

REVIEW

Core concepts in pharmacoepidemiology: Fundamentals of the cohort and case-control study designs

Anton Pottegård 

Clinical Pharmacology, Pharmacy, and Environmental Medicine, Department of Public Health, University of Southern Denmark, Odense, Denmark

Correspondence

Anton Pottegård, Clinical Pharmacology, Pharmacy and Environmental Medicine, University of Southern Denmark, JB Winsløvsvej 19, 2 DK-5000 Odense C, Denmark.
Email: apottegaard@health.sdu.dk

Abstract

In this review paper, I outline the principles of the cohort as a sampling frame and provide a basic introduction to the cohort study design and the case-control study design, two of the most important designs in the pharmacoepidemiologist's toolbox. Further, I discuss when to prefer one design over the other. The paper is intended as a primer for people new to the field of pharmacoepidemiology and contains a range of suggestions for additional reading regarding the study designs and related epidemiological topics.

KEYWORDS

case-control study, cohort study, pharmacoepidemiology

Key points

- The cohort and case-control study designs are the two most central designs in pharmacoepidemiology. The proper use of both these designs requires an understanding of the cohort as a sampling frame.
- The cohort study compares the rate of outcome events (or other outcome metrics) among those exposed to a drug (or other exposure) to the rate among individuals not exposed to the drug or exposed to a comparator drug.
- The case-control study design compares the use of a drug (or other exposure) among those with a disease (cases) to the use of the drug among controls. The controls represent the background use of the drug in the population from which cases arise.
- The cohort and case-control study designs have distinct advantages that makes them particularly useful in different scenarios. Importantly, however, when properly designed and interpreted, the two designs will yield similar results and should be considered equal.

1 | INTRODUCTION

Pharmacoepidemiology is the study of use and effects of medications on a population basis.¹ Effects of medications can be assessed using a wide variety of formalized study designs spanning from the cross-sectional and ecological designs over the cohort and case-control designs to the large and growing family of self-controlled designs.² New designs as well as extensions of existing ones are continuously added to the pharmacoepidemiologist's armamentarium, such as the

trend-in-trend design³ or the case-time-control extension⁴ of the self-controlled case-crossover design.⁵ In parallel, emerging methods enables the pharmacoepidemiologist to not only assess associations but also to provide direct estimates of the causal link between an exposure and an outcome.⁶ All of this allows pharmacoepidemiologists to tread new ground. However, to utilize these new opportunities, it is essential to establish and maintain an understanding of the core concepts in pharmacoepidemiology⁷ upon which they are built. In this paper, I will therefore outline the

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Author. *Pharmacoepidemiology and Drug Safety* published by John Wiley & Sons Ltd.

principles of the cohort as a sampling frame and give a basic introduction to the two most common pharmacoepidemiological study designs: The cohort study design and the case-control study design.

2 | WHAT IS A COHORT?

A cohort denotes a group of individuals fulfilling a set of cohort-defining criteria. As examples, a cohort could comprise initiators of a specific antidiabetic drug within a given time period, males using two or more antihypertensive drugs, or inhaled steroid users that do not have a diagnosis of chronic obstructive pulmonary disease (COPD).

2.1 | Types of cohorts

In some cohorts, each individual is able to enter only once. This is sometimes referred to as a “closed” (or static) cohort and could, for example, be a cohort of children born to mothers using psychotropics, who are followed for the first year of their life. In other cohorts, individuals can go in and out of the cohort. This is sometimes referred to as an “open” (or dynamic) cohort and could for example be a cohort of current users of the anticoagulant drug warfarin. Warfarin use can be episodic, and as such the same individual might join the cohort for example a year, after which use is stopped and the individual is therefore censored from the cohort, only to rejoin the cohort later upon reinitiating warfarin. Another definition of “open cohorts” is that more individuals can join over time, while “closed cohorts” only include participants that were included when initial follow-up begins (e.g., the Framingham Heart Study⁸). No real consensus exists on the nomenclature around closed/open or static/dynamic cohorts. As such, these terms can reasonably be avoided in favor of clear and transparent descriptions of how the boundaries of one's cohort are defined.

2.2 | Person-time

One of the main epidemiological units of interest is the concept of person-time. Person-time, or follow-up time, refers to the time that individuals contribute to an analysis. Person-time is often measured in units of person-years but can also be counted as person-months or -days or any other time unit. An analysis that includes 10 person-years of follow-up can stem from one individual followed for 10 years, 10 individuals each followed for 1 year, or 872 individuals followed on average 4.2 days. While a cohort was defined as a group of people, the focus is most often on the amount of person-time rather than the number of individuals. Consider an analysis of the occurrence of an adverse event with the use of a drug. This could for example be hypoglycemia with use of an antidiabetic medication. To allow comparison with other studies or another cohort, we should standardize the number of events to the amount of time spent using the drug. If we observe more users or observe each user for longer time, the finding that more events occur is neither surprising nor informative. Rather,

we are interested in the rate of events, that is, the number of events per unit of person-time. If events are confined to occur within a narrow time window, however, we would be more interested in the risk within that time window, rather than the rate over an extended time period. In a study of the occurrence of congenital malformations after drug exposure during pregnancy, the measure of interest would be the proportion of children with a malformation (at birth) and not the rate of malformations per year, because the time window for a malformation to occur is narrow and after it ends no more malformation can occur. Similarly, in a study of the 30-day risk of seizures following vaccination, we would be interested in the proportion experiencing the outcome within the 30-day risk window (cumulative risk proportion) and not the rate of the outcome. Nevertheless, a thorough understanding of the concept of person-years and how an individual's experience can be mapped over time, is central to the understanding of epidemiology and key to avoid some of the most important and common biases in epidemiological studies.^{9,10}

3 | THE COHORT STUDY DESIGN

The cohort study design is the most commonly used study design in pharmacoepidemiology. The cohort design is a design, in which you compare the rate or risk of events, or some other relevant outcome metric, between two or more cohorts. This way, you obtain a measure of increased, decreased or unaffected risk with use of a given drug or other intervention. As an example, this could be comparison of the rate of bleedings among users of two distinct anticoagulants or between high- and low-dose users of the same drug. In pharmacoepidemiology, cohorts are often defined by initiation of a given drug. However, non-user cohorts are also commonly used as reference, and so is “previous use,” that is, comparison of person-time exposed to a drug to person-time from previous-but-not-current users of the drug. The choice of what comparator to use is arguably the most important decision in the design of a cohort study, with the optimal choice being heavily dependent on the study question under scrutiny and requiring insight into the clinical practice within the specific condition being studied.

3.1 | Follow-up

The first step to the design (and reporting) of a cohort study is to outline exactly which individuals are followed from when to when, that is, to define the cohort or cohorts. First, the cohort entry date is defined by the date upon which an individual meets all the cohort defining criteria, which as mentioned is often anchored on the date that an individual begins using a given drug. It can, however, also be January 1st of a given year among prevalent users of a drug, the date an individual fills his/her second prescription for a given drug, the date of diagnosis of a particular disease, or any other definition for that matter. Among those meeting the cohort entry criteria, some will be excluded because they meet various exclusion criteria. This often,

although not always, involves exclusion of individuals that have previously experienced the outcome under study. Study patients are followed from their entry date until the earliest of a number of stopping criteria. This stopping date could be the last date of the data that is available, developing the study outcome, meeting an exclusion or censoring criterion, migrating, or dying.

Note that the cohort entry date and “start of follow-up” do not have to be identical. In a study of cancer as a side-effect to drug treatment, it would be reasonable to only include follow-up time starting several years after the drug is initiated, as it is unlikely that any increased risk would manifest shortly after drug initiation. Here, an individual could enter the cohort upon drug initiation, but only contribute person-time to the rates measured in the study from for example year three after start of treatment and onwards. Conversely, in a study of analgesics and risk of gastrointestinal bleeding, the increased risk is immediate, and follow-up should thus start upon treatment initiation and further be stopped again shortly after treatment is stopped, as it is unlikely that the effect of the drug will persist after the treatment is stopped. Overall, ensuring a pharmacologically and clinically relevant exposure assessment and definition of follow-up is central to the design of any cohort study.

3.2 | Outcome metrics

Once the cohort and follow-up definitions are settled, the actual outcomes and the total amount of follow-up can be tallied. What outcome metric is used will depend on the given study, but often it is incidence rates, that is, the rate of events per person-time, in each of the groups you are comparing. It could also be risk proportions or any other measure of frequency. How to obtain the actual measures of association, that is, what statistical models or assumptions are invoked and which are more appropriate in the context of a given study, is beyond the scope of this paper. However, with the cohort design you will generally be able to estimate measures of relative risk increases, such as hazard ratios (HR) or incidence rate ratios (IRR). You can usually also estimate absolute risk increases, for example the incidence rate difference (IRD) or risk difference, which are important measures, as they are generally more relevant to clinical decisions and public health planning.¹¹

3.3 | Confounding

Confounding is the confusion of causes and is best illustrated with an example. Consider a comparison between use of two antidiabetics on the risk of developing heart disease. If one antidiabetic is preferred for older patients, and as old age is a risk factor for heart disease, it will result in a spurious association between use of this drug and cardiovascular disease if age is not taken into account in the comparison. Here, age act as a confounder.

Confounding constitutes a key challenge in all observational research. A full account of the mechanisms through which

confounding can arise is outside the scope of this paper. A useful introduction to confounding is provided by Gerhard.⁹ Numerous methods can be applied to handle confounding in cohort studies, examples of which include (i) statistical modeling (adjusting for potential confounders as part of a regression), (ii) methods using summary confounding variables (most often either matching or weighting on the “propensity score”¹²), (iii) use of active comparators¹³ or (iv) restriction or stratification.¹⁴ Proper consideration and handling of confounders, either through these common techniques or other creative solutions, is central to ensure the validity of any cohort study.

3.4 | Immortal time bias

In defining the follow-up for the individual, care must be taken not to “condition on the future.” Consider, as an example, a study on the effect of a medication initiated after hospital discharge. The patients can be divided into users and non-users of this drug based on whether they fill or do not fill a prescription within 30 days from discharge. However, if they are divided as such yet are followed from the date of hospital discharge, you will have conditioned on the future. Consider a patient that fills his first prescription on day 20. If this individual is followed as an exposed individual from the day of discharge, the first 19 days are incorrectly classified as being exposed. The bias that arises from such misclassification is often referred to as “immortal time bias,” as the person-time that is classified based on the patient doing something in the future can be considered “immortal” (had the patient died, the future exposure would not have happened). This dilution of one cohort with “immortal time” confers a strong bias, most often in favor of the treatment.

Time-related biases also exist in other forms outside the specific concept of immortal time bias. Such biases, for example conditioning on the future, are common and can result in strongly biased estimates. For a detailed introduction to time-related biases, see Suissa and Dell'Aniello.¹⁵

3.5 | Examples

Here, I provide four applied examples of cohort studies that illustrate some of the issues outlined above.

3.5.1 | Methylphenidate during pregnancy and malformations

To investigate the potential association between use of methylphenidate during the first trimester of pregnancy and risk of congenital malformations in the offspring, a cohort study was conducted.¹⁶ Data was primarily obtained from the Danish Medical Birth Registry¹⁷ and the Danish National Prescription Registry.¹⁸ Women using methylphenidate during first trimester, as evaluated based on prescription fills, were compared with women not using methylphenidate during

pregnancy, and in a sensitivity analysis with women that had previously used methylphenidate. Malformations were assessed at birth, and no follow-up time for the infant was defined. Potential confounding was addressed using propensity score matching to increase comparability between methylphenidate users and non-users. Methylphenidate was not found to be associated with malformations.

3.5.2 | Digoxin use and breast cancer risk

To investigate whether use of the antiarrhythmic drug digoxin increases the risk of breast cancer, a cohort study was carried out.¹⁹ Data was obtained from the Nurses' Health Study,²⁰ a prospective established cohort. Exposure to digoxin was collected via telephone interviews and self-reported breast cancer outcomes was confirmed by review of medical records. The study identified 5004 digoxin users during 1.05 million years of person-time. Use of digoxin was associated with a 45% increased rate of breast cancer compared with person-time among individuals having never used digoxin.

3.5.3 | Antidiabetics and risk of gout

To investigate whether the use of SGLT2 inhibitor antidiabetics was associated with risk of gout, a cohort study was conducted.²¹ Data were obtained from IBM MarketScan,²² a US nationwide commercial insurance database. New users of SGLT2 inhibitors were compared with new users of another antidiabetic (GLP1 agonists). The use of an active comparator, indicated at a similar stage of disease, alleviated a central confounder, as diabetes is in itself associated with risk of gout. Comparison of users of SGLT2 inhibitors with non-users would thus result in a confounded comparison. Further comparability was achieved using propensity score matching. Each individual was followed from treatment initiation until a study outcome, end of study period, death, or discontinuation of index medication. SGLT2 inhibitor use was associated with a 36% decreased relative risk of gout.

3.5.4 | Warfarin switching and risk of bleeding

To investigate a safety signal that switching from branded to generic warfarin was association with an increased risk of bleeding, a cohort study was conducted.²³ Data was primarily obtained from the Danish National Patient Registry²⁴ and the Danish National Prescription Registry.¹⁸ First, a cohort was constructed comprising all warfarin users during a specific time-period. As the safety signal specifically related to "recent switch from branded to generic warfarin," each warfarin user had his/her person-time mapped out regarding whether switches occurred. The main analysis compared person-time classified as "recent switch to generic warfarin", defined as 60 days following a switch, to person-time classified as "continuous use of branded warfarin." No increased risk of bleeding was observed with switching to generic warfarin.

3.6 | Cohort study design visualization

The design of a cohort study can be visualized using the design diagrams developed by Schneeweiss et al²⁵ made easily accessible via an online tool developed by Lund et al.²⁶ An example of such a diagram is provided in Figure 1, outlining the design of the cohort study investigating the association between use of antidiabetics and risk of gout described above.

3.7 | Suggested further reading

For further reading on the cohort study design, I refer to the paper by Lund et al¹³ on the active comparator, new user cohort study design. For an introduction to the "target trial" approach, that is, where cohort studies are explicitly designed to mirror the corresponding trial, I refer to the paper by Dickerman et al.²⁷

4 | THE CASE-CONTROL STUDY DESIGN

The case-control study design is another central study design in pharmacoepidemiology. The case-control design is based on a comparison of the use of a drug among those with the disease (cases) to the use of the drug among people without the disease (controls). This could for instance be the use of anticoagulants among individuals with bleeding (cases) compared with the use of anticoagulants among population controls. Importantly, while use of medicine is compared between cases and controls, the underlying comparison is identical to that of the corresponding cohort study, that is, a comparison between use of a given drug and either non-use of the drug or use of another drug. In fact, a case-control study can in many ways be considered equivalent to a cohort study, with the controls representing a sample of the person-time from the underlying cohort study.

Originally, the case-control study was conceptualized due to its efficiency, allowing analyses where cohort studies could not be done, for example, due to lack of the data availability, costs related to data collection (especially in studies with long follow-up), or computational constraints. With emergence of better data sources and increases in computational power, some of these advantages are diminishing. However, the case-control study design has evolved since its early years and still has a place in pharmacoepidemiology.²⁸

4.1 | The underlying cohort

When thinking about and designing case-control studies, one should always keep in mind what the corresponding cohort study would look like. If a cohort study investigates whether the risk of prostate cancer is associated with use of warfarin compared with non-use of warfarin, then the corresponding case-control study should compare proportions of warfarin use among prostate cancer patients (cases) with those among population controls. If the intention is to look at long-

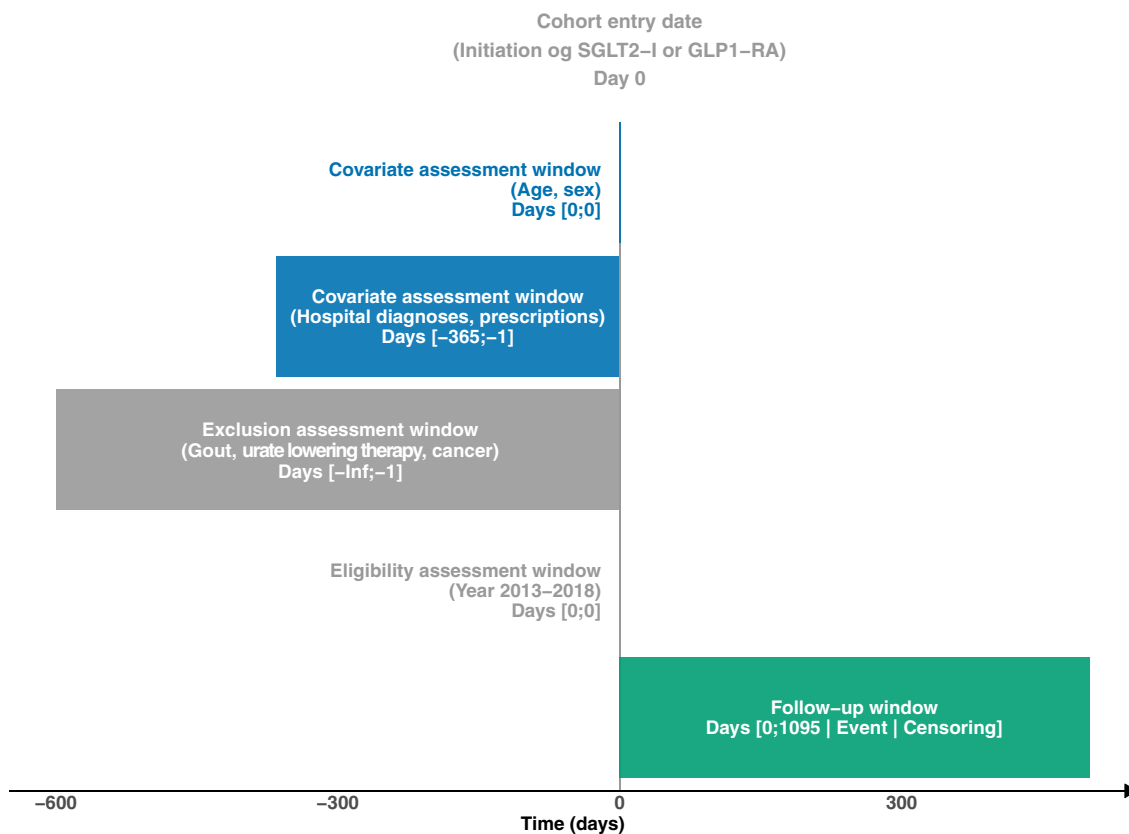


FIGURE 1 A graphical representation of the study design of the cohort study “Assessing the Risk for Gout With Sodium-Glucose Cotransporter-2 Inhibitors in Patients With Type 2 Diabetes: A Population-Based Cohort Study.” by Fralick et al.²¹ The design diagram follows the concepts outlined by Schneeweiss et al.²⁵ and it is created using the online tool provided by Lund et al.²⁶

term effects only, the cohort study would include individuals into the exposed cohort at the time of having accumulated for example 5 years of warfarin use. In that case, the use metric in the corresponding case-control study should change accordingly to the proportion with 5+ years of use of warfarin among cases and controls. Finally, if the intent is to compare long-term users with short-term users (e.g., to handle confounding), the cohort study would only include individuals that have at some point used warfarin, and correspondingly, in the case-control study both cases and controls should be sampled among warfarin ever users, comparing the proportion with long-term use among cases and controls. These examples all describe the concept of the underlying cohort.

When, as in the last example, cases and controls are sampled from a well-defined population, the case-control study is referred to as “nested” (here within warfarin ever users). Technically, all case-control studies can be viewed as being nested within a source population although this source population is not always explicitly defined nor fully available. However, specifically nesting a case-control study for example within ever users of a drug or individuals with a specific diagnosis can be used to increase the comparability between exposed and unexposed individuals (no among cases and controls) and thus counter potential confounding. Another feature of nested case-control studies, although less often used, is that it is possible to also estimate measures of absolute effects if the full underlying cohort is available. In brief, this

can be done by estimating the rate of the outcome under scrutiny among unexposed individuals and then multiplying this by the observed odds ratio (OR). If the rate of an event among unexposed individuals is 2.0 per 1000 person-years and the observed OR is 3, the rate must thus be 6.0 per 1000 person-years among the exposed and thus the absolute effect size is 4.0 per 1000 person-years. Additional methods to estimate absolute effects in nested case-control studies has been described in more detail by Hallas et al.²⁹

4.2 | Case and control selection

Cases are defined by a given outcome of interest, for example, patients receiving a prostate cancer diagnosis, admitted due to upper gastrointestinal bleeding, or dying. The date of experiencing this outcome (e.g., the date of admission during which the diagnosis is received) is considered the index date. Cases should represent outcome events from the underlying cohort, that is, events that would be included as outcomes in the corresponding cohort study. As such, any criteria that would either lead to exclusion or censoring from the cohort study should also act as exclusion criteria for cases in a case-control study, evaluated at the index date.

A common misconception is that controls should be considered as the opposite of cases, that is, that they represent the individuals

without the disease (non-cases). This is incorrect. Rather, controls represent the background population from which cases can arise. Controls should therefore be sampled so that they are a representative sample of the underlying cohort. Importantly, this means that individuals should be eligible for sampling as controls at a given date even if they later become cases, as they, before meeting the case definition, represent the underlying cohort from which cases arise. Further, this means that the exact same in- and exclusion criteria should be applied to both cases and controls, with the additional criteria that previous history of the case-defining event should also lead to exclusion of controls.

Ensuring that controls are sampled so that they can be considered a representative sample of the underlying cohort can be achieved in different ways. The simplest approach is to provide all individuals in the study with a randomly selected index date during the study period. All controls are then evaluated for eligibility at this date and those found ineligible are excluded from the study. Another approach involves matching of controls to cases on for example sex and birth year. The index date for the controls is inherited from the corresponding case, thus ensuring matching on calendar time as well. After eligibility criteria are applied, a random subset of eligible controls is selected, for example 5 or 10 for each case. This approach is typically referred to as risk-set sampling. For both of these approaches, the probability of being selected as a control correlates with the individual's contribution of person-time to the underlying cohort.

4.3 | The table 1 fallacy

Most readers of medical literature are used to reading papers reporting on randomized controlled trials. In such studies, the classic table 1 describes the baseline characteristics of the two arms, to allow the reader to judge that the randomization has worked as intended, that is, has left the two groups comparable. In cohort studies, a similar table 1 can be presented showing the characteristic of the groups that are compared. In a case-control study, however, table 1 usually describes the characteristics of cases and controls. Importantly, these are not expected to be similar, nor is similarity a prerequisite for valid inferences in the study. In fact, as cases represent those with the outcome of interest, one should expect any risk factor for the disease to be more prevalent among the cases than among the controls. This at times lead to misunderstandings among readers unfamiliar with the case-control design and can thus constitute a barrier to the presentation of such studies. One potential solution is to add two additional columns to the table 1 describing exposed and non-exposed controls, which corresponds to the underlying comparison that is made.

4.4 | Outcome metrics

The outcome measure from a case-control study is an odds ratio (OR) obtained for example using logistic regression. When using one of the two sampling methods outlined above, or any other method

that satisfies the criteria that the controls represent the underlying background population, this OR can be considered an unbiased estimate of the IRR that would have been obtained in the corresponding cohort study.³⁰ When matching is used (e.g., by sex and age), conditional logistic regression is often applied, which restricts comparisons to within each riskset (i.e., a set of a case and its corresponding controls), which effectively controls for confounding from these parameters (see below), however, at the same time precluding assessment of the association between these variables and the disease under study.

4.5 | Confounding

The most common approach to handling confounding in a case-control study is adjustment for confounding variables in a regression model. Conditional regression can be used to effectively control for variables also used for matching. Considerations for what to adjust for, that is, the identification and adjustment for potential confounders, follow the same principles as outlined for cohort studies above. Further, as also noted above, nesting a case-control study within a specific subpopulation, either to individuals with a specific disease (e.g., individuals with diabetes or thyroid disorders) or those using either the drug under scrutiny or an active comparator (effectively resulting in an active comparator case-control study), can be very effective in handling confounding and should be considered whenever possible. Finally, case-control counterparts for more specialized approaches to confounder adjustment used in cohort studies also exist but are generally less frequently used, for example use of so-called "disease risk scores"^{31,32} which offer somewhat similar advantages as propensity score models does in cohort studies.

4.6 | Benefits and risks with prevalent sampling

Characteristics of cases and controls are most often assessed at index date, that is, at the date of outcome for cases and the date of sampling for controls. This is at first glance different from cohort studies, in which characteristics are most often assessed at time of cohort entry. Assessing characteristics at time of sampling is, however, identical to using continuously updated (i.e., time-varying) assessment of both covariates and exposure status in the corresponding cohort study. Using time varying exposure and confounder assessment comes with both benefits and risks. The main benefit is that continuously updated covariates allow for a more detailed confounder assessment and thus better confounder adjustment. If some patients are followed for up to for example 15 years after treatment initiation, using baseline characteristics assessed at treatment initiation will in most cases lead to significant misclassification of actual patient characteristics. The benefit of using information assessed later is particularly pronounced in studies with long study periods where some patients contribute with very long follow-up. However, time-varying assessment of covariates also comes with risks. If the treatment under scrutiny in itself affects some covariates (e.g., statin treatment

affecting cholesterol levels), assessing these after treatment initiation, and thus adjust for on-treatment covariates, will induce bias, in that the effect of the intervention will be indirectly adjusted for and thus all estimates will be biased toward seeing no difference. Use of time-varying covariates in a cohort study, or the equivalent use of covariates assessed at index date in a case-control study, should be done with due consideration as to whether the covariates are influenced by the treatment. In such cases, case-control studies can be nested among ever users of the drug or drugs being studied, and covariates can be assessed at treatment initiation instead of at index date.

4.7 | Examples

Here, I provide four applied examples of case-control studies that illustrate some of the issues outlined above.

4.7.1 | Antithrombotic drug use and risk of upper gastrointestinal bleeding

To investigate the association between use of various classes of antithrombotics and risk of upper gastrointestinal bleeding, a case-control study was carried out.³³ Cases were identified from hospital discharge diagnoses and manually validated and matched to population controls. Data on use of medicines were obtained from a regional prescription database.³⁴ Risks with use of individual drugs and combination treatment were identified, finding markedly increased risks in particular with combination treatment. As the study was nested in a geographically well-defined population (residents in Funen county, Denmark), background rates of gastrointestinal bleedings could be calculated, and from that the observed relative effect sizes (ORs) could be translated into absolute risks with use of antithrombotics.

4.7.2 | Corticosteroid use and risk of orofacial clefts

To investigate the association between use of corticosteroids during early pregnancy and the risk of orofacial clefts in the children, a case-control study was carried out.³⁵ The study was based on the National Birth Defect Prevention Study, a manually assembled case-control material with children with malformations (cases) and children without malformations (controls). Exposure data to medicines as well as other exposures was assessed via telephone interviews. No increased risk of orofacial clefts was identified with use of corticosteroids.

4.7.3 | Statins and risk of rheumatoid arthritis

To investigate the putative protective effect of statins on risk of rheumatoid arthritis, a case-control study was carried out.³⁶ Data were obtained from the UK Clinical Practice Research Datalink.³⁷ The study

was nested within a cohort of individuals initiating statin treatment. Within this cohort, cases were defined as individuals receiving a new rheumatoid arthritis diagnosis, with each such case matched to 10 controls on age, sex and calendar year, with subsequent adjustment for among other things smoking, history of cardiovascular disease, and cholesterol levels, all assessed prior to statin initiation (not index date). The main exposure was high-intensity statin use which was compared with low-intensity treatment. The study found that such high-intensity statin use was associated with a decreased risk of rheumatoid arthritis.

4.7.4 | Hypothesis-free screening for drug-cancer associations

To identify new signals for carcinogenic or chemopreventive effects of prescription medicines, an exploratory high-throughput case-control study was conducted.³⁸ Data was primarily obtained from the Danish Cancer Registry³⁹ and the Danish National Prescription Registry.¹⁸ For each of 99 different cancer outcomes, cases were identified, and controls were sampled 1:10. For each individual outcome, analyses were carried out for all individual drugs or drug classes with more than 10 exposed cases, ultimately resulting in 22 125 individual analyses being carried out. Based on pre-defined criteria regarding signal strength and dose-response assessments, signals were identified for further scrutiny.

4.8 | Case-control study design visualization

Similarly as for cohort studies, the design of a case-control study can be visualized using design diagrams^{25,26} An example of such a diagram is provided in Figure 2, outlining the design of the case-control study investigating the association between statin use and risk of rheumatoid arthritis described above.

4.9 | Suggested further reading

For further reading on the case-control study design, I refer to the paper by Vandenbroucke and Pearce⁴⁰ giving an in-depth presentation of the case-control methodology. For an introduction to the strengths of the nested case-control approach, Essebagg et al⁴¹ gives very illustrative examples from studies within cardiology.

5 | THE COHORT VERSUS THE CASE-CONTROL DESIGN

When should one prefer the cohort study design over the case-control design or vice-versa? Unsurprisingly, this depends on the study question under scrutiny and the data sources available. As argued above, the two designs provide estimates of associations that

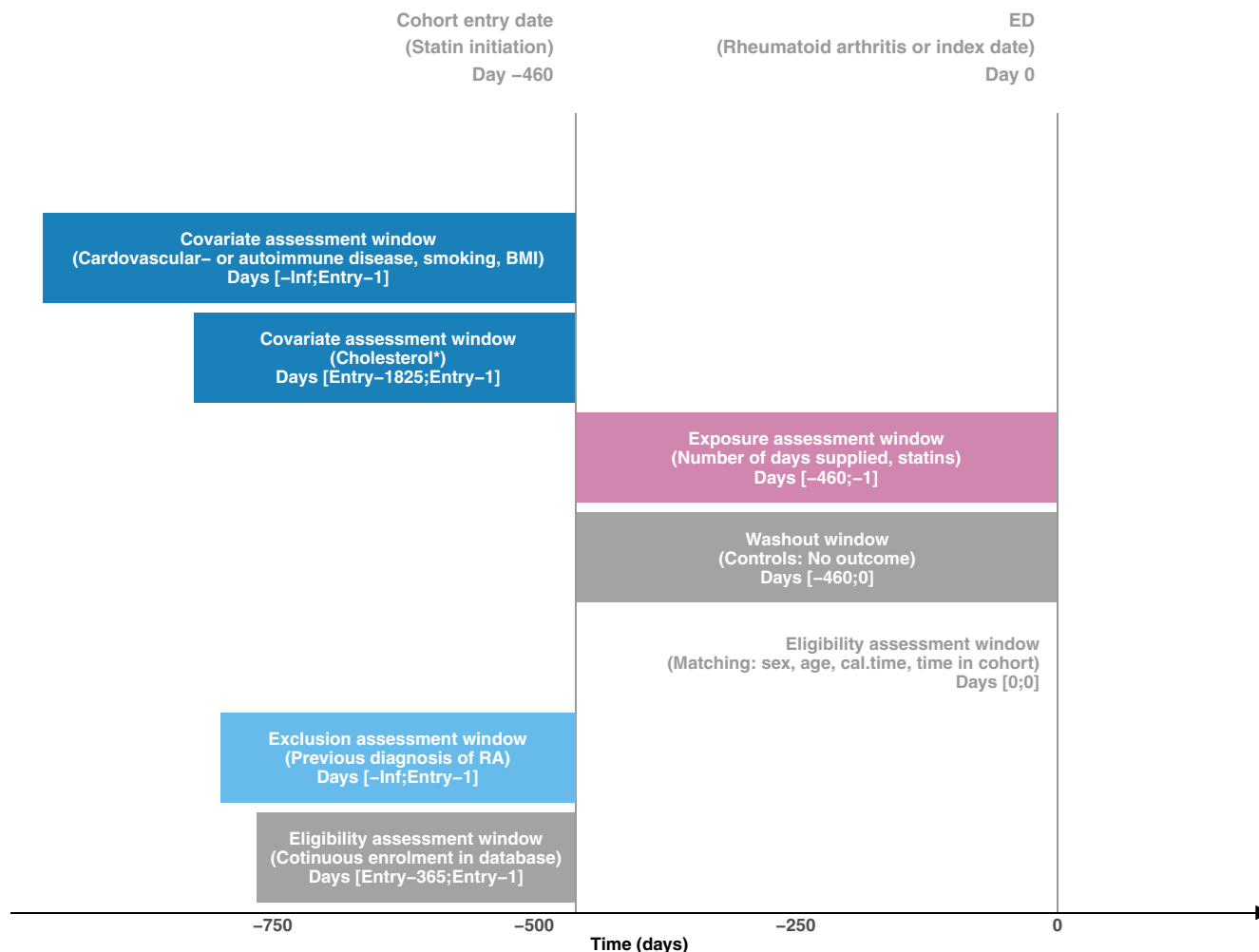


FIGURE 2 A graphical representation of the study design of the case-control study “Statins and Risk of Rheumatoid Arthritis: A Nested Case-Control Study.” by Tascilar et al.³⁶ The design diagram follows the concepts outlined by Schneeweiss et al.²⁵ and it is created using the online tool provided by Lund et al.²⁶

should in most situations be regarded as equally robust. However, some factors might be worthwhile considering when deciding what design to use. The case-control study is particularly convenient if studying many different exposures or exposure definitions, as its anchoring on the outcome event allows the researcher to easily apply very many different exposure metrics. Further, if the outcome is rare, and in particular if manual data collection is considered (e.g., to obtain data on important confounders not available in registries), the case-control design would often allow for a more efficient data collection. Finally, in the rare situations where computer power can be a limitation, case-control studies are usually more computationally efficient. Conversely, if studying multiple different outcomes, the cohort study design is often preferable, as its anchoring on the exposure allows the researcher to investigate multiple different outcomes without resampling the cohorts. Further, if the exposure is rare a cohort study would allow for more efficient data collection as data collection can be focused on those exposed. Finally, and very importantly, the cohort study more readily lends itself to use of active comparators and more easily provides measures of absolute risk.

A final but important advantage of the cohort design is that it is generally more widely accepted and also more easily explained, making the results more cogent. In situations where no factors speak specifically in favor of using the case-control design, the cohort design will for this reason alone likely be the preferred design for most pharmacoepidemiologists. In some cases, the complementary strengths of the two designs might support using both within the same study to shed light on different aspects of a given hypothesis.^{42,43} After all, the cohort and case-control designs constitute two of the most important tools in the pharmacoepidemiologist's toolbox.

ACKNOWLEDGMENTS

The manuscript has been drafted and shaped through considerable input from both pharmacoepidemiologists in training (pregraduate and graduate level) and very experienced pharmacoepidemiologists. I would like to extend my sincerest gratitude to all of these individuals for helping me improve the manuscript. In particular, I would like to thank (listed alphabetically): Alaa Burghle, Amalie Timmermann, Blánaid Hicks, Carina Lundby, Emma Bjørk, Freja B. Hermansen, Jan

Vandenbroucke, Jesper Hallas, Ken Rothman, Lærke K. Overgaard, and Mikkel Søby. Further I would like to acknowledge the very constructive input from handling editor Jennifer L. Lund and the anonymous peer reviewers.

FUNDING INFORMATION

No funding was obtained for the present study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ETHICS STATEMENT

No ethical approval was required for the present study.

ORCID

Anton Pottegård  <https://orcid.org/0000-0001-9314-5679>

REFERENCES

- Strom BL, Kimmel SE, Hennessy S. *Pharmacoepidemiology*. 5th ed. Wiley-Blackwell; 2012.
- Hallas J, Pottegård A. Use of self-controlled designs in pharmacoepidemiology. *J Intern Med*. 2014;275:581-589. doi:10.1111/joim.12186
- Ji X, Small DS, Leonard CE, Hennessy S. The trend-in-trend research Design for Causal Inference. *Epidemiology*. 2017;28:529-536. doi:10.1097/EDE.0000000000000579
- Suissa S. The case-time-control design. *Epidemiology*. 1995;6:248-253.
- Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol*. 1991;133:144-153.
- Pearl & Mackenzie. *The Book of why - the New Science of Cause and Effect*. Basic Books; 2018.
- Public Policy Committee, International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practice (GPP). *Pharmacoepidemiol Drug Saf*. 2016;25:2-10. doi:10.1002/pds.3891
- Tsao CW, Vasan RS. Cohort profile: the Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol*. 2015;44:1800-1813. doi:10.1093/ije/dyv337
- Gerhard T. Bias: considerations for research practice. *Am J Health Syst Pharm*. 2008;65:2159-2168. doi:10.2146/ajhp070369
- Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol*. 2008;167:492-499. doi:10.1093/aje/kwm324
- Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest*. 2007;8:53-96. doi:10.1111/j.1539-6053.2008.00033.x
- Stürmer T, Wyss R, Glynn RJ, Brookhart MA. Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs. *J Intern Med*. 2014;275:570-580. doi:10.1111/joim.12197
- Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep*. 2015;2:221-228. doi:10.1007/s40471-015-0053-5
- Schneeweiss S, Patrick AR, Stürmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care*. 2007;45:S131-S142. doi:10.1097/MLR.0b013e318070c08e
- Suissa S, Dell'Aniello S. Time-related biases in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. 2020;29:1101-1110. doi:10.1002/pds.5083
- Pottegård A, Hallas J, Andersen JT, et al. First-trimester exposure to methylphenidate: a population-based cohort study. *J Clin Psychiatry*. 2014;75:e88-e93. doi:10.4088/JCP.13m08708
- Bliddal M, Broe A, Pottegård A, Olsen J, Langhoff-Roos J. The Danish medical birth register. *Eur J Epidemiol*. 2018;33:27-36. doi:10.1007/s10654-018-0356-1
- Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data resource profile: the Danish National Prescription Registry. *Int J Epidemiol*. 2017;46:798-798f. doi:10.1093/ije/dyw213
- Ahern TP, Tamimi RM, Rosner BA, Hankinson SE. Digoxin use and risk of invasive breast cancer: evidence from the Nurses' Health Study and meta-analysis. *Breast Cancer Res Treat*. 2014;144:427-435. doi:10.1007/s10549-014-2886-x
- Colditz GA, Manson JE, Hankinson SE. The Nurses' Health Study: 20-year contribution to the understanding of health among women. *J Womens Health*. 1997;6:49-62. doi:10.1089/jwh.1997.6.49
- Fralick M, Chen SK, Patorno E, Kim SC. Assessing the risk for gout with sodium-glucose Cotransporter-2 inhibitors in patients with type 2 diabetes: a population-based cohort study. *Ann Intern Med*. 2020;172:186-194. doi:10.7326/M19-2610
- Butler AM, Nickel KB, Overman RA, Brookhart MA. IBM MarketScan research databases. In: Sturkenboom M, Schink T, eds. *Databases for Pharmacoepidemiological Research*. Springer International Publishing; 2021:243-251. doi:10.1007/978-3-030-51455-6_20
- Hellfritsch M, Rathe J, Stage TB, et al. Generic switching of warfarin and risk of excessive anticoagulation: a Danish nationwide cohort study. *Pharmacoepidemiol Drug Saf*. 2016;25:336-343. doi:10.1002/pds.3942
- Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449-490. doi:10.2147/CLEP.S91125
- Schneeweiss S, Rassen JA, Brown JS, et al. Graphical depiction of longitudinal study designs in health care databases. *Ann Intern Med*. 2019;170:398-406. doi:10.7326/M18-3079
- Lund LC, Hallas J, Wang SV. Online tool to create publication ready graphical depictions of longitudinal study design implemented in healthcare databases. *Pharmacoepidemiol Drug Saf*. 2021;30:982. doi:10.1002/pds.5241
- Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. *Nat Med*. 2019;25:1601-1606. doi:10.1038/s41591-019-0597-x
- Schneeweiss S, Suissa S. Discussion of Schuemie et al: "A plea to stop using the case-control design in retrospective database studies". *Stat Med*. 2019;38:4209-4212. doi:10.1002/sim.8320
- Hallas J, Christensen RD, Stürmer T, Pottegård A. Measures of "exposure needed for one additional patient to be harmed" in population-based case-control studies. *Pharmacoepidemiol Drug Saf*. 2014;23:868-874. doi:10.1002/pds.3635
- Rothman KJ. *Epidemiology: an Introduction (2nd Edition)*. Oxford University Press; 2012.
- Tadrous M, Gagne JJ, Stürmer T, Cadarette SM. Disease risk score as a confounder summary method: systematic review and recommendations: DRS as a confounder summary method. *Pharmacoepidemiol Drug Saf*. 2013;22:122-129. doi:10.1002/pds.3377
- Desai RJ, Glynn RJ, Wang S, Gagne JJ. Performance of disease risk score matching in nested case-control studies: a simulation study. *Am J Epidemiol*. 2016;183:949-957. doi:10.1093/aje/kwv269
- Hallas J, Dall M, Andries A, et al. Use of single and combined anti-thrombotic therapy and risk of serious upper gastrointestinal bleeding: population based case-control study. *BMJ*. 2006;333:726. doi:10.1136/bmj.38947.697558.AE
- Hallas J, Hellfritsch M, Rix M, Olesen M, Reilev M, Pottegård A. Odense Pharmacoepidemiological database: a review of use and

- content. *Basic Clin Pharmacol Toxicol*. 2017;120:419-425. doi:[10.1111/bcpt.12764](https://doi.org/10.1111/bcpt.12764)
35. Skuladottir H, Wilcox AJ, Ma C, et al. Corticosteroid use and risk of orofacial clefts. *Birt Defects Res A Clin Mol Teratol*. 2014;100:499-506. doi:[10.1002/bdra.23248](https://doi.org/10.1002/bdra.23248)
36. Tascilar K, Dell'Aniello S, Hudson M, Suissa S. Statins and risk of rheumatoid arthritis: a nested case-control study. *Arthritis Rheumatol*. 2016;68:2603-2611. doi:[10.1002/art.39774](https://doi.org/10.1002/art.39774)
37. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015;44:827-836. doi:[10.1093/ije/dyv098](https://doi.org/10.1093/ije/dyv098)
38. Pottegård A, Friis S, de Pont CR, Habel LA, Gagne JJ, Hallas J. Identification of associations between prescribed medications and cancer: a nationwide screening study. *EBioMedicine*. 2016;7:73-79. doi:[10.1016/j.ebiom.2016.03.018](https://doi.org/10.1016/j.ebiom.2016.03.018)
39. Gjerstorff ML. The Danish cancer registry. *Scand J Public Health*. 2011;39:42-45. doi:[10.1177/1403494810393562](https://doi.org/10.1177/1403494810393562)
40. Vandenbroucke JP, Pearce N. Case-control studies: basic concepts. *Int J Epidemiol*. 2012;41:1480-1489. doi:[10.1093/ije/dys147](https://doi.org/10.1093/ije/dys147)
41. Essebag V, Genest J Jr, Suissa S, Pilote L. The nested case-control study in cardiology. *Am Heart J*. 2003;146:581-590. doi:[10.1016/S0002-8703\(03\)00512-X](https://doi.org/10.1016/S0002-8703(03)00512-X)
42. Christiansen CF, Pottegård A, Heide-Jørgensen U, et al. SARS-CoV-2 infection and adverse outcomes in users of ACE inhibitors and angiotensin-receptor blockers: a nationwide case-control and cohort analysis. *Thorax*. 2020;76:370-379. doi:[10.1136/thoraxjnl-2020-215768](https://doi.org/10.1136/thoraxjnl-2020-215768)
43. Liu P, McMenamin ÚC, Johnston BT, et al. Use of proton pump inhibitors and histamine-2 receptor antagonists and risk of gastric cancer in two population-based studies. *Br J Cancer*. 2020;123:307-315. doi:[10.1038/s41416-020-0860-4](https://doi.org/10.1038/s41416-020-0860-4)

How to cite this article: Pottegård A. Core concepts in pharmacoepidemiology: Fundamentals of the cohort and case-control study designs. *Pharmacoepidemiol Drug Saf*. 2022;1-10. doi:[10.1002/pds.5482](https://doi.org/10.1002/pds.5482)