Int J Clin Pharm (2016) 38:714–723
DOI 10.1007/s11096-016-0299-0

REVIEW ARTICLE

Methods to control for unmeasured confounding in pharmacoepidemiology: an overview

Md. Jamal Uddin[1,2] · Rolf H. H. Groenwold[3] · Mohammed Sanni Ali[3] · Anthonius de Boer[2] · Kit C. B. Roes[3] · Muhammad A. B. Chowdhury[4] · Olaf H. Klungel[2]

Received: 15 July 2015 / Accepted: 4 April 2016 / Published online: 18 April 2016

Abstract *Background* Unmeasured confounding is one of the principal problems in pharmacoepidemiologic studies. Several methods have been proposed to detect or control for unmeasured confounding either at the study design phase or the data analysis phase. *Aim of the Review* To provide an overview of commonly used methods to detect or control for unmeasured confounding and to provide recommendations for proper application in pharmacoepidemiology. *Methods/Results* Methods to control for unmeasured confounding in the design phase of a study are case only designs (e.g., case-crossover, case-time control, self-controlled case series) and the prior event rate ratio adjustment method. Methods that can be applied in the data analysis phase include, negative control method, perturbation variable method, instrumental variable methods, sensitivity analysis, and ecological analysis. A separate group of methods are those in which additional information on confounders is collected from a substudy. The latter group includes external adjustment, propensity score calibration, two-stage sampling, and multiple imputation. *Conclusion* As the performance and application of the methods to handle unmeasured confounding may differ across studies and across databases, we stress the importance of using both statistical evidence and substantial clinical knowledge for interpretation of the study results.

Keywords Observational studies · Pharmacoepidemiology · Residual confounding · Review · Statistical methods · Unmeasured confounding · Unobserved confounding

Impacts on practice

- Unmeasured confounding is a potential source of bias in pharmacoepidemiologic studies, which can be detected and controlled.
- The method to detect and control for unmeasured confounding in applied pharmacoepidemiologic research should be carefully selected, and depends on the study and databases.

Introduction

Randomized controlled trials are considered the gold standard to estimate the effect of a pharmacological treatment [1–3]. However, randomized trials may be of limited use or not feasible in several situations, for instance, in the case of rare adverse drug events, or when the outcome of interest is far ahead in the future [1–5]. In such situations, observational pharmacoepidemiologic studies may be viable alternatives to provide important evidence on the comparative safety and effectiveness of drugs [1, 2, 5–8].

✉ Md. Jamal Uddin
jamal-sta@sust.edu

[1] Department of Statistics (Biostatistics and Epidemiology), Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh

[2] Division of Pharmacoepidemiology and Clinical Pharmacology, University of Utrecht, Utrecht, The Netherlands

[3] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

[4] Department of Biostatistics, Robert Stempel College of Public Health and Social Work, Florida International University, Miami, FL, USA

In observational studies, treatment assignment is outside the control of the investigator. In daily clinical practice, treatment assignment is generally based on the physician's perception of a patient's risk of a particular outcome [6, 9]. Because of the absence of randomization, prognostic patient characteristics are typically unevenly distributed among the treatment groups. Hence, a direct comparison between treatment groups will likely be biased due to confounding [6, 9–11].

A confounder is a risk factor for the outcome and is associated with the exposure, yet does not lie on the causal pathway between exposure and the outcome. Moreover, a confounder may bias the observed effect of the exposure on the outcome [12, 13]. For example in Fig. 1, the variable pre-treatment blood pressure is associated with treatment status, angiotensin-converting-enzyme (ACE) inhibitor use, and an independent risk factor for the outcome, myocardial infarction. There may also be unmeasured or poorly measured risk factors (e.g., smoking) of the outcome that are also associated with the exposure. These are referred to as unmeasured, unobserved, or residual confounders [12].

Several methods have been proposed to detect or control for measured and unmeasured confounding either in the study design or the data analysis phase [8, 9, 14–19]. Reviews of these methods often focus on methods to control for measured confounding [9, 20, 21].

Aim of the review

In this review, we focus on methods that are mostly used to detect or control for unmeasured confounding. Hence, the objective of this review is to provide an overview of commonly used methods to detect or control for unmeasured confounding in observational studies and to provide recommendations for proper applications in pharmacoepidemiology.

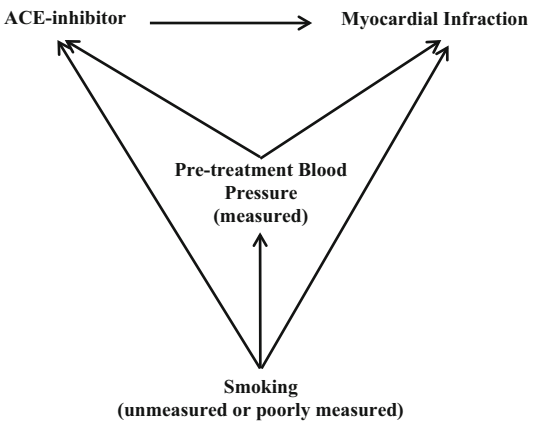

Fig. 1 Directed acyclic graph of observational studies to illustrate the concept of confounding. *Treatment/Exposure* ACE-inhibitor (angiotensin converting enzyme inhibitor), *Outcome* Myocardial Infarction, *Measured confounder* Pre-treatment blood pressure, *Unmeasured/poorly measured confounder* Smoking

Methods

Sources of confounding in pharmacoepidemiology

Pharmacoepidemiologic studies to detect unintended and intended effects of drugs are often conducted using electronic healthcare databases (e.g., pharmacy, general practice, and hospital records). These data are collected for reasons unrelated to a particular research question and, therefore, may have missing, limited, or inaccurate information on potential confounding variables such as smoking, alcohol consumption, body-mass index, frailty, and disease severity [6, 22]. In addition, there may be variables for which it is unknown that they are risk factors for the outcome (e.g., genetic variants). However, it seems unlikely that these unknown variables will affect treatment decisions made by the physician (for the reason that the physician does not know that these are risk factors for the outcome). Hence, unknown risk factors for the outcome may be less likely of a problem in pharmacoepidemiologic studies than known (such as smoking), yet unmeasured, risk factors for the outcome.

Methods to control for measured confounding

Confounding can be controlled by design or in the analysis phase of a study. For an overview of the methods that control for measured confounding, we refer to the review by Klungel et al. [9]. In short, by design, restriction and matching can be used to control for measured confounding. Stratification, standardization, multivariable regression, and propensity score methods can be applied in the analysis phase to control for measured confounders. In the following sections, we will discuss methods to detect or control unmeasured confounding in the context of pharmacoepidemiologic research. We will use the terms "treatment" and "exposure" interchangeably.

Methods to detect or control for unmeasured confounding in the design phase

Case only designs

In case-only designs, only those subjects who experience the outcome of interest are included (i.e., 'cases only') and all subjects act as their own control. Since comparisons are made within individuals, confounding by characteristics that are constant over time, such as sex, is eliminated. Different types of case-only designs include case-

crossover, case-time control, and self-controlled case series. Brief descriptions of these study designs are given below.

Case-crossover design The case-crossover design (CCO) was introduced by Maclure [23] in 1991 to study the short-term effects of intermittent or time-varying exposures on the risk of acute outcomes where a conventional case–control study may not be feasible [23–26]. The CCO design has been used in multiple pharmacoepidemiologic studies to estimate drug effects [26–29]. This design adopts a case–control perspective through a comparison of treatment status before the occurrence of the outcome and treatment status during one or more reference period(s) of the same subjects, in which the outcome did not occur. Data are usually analyzed using a conditional logistic regression model, which accounts for the matched nature of the data. Subjects who chronically use a certain treatment (i.e., in which treatment status does not change) do not contribute to the analysis [30]. Details on the analysis of the CCO design (using e.g., SAS) can be found in Wang et al. [31]. The schematic representation of the CCO design is shown in Fig. 2. In a case-crossover design, exposure status at the time of the outcome (i.e., the hazard period, B) is contrasted with exposure status during a control period (A).

In the CCO design, the choice of exposure time windows is crucial, as well as the timing of the control period [30]. Moreover, exposure misclassification and recall bias are likely if exposure status is assessed, e.g., by questionnaire, as patients may not remember the drugs they have taken, especially for periods long before the outcome occurred. On the other hand, bias can result from temporal changes in confounding variables, e.g., occurrence of a disease. Time-varying confounding can be accounted for in the analysis, provided that the confounders are measured.

Case-time control design In 1995, Suissa [30] proposed the case-time control (CTC) design, also known as case–control-crossover design, which is an extension of the CCO design and uses exposure information from a historical control group [24, 30]. The CTC odds ratio is the case-crossover odds ratio (from the cases) divided by the time trend odds ratio (from the controls) (Fig. 3) [32]. Like the CCO, it allows studying intermittent exposures with

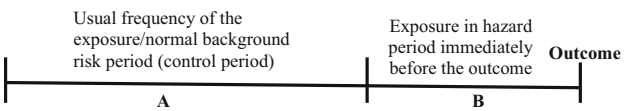

Fig. 2 Schematic representation of the case-crossover design. In a case-crossover design, exposure status at the time of the outcome (i.e., the hazard period, *B*) is contrasted with exposure status during a control period (*A*). The x-axis represents time

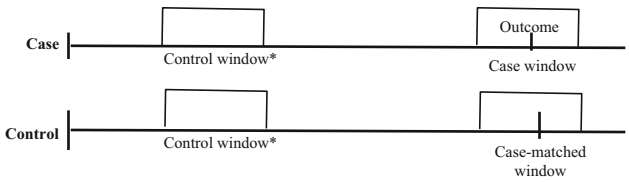

Fig. 3 Schematic representation of the case-time control design. The x-axis represents the time. *Asterisk* Historical control group

transient effects, while accounting for confounding by variables that do not change over time. CTC may be considered if there are strong time trends in treatment prescription [25]. The main advantage of the CTC over CCO is that it adjusts for natural time trends in drug utilization. However, it is less precise than the conventional design and does not deal with time to event and death as an outcome. The Schematic representation of the CTC is shown in Fig. 3. More information about this design can be found in Schneeweiss et al. [25], Hernández-Díaz et al. [32], and Schneider et al. [33].

Self-controlled case-series design Farrington et al. proposed the self-controlled case-series (SCCS) design in 1995 to assess post-licensure adverse events related to vaccines, and more generally associations between acute outcomes and transient exposures [4]. The SCCS design adopts a cohort perspective through a comparison of the rate of the outcome of interest between exposed and unexposed time periods for each individual. Although the SCCS design was originally developed to evaluate vaccine safety, it has been increasingly used in pharmacoepidemiologic studies using healthcare databases [24, 34, 35]. In this design, individual follow-up time is divided into exposed and unexposed periods, which may vary in length of duration. A comparison is made within subjects, between the incidence rates of the outcome of interest in these periods and thus controls for time-fixed confounders (measured or unmeasured) (Fig. 4) [36]. The key assumptions of this design are: (1) events are possibly recurrent and occur randomly over the study period; (2) the occurrence of an event must not alter the probability of subsequent exposure; and (3) the occurrence of an event should not censor the observation period [34–36]. The procedure for implementation of the SCCS design in the statistical software (e.g., SAS, R, and STATA) can be found in Whitaker et al. [36] or in the Ref. [37].

Prior event rate ratio adjustment

The recently proposed prior event rate ratio (PERR) adjustment method [38–42], is a type of self-controlled design in which the exposure effect is estimated by the

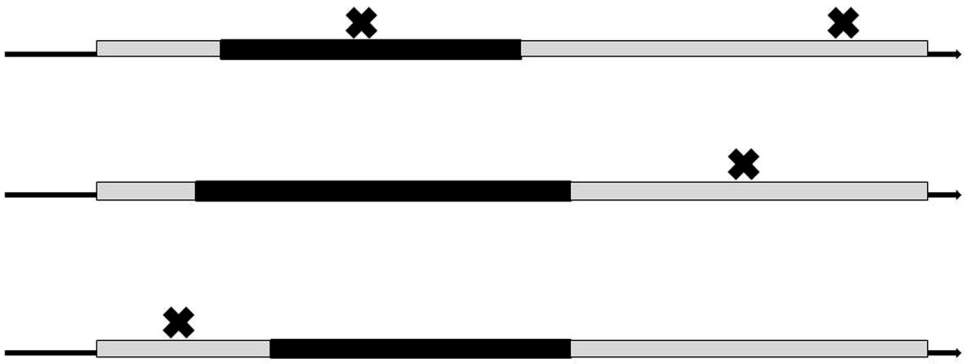

Fig. 4 Schematic representation of the self-controlled case-series design. The x-axis represents time. *Grey* periods indicate baseline (unexposed) periods. *Black* periods indicate exposed time. "*X*" indicates outcome events

ratio of two rate ratios (RRs): RR after the initiation of an exposure ($RR._{post}$) and the RR prior to initiation of that exposure ($RR._{prior}$) (Fig. 5) [15, 39, 43]. The PERR adjustment method can be used in a setting where neither the exposed nor unexposed patients are treated with the study drugs before the start of the follow-up [39]. In this setting, the rate ratio (RR) observed in the prior period (i.e., before initiation of the exposure) is due to differences in patient characteristics between the two study groups (of future users and future non-users). The RR in the post period, after initiation of the exposure, is due to those differences in confounders as well as the treatment under study.

The treatment effect in the post period can be separated from the effect due to differences in confounders by taking the ratio of the RRs observed in the two periods:

$$PERR = \frac{Rate\ ratio\ during\ post\ period}{Rate\ ratio\ during\ prior\ period} = \frac{RR_{post}}{RR_{prior}} \quad (1)$$

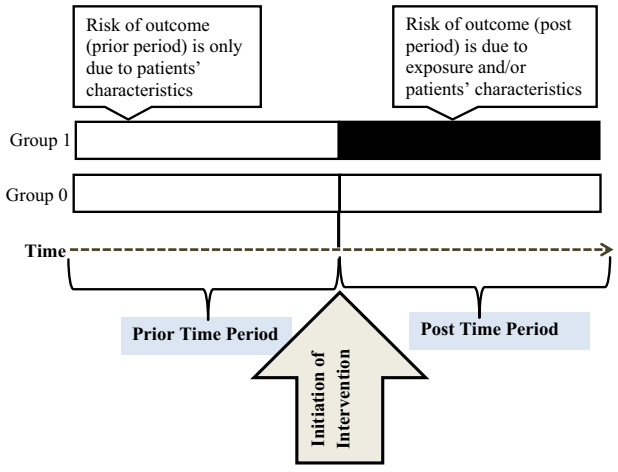

Fig. 5 Prior event rate ratio adjustment method. (Adapted from Uddin et al. [43]). The x-axis represents time. The *prior time period* indicates the time period before an intervention is initiated; the *post time period* refers to the period after initiation of the intervention. The *black bar* indicates exposed subjects. The *white bars* indicate unexposed subjects

PERR can be estimated using incidence rate ratios or hazard ratios [40]. Confidence intervals can be obtained by bootstrapping as it is difficult to estimate the covariance between the RRs of the prior and post periods [38, 39, 41]. This method requires assumptions: constant temporal effects (i.e., confounding effects are constant across prr and post exposure initiation periods), there is no confounder-by-treatment interaction, and outcomes are non-lethal events [14, 39, 42].

Methods to detect or control for unmeasured confounding in the analysis phase

Negative control

Negative controls (NC) have been proposed to detect unmeasured confounding in epidemiological studies [44]. There are two types of negative controls: exposure controls and outcome controls (Fig. 6) [44, 45]. A negative exposure control is an exposure that is known to be unrelated to the outcome under study, whereas a negative outcome control is an outcome that is known to be unrelated to the exposure under study. An example can be found in a study of annual influenza vaccination [46], in which an expected

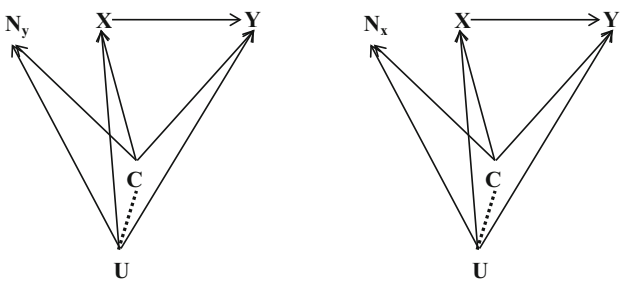

Fig. 6 Directed acyclic graph of an ideal negative control outcome (N_y) and negative control exposure (N_x). Here *X* represents exposure, *Y* outcome, *C* a measured confounder, and *U* represents an unmeasured confounder. *Dashed line* represents a possible correlation between confounders

null-effect of the vaccine during summer periods (when hardly any influenza virus is circulating) served as a control on the assumption of no unmeasured confounding. If the NC is empirically associated with the exposure (or outcome) after adjustment for measured confounders, the observed association may be the result of unmeasured confounding [44, 47, 48]. However, a non-null association will only suggest unmeasured confounding for the relation between the NC exposure and the outcome (or exposure and NC outcome) and does not necessarily indicate unmeasured confounding of the exposure–outcome relation of interest [49]. Hence subject matter knowledge is essential for choosing a negative control [44]. Detailed explanations of the NC can be found in Lipsitch et al. [44] and Flanders et al. [47].

Instrumental variable

Instrumental variable (IV) methods were invented over 70 years ago [18]. They have primarily been used in econometrics and social sciences, but appeared in epidemiologic research over the last decades, because of the potential to control for unmeasured confounding [7, 50–52]. An IV is a variable that satisfies three key assumptions: (1) the IV is associated with the treatment/exposure under study, (2) the IV affects the outcome only through the exposure (exclusion restriction), and (3) the IV is independent of confounders [53–55]. In observational studies, IV analysis tries to mimic a randomized study in which treatment assignment is related to the actual treatment received and treatment assignment only affects outcome through the treatment received (hence, the term pseudo-randomisation that is sometimes used for IV analysis) [55]. As an IV is assumed to affect the outcome only through the treatment/exposure under study and an IV is independent of confounders, this implies that all measured and unmeasured confounders are equally distributed among IV categories (similar to a randomized controlled trial, where measured and unmeasured confounders are equally distributed between treatment arms).

An example of an IV in an observational study is illustrated in Fig. 7 [56]. The idea behind this IV, physician prescribing preference, is that physicians differ with respect to their preference for conventional versus atypical antipsychotic medication: to a similar patient, one physician will prescribe conventional antipsychotic medication while another physician will prescribe an atypical antipsychotic medication. Furthermore, it is assumed that the preference is only influencing mortality (the outcome) through the exposure (conventional or atypical antipsychotic medication) and is independent of patient characteristics [56]. A possible limitation of IV analysis is that of weak instruments (i.e., IV is weakly associated with the exposure), which leads to decreased statistical efficiency and biased IV estimates [53, 57]. An overview of IV analysis with statistical software code is given by Baiocchi et al. [7]. Moreover, more statistical software for IV analyses can be found in Brookhart et al. [58], or in the Ref. [59].

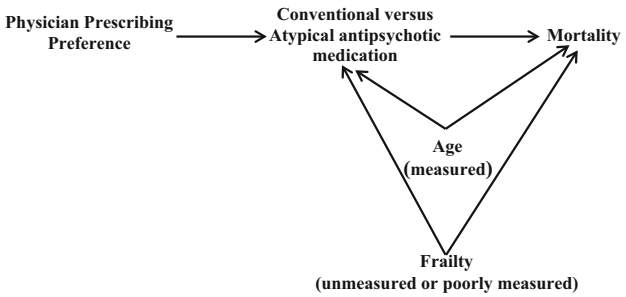

Fig. 7 Directed acyclic graph of an observational study of antipsychotic medication and mortality to illustrate the concept of instrumental variable analysis. *Instrumental variable* Physician prescribing preference, *Treatment/Exposure* Conventional versus atypical antipsychotic medication, *Outcome* Mortality, *Measured confounder* Age, *Unmeasured/poorly measured confounder* Frailty

Sensitivity analysis

In 1959, Cornfield et al. first introduced the idea of sensitivity analysis for unmeasured confounding. In a sensitivity analysis, one can not control for unmeasured confounding, but can merely evaluate what the potential impact of unmeasured confounding can be on the estimate of the treatment–outcome association [17]. Specifically, a sensitivity analysis of unmeasured confounding is a data-driven method where we do need to specify the sensitivity parameters with their distributions [60, 61]. Moreover, it assesses how sensitive an estimated treatment effect is to unmeasured confounding and when conclusions from a study will change [60, 61] (and—importantly—what the unmeasured confounding then should look like). If unmeasured confounding should be unrealistically large (e.g., very strong association between unmeasured confounders and outcome) in order to nullify the observed treatment–outcome relation, researchers may argue that it is unlikely that the observed relation is fully attributable to unmeasured confounding. Sensitivity analyses allow us for making a comparison of different scenarios and there is no restriction on the distribution of the unmeasured confounder(s). For details of sensitivity analysis methods, we refer to the studies by Lin et al. [60], Greenland [62], and Schneeweiss [17].

Perturbation variable

The perturbation variable (PV) analysis is a data-mining approach and has been recommended for use in detecting

(via the perturbation test) and controlling (via adjustment for the PVs) for unmeasured confounding in observational studies [63]. A PV is a variable that should be easy to find or that can be measured easily and is potentially associated, though perhaps only very weakly, with unmeasured confounders (Fig. 8) [44, 47, 63]. The PV method relies on collecting as many PVs as possible, which then indirectly provide information on the unmeasured confounders. When the number of PVs increases, the power of the perturbation test increases and the bias due to unmeasured confounding after PV adjustment decreases [63]. However, this method is data driven and different from the sensitivity analysis of unmeasured confounding where we do need to specify the sensitivity parameters or assume distributions for them [64]. Details on this data-driven method can be found in Lee [63].

Ecological analysis

Wen and Kramer proposed the ecologic analysis as a method to handle confounding by indication at the individual level when a study investigates the intended effects of drugs [65, 66]. This method is based on differences in e.g., exposure patters between regions or countries. Even if confounding acts at the individual level within regions, a comparison across regions controls for that confounding, if confounders are equally distributed across the regions. An example of an ecological analysis is a comparison between, e.g., countries that differ in treatment profiles. If countries differ in terms of occurrence of the outcome, this may be due to differences in treatment. However, countries may differ in more respects than only treatment rates, thus leading to the so-called 'ecological fallacy' [9]. Moreover, it has low statistical power by the reduced number of experimental units and provides little information regarding the individuals in the compared groups [9, 67].

Sub-studies to control for unmeasured confounding

Instead of controlling for unmeasured confounding in the design or analysis phase of a pharmacoepidemiologic study, one could collect additional information on unmeasured confounding in a subset of the study and incorporate this additional information in the overall analysis of the entire study, in order to control for unmeasured confounding. Different types of methods include external adjustment, propensity score calibration, two-stage sampling, and multiple imputations. Brief descriptions of these methods are given below.

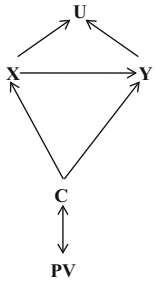

Fig. 8 Relations between exposure (*X*), outcome (*Y*), confounding variable (*C*), perturbation variable (*PV*), and a unmeasured confounder (*U*)

External adjustment

In the external adjustment method, extra information on clinical risk factors is collected on a subsample of individuals and that are used to correct for unmeasured confounders in the main study [68]. Requirements are that the subsample must be a representative sample and all confounders should be identified [68]. A limitation of this method is that it does not consider the joint effect of multiple confounders that may add up or cancel out [68].

Propensity score calibration

Propensity score analysis, proposed by Rosenbaum in and Rubin [69], is a commonly used approach to control for measured confounding. The propensity score (PS) is first estimated as the probability of receiving a treatment versus no (another) treatment conditional on measured confounders. Next, the PS is used as a matching, stratifying, a weighting variable, or included as a covariate in a model regressing the outcome on the exposure. Conventional PS methods can only control for measured confounding [70]. Propensity score calibration (PSC) was developed by Stürmer et al. [71] to control for multiple confounders that are measured in a subset of the study only. Using this method, researchers can control for the joint effect of multiple unmeasured confounders in the main study by using variables that are observed in the validation substudy [68, 71–73]. PSC combines propensity scores and regression calibration to control for unmeasured confounding variable [68, 71–73]. First, a propensity score (PS) is estimated in the main study similar to the conventional PS analysis. This PS can be viewed as a variable measured with error ("error-prone" PS) when additional confounders are unobserved, either due to the lack of information on important predictors of the exposure (unmeasured variables) or due to imperfectly measured predictors. Second, in a data-rich subset, an error-prone PS (identical in its determinants to the error-prone PS in the main study) and a 'gold-standard' PS are estimated. This subset includes variables that are not measured (or measured with error) in the main study. The 'gold-standard' PS includes additional important determinants of the exposure that were unobserved in the main study. Third, the error in the main study PS is estimated in the validation study by directly

comparing the PS containing the same information as the main study PS (i.e., error-prone PS) with the gold-standard PS. Last, the regression calibration is used to adjust the main study PS (i.e., error-prone PS) for that error. PSC does not require outcome information to be available in the validation study, but it helps to check the assumptions of PSC. A key assumption is the so-called surrogacy assumption, i.e., the error-prone PS is a valid surrogate for the gold-standard PS [68, 72].

Two-stage sampling

Two-stage sampling (TSS) designs, similar to the PSC, rely on an internal validation study to collect information on covariates that were not measured in the main study [68]. As such, TSS is an efficient method to collect information on initially unobserved confounders and predominantly used in case–control studies [16, 74]. This means information on unmeasured confounders is collected in a sample of the study population and this information is extrapolated to the entire study population. Hence, confounding can be adjusted for by confounder information that was only collected in a sample of the study population. Several methods of sampling have been described, for example, obtaining information on cases only, or on samples of cases and controls [75]. The optimal size of the sample depends on the true disease model, and the association between the information in the dataset and the information in the sample [75]. A priori, such information is typically unknown and it is, therefore, difficult to estimate the optimal size before data is collected. TSS may become problematic in anonymized databases in which name and postal addresses of patients are lacking and, hence, questionnaires cannot be sent to these persons.

Multiple imputations

Multiple imputations (MI) has been used to overcome bias due to unmeasured confounding in pharmacoepidemiologic studies [73]. It requires a sub-study or validation sample, similar to the PSC and two-stage sampling. When the sub-study contains outcome information as well as information on confounders that are unmeasured for the rest of the study population, MI can be used to impute the missing information of those confounders [73]. One of the advantages of the MI over PSC is that it does not require the surrogacy assumption of regression calibration. Both methods require that missing confounder values can be predicted using observed information, i.e., that missing data are missing at random [76]. MI does not perform well when the majority of the data is missing, which may be realistic in pharmacoepidemiologic studies [77]. The statistical software for multiple imputation can be found in the Refs. [78–80].

Conclusions

Unmeasured confounding is an important source of bias in pharmacoepidemiologic studies. Here, we provided an overview of methods to detect or control for unmeasured confounding in observational studies, highlighting their strengths and limitations for pharmacoepidemiologic studies. Choosing between methods to reduce bias due to unmeasured confounding is a challenge.

Before applying any method to detect or control for unmeasured confounding in a study, we recommend using clinical knowledge combined with empirical evidence to argue the extent to which unmeasured confounding is a relevant threat to the validity of the study. For example, when studies involve intended drug effects, confounding by indication may induce a large amount of bias in the exposure effects (i.e., there is a big concern about unmeasured confounding) [2, 7, 81]. In that case, case-only designs or IV analysis could be viable options provided that the assumptions of the methods are fulfilled. On the other hand, in studies of adverse drugs effects in which individual risks of particular outcomes (e.g., immunological adverse effects) [2] is often unknown and does not influence the prescription of drugs [2], the impact of unmeasured confounding may be small and thus ordinary methods to control for confounding may be sufficient (or even better) to estimate exposure effects [82]. Nevertheless, when an adverse effect is well known (e.g., relation between aspirin use and risk of gastrointestinal bleeding) [81, 82], individual risks are often taken into account when prescribing (confounding by contraindication) and hence strong confounding may be present and the methods that make the assumption of no unmeasured confounding (e.g., propensity scores, regression adjustment) may yield biased results.

If there is a strong concern about unmeasured confounding in a study, it may be worthwhile to replace the assumption of no unmeasured confounding by perhaps more plausible assumptions required for the methods to control for unmeasured confounding. Importantly, some of these assumptions (e.g., in IV analysis the assumption that the IV affects the outcome only through the exposure) are not verifiable using the data; hence, expert knowledge is always necessary to justify them.

The type of data available for a pharmacoepidemiologic study, exposure or outcome under study may guide the choice for a method to handle unmeasured confounding. For instance, IV analysis can be applied in a cohort/nested case–control study and for any type of outcome. However,

for small samples with rare outcomes, or in case of time-varying exposure and confounders, IV analysis may provide biased results [54, 57]. Moreover, the PERR method requires that data are available on two periods: i.e., before and after initiation of the exposure and it is less useful when studying e.g., mortality [39, 43]. Furthermore, the case-only design uses information of cases only and also outcome specific. In addition, two-stage sampling is used predominantly in case–control studies. Hence, the possibility to apply any of these (and the other discussed) methods depends on the research question and/or availability of the data.

Finally, there are several fundamental differences between methods that control for unmeasured confounding. All methods make several strong assumptions to estimate exposure effects and have several pros and cons in empirical settings. The performance and application of the methods strongly depend on the nature of confounding, exposure, type of outcome, size of the sample in a particular clinical example. Therefore, we stress the importance of a complete understanding of the methods before applying them and a routine evaluation of the underlying assumptions. As the performance and application of the methods may differ across studies and across databases, we also stress the importance of using both statistical evidence and substantial clinical knowledge for interpretation of the study results.

Funding None.

Conflicts of interest There are no financial, personal, political, academic or other relations that could lead to a conflict of interest.

References

1. Grobbee DE, Hoes AW. Clinical epidemiology: principles, methods, and applications for clinical research. Jones & Bartlett Learning, ISBN: 1449674321; 2009.
2. Vandenbroucke JP. When are observational studies as credible as randomised trials? Lancet. 2004;363(9422):1728–31.
3. Berger ML, Martin BC, Husereau D, Worley K, Allen JD, Yang W, et al. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC good practice task force report. Value Health. 2014;17(2):143–56.
4. Feinstein AR. Current problems and future challenges in randomized clinical trials. Circulation. 1984;70(5):767–74.
5. Black N. Why we need observational studies to evaluate the effectiveness of health care. Br Med J. 1996;312(7040):1215–8.
6. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. Med Care. 2010;48(6 SUPPL.):S114–20.
7. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Stat Med. 2014;33(13):2297–340.
8. McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. Pharmacoepidemiol Drug Saf. 2003;12(7):551–8.
9. Klungel OH, Martens EP, Psaty BM, Grobbee DE, Sullivan SD, Stricker BHC, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. J Clin Epidemiol. 2004;57(12):1223–31.
10. Bosco JLF, Silliman RA, Thwin SS, Geiger AM, Buist DSM, Prout MN, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol. 2010;63(1):64–74.
11. Groenwold RHH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. J Clin Epidemiol. 2009;62(1):22–8.
12. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, editors. Developing a protocol for observational comparative effectiveness research: a user's guide. AHRQ Publication No. 12(13)-EHC099. Rockville, MD: Agency for Healthcare Research and Quality; January 2013. www.effectivehealthcare.ahrq.gov/Methods-OCER.cfm.
13. VanderWeele TJ, Shpitser I. On the definition of a confounder. Ann Stat. 2013;41(1):196–220.
14. Alemayehu D, Alvir JM, Jones B, Willke RJ. Statistical issues with the analysis of nonrandomized studies in comparative effectiveness research. J Manag Care Pharm. 2011;17(9 Suppl A):S22–6.
15. Schneeweiss S, Seeger JD, Smith SR. Methods for developing and analyzing clinically rich data for patient-centered outcomes research: an overview. Pharmacoepidemiol Drug Saf. 2012;21(Suppl. 2):1–5.
16. Groenwold RHH, Hoes A, Nichol KL, Hak E. Quantifying the potential role of unmeasured confounders: the example of influenza vaccination. Int J Epidemiol. 2008;37(6):1422–9.
17. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf. 2006;15(5):291–303.
18. 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.
19. Lu CY. Observational studies: a review of study designs, challenges and strategies to reduce confounding. Int J Clin Pract. 2009;63(5):691–7.
20. Bradbury BD, Gilbertson DT, Brookhart MA, Kilpatrick RD. Confounding and control of confounding in nonexperimental studies of medications in patients with CKD. Adv Chronic Kidney Dis. 2012;19(1):19–26.
21. Wunsch H, Linde-Zwirble WT, Angus DC. Methods to adjust for bias and confounding in critical care health services research involving observational data. J Crit Care. 2006;21(1):1–7.
22. Stürmer 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 SUPPL. 2):S158–65.
23. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol. 1991;133(2):144–53.
24. Nordmann S, Biard L, Ravaud P, Esposito-Farèse M, Tubach F. Case-only designs in pharmacoepidemiology: a systematic review. PLoS One. 2012;7(11):e49444.
25. Schneeweiss S, Stürmer T, Maclure M. Case-crossover and case-time-control designs as alternatives in pharmacoepidemiologic research. Pharmacoepidemiol Drug Saf. 1997;6(SUPPL. 3):S51–9.
26. Delaney JA, Suissa S. The case-crossover study design in pharmacoepidemiology. Stat Methods Med Res. 2009;18(1):53–65.
27. Aberra FN, Brensinger CM, Bilker WB, Lichtenstein GR, Lewis JD. Antibiotic use and the risk of flare of inflammatory bowel disease. Clin Gastroenterol Hepatol. 2005;3(5):459–65.

28. Ray WA, Fought RL, Decker MD. Psychoactive drugs and the risk of injurious motor vehicle crashes in elderly drivers. Am J Epidemiol. 1992;136(7):873–83.
29. Hebert C, Delaney J, Hemmelgarn B, Lévesque LE, Suissa S. Benzodiazepines and elderly drivers: a comparison of pharmacoepidemiological study designs. Pharmacoepidemiol Drug Saf. 2007;16(8):845–9.
30. Suissa S. The case-time-control design. Epidemiology. 1995;6(3):248–53.
31. Wang SV, Coull BA, Schwartz J, Mittleman MA, Wellenius GA. Potential for bias in case-crossover studies with shared exposures analyzed using SAS. Am J Epidemiol. 2011;174(1):118–24.
32. 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.
33. Schneider MF, Gange SJ, Margolick JB, Detels R, Chmiel JS, Rinaldo C, et al. Application of case-crossover and case–time–control study designs in analyses of time-varying predictors of T-cell homeostasis failure. Ann Epidemiol. 2005;15(2):137–44.
34. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. Biometrics. 1995;51(1):228–35.
35. Whitaker HJ, Hocine MN, Farrington CP. The methodology of self-controlled case series studies. Stat Methods Med Res. 2009;18(1):7–26.
36. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. Stat Med. 2006;25(10):1768–97.
37. The Self Controlled Case Series Method (internet). The Open University, UK (updated 2010; cited 2015 Dec 4). http://statistics.open.ac.uk/sccs/index.htm.
38. Yu M, Xie D, Wang X, Weiner MG, Tannen RL. Prior event rate ratio adjustment: numerical studies of a statistical method to address unrecognized confounding in observational studies. Pharmacoepidemiol Drug Saf. 2012;21(Suppl. 2):60–8.
39. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. BMJ. 2009;338(7691):395–9.
40. Tannen R, Xie D, Wang X, Yu M, Weiner MG. A new "Comparative Effectiveness" assessment strategy using the THIN database: comparison of the cardiac complications of pioglitazone and rosiglitazone. Pharmacoepidemiol Drug Saf. 2013;22(1):86–97.
41. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the 'prior event rate ratio' to adjust for unmeasured confounding by indication. Pharmacoepidemiol Drug Saf. 2008;17(7):671–85.
42. Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. Pharmacoepidemiol Drug Saf. 2008;17(7):661–70.
43. Uddin MJ, Groenwold RHH, Van Staa TP, De Boer A, Belitser SV, Hoes AW, et al. Performance of prior event rate ratio adjustment method in pharmacoepidemiology: a simulation study. Pharmacoepidemiol Drug Saf. 2015;24(5):468–77.
44. Lipsitch M, Tchetgen E Tchetgen, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21(3):383–8.
45. Schuemie MJ, Ryan PB, Dumouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. Stat Med. 2014;33(2):209–18.
46. Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. Int J Epidemiol. 2006;35(2):337–44.
47. Flanders WD, Klein M, Darrow LA, Strickland MJ, Sarnat SE, Sarnat JA, et al. A method for detection of residual confounding in time-series and other observational studies. Epidemiology. 2011;22(1):59–67.
48. Tchetgen E Tchetgen. The control outcome calibration approach for causal inference with unobserved confounding. Am J Epidemiol. 2014;179(5):633–40.
49. Groenwold RHH. Falsification end points for observational studies. JAMA. 2013;309(17):1769–70.
50. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. J Clin Epidemiol. 2009;62(12):1226–32.
51. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. J Clin Epidemiol. 2011;64(6):687–700.
52. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. Epidemiology. 2013;24(3):363–9.
53. Martens EP, Pestman WR, De Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology. 2006;17(3):260–7.
54. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006;17(4):360–72.
55. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000;29(4):722–9.
56. Schneeweiss S, Setoguchi S, Brookhart A, Dormuth C, Wang PS. Risk of death associated with the use of conventional versus atypical antipsychotic drugs among elderly patients. Can Med Assoc J. 2007;176(5):627.
57. Uddin MJ, Groenwold RHH, de Boer A, Belitser S, Roes KCB, Hoes AW, et al. Performance of instrumental variable methods in cohort and nested case–control studies: a simulation study. Pharmacoepidemiol Drug Saf. 2014;23:165–77.
58. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiol Drug Saf. 2010;19(6):537–44.
59. Instrumental Variables (internet). Econometrics Academy (updated 2015; cited 2015 Dec 4). https://sites.google.com/site/econometricsacademy/econometrics-models/instrumental-variables.
60. Lin NX, Logan S, Henley WE. Bias and sensitivity analysis when estimating treatment effects from the cox model with omitted covariates. Biometrics. 2013;69(4):850–60.
61. Groenwold RH, Nelson DB, Nichol KL, Hoes AW, Hak E. Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. Int J Epidemiol. 2010;39(1):107–17.
62. Greenland S. Basic methods for sensitivity analysis of biases. Int J Epidemiol. 1996;25(6):1107–16.
63. Lee WC. Detecting and correcting the bias of unmeasured factors using perturbation analysis: a data-mining approach. BMC Med Res Methodol. 2014;14(1):18.
64. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology. 2011;22(1):42–52.
65. Wen SW, Kramer MS. Uses of ecologic studies in the assessment of intended treatment effects. J Clin Epidemiol. 1999;52(1):7–12.
66. Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. Annu Rev Public Health. 1995;16:61–81.
67. Groome PA, Mackillop WJ, Naylor CD. Uses of ecologic studies in the assessment of intended treatment effects multiple letters. J Clin Epidemiol. 1999;52(9):903–4.
68. Schneeweiss S. Developments in post-marketing comparative effectiveness research. Clin Pharmacol Ther. 2007;82(2):143–56.

69. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.
70. Ali MS, Groenwold RH, Belitser SV, Pestman WR, Hoes AW, Roes KC, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. J Clin Epidemiol. 2015;68(2):112–21.
71. Stürmer 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.
72. Sturmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration—a simulation study. Am J Epidemiol. 2007;165(10):1110–8.
73. Lunt M, Glynn RJ, Rothman KJ, Avorn J, Stürmer T. Propensity score calibration in the absence of surrogacy. Am J Epidemiol. 2012;175(12):1294–302.
74. 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.
75. Hanley JA, Dendukuri N. Efficient sampling approaches to address confounding in database studies. Stat Methods Med Res. 2009;18(1):81–105.
76. Stürmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Comparison of performance of propensity score calibration (PSC) and multiple imputation (MI) to control for unmeasured confounding using an internal validation study. Pharmacoepidemiol Drug Saf. 2006;15:S39–40 **(abstract)**.
77. Faries D, Peng X, Pawaskar M, Price K, Stamey JD, Seaman JW. Evaluating the impact of unmeasured confounding with internal validation data: an example cost evaluation in type 2 diabetes. Value Health. 2013;16(2):259–66.
78. Yao X, Lix L. A flexible method to apply multiple imputation using SAS/IML® studio. SAS Global Forum 2013; Paper 283-2013.
79. Soley-Bori M. Dealing with missing data: key assumptions and methods for applied analysis 2013.
80. Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing variables. Am Stat. 2001;55(3):244–54.
81. Miettinen OS. The need for randomization in the study of intended effects. Stat Med. 1983;2(2):267–71.
82. Joseph K, Mehrabadi A, Lisonkova S. Confounding by indication and related concepts. Curr Epidemiol Rep. 2014;1(1):1–8.