance, who may have been exposed five hours before ventricular tachycardia, and recalculating the appropriate exposure probabilities, the relative risk becomes 5.0 (95% CI 2.0–12.2). The difference in the magnitude of the risk and the corresponding statistical significance between the various scenarios is indicative of the importance of the need for an accurate specification of the length of the effect period.

This method is valuable when studying an acute adverse event that is alleged to be the result of a transient drug effect. Consequently, it excludes the study of drugs with regular patterns of use that vary minimally within individuals or adverse events which can only result from long extended exposure. Moreover, the case–crossover design requires precise knowledge about the effect period (or time window of effect), although the latter can be varied to investigate the optimum window to use. The design is also very useful when the selection of controls in the usual sense is uncertain. A significant advantage of this design is that it elimi-

Table 43.1 Hypothetical data for 10 subjects with ventricular tachycardia included in a case–crossover study of the risk of ventricular tachycardia in asthma associated with the four-hour period after beta-agonist exposure.

Case #	Beta-agonist use in last 4 hours[a] (E_i)	Usual beta-agonist use in last year	Periods of exposure (N_{1i})	Periods of no exposure (N_{0i})
1	0	1/day	365	1825
2	1	6/year	6	2184
3	0	2/day	730	1460
4	1	1/month	12	2178
5	0	4/week	208	1982
6	0	1/week	52	2138
7	0	1/month	12	2178
8	1	2/month	24	2166
9	0	2/day	730	1460
10	0	2/week	104	2086

[a] Inhalations of 200 μg: 1 = yes, 0 = no.

Note: Rate ratio estimator is $(\Sigma\ E_i\ N_{0i})\ /\ (\Sigma\ (1-\ E_i)\ N_{1i})$.

nates the problem of confounding by factors that do not change over time. It cannot, however, easily address the problem of confounding by factors that do change over time. In this instance, time-dependent data will be required for such confounders, a possibly difficult task.

The case–crossover design is automatically free of control selection bias, which occurs when controls are not representative of the base population from which the cases arose. However, the case was inevitably different during the time period when they took the drug, from the time period when they did not take the drug. Thus, in this design, confounding by indication (see Chapter 33) can be severe. Although such control selection bias (in the usual control sense) is eliminated, case selection bias could be present if case selection is related to the exposure under study. Information bias resulting from the differential quality of recent and past drug exposure data can be a concern but less so if one uses drug exposure data from computerized databases. However, this design requires very precise knowledge of when a drug was actually taken, often a very difficult task in computerized databases, especially with drugs that are taken intermittently, exactly when this design could be useful.

Finally, the case–crossover design is intended to be used with transient exposures; otherwise estimates will be biased towards the null, as was shown empirically in a case–crossover study of the effects of long half-life benzodiazepines and the risk of motor vehicle crashes (MVC) in the elderly [40]. There were 5579 cases of MVC identified from the Province of Quebec, Canada, computerized databases. The case–crossover approach applied to all cases did not show any effect (OR 0.99; 95% CI 0.83–1.19). However, among the cases restricted to subjects with four or fewer prescriptions filled in the previous year (transient use), the odds ratio was 1.53 (95% CI 1.08–2.16]. Thus, it is important to verify this assumption of transient exposure, which may not be met in practice for drug therapies that are given for chronic conditions. This approach has been used successfully in several studies [41–45]. It has also been adapted for application to the risk assessment of vaccines (see Chapter 20) [46].

Case–Time–Control Design

One of the limitations of the case–crossover design is the assumption of the absence of a time trend in the exposure prevalence. An approach that adjusts for such time trends is the case–time–control method. By using cases and controls of a conventional case–control study as their own referents, the *case–time–control design* attempts to limit the biasing effect of unmeasured confounding factors, such as drug indication, while addressing the time trend assumption [47]. The method is an extension of the case–crossover analysis that uses, in addition to the case series, a series of control subjects to adjust for exposure time trends.

The approach is illustrated with data from the Saskatchewan Asthma Epidemiologic Project [14], a study conducted to investigate the risks associated with the use of inhaled beta-agonists in the treatment of asthma. Using a cohort of 12 301 asthmatics followed during 1980–87, 129 cases of fatal or near-fatal asthma and 655 controls were identified. The amount of beta-agonist used in the year prior to the index date was used for exposure. Table 43.2 displays the data comparing low (12 or fewer canisters per year) with high (more than 12) use of beta-agonists. The crude odds ratio for high beta-agonist use was 4.4 (95% CI 2.9–6.7). Adjustment for all available markers of severity, such as oral corticosteroids and prior asthma hospitalizations as confounding factors, lowers the odds ratio to 3.1 (95% CI 1.8–5.4), the "best" estimate one can derive from these case–control data using conventional tools.

To apply the case–time–control design, exposure to beta-agonists was obtained for the one-year current period and the one-year reference period prior to the current period. First, a

Table 43.2 Illustration of a case–time–control analysis of data from a case–control study of 129 cases of fatal or near-fatal asthma and 655 matched controls, and current beta-agonist use.

	Controls					
Cases	**High**	**Low**	**High**	**Low**	**OR**	**95% CI**
Current beta-agonist use (case–control)	93	36	241	414	3.1[b]	1.8–5.4
Discordant[a] use (case–crossover)	29	9			3.2	1.5–6.8
Discordant[a] use (control–crossover)			65	25	2.6	1.6–4.1
Case–time–control	29	9	65	25	1.2	0.5–3.0

[a] Discordant from exposure level during reference time period.
[b] Adjusted estimate from case–control analysis.
CI, confidence interval; OR, odds ratio.

case–crossover analysis was performed using the discordant subjects among the 129 cases, namely the 29 who were current high users of beta-agonists and low users in the reference period, and the nine cases who were current low users of beta-agonist and high users previously. This analysis is repeated for the 655 controls, of which there were 90 discordant in exposure; that is, 65 were current high users of beta-agonists and low users in the reference period, and 25 were current low users of beta-agonists and high users previously. The case–time–control odds ratio, using these discordant pairs frequencies for a paired-matched analysis, is given by (29/9)/(65/25) = 1.2 (95% CI 0.5–3.0). This estimate, which minimizes the effect of unmeasured confounding by disease severity, indicates a very small risk for these drugs.

The case–time–control approach can provide an unbiased estimate of the odds ratio in the presence of confounding by indication, despite the fact that the indication for drug use (in our example, intrinsic disease severity) is not measured, because of the within-subject analysis. It also controls for time trends in drug use. Nevertheless, its validity is subject to several assumptions, including the absence of time-dependent confounders, such as increasing asthma severity over time (an important problem, since new drugs may be more likely to be implemented when disease is most severe), so that caution is recommended in its use [48,49]. This approach has been used successfully in several studies [50–55].

Analytic Approaches for Improved Confounding Control

Balancing Patient Characteristics

Confounding caused by imbalance of patient risk factors between treatment groups is a known threat to validity in nonrandomized studies of treatment effects. A litany of options for reducing confounding is available to epidemiologists [56,57]. Several approaches fit key characteristics of longitudinal healthcare databases well and address important concerns in pharmacoepidemiologic analyses.

Propensity Score Analyses

Propensity score analysis has emerged as a convenient and effective tool for adjusting large numbers of confounders. In an incident user cohort design, a propensity score (PS) is the estimated probability of starting medication A versus starting medication B, conditional on all observed pretreatment patient characteristics. Such prediction of treatment choice based on preexisting patient characteristics fits the structure of the incident user cohort design.

Propensity scores are known as a multivariate balancing tool that balance large numbers of covariates in an efficient way even if the study outcome is rare, which is frequent in pharmacoepidemiology [58]. Estimating the propensity score using logistic regression is uncomplicated. Strategies for variable selection (i.e., the variables to include in the logistic regression model to estimate the propensity score) are well described [59]. Variables that are only predictors of treatment choice but are not independent predictors of the study outcome will lead to less precise estimates and in some extreme situations to bias [60]. Selecting variables based on *P* values is not helpful as this strategy depends on study size, and different variables would be selected or unselected for confounding adjustment if the study size changes, although the confounding effect of each variable may remain unchanged. Once a propensity score is estimated based on observed baseline covariates, there are several options to utilize it in a second step to adjust confounding. Typical strategies include adjustment for quintiles or deciles of the score with or without trimming, regression modeling of the PS, or matching on propensity scores [61]. Matching illustrates the working of propensity scores well.

Fixed ratio matching on propensity scores like 1:1 matching has several advantages that may outweigh its drawback of not utilizing the full dataset in situations where not all eligible patients match. Several matching algorithms are frequently used [62]. They have in common that for each exposed patient with a specific propensity score, one or multiple comparator patients will be picked with a propensity score that is similar within a defined caliper [63]. They vary in how they identify the best matches. Such matching will exclude patients in the extreme PS ranges where there is little clinical ambivalence in treatment choice; we therefore see little or no overlap in data (Figure 43.7). The tails of the PS distribution often harbor extreme patient scenarios caused by unobserved patient characteristics often in patients who are not representative for the majority in clinical practice. Keeping them in the analyses may lead to clinically less relevant

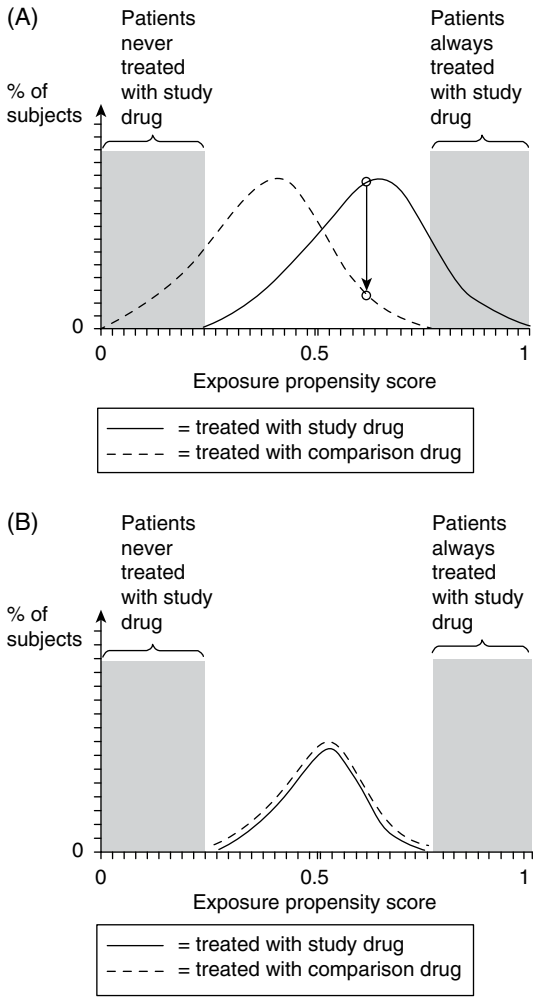

Figure 43.7 Two hypothetical propensity score distributions before and after matching. (A) Before matching: two propensity score distributions partially overlap, indicating some similarities between the comparison groups in a multivariate parameter space. (B) After 1:1 matching on propensity score: not all patients found matches that were similar enough in their multivariable characteristics. Areas of nonoverlap between PS distributions drop out entirely.

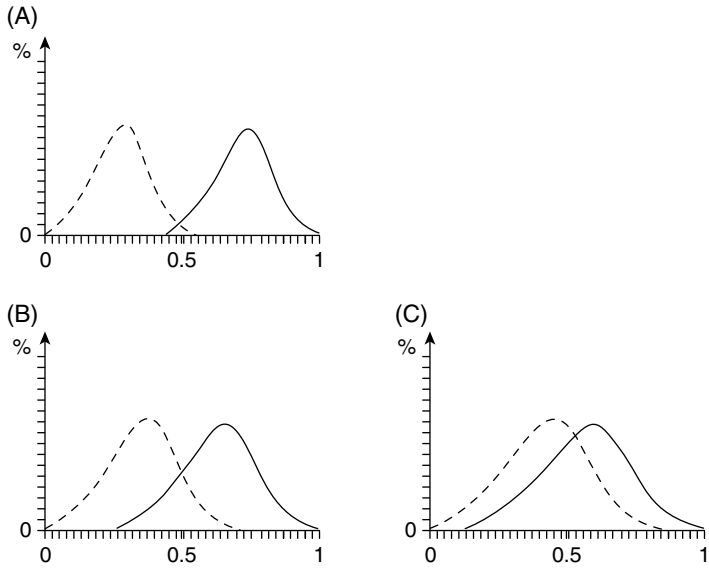

Figure 43.8 Multivariate propensity score distributions with varying degrees of overlap (A, low; B, moderate; C, high) as a diagnostic tool.

findings [64,65]. Trimming the extremes of the propensity score distributions is a data restriction strategy that generally will increase internal validity of findings [66]. Another advantage is that the multivariate balance of potential confounders can be demonstrated by cross-tabulating observed patient characteristics by actual exposure status after fixed ratio matching. 1:1 matching in cohort studies does not require matched analyses, which simplifies the effect estimation to a bivariate analysis. 1:1 matching allows inclusion of all overlapping comparator patients within a defined caliper in the analysis. However, in a variable ratio matching design the matching sets need to be preserved in the analysis to avoid bias. Analytic techniques that condition on the matching sets and may be used in this setting include conditional logistic regression or stratified Cox regression, depending on the data model.

It has been shown that on average, multivariate covariate balance will be achieved between treatment groups when matching on propensity score [67]. If a rational treatment decision process can be modeled well with observed patient characteristics, a resulting propensity score may lead to substantial or even full separation of treated and untreated patients (Figure 43.8A) [68]. This means that for patients initiated on a study drug, very few patients initiated on a comparison drug could be identified who had the same propensity for treatment given the observed patient characteristics. This would leave few comparable patients for analysis. In other words, treatment choice would be almost deterministic; little random treatment choice or empirical equipoise would be left in the prescribing decision that could be exploited for inference about the drug effect.

Consider the comparison of a fixed combination of ezetimibe and simvastatin versus simvastatin alone and their effect on coronary events as an example of such a situation. Assume that a health plan that provides the study data covers the ezetimibe/simvastatin combination only if LDL and HDL levels have crossed certain thresholds: every patient below those thresholds will use simvastatin alone. The LDL and HDL levels therefore become strong if not perfect determinants of treatment choice, and

including them in the propensity score estimation will lead to substantial or complete separation of the PS distributions of the two treatment groups. As the ezetimibe/simvastatin combination continues to be marketed, it will be used less selectively by more and more patients. Consequently, as the prescribers' treatment decisions are less disease state determined (e.g., not driven by LDL/HDL levels in the ezetimibe vs simvastatin example) and increasingly preference based, the propensity score distributions will overlap more and more as a sign that more patients are subject to treatment equipoise (Figure 43.8B,C).

If strong separation of PS distributions is observed, it indicates that the specific comparison cannot be made validly in the study population. In the above example, all ezetimibe and simvastatin users have high LDL level and hardly any simvastatin users have a comparable LDL level. Therefore, very few comparable patients are available for valid inference. This is not a limitation of the method, but rather a very insightful multivariate diagnostic describing the limitations inherent in a study population. The corresponding effect estimates from conventional multivariate outcome models will have substantial imprecision, reflecting the fact that few patients contribute to the estimation despite a large study size. Investigators may want to reconsider the comparison agent and choose a more comparable drug or use another study population where there is less treatment separation in clinical practice.

In summary, propensity score analyses are convenient tools to adjust for many covariates when study outcomes are rare. Extensive confounding adjustment is central in most pharmacoepidemiologic applications and in secondary healthcare databases we can often define many covariates in an effort to reduce the limitation of unobserved or misclassified patient characteristics. As such, PS analyses fit the needs of pharmacoepidemiologists working with longitudinal claims data well. In contrast to traditional outcome models, PS analyses allow the investigator to demonstrate the covariate balance achieved in the final study sample. Postmatching c-statistics or standardized differences of covariates have gained popularity in PS matching analyses [63,69]. PS estimation is well developed for comparing two agents using logistic regression to predict treatment choice. When more than two agents or several dose categories are compared, polytomous regression models are used to estimate the propensity score [70] and either pragmatic pairwise matching to a common reference group or multidimensional matching is applied [71]. Of importance, PS analyses adjust for measured variables, although they can be used to adjust for many at the same time some of which will be proxies for unobserved confounders [72]. Further, one loses the ability to see the effects of adjusting for one variable at a time.

In situations where exposure is rare, disease risk scores, an alternative to propensity score analysis, might be more suitable [73,74]. They estimate the association between patient factors and the study outcome in an unexposed population using multivariate regression and summarize the relationship in each patient's estimated probability of the outcome independent of exposure.

Focusing on the Analysis of Comparable Patients

Restriction is a common and effective analytic tool to make drug user groups more comparable by making populations more homogeneous, which leads to less residual confounding. Some restrictions are quite obvious since they are made by explicit criteria, for example, limiting the study population to elderly patients with dementia to study the safety of antipsychotic medications used to control behavioral disturbances in this population. Other restrictions, like PS matching, are more implicit and blur the line between design choices and analytic strategies to reduce confounding. It is important for pharmacoepidemiologists to understand the

reasons for specific restrictions and their implications for the generalizability of findings.

Choice of Comparator Group

Picking a comparator group is arguably the most fundamental choice in a pharmacoepidemiologic study design and may influence results substantially. Ideally, we want to restrict the comparison population to patients who have the identical indication as the users of the study agent in routine care. Rosiglitazone and pioglitazone are such a medication pair. They were marketed around the same time, were both indicated for second-line treatment of diabetes, come from the same class of compound, and in the early marketing phase were thought to have similar effectiveness and safety profiles. This should make treatment choice largely random with regard to patient characteristics and treatment groups comparable by design, resulting in almost overlapping propensity score distributions and little confounding (see Figure 43.8C). In individual situations, it may be that rosiglitazone-preferring physicians may treat less sick patients or independently produce better health outcomes in comparable patients. However, these physicians may or may not average out with similar pioglitazone-preferring physicians in this setting of treatment equipoise. As indications are usually recorded unreliably and frequently go beyond the labeled indications, picking a comparison drug that implicitly has the identical indication, if available, is usually more fruitful.

Limiting to Incident Users

By restricting the study population to new users of the study agent or a comparator agent, we implicitly require that both groups have been recently evaluated by a physician. Based on this evaluation, the physician has decided that the indicating condition has reached a state where a pharmacologic treatment should be initiated. Therefore, such patients are likely to be more similar in observable and unobservable characteristics than comparing incident users versus nonusers or versus ongoing users of another drug.

Matching on Patient Characteristics

Multivariate propensity scores demonstrate areas of nonoverlap where no referent patients with comparable baseline characteristics can be identified. It is recommended to remove those patients from the analysis as they are not contributing to the estimation and may introduce bias. Such a restriction can be achieved by trimming these patients from the study population [66] or by matching patients on the propensity score or on specific key patient characteristics of importance.

While restriction is an important tool to improve internal validity, it will reduce generalizability of findings. However, in pharmacoepidemiology we usually place higher value on internal validity even if that comes at the price of reduced external validity. Investigators will need to be aware of this trade-off and justify their choices accordingly.

Unobserved Patient Characteristics and Residual Confounding

Once a study is implemented, strategies to reduce confounding further are limited to observable disease risk factors. Secondary data, like electronic healthcare databases, often lack critical details on health state and risk factors, which leads to residual confounding if left unadjusted.

Proxy Adjustment

Longitudinal electronic healthcare databases are as much a description of medical sociology under financial constraints as they are records of delivered healthcare and can be analyzed as a set of proxies that indirectly describe the health status of patients [75]. This status is presented through the lenses of healthcare providers recording their findings and interventions with or without the help of professional coders and operating under the constraints of a specific healthcare system. On several steps along the

way, weighing of medical evidence and treatment options occurred; these are not observable in claims data but collectively resulted in a measurable action. A measured action like the filling of a medication has a clear interpretation but such interpretations are not always possible. In fact, in most cases we cannot determine the exact interpretation, but an exact interpretation may not be required for effective confounder adjustment. For example, old age serves as a proxy for many factors including co-morbidity, frailty, and cognitive decline; use of an oxygen canister is a sign of frail health; having regular annual check-ups is indicative of a health-seeking lifestyle and increased adherence. Adjusting for a perfect surrogate of an unmeasured factor is equivalent to adjusting for the factor itself [76].

The degree to which a surrogate is related to an unobserved or imperfectly observed confounder is proportional to the degree to which adjustment can be achieved [77,78]. Frequently used proxies in pharmacoepidemiologic analyses are the number of prescription drugs dispensed, the number of physician visits, and hospitalizations before the index drug exposure. Such measures of healthcare utilization intensity are useful proxies for general health, access to care, and surveillance. They have been shown to meaningfully help adjust for confounding [79].

Proxy adjustment can be exploited by algorithms that systematically search through recorded codes for diagnoses, procedures, equipment purchases, and drug dispensings to identify potential confounders or proxies thereof [72]. The hundreds of proxies that will be identified can then be adjusted for in a large propensity score model. Collinearity may likely occur in such large models. It will not affect estimation validity as the individual parameters estimated in the large propensity score regression will not be interpreted but only used for predicting treatment [58]; however, it may reduce precision [60]. This high-dimensional propensity score approach has been empirically shown to improve confounding adjustment in many settings over and above investigator-selected covariates [72,80–83].

While the semi-automated high-dimensional PS approach is remarkably robust, issues may arise in small studies with few exposed and rare outcomes [84,85]. Generally in such settings PS stratification is more robust [86] and variance estimates may be inflated [87]. Although adjusting for variables that are only related to the exposure and not to the outcome (an instrumental variable) could theoretically increase bias [60], in practical scenarios the advantage of adjusting for potential confounders outweighs the risk of adjusting for the rare instrument according to a recent simulation study [88]. A challenge remains that, empirically, it is very difficult to know with enough certainty whether a variable is a confounder or an instrument.

Exploiting Random Aspects in Treatment Choice Via Instrumental Variable Analysis

As explained earlier, we are interested in identifying residual random exposure variation after adjusting for observable confounders in order to more completely account for residual confounding. However, in secondary data such as longitudinal claims databases, electronic medical records, or registries, not all clinically relevant risk factors of the outcome may be recorded. To attempt to address this limitation, we can try to identify naturally occurring quasi-random treatment choices in routine care. Factors that determine such quasi-random treatment choices are called instrumental variables (IVs), and IV analyses can result in unbiased effect estimates even without observing all confounders if several assumptions are fulfilled (discussed further later).

An instructive example of an instrument is a hospital drug formulary. Some hospitals list only drug A for a given indication and other hospitals list only drug B. It is a reasonable assumption that patients do not choose their preferred hospital based on its formulary but rather on location and recommendation. Therefore, the

choice of drug A versus drug B should be independent of patient characteristics in the hospitals with these restricted formularies. Thus, comparing patient outcomes from drug A hospitals with patient outcomes from drug B hospitals should result in unbiased effects of drug A versus drug B, using the appropriate analytic tools. An example of such a study is one on the risk of death from aprotinin, an antifibrinolytic agent given to reduce bleeding during cardiac surgery [89]. The study identified surgeons who always used aprotinin and compared their outcomes to surgeons who always used aminocaproic acid, an alternative drug. If physician skill level and performance are on average equal between institutions, independent of drug use, this will result in valid findings. On the other hand, of course, such an assumption may not be valid, for example if academic hospitals allow less restrictive formularies, are more likely to see sicker patients, and have skilled physicians, all of which may be true.

Instrumental variable analyses rely on the identification of a valid instrument, a factor that is assumed to be related to treatment, but neither directly nor indirectly related to the study outcome. As such, an IV is an observed variable that causes (or is a marker of) variation in the exposure similar to random treatment choice. Typically, the following three assumptions need to be fulfilled for valid IV estimation: 1) an IV should affect treatment or be associated with treatment choice by sharing a common cause – the strength of this association is also referred to as the instrument strength; 2) an IV should be a factor that is as good as randomly assigned, so that it is unrelated to patient characteristics; and 3) an IV should not be related to the outcome other than through its association with treatment. As such, an IV analysis sounds very much like a randomized trial with noncompliance. The flip of a coin determines the instrument status (treat with A vs treat with B) and the amount of random noncompliance determines the strength of the instrument. In nonrandomized research, however, identifying valid instruments is difficult and successful IV analyses are infrequent. In principle, treatment preference can be influenced by time if treatment guidelines change rapidly and substantially. A comparison of patient outcome before versus after a sudden change in treatment patterns may then be a reasonable instrument [90,91]. Table 43.3 summarizes a list of some published IV analyses in healthcare.

More commonly, IV analyses utilize individual, local, or regional treatment preferences. For example, Brookhart *et al*. used physician prescribing preference to study the effect of analgesic treatment with COX-2 selective inhibitors (coxibs) versus nonselective NSAIDs (nsNSAID) on the risk of upper gastrointestinal (GI) bleed [93]. Many variations in defining this preference were tested and a reasonable instrument implementation turned out to be the same physician's analgesic prescription (IV status = coxib vs nsNSAID) to the previous patient who needed an analgesic [97]. The authors could demonstrate that such preference is a fairly strong instrument compared to instruments often used in economics. However, despite additional adjustment for observed patient characteristics and general quality of care [98,99], sicker patients may still cluster in coxib-preferring practices and be associated with GI bleed, which would invalidate the IV analysis. Stuckel et al. [94] used regional variation in the rate of cardiac catheterization (IV status = high vs low rate) to estimate its effect on post-MI mortality. While this regional preference instrument was weaker than the physician prescribing preference, it was argued that the instrument was more valid as it is less likely that patients would move to another region to receive the preferred care rather than simply switching their physician.

An IV analysis is technically fairly straightforward once all IV assumptions are fulfilled. In the case of a dichotomous instrument (Z) and exposure (X), the classic IV estimator is:

$$\beta_{IV} = \frac{E[Y \mid Z=1] - E[Y \mid Z=0]}{E[X \mid Z=1] - E[X \mid Z=0]}$$

Table 43.3 Selected examples of instrumental variable analyses in clinical epidemiology.

Instrument group	Instrument type	Examples
Sudden changes in treatment preference over time	Regulatory or coverage interventions	Johnston *et al.:* Beta-blocker use after heart failure hospitalization before and after 1998 [90]
	Innovations and rapid adoption	Juurlink *et al.:* Triamterene use in patients with hypertension before and after the RALES trial [91]
Provider treatment preference	Distance to specialist provider	McClellan *et al.:* Distance to cardiac cath lab facility in patients with acute myocardial infarction (MI) [92]
	Physician prescribing preference (PPP)	Brookhart *et al.:* Physician's treatment initiation choice to the preceding patient [93]
	Regional treatment preference	Stukel *et al.:* Variation of cardiac catheterization rates in 530 US regions in patients with MI [94]
	Hospital formulary/ surgeon treatment preference	Schneeweiss *et al.:* Cardiac surgeons who always use aprotinin as antifibrinolytic agent [89]
	Medication co-payment level	Cole *et al.:* Medication co-payment level in patients with heart failure and adherence [95]
	Dialysis center preference	Thamer *et al.:* Epo dosing by nonprofit vs for-profit dialysis centers [96]

where Y is the study outcome and β is a measure of the effect of X on Y [100]. The numerator of this estimator is the effect of the instrument status (coxib-preferring physician vs not) on the outcome measured as a risk difference. The denominator is the association between instrument status and actual treatment and is a measure of the strength of an instrument. In the case where the instrument perfectly predicts the treatment (e.g., in the example of a restrictive hospital formulary), then the denominator is 1 and the IV estimator will be identical to the naive risk difference estimate. As the instrument gets weaker, the denominator shrinks and the IV estimator increases relative to the naive risk difference estimate. The denominator is sometimes called a rescaling parameter as it scales up the naive risk difference estimate.

In practice, IV analyses use two-stage regression models that allow additional adjustment for multiple observed characteristics. These can be linear models to estimate risk differences or nonlinear models for risk ratio estimation [101]. Brookhart *et al.* have suggested several empirical tests to investigate the quality of an instrument in healthcare effectiveness research [4]. However, such strategies cannot test all assumptions and only help to rule out unsatisfactory IVs rather than confirm valid IVs. Fundamentally, the price of potentially unbiased estimation in IV analyses is the ultimately untestable assumptions that the authors will have to argue based on substantive knowledge and some empirical data. Because of the two-stage estimation, IV analyses are generally less precise which can, in some situations, severely reduce their utility for decision making. Users should also be cautioned that IV inference is based on those "marginal" patients whose treatment decision is influenced by the IV status. This concept is somewhat similar to propensity score analyses where only patients in the overlapping area of propensity score distributions contribute to the multivariate analysis. The IV analyses make an assumption of random treatment choice based on the nature of the healthcare system while

propensity score estimation is trying to utilize unexplained random treatment variation that is left after adjusting for all measured confounders.

Supplementing Database Studies with Clinically Rich Data on Potential Confounders

Resources and time permitting, another strategy to mitigate residual confounding is to identify a subsample and observe among a small number of patients detailed information on potential confounders (see sections earlier). A common version thereof is the nested case–control design or the case–cohort design where only a sample of controls or a sample of exposed and unexposed will be used to collect detailed confounder information. Eng *et al*. demonstrated the use of a case–cohort design embedded in a much larger claims-based analysis [102]. The two-stage sampling approach samples patients according to their exposure and outcome status simultaneously and then reweights findings [103]. Collet *et al*. demonstrated two-stage sampling in a Canadian healthcare database [104]. Increasingly, it is possible to link information-rich electronic health records or registry data in subsets of patients of large claims data studies. It is used to demonstrate that balance had been achieved in patient characteristics that were not observable in claims data [102]. In a new-user cohort study of oral antidiabetic medications with propensity score matching, it was demonstrated that laboratory test results, BMI, and duration of diabetes were well balanced although these parameters were only observable in the subset of EHR-linked patients and not part of the claims data analysis [105]. Such a process is less resource intensive and can be routinely applied in the right data environment.

From the perspective of secondary database studies, all these approaches can be described as internal validation studies, as patients are identified within the underlying study cohort and then contacted to retrieve more details on patient characteristics [106]. The advantage of these approaches is that they are tailored towards the specific question at hand, that is, the sampling as well as the confounder information of interest can be defined by the investigator. However, these approaches are operationally not necessarily efficient ways to collect information. They are often time-consuming since patients need to be identified and information needs to be collected.

An alternative approach is to utilize detailed confounder information that was already collected and then can be tied into the adjusted analysis of the main study cohort. If additional information is available elsewhere, such as a routinely conducted survey of a representative sample of the main database study, such external data sources can be used for reducing residual confounding under certain assumptions [107,108]. For example, each year the Medicare Current Beneficiary Survey routinely studies a representative sample of Medicare beneficiaries to measure a wide variety of characteristics that are not captured in Medicare claims data, for example limitations in activities of daily living [109], cognitive impairment, and physical impairments [110]. If such surveys are truly representative of the study cohort and data are already collected then such external adjustment has the advantage of being much faster and less costly. As the exact study question is not known when the external survey is conducted, it is recommendable to include a wide battery of patient characteristics in the questionnaire.

The available algebraic methods for such external adjustment [108] are limited to single binary confounders and cannot consider the joint confounding arising from several factors. These methods were recently extended to adjustment for multiple confounders of any scale using propensity score calibration (PSC) [111]. The basic concept of PSC is to estimate two multivariate propensity scores in the information-rich survey. One PS mimics the information available in the main study and is seen as an error-prone PS. The second PS uses all available information and is called the complete PS. By regressing the error-prone PS on the complete PS, a calibration factor can be estimated.

With this factor, the error-prone PS-adjusted result in the main study will be calibrated to produce results that are adjusted for the additional factors only available in the more detailed survey data using established regression calibration techniques [112]. Simulation studies have demonstrated good performance of PSC assuming that the relevant confounders were captured in the survey and the survey is representative of the main study [113]. PSC methods can be extended to other than survey data, including electronic medical records or disease registries.

Sensitivity Analyses

A series of sensitivity analyses can help investigators to better understand how robust a study's findings are to implicit and explicit assumptions. Some of the sensitivity analyses suggested below are generic and others are specific to database analyses.

An important but underutilized diagnostic tool for the impact of unobserved confounders on the validity of findings in nonrandomized studies is quantitative sensitivity analyses. Basic sensitivity analyses of residual confounding try to determine how strong and how imbalanced a confounder would have to be among drug categories to explain the observed effect. Such an "externally" adjusted relative risk (RR_{adj}) can be expressed as a function of the unadjusted relative risk (RR_{unadj}), the independent RR of the unmeasured confounder on the disease outcome (RR_{CD}), and the prevalence of the confounder in both drug exposure categories ($P_{C|E}$) [16]:

$$RR_{adj.} = \frac{RR_{unadj}}{\left[\frac{P_{C|E=1}\left(RR_{CD}-1\right)+1}{P_{C|E=0}\left(RR_{CD}-1\right)+1}\right]}$$

A recent cohort study could not find the expected association between use of TNF-alpha inhibitors, an immunomodulating agent, in treating rheumatoid arthritis, and the incidence of serious bacterial infections. There was a concern that physicians may have prescribed the agent selectively in patients with more progressive disease. A sensitivity analysis demonstrated the direction and strength of any such bias and concluded that it would be unlikely to change the clinical implications of the study [114]. This type of sensitivity analysis is particularly helpful in database studies, but is underutilized. Spreadsheet software is available for easy implementation of such sensitivity analyses (drugepi.org) [115]. Lash and Fink proposed an approach that considers several systematic errors simultaneously, allowing sensitivity analyses for confounding, misclassification, and selection bias in one process [116].

When using retrospective databases, it is usually cumbersome or impossible to contact patients and ask when they began using a drug for the first time in order to implement an incident user cohort design. Therefore, incident users are identified empirically by a drug dispensing that was not preceded by a dispensing of the same drug for a defined time period. This washout period is identical for all patients. A typical length is six months. In sensitivity analyses, this interval could be extended to nine and 12 months. In a study on the comparative safety of antidepressant agents in children in British Columbia, this interval was extended from one year to three years to ensure that the children in the study were treatment naive before their first use, which helped balance comparison groups and reduce confounding [117]. Although increasing the length of the washout increases the likelihood that patients are truly incident users, it also reduces the number of patients eligible for the study. This trade-off is particularly worth noting in health plans with high enrollee turnover.

There is often uncertainty about the correct definition of the exposure risk window based on the clinical pharmacology of the study agent. This is further complicated in healthcare databases, since the discontinuation date is imputed through the days' supply of the last dispensing/prescription. Varying the exposure risk window

is therefore insightful and easy to accomplish in cohort studies [118].

Another set of sensitivity analyses concerns the potential for informative censoring. Patients change and discontinue treatment because they lack a treatment effect or experience early signs of a side effect. The more strongly such nonadherence (i.e., drug switching or discontinuation) is associated with the outcome, the more an as-treated analysis, which censors at the point of discontinuation, will be biased. A cumulative risk analysis follows all patients for a fixed time period, carrying forward the initial exposure status and disregarding any changes in treatment status over time. Because this analysis disregards informative nonadherence, it will not suffer bias as a consequence of censoring, but it will suffer bias as a consequence of exposure misclassification. Such misclassification increases with a longer follow-up period and a shorter average time to discontinuation. In most cases, though not all, such misclassification will bias effects towards the null, similar to intention-to-treat analyses in randomized trials. Viewed separately, these two analysis types trade different biases but together, they give a range of plausible effect estimates. Adjusting for nonadherence in an analysis of a drug effect requires information about the predictors of treatment discontinuation [119,120], which is often not available with sufficient accuracy in pharmacoepidemiologic studies.

The Future

Minimizing confounding in nonrandomized pharmacoepidemiologic research is an ongoing development. While great progress has been made in analyzing longitudinal healthcare databases, much remains to be improved in order to reliably achieve unbiased estimates that will carry the weight of medical decision making. Several developments are promising. One is the use of instrumental variable analyses utilizing the multilevel structure of healthcare systems. Another is the expanded use of propensity score methods, including its combination with data-mining activities for high-dimensional proxy adjustment. A development that is gaining importance is the enrichment of existing data environments with supplemental clinical data linked from electronic medical records, disease registries, patient surveys, and/or laboratory test result repositories. While this information will provide an opportunity for improved confounding adjustment, it comes with equally large methodologic challenges, as information is collected in routine care and may have been requested/recorded selectively in patients who were thought to benefit most. Clearly, there is still plenty of work to be done to find satisfactory solutions for the control of confounding in the broad range of pharmacoepidemiologic research.

References

1 Walker AM. Confounding by indication. *Epidemiology* 1996; **7**(4): 335–6.

2 Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003; **158**(9): 915–20.

3 Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010; **19**(8): 858–68.

4 Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf* 2010; **19**(6): 537–54.

5 Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009; **20**(4): 488–95.

6 Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol* 2008; **167**(4): 492–9.
7 Moride Y, Abenhaim L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol* 1994; **47**: 731–7.
8 Schneeweiss S, Patrick AR, Sturmer T, *et al*. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care* 2007; **45**(10 Suppl 2): S131–S142.
9 Psaty BM, Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. *JAMA* 2010; **304**(8): 897–8.
10 Azoulay L, Assimes TL, Yin H, Bartels DB, Schiffrin EL, Suissa S. Long-term use of angiotensin receptor blockers and the risk of cancer. *PloS One* 2012; **7**(12): e50893.
11 Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973; **29**(3): 479–86.
12 Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis appraisal by application to asbestos mining. *J Roy Stat Soc A* 1977; **140**: 469–91.
13 Lubin JH, Gail MH. Biased selection of controls for case–control analyses for cohort studies. *Biometrics* 1984; **40**: 63–75.
14 Spitzer WO, Suissa S, Ernst P, *et al*. The use of beta-agonists and the risk of death and near death from asthma. *N Engl J Med* 1992; **326**(8): 501–6.
15 Breslow N, Day NE. Statistical Methods in Cancer Research. Volume I – The Analysis of Case–Control Studies, 2nd edn. Lyon: International Agency for Research on Cancer, 1980.
16 Rothman KJ, Greenland S. *Modern Epidemiology*, 2nd edn. Hagerstown: Lippincott-Raven, 1998.
17 Breslow N, Day NE. Statistical Methods in Cancer Research. Volume II – The Design and Analysis of Cohort Studies, 2nd edn. Lyon: International Agency for Research on Cancer, 1987.
18 Suissa S, Ernst P, Kezouh A, Bitton A, Hudson M. Newer disease modifying anti-rheumatic drugs (DMARDs) and the risk of serious hepatic adverse events in rheumatoid arthritis. *Am J Med* 2004; **117**(2): 87–92.
19 Suissa S, Hudson M, Ernst P. Leflunomide use and the risk of interstitial lung disease in rheumatoid arthritis. *Arthritis Rheum* 2006; **54**(5): 1435–9.
20 Suissa S, Edwardes MD, Boivin JF. External comparisons from nested case–control designs. *Epidemiology* 1998; **9**(1): 72–8.
21 Rothman KJ. Modern Epidemiology, 3rd edn. Philadelphia: Lippincott Williams & Wilkins, 2008.
22 Suissa S. The Quasi-cohort approach in pharmacoepidemiology: upgrading the nested case–control. *Epidemiology* 2015; **26**(2): 242–6.
23 Qian CJ, Coulombe J, Suissa S, Ernst P. Pneumonia risk in asthma patients using inhaled corticosteroids: a quasi-cohort study. *Br J Clin Pharmacol* 2017; **83**(9): 2077–86.
24 Anonymous. Risks of agranulocytosis and aplastic anemia. A first report of their relation to drug use with special reference to analgesics. The International Agranulocytosis and Aplastic Anemia Study. *JAMA* 1986; **256**(13): 1749–57.
25 Suissa S, Dell'aniello S, Martinez C. The multitime case–control design for time-varying exposures. *Epidemiology* 2010; **21**(6): 876–83.
26 Hutchinson GB. Leukemia in patients with cancer of the cervix uteri treated with radiation. *J Natl Cancer Inst* 1968; **40**: 951–82.
27 Prentice RL. A case–cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; **73**: 1–11.
28 Bergkvist L, Adami HO, Persson I, Hoover R, Schairer C. The risk of breast cancer after estrogen and estrogen–progestin replacement. *N Engl J Med* 1989; **321**: 293–27.

29 van der Klauw MM, Stricker BH, Herings RM, Cost WS, Valkenburg HA, Wilson JH. A population based case–cohort study of drug-induced anaphylaxis. *Br J Clin Pharmacol* 1993; **35**(4): 400–8.

30 Moulton LH, Wolff MC, Brenneman G, Santosham M. Case–cohort analysis of case–coverage studies of vaccine effectiveness. *Am J Epidemiol* 1995; **142**(9): 1000–6.

31 Strom BL, Schinnar R, Bilker WB, Feldman H, Farrar JT, Carson JL. Gastrointestinal tract bleeding associated with naproxen sodium vs ibuprofen. *Arch Intern Med* 1997; **157**(22): 2626–31.

32 Wacholder S. Practical considerations in choosing between the case–cohort and nested case–control designs. *Epidemiology* 1991; **2**: 155–8.

33 Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case–cohort designs. *J Clin Epidemiol* 1999; **52**(12): 1165–72.

34 Wacholder S, Boivin JF. External comparisons with the case–cohort design. *Am J Epidemiol* 1987; **126**(6): 1198–209.

35 Hemmelgarn B, Suissa S, Huang A, Boivin JF, Pinard G. Benzodiazepine use and the risk of motor vehicle crash in the elderly. *JAMA* 1997; **278**(1): 27–31.

36 Schneeweiss S, Doherty M, Zhu S, *et al.* Topical treatments with pimecrolimus, tacrolimus and medium- to high-potency corticosteroids, and risk of lymphoma. *Dermatology* 2009; **219**(1): 7–21.

37 Wacholder S. Practical considerations in choosing between the case–cohort and nested case–control designs. *Epidemiology* 1991; **2**(2): 155–8.

38 Suissa S, Moodie EE, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf* 2017; **26**(4): 459–68.

39 Maclure M. The case–crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991; **133**: 144–53.

40 Hebert C, Delaney JA, Hemmelgarn B, Levesque LE, Suissa S. Benzodiazepines and elderly drivers: a comparison of pharmacoepidemiological study designs. *Pharmacoepidemiol Drug Saf* 2007; **16**(8): 845–9.

41 Ray WA, Fought RL, Decker MD. Psychoactive drugs and the risk of injurious motor vehicle crashes in elderly drivers. *Am J Epidemiol* 1992; **136**: 873–83.

42 Sturkenboom MC, Middelbeek A, de Jong van den Berg LT, van den Berg PB, Stricker BH, Wesseling H. Vulvo-vaginal candidiasis associated with acitretin. *J Clin Epidemiol* 1995; **48**(8): 991–7.

43 Meier CR, Derby LE, Jick SS, Vasilakis C, Jick H. Antibiotics and risk of subsequent first-time acute myocardial infarction. *JAMA* 1999; **281**(5): 427–31.

44 Barbone F, McMahon AD, Davey PG, *et al.* Association of road-traffic accidents with benzodiazepine use. *Lancet* 1998; **352**(9137): 1331–6.

45 Confavreux C, Suissa S, Saddier P, Bourdes V, Vukusic S. Vaccinations and the risk of relapse in multiple sclerosis. Vaccines in Multiple Sclerosis Study Group. *N Engl J Med* 2001; **344**(5): 319–26.

46 Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *Am J Epidemiol* 1996; **143**(11): 1165–73.

47 Suissa S. The case–time–control design. *Epidemiology* 1995; **6**(3): 248–53.

48 Greenland S. Confounding and exposure trends in case–crossover and case–time–control design. *Epidemiology* 1996; **7**: 231–9.

49 Suissa S. The case–time–control design: further assumptions and conditions. *Epidemiology* 1998; **9**(4): 441–5.

50 Hernandez-Diaz S, Hernan MA, Meyer K, Werler MM, Mitchell AA. Case–crossover and case–time–control designs in birth

defects epidemiology. *Am J Epidemiol* 2003; **158**(4): 385–91.

51 Corrao G, Zambon A, Faini S, Bagnardi V, Leoni O, Suissa S. Short-acting inhaled beta-2-agonists increased the mortality from chronic obstructive pulmonary disease in observational designs. *J Clin Epidemiol* 2005; **58**(1): 92–7.

52 Kjaer D, Horvath-Puho E, Christensen J, *et al*. Use of phenytoin, phenobarbital, or diazepam during pregnancy and risk of congenital abnormalities: a case–time–control study. *Pharmacoepidemiol Drug Saf* 2007; **16**(2): 181–8.

53 Zambon A, Polo FH, Contiero P, Corrao G. Effect of macrolide and fluoroquinolone antibacterials on the risk of ventricular arrhythmia and cardiac arrest: an observational study in Italy using case–control, case–crossover and case–time–control designs. *Drug Saf* 2009; **32**(2): 159–67.

54 Park-Wyllie LY, Mamdani MM, Li P, Gill SS, Laupacis A, Juurlink DN. Cholinesterase inhibitors and hospitalization for bradycardia: a population-based study. *PLoS Med* 2009; **6**(9): e1000157.

55 Risselada R, Straatman H, van KF, *et al*. Platelet aggregation inhibitors, vitamin K antagonists, and risk of subarachnoid haemorrhage. *J Thromb Haemost* 2011; **9**(3): 517–23.

56 McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 2003; **12**(7): 551–8.

57 Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005; **58**(4): 323–7.

58 Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; **127**(8 Pt 2): 757–63.

59 Brookhart MA, Sturmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care* 2010; **48**(6 Suppl): S114–S120.

60 Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006; **163**(12): 1149–56.

61 Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol* 2006; **98**(3): 253–9.

62 Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 2): 69–80.

63 Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 2009; **28**(25): 3083–107.

64 Kurth T, Walker AM, Glynn RJ, *et al*. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 2006; **163**(3): 262–70.

65 Lunt M, Solomon D, Rothman K, *et al*. Different methods of balancing covariates leading to different effect estimates in the presence of effect modification. *Am J Epidemiol* 2009; **169**(7): 909–17.

66 Sturmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution – a simulation study. *Am J Epidemiol* 2010; **172**(7): 843–54.

67 Rosenbaum PR, Rubin DB. The bias due to incomplete matching. *Biometrics*1985; **41**(1): 103–16.

68 Sekula P, Caputo A, Dunant A, *et al*. An application of propensity score methods to estimate the treatment effect of corticosteroids in patients with severe cutaneous adverse reactions.

Pharmacoepidemiol Drug Saf 2010; **19**(1): 10–18.

69 Franklin JM, Rassen JA, Ackermann D, Bartels DB, Schneeweiss S. Metrics for covariate balance in cohort studies of causal effects. *Stat Med* 2014; **33**(10): 1685–99.

70 Cadarette SM, Gagne JJ, Solomon DH, Katz JN, Sturmer T. Confounder summary scores when comparing the effects of multiple drug exposures. *Pharmacoepidemiol Drug Saf* 2010; **19**(1): 2–9.

71 Solomon DH, Rassen JA, Glynn RJ, *et al*. The comparative safety of opioids for nonmalignant pain in older adults. *Arch Intern Med* 2010; **170**(22): 1979–86.

72 Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009; **20**(4): 512–22.

73 Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res* 2009; **18**(1): 67–80.

74 Arbogast PG, Kaltenbach L, Ding H, Ray WA. Adjustment for multiple cardiovascular risk factors using a summary risk score. *Epidemiology* 2008; **19**(1): 30–7.

75 Schneeweiss S. Understanding secondary databases: a commentary on "Sources of bias for health state characteristics in secondary databases". *J Clin Epidemiol* 2007; **60**(7): 648–50.

76 Wolldridge JM. Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press, 2001.

77 Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980; **112**(4): 564–9.

78 Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985; **122**(3): 495–506.

79 Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol* 2001; **154**(9): 854–64.

80 Garbe E, Klos, S, Suling M, Pigeot I, Schneeweiss S. High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications. *Eur J Clin Pharmacol* 2013; **69**(3), 549–57.

81 Guertin JR, Rahme E, LeLorier J. Performance of the high-dimensional propensity score in adjusting for unmeasured confounders. *Eur J Clin Pharmacol* 2016; **72**(12): 1497–505.

82 Hallas J, Pottegard A. Performance of the high-dimensional propensity score in a Nordic healthcare model. *Basic Clin Pharmacol Toxicol* 2017; **120**(3), 312–17.

83 Rassen JA, Schneeweiss S. Using high-dimensional propensity scores to automate confounding control in a distributed medical product safety surveillance system. *Pharmacoepidemiol Drug Saf* 2012; **21**(Suppl 1): 41–9.

84 Patorno E, Glynn RJ, Hernandez-Diaz S, Liu J, Schneeweiss S. Studies with many covariates and few outcomes: selecting covariates and implementing propensity-score-based confounding adjustments. *Epidemiology* 2014; **25**(2): 268–78.

85 Rassen JA, Glynn RJ, Brookhart MA, Schneeweiss S. Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am J Epidemiol* 2011; **173**(12): 1404–13.

86 Desai RJ, Rothman KJ, Bateman BT, Hernandez-Diaz S, Huybrechts KF. A propensity-score-based fine stratification approach for confounding adjustment when exposure is infrequent. *Epidemiology* 2017; **28**(2): 249–57.

87 Schuster T, Lowe,WK, Platt RW. Propensity score model overfitting led to inflated variance of estimated odds ratios. *J Clin Epidemiol* 2016; **80**: 97–106.

88 Myers JA, Gagne JJ, Rassen JA, *et al*. [Abstract] Effects of adjusting for instrumental variables on bias and precision of effect estimates. International Conference for Pharmacoepidemiology, 2011.

89 Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med* 2008; **358**(8): 771–83.

90 Johnston KM, Gustafson P, Levy AR, Grootendorst P. Use of instrumental variables in the analysis of generalized linear models in the presence of unmeasured confounding with applications to epidemiological research. *Stat Med* 2008; **27**(9): 1539–56.

91 Juurlink DN, Mamdani MM, Lee DS, *et al*. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med* 2004; **351**(6): 543–51.

92 McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994; **272**(11): 859–66.

93 Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006; **17**(3): 268–75.

94 Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. 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**(3): 278–85.

95 Cole JA, Norman H, Weatherby LB, Walker AM. Drug copayment and adherence in chronic heart failure: effect on cost and outcomes. *Pharmacotherapy* 2006; **26**(8): 1157–64.

96 Thamer M, Zhang Y, Kaufman J, Cotter D, Dong F, Hernan MA. Dialysis facility ownership and epoetin dosing in patients receiving hemodialysis. *JAMA* 2007; **297**(15): 1667–74.

97 Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application – in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol* 2009; **62**(12): 1233–41.

98 Schneeweiss S, Solomon DH, Wang PS, Rassen J, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: an instrumental variable analysis. *Arthritis Rheum* 2006; **54**(11): 3390–8.

99 Brookhart MA, Rassen JA, Wang PS, Dormuth C, Mogun H, Schneeweiss S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Med Care* 2007; **45**(10 Suppl 2): S116–S122.

100 Angrist J.D, Imbens G.W, Rubin B. Identification of Casual effects Using Instrumental Variables. Journal of the American Statistical Association 2002.

101 Rassen JA, Schneeweiss S, Glynn RJ, Mittleman MA, Brookhart MA. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *Am J Epidemiol* 2009; **169**(3): 273–84.

102 Eng PM, Seeger JD, Loughlin J, Clifford CR, Mentor S, Walker AM. Supplementary data collection with case–cohort analysis to address potential confounding in a cohort study of thromboembolism in oral contraceptive initiators matched on claims-based propensity scores. *Pharmacoepidemiol Drug Saf* 2008; **17**(3): 297–305.

103 Walker AM. Anamorphic analysis: sampling and estimation for covariate effects when both exposure and disease are known. *Biometrics* 1982; **38**(4): 1025–32.

104 Collet JP, Schaubel D, Hanley J, Sharpe C, boivin JF. Controlling confounding when studying large pharmacoepidemiologic databases: a case study of the two-stage sampling design. *Epidemiology* 1998; **9**(3): 309–15.

105 Patorno E, Gopalakrishnan C, Franklin JM, *et al.* Claims-based studies of oral glucose-lowering medications can achieve balance in critical clinical parameters only observed in electronic health records. *Diabetes Obes Metab* 2018; **20**(4): 974–84.

106 Sturmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Med Care* 2007; **45**(10 Supl 2): S158–S165.

107 Suissa S, Edwardes MD. Adjusted odds ratios for case–control studies with missing confounder data in controls. *Epidemiology* 1997; **8**(3): 275–80.

108 Schneeweiss S, Glynn RJ, Tsai EH, Avorn J, Solomon DH. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. *Epidemiology* 2005; **16**(1): 17–24.

109 Saliba D, Orlando M, Wenger NS, Hays RD, Rubenstein LZ. Identifying a short functional disability screen for older persons. *J Gerontol A Biol Sci Med Sci* 2000; **55**(12): M750–M756.

110 Eppig FJ, Chulis GS. Matching MCBS (Medicare Current Beneficiary Survey) and Medicare data: the best of both worlds. *Health Care Financ Rev* 1997; **18**(3): 211–29.

111 Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol* 2005; **162**(3): 279–89.

112 Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol* 1990; **132**(4): 734–45.

113 Sturmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Propensity score calibration and its alternatives. *Am J Epidemiol* 2007; **165**(10): 1122–3.

114 Psaty BM, Koepsell TD, Lin D, *et al.* Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc* 1999; **47**(6): 749–54.

115 Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf* 2006; **15**(5): 291–303.

116 Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology* 2003; **14**(4): 451–8.

117 Schneeweiss S, Patrick AR, Solomon DH, *et al.* Comparative safety of antidepressant agents for children and adolescents regarding suicidal acts. *Pediatrics* 2010; **125**(5): 876–88.

118 McMahon AD, Evans JM, McGilchrist MM, McDevitt DG, MacDonald TM. Drug exposure risk windows and unexposed comparator groups for cohort studies in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 1998; **7**(4): 275–80.

119 Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**(3): 779–88.

120 Toh S, Hernandez-Diaz S, Logan R, Rossouw JE, Hernan MA. Coronary heart disease in postmenopausal recipients of estrogen plus progestin therapy: does the increased risk ever disappear? A randomized trial. *Ann Intern Med* 2010; **152**(4): 211–17.