

Chapter 2. Study Design Considerations

Abstract

The choice of study design often has profound consequences for the causal interpretation of study results. The objective of this chapter is to provide an overview of various study design options for non-experimental comparative effectiveness research (CER), their relative advantages and limitations, and provide information to guide the selection of an appropriate study design for a research question of interest. We begin the chapter by reviewing the potential for bias in non-experimental studies and the central assumption needed for non-experimental CER, i.e., that treatment groups compared have the same underlying risk for the outcome within subgroups definable by measured covariates (no unmeasured confounding). Commonly used cohort and case-control study designs are then described, along with other designs relevant to CER such as case cohort designs (selecting a random sample of the cohort and all cases), case-crossover designs (using prior exposure history of cases as their own controls), case-time controlled designs (dividing the case-crossover odds ratio by the equivalent odds ratio estimated in controls to account for calendar time trends), and self-controlled case series (estimating the immediate effect of treatment in those treated at least once). Selecting the appropriate data source, patient population, inclusion/exclusion criteria, and comparators are discussed as critical design considerations. Employing a new user design, which allows adjustment for confounding at treatment initiation without the concern of mixing confounding with selection bias during follow-up, and recognizing and avoiding immortal time bias, which is introduced by defining the exposure during the follow-up time versus prior to follow-up, are also described. The chapter concludes with a checklist for the development of the study design section of a CER protocol or proposal, emphasizing the provision of a rationale for study design selection and the need for clear definitions of inclusion/exclusion criteria, exposures (treatments), outcomes, confounders, and start of follow-up or risk period.

Introduction

The objective of this chapter is to provide an overview of various study design options for non-experimental comparative effectiveness research (CER) and their relative advantages and limitations. Out of the multitude of epidemiologic design options, we will focus on observational designs that compare two or more treatment options with respect to an outcome of interest where treatments are not assigned by the investigator but according to routine medical practice. We will not cover experimental or quasi-experimental designs, such as interrupted time series¹, designed delays², cluster randomized trials, individually randomized trials, pragmatic trials, or adaptive trials. These designs also have important roles in CER; however the focus of this guide is on non-experimental approaches to directly compare treatment options.

The choice of study design often has profound consequences for the causal interpretation of study results that are irreversible in many settings. Study design decisions must therefore be considered even more carefully than analytic decisions, which can often be changed and adapted at later stages of the research project. Those unfamiliar with non-experimental design options are thus strongly encouraged to involve experts in the design of non-experimental treatment comparisons, e.g., epidemiologists, especially ones that are familiar with comparing medical treatments (e.g., pharmacoepidemiologists) during the planning stage of CER studies and throughout the project. In the planning stage of a CER study, researchers need to determine

whether the research question should be studied using non-experimental or experimental methods (or a combination thereof, e.g., 2-stage RCTs).^{3,4} Feasibility may determine whether an experimental or a non-experimental design is most suitable; situations may arise where neither approach is feasible.

Issues of Bias in Observational CER

In observational CER, the exposures or treatments are not assigned by the investigator but rather by mechanisms of routine practice. Although the investigator can (and should) speculate on the treatment assignment process or mechanism, the actual process will be unknown to the investigator. The non-random nature of treatment assignment leads to the major challenge in non-experimental CER studies, ensuring internal validity. Internal validity is defined as the absence of bias; biases may be broadly classified as selection bias, information bias, and confounding bias. Epidemiology has advanced our thinking about these biases for more than 100 years, and many papers describing the underlying concepts and approaches to bias reduction have been published. For a comprehensive description and definition of these biases we suggest the book *Modern Epidemiology*.⁵ Ensuring a study's internal validity is a prerequisite for its external validity or generalizability. The limited generalizability of findings from randomized controlled trials (RCTs), e.g., to older adults, patients with co-morbidities or co-medications, is one of the major drivers for the conduct of non-experimental CER.

The central assumption needed for non-experimental CER is that the treatment groups compared have the same underlying risk for the outcome within subgroups definable by measured covariates. Until recently this “no unmeasured confounding” assumption was deemed plausible only for unintended (usually adverse) effects of medical interventions, i.e., safety studies. The assumption was considered to be less plausible for intended effects of medical interventions (effectiveness) because of intractable confounding by indication.^{6,7} Confounding by indication leads to higher propensity for treatment or more intensive treatment in those with the most severe disease. A typical example would be a study on the effects of beta-agonists on asthma mortality in patients with asthma. The association between treatment (intensity) with beta-agonists and asthma mortality would be confounded by asthma severity. The direction of the confounding by asthma severity would tend to make the drug look bad (as if it is “causing” mortality). The study design challenge in this example would not be the confounding itself, but that it is hard to control for asthma severity because it is difficult to measure precisely. Confounding by frailty has been identified as another potential bias when assessing preventive treatments in population-based studies, particularly those among older adults.^{8,9, 10,11} Because frail persons (close to death) are less likely to be treated with a multitude of preventive treatments⁸, frailty would lead to confounding which would bias the association between preventive treatments and outcomes associated with frailty (e.g., mortality). Since the bias would be that the untreated cohort has a higher mortality irrespective of the treatment, this would make the drug's effectiveness look too good. Here again the crux of the problem is that frailty is hard to control for because it is difficult to measure.

Basic Epidemiologic Study Designs

The general principle of epidemiologic study designs is to compare the distribution of the outcome of interest in groups characterized by the exposure/treatment/intervention of interest. The association between the exposure and outcome is then assessed using measures of

association. The causal interpretation of these associations is dependent on additional assumptions, most notably that the risk for the outcome is the same in all treatment groups compared (before they receive the respective treatments), also called exchangeability.^{12,13} Additional assumptions for a causal interpretation, starting with the Hill criteria¹⁴, are beyond the scope of this chapter, although most of these are relevant to many CER settings (i.e., when treatment effects are heterogeneous, see chapter 3).

The basic epidemiologic study designs are usually defined by whether study participants are sampled based on their exposure or outcome of interest. In a cross sectional study, participants are sampled independent of exposure and outcome and prevalence of exposure and outcome are assessed at the same point in time. In cohort studies, participants are sampled according to their exposures and followed over time for the incidence of outcomes. In case-control studies, cases and controls are sampled based on the outcome of interest and the prevalence of exposure in these two groups is then compared. Because the cross-sectional study design usually does not allow the investigator to define whether the exposure preceded the outcome, one of the prerequisites for a causal interpretation, we will focus on cohort and case-control studies as well as some more advanced designs with specific relevance to CER.

Definitions of some common epidemiologic terms are presented in Table 2.1. Given the space constraints and the intended audience, these definitions do not capture all nuances.

Table 2.1. Definition of epidemiologic terms

Term	Definition	Comments
Incidence	Occurrence of the disease outcome over a specified time period. Incidence is generally assessed as a risk/proportion over a fixed time-period (e.g., risk for 1-year mortality) or as a rate defined by persons and time (e.g., mortality rate per person-year). Incidence is often defined as first occurrence of the outcome of interest which requires prior absence of the outcome.	Etiologic studies are based on incidence of the outcome of interest rather than prevalence because prevalence is a function of disease incidence and duration of disease.
Prevalence	Proportion of persons with the exposure/outcome at a specific point in time. Because prevalence is a function of the incidence and the mean duration of the disease, incidence is generally used to study etiology.	

Term	Definition	Comments
Measures of association	Needed to compare outcomes across treatment groups. The main epidemiologic measures of association are ratio measures (risk ratio, incidence rate ratio, odds ratio, hazard ratio) and difference measures (risk difference, incidence rate difference).	Difference measures have some very specific advantages over ratio measures, including the possibility to calculate numbers needed to treat (or harm) and providing a biologically more meaningful scale to assess heterogeneity. ⁵ Ratio measures nevertheless abound in medical research. All measures of association should be accompanied by a measure of precision, e.g., a confidence interval.
Confounding	Mixing of effects; the effect of the treatments is mixed with the effect of the underlying risk for the outcome being different in the treatment groups compared.	Leads to biased treatment effect estimates unless controlled for by design (randomization, matching, restriction) or analysis (stratification, multivariable models).
Selection bias	Distortion of treatment effect estimate as a result of procedures used to select subjects and from factors that influence study participation.	While procedures to select subjects usually lead to confounding that can be controlled for, factors affecting study participation cannot be controlled for; factors affecting study participation are referred to as selection bias throughout this chapter to differentiate it from confounding.
Information bias	Distortion of treatment effect estimate as a result of measurement error in any variable used in a study, i.e., exposure, confounder, outcome.	Often measurement error is used for continuous variables and misclassification for categorical variables; it is important to separate non-differential from differential measurement error; non-differential measurement error in exposures and outcomes tends to bias treatment effect estimates towards the null (no effect); non-differential measurement error in confounders leads to residual confounding (any direction); differential measurement error leads to bias in any direction.

Cohort Study

Description

Cohorts are defined by their exposure at a certain point in time (baseline date) and are followed over time after baseline for the occurrence of the outcome. For the usual study of first occurrence of outcomes, cohort members with the outcome prevalent at baseline need to be excluded.

Cohort entry (baseline) is ideally defined by a meaningful event (e.g., initiation of treatment; see section on new user design) rather than convenience (prevalence of treatment) although this may not always be feasible or desirable.

Advantages

The main advantage of the cohort design is that it has a clear timeline separating potential confounders from the exposure and the exposure from the outcome. Cohorts allow the estimation of actual incidence (risk or rate) in all treatment groups and thus the estimation of risk or rate differences. Cohort studies allow investigators to assess multiple outcomes from given treatments. The cohort design is also easy to conceptualize and readily compared to the RCT, a design with which most medical researchers are very familiar.

Limitations

If participants need to be recruited and followed over time for the incidence of the outcome, the cohort design quickly becomes inefficient when the incidence of the outcome is low. This has led to the widespread use of case-control designs (see below) in pharmacoepidemiologic studies using large automated databases. With the IT revolution over the past 10 years, lack of efficiency is rarely, if ever, a reason not to implement a cohort study even in the largest healthcare databases if all data have already been collected.

Important considerations

Patients can only be excluded from the cohort based on information available at start of follow-up (baseline). Any exclusion of cohort members based on information accruing during follow-up, including treatment changes, has a strong potential to introduce bias. The idea to have a “clean” treatment group usually introduces selection bias, e.g., by removing the sickest, those with treatment failure, or those with adverse events, from the cohort. The fundamental principle of the cohort is the enumeration of people at baseline (based on inclusion and exclusion criteria) and reporting losses to follow-up for everyone enrolled at baseline. Clinical researchers may also be tempted to assess the treatments during the same time period the outcome is assessed (i.e., during follow-up) instead of prior to follow-up. Another fundamental of the cohort design is, however, that the exposure is assessed prior to the assessment of the outcome, thus limiting the potential for reverse causality. This general principle also applies to time-varying treatments for which the follow-up time needs to start new after treatment changes rather than from baseline.

Cadarette et al.¹⁵ employed a cohort design to investigate the comparative effectiveness of 4 alternative treatments to prevent osteoporotic fractures. The four cohorts were defined by the initiation of the four respective treatments (baseline date). Cohorts were followed from baseline to the first occurrence of a fracture at various sites. Statistical analyses adjusted for risk factors for fractures assessed at baseline to minimize bias. As discussed, the cohort design provided a clear timeline, differentiating exposure from potential confounders and the outcomes.

Case-control Study

Description

Nested within an underlying cohort, the case-control design identifies all incident cases that develop the outcome of interest and compares their exposure history with the exposure history of controls sampled at random from everyone within the cohort that is still at risk for developing the outcome of interest. Given proper sampling of controls from the risk set, the estimation of the

odds ratio in a case-control study is a computationally more efficient way to estimate the otherwise identical incidence rate ratio in the underlying cohort.

Advantages

The oversampling of persons with the outcome increases efficiency compared with the full underlying cohort. As outlined above, this efficiency advantage is of minor importance in many CER settings. Efficiency is of major importance, however, if additional data (e.g., blood levels, biologic materials, validation data) need to be collected. It is straightforward to assess multiple exposures, although this will quickly become very complicated when implementing a new user design.

Limitations

The case-control study is difficult to conceptualize. Some researchers do not understand, for example, that matching does not control for confounding in a case-control study, whereas it does in a cohort study.¹⁶ Unless additional information from the underlying cohort is available, risk or rate differences cannot be estimated from case-control studies. Because the timing between potential confounders and the treatments is often not taken into account, current implementations of the case control design assessing confounders at the index date rather than prior to treatment initiation will be biased when controlling for covariates that may be affected by prior treatment. Thus, implementing a new user design with proper definition of confounders will often be difficult, although not impossible. If information on treatments needs to be obtained retrospectively, e.g., from an interview with study participants identified as cases and controls, there is the potential that treatments will be assessed differently for cases and controls which will lead to bias (often referred to as recall bias).

Important considerations

Controls need to be sampled from the “risk set”, i.e., all patients from the underlying cohort who remain at risk for the outcome at the time a case occurs. Sampling of controls from all those who enter the cohort (i.e., at baseline) may lead to biased estimates of treatment effects if treatments are associated with loss to follow-up or mortality. Matching on confounders can improve the efficiency of estimation of treatment effects, but does not control for confounding in case control studies. Matching should only be considered for strong risk factors for the outcome; however, the often small gain in efficiency must be weighed against the loss of the ability to estimate the effect of the matching variable on the outcome (which could, for example, be used as a positive control to show content validity of an outcome definition).¹⁷ Matching on factors strongly associated with treatment often reduces efficiency of case control studies (overmatching). Generally speaking, matching should not routinely be performed in case-control studies but be carefully considered, ideally after some study of the expected efficiency gains.^{16,18}

Martinez et al.¹⁹ conducted a case-control study employing a new user design. The investigators compared venlafaxine and other anti-depressants and risk of sudden cardiac death or near death. An existing cohort of new users of anti-depressants was identified (“new” users were defined as subjects without a prescription for the medication in the year prior to cohort entry). Nested within the underlying cohort, cases and up to 30 randomly selected matched controls were identified. Potential controls were assigned an “index date” corresponding to the same follow-up time to event as the matched case. Controls were only sampled from the “risk set”; i.e., controls

had to be at risk for the outcome on their index date, thus ensuring that bias was not introduced via the sampling scheme.

Case Cohort Study

In the case-cohort design, cohorts are defined as in a cohort study, and all cohort members are followed for the incidence of the outcomes. Additional information required for analysis (e.g., blood levels, biologic materials for genetic analyses) is collected for a random sample of the cohort and for all cases (note that the random sample may contain cases). This sampling needs to be accounted for in the analysis²⁰, but otherwise this design offers all the advantages and possibilities of a cohort study. The case-cohort design is intended to increase efficiency compared with the nested case-control design when selecting participants for whom additional information needs to be collected or when studying more than one outcome.

Other Epidemiological Study Designs Relevant to CER

Case-crossover Design

Faced with the problem of selection of adequate controls in a case-control study of triggers of myocardial infarction, Maclure proposed to use prior exposure history of cases as their own controls.²¹ For this study design, only patients with the outcome (cases) who have discrepant exposures during the case and the control period contribute information. A feature of this design is that it is self-controlled, which removes the confounding effect of any characteristic of subjects that are stable over time (e.g., genetics). For CER, the latter property of the case-crossover design is a major advantage because measures of stable confounding factors (to address confounding) are not needed. The former property or initial reason to develop the case-crossover design, i.e. its ability to assess triggers of (or immediate, reversible effects of e.g., treatments on) outcomes may also have specific advantages for CER. The case crossover design is thought to be appropriate for studying acute effects of transient exposures.

Description

Exactly as in a case-control study, the first step is to identify all cases with the outcome and assess the prevalence of exposure during a brief time window before the outcome occurred. Instead of sampling controls, we create a separate observation for each case that contains all the same variables except for the exposure, which is defined for a different time period. This “control” time period has the same length as the case period and needs to be carefully chosen to take e.g., seasonality of exposures into account. The dataset is then analyzed as an individually matched case-control study.

Advantages

The lack of need to select controls, the ability to assess short-term reversible effects, the ability to inform about the time window for this effect using various intervals to define treatment, and the control for all, even unmeasured factors that are stable over time are the major advantages of the case-crossover design. The design can also be easily added to any case-control study with little (if any) cost.

Limitations

Because only cases with discrepant exposure history contribute information to the analysis, the case-crossover design is often not very efficient. This may not be a major issue if the design is used in addition to the full case-control design. While the design avoids confounding by factors that are stable over time, it can still be confounded by factors that vary over time. The possibility of time-varying conditions leading to changes in treatment and increasing the risk for the outcome (i.e., confounding by indication) would need to be carefully considered in CER studies.

The causal interpretation changes from the effect of treatment versus no treatment on the outcome to *the short term effect of treatment in those treated*. Thus, it can be used to assess the effects of adherence/persistence with treatment on outcomes in those who have initiated treatment.²²

Case-time Controlled Design

One of the assumptions behind the case-crossover design is that the prevalence of exposure stays constant over time in the population studied. While plausible in many settings, this assumption may be violated in dynamic phases of therapies (after market introduction or safety alerts). To overcome this problem, Suissa proposed the case-time controlled design.²³ This approach divides the case-crossover odds ratio by the equivalent odds ratio estimated in controls. Greenland has criticized this design because it can re-introduce confounding thus detracting from one of the major advantages of the case-crossover design.²⁴

Description

This study design tries to adjust for calendar time trends in the prevalence of treatments which can introduce bias in the case-crossover design. To do so, the design uses controls as in a case-control design but estimates a case-crossover odds ratio (i.e., within individuals) in these controls. The case-crossover odds ratio (in cases) is then divided by the case-crossover odds ratio in controls.

Advantages

This design is the same as case-crossover design (with the caveat outlined by Greenland) with the additional advantage of not being dependent on the assumption of no temporal changes in the prevalence of the treatment.

Limitations

The need for controls removes the initial motivation for the case-crossover design and adds complexity. The control for the time trend can introduce confounding although the magnitude of this problem for various settings has not been quantified.

Self-controlled Case-series

Some of the concepts of the case-crossover design have also been adapted to cohort studies. This design, called self-controlled case-series²⁵, shares most of the advantages with the case-crossover design, but requires additional assumptions.

Description

As with the case-crossover design, the self-controlled case series estimates the immediate effect of treatment in those treated at least once. It is similarly dependent on cases that have changes in treatment during a defined period of observation time. This observation time is divided into treated person-time, a washout period of person-time, and untreated person-time. A conditional Poisson regression is used to estimate the incidence rate ratio within individuals. A SAS macro with software to arrange the data and to run the conditional Poisson regression is available.^{26,27}

Advantages

The self-controlled design controls for factors that are stable over time. The cohort design, using all the available person-time information, has the potential to increase efficiency compared with the case-crossover design. The design was originally proposed for rare adverse events in vaccine safety studies for which it seems especially well suited.

Limitations

The need for repeated events or, alternatively, a rare outcome, and the apparent need to assign person-time for treatment even after the outcome of interest occurs, limits the applicability of the design in many CER settings. The assumption that the outcome does not affect treatment will often be implausible. Furthermore, it precludes the study of mortality as an outcome. The reason why treatment information after the outcome is needed is not obvious to us and this issue needs further study. More work is needed to understand the relation of the self-controlled case-series with the case-crossover design and to delineate relative advantages and limitations of these designs for specific CER settings.

Study Design Features

Study Setting

One of the first decisions with respect to study design is consideration of the population and data source(s) from which the study subjects will be identified. Usually, the general population or a population-based approach is preferred but selected populations (e.g., a drug/device or disease registry) may offer advantages such as availability of data on covariates in specific settings. Availability of existing data and their scope and quality will determine whether a study can be done using existing data or whether additional new data need to be collected (see Chapter 8 for a full discussion of data sources). Researchers should start with a definition of the treatments and outcomes of interest, as well as the predictors of outcome risk potentially related to choice of treatments of interest (i.e., potential confounders). Once these have been defined, availability and validity of information on treatments, outcomes, and confounders in existing databases should be weighed against the time and cost involved in collecting additional or new data. This process is iterative insofar that availability and validity of information may inform the definition of treatments, outcomes, and potential confounders. We need to point out that we do not make the distinction between retrospective and prospective studies here because this distinction does not affect the validity of the study design. The only difference between these general options of how to implement a specific study design lies in the potential to influence what kind of data will be available for analysis.

Inclusion and Exclusion Criteria

Every CER study should have clearly defined inclusion and exclusion criteria. The definitions need to include details about the study time period and dates used to define these criteria. Great care should be taken to use uniform periods to define these criteria for all subjects. If this cannot be achieved, then differences in periods between treatment groups need to be carefully evaluated because such differences have the potential to introduce bias. Inclusion and exclusion criteria need to be defined based on information available at baseline and cannot be updated based on accruing information during follow-up (see discussion of immortal time below).

Inclusion and exclusion criteria can also be used to increase internal validity of non-experimental studies. Consider an example where an investigator suspects that an underlying comorbidity is a confounder of the association under study. A diagnostic code with a low sensitivity but high specificity for the underlying comorbidity exists (i.e., many subjects with the comorbidity aren't coded; however, for patients that do have the code, nearly all have the comorbidity). In this example, the investigator's ability to control for confounding by the underlying comorbidity would be hampered by the low sensitivity of the diagnostic code (as there are potentially many subjects with the comorbidity that are not coded). In contrast, restricting the study population to those with the diagnostic code removes confounding by the underlying condition due to the high specificity of the code.

It should be noted that inclusion and exclusion criteria also affect generalizability of results. If in doubt, potential benefits in internal validity will outweigh any potential reduction in generalizability.

Choice of Comparators

Both confounding by indication and confounding by frailty may be strongest and most difficult to adjust for when comparing treated with untreated persons. One way to reduce the potential for confounding is to compare the treatment of interest with a different treatment for the same indication or an indication with a similar potential for confounding.²⁸ A comparator treatment within the same indication is likely to reduce the potential for bias from both confounding by indication and confounding by frailty. This opens the door for using non-experimental methods to study intended effects of medical interventions (effectiveness). Comparing different treatment options for a given patient (i.e., the same indication) is at the very core of CER. Thus both methodological and clinical relevance considerations lead to the same principle for study design.

Another beneficial aspect of choosing an active comparator group comprised of a treatment alternative for the same indication is the identification of the point in time when the treatment decision is made, so that all subjects may start follow-up at the same time, "synchronizing" both the timeline and the point at which baseline characteristics are measured. This reduces the potential for various sources of confounding and selection bias, including by barriers to treatment (e.g., frailty).^{8,29} A good source for active comparator treatments are current treatment guidelines for the condition of interest.

Other Study Design Considerations

New User Design

It has long been realized that the biologic effects of treatments may change over time since initiation.³⁰ Guess used the observed risk of angioedema after initiation of angiotensin converting enzyme inhibitors, which is orders of magnitude higher in the first week after initiation compared with subsequent weeks³¹, to make the point. Non-biologic changes of treatment effects over time since initiation may also be caused by selection bias.^{8,29,32} For example, Dormuth et al.³¹ examined the relationship between adherence to statin therapy (more adherent versus less adherent) and a variety of outcomes thought to be associated with and not associated with statin use. The investigators found that subjects classified as more adherent were less likely to experience negative health outcomes unlikely to be caused by statin treatment.

Poor health, e.g., frailty, is also associated with non-adherence in RCTs³³ and thus those adhering to randomized treatment will appear to have better outcomes, including those adhering to placebo.³³ This selection bias is most pronounced for mortality³⁴, but extends to a wide variety of outcomes, including accidents.³¹ The conventional prevalent user design thus is prone to suffer from both confounding and selection bias. While confounding by measured covariates can usually be addressed by standard epidemiologic methods, selection bias cannot. An additional problem of studying prevalent users is that covariates that act as confounders may also be influenced by prior treatment (e.g., blood pressure, asthma severity, CD4 count); in such a setting, necessary control for these covariates to address confounding will introduce bias because some of the treatment effect is removed.

The new user design^{6,30,31,35,36} is the logical solution to the problems resulting from inclusion of persons who are persistent with a treatment over prolonged periods because researchers can adjust for confounding at initiation without the concern of selection bias during follow-up. Additionally, the approach avoids the problem of confounders potentially being influenced by prior treatment, and provides approaches for structuring comparisons which are free of selection bias, such as first treatment carried forward or intention to treat. These and other considerations are covered in further detail in Chapter 5. In addition, the new user design offers a further advantage in anchoring the time scale for analysis at time since initiation of treatment for all subjects under study. Advantages and limitations of the new user design are clearly outlined in the paper by Ray.³⁶ Limitations include the reduction in sample size leading to reduced precision of treatment effect estimates and the potential to lead to a highly selected population for treatments often used intermittently (e.g., pain medications).³⁷ Given the conceptual advantages of the new user design to address confounding and selection bias, it should be the default design for CER studies; deviations should be argued for and their consequences discussed.

Immortal Time Bias

While the term “immortal time bias” was introduced by Suissa in 2003³⁸, the underlying bias introduced by defining the exposure during the follow-up time rather than before follow-up was first outlined by Gail.³⁹ Gail noted that the survival advantage attributed to getting a heart transplant in two studies enrolling cohorts of potential heart transplant recipients was a logical consequence of the study design. The studies compared survival in those that later got a heart transplant with those that did not, starting from enrollment (getting on the heart transplant list). As one of the conditions to get a heart transplant is survival until the time of surgery, this survival time prior to the exposure classification (heart transplant or not) should not be attributed

to the heart transplant and is described as “immortal”. Any observed survival advantage in those who received transplants cannot be clearly ascribed to the intervention if time prior to the intervention is included because of the bias introduced by defining the exposure at a later point during follow-up. Suissa³⁸ showed that a number of pharmacoepidemiologic studies assessing the effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease were also affected by immortal time bias. While immortal person time and the corresponding bias is introduced whenever exposures (treatments) are defined during follow-up, immortal time bias can also be introduced by exclusion of patients from cohorts based on information accrued after the start of follow-up, i.e., based on changes in treatment or exclusion criteria during follow-up.

It should be noted that both the new user design and the use of comparator treatments reduce the potential for immortal time bias. These design options are no guarantee against immortal time bias, however, unless the corresponding definitions of cohort inclusion and exclusion criteria are based exclusively on data available at start of follow-up (i.e. at baseline).⁴⁰

Conclusion

This chapter provides an overview of advantages and limitations of various study designs relevant to CER. It is important to realize that many see the cohort design as more valid than the case-control design. Although the case-control design may be more prone to potential biases related to control selection and recall in *ad hoc* studies, if a case-control study is nested within an existing cohort (e.g., based within a large healthcare database) its validity is equivalent to the one of the cohort study under the condition that the controls are sampled appropriately and the confounders are assessed during the relevant time period (i.e., before the treatments). Because the cohort design is generally easier to conceptualize, implement, and communicate, and computational efficiency will not be a real limitation in most settings, the cohort design will be preferred when data have already been collected. The cohort design has the added advantage that absolute risks or incidence rates can be estimated and therefore risk or incidence rate differences can be estimated, which have specific advantages as outlined above. While we would always recommend including an epidemiologist in the early planning phase of a CER study, an experienced epidemiologist would be a prerequisite outside of these basic designs.

Some additional study designs have not been discussed. These include hybrid designs like 2-stage studies⁴¹, validation studies⁴², ecologic designs arising from natural experiments, interrupted time series, adaptive designs and pragmatic trials. Many of the issues that will be discussed in the following chapters about how to deal with treatment changes (stopping, switching, and augmenting) will also need to be addressed in pragmatic trials because their potential to introduce selection bias will be the same in both experimental and non-experimental studies.

Knowledge of study designs and design options is essential to increase internal and external validity of non-experimental CER studies. An appropriate study design is a prerequisite to reduce the potential for bias. Biases introduced by suboptimal study design cannot usually be removed during the statistical analysis phase. Therefore the choice of an appropriate study design is at least as important, if not more important, than the approach to statistical analysis.

Checklist: Guidance and Key Considerations for Study Design for an Observational CER protocol or proposal

Guidance	Key Considerations	Check
Provide rationale for study design choice and describe key design features	<ul style="list-style-type: none"> - Cohort study proposals should clearly define cohort entry date (baseline date), employ a new user design (or provide rationale for including prevalent users), and plans for reporting losses to follow-up - Case-control study proposals should clearly describe the control sampling method, employ a new user design (or provide a rationale for assessing confounders at index date), and assess potential for recall bias (if applicable) - Case-cohort study proposals should include how the sampling scheme will be accounted for during analysis - Case-crossover study proposals should discuss the potential for confounding by time-varying factors, and clearly state how the resulting effect estimate can be interpreted - Case-time controlled study proposals should clearly weigh the pros and cons of accounting for calendar trends in the prevalence of exposure 	<input type="checkbox"/>
Define start of follow-up (baseline)	<ul style="list-style-type: none"> - The time point for start of follow-up should be clearly defined and meaningful, ideally anchored to the time of a medical intervention (e.g., initiation of drug use) - If alternative approaches are proposed, the rationale should be provided and implications discussed 	<input type="checkbox"/>
Define inclusion and exclusion criteria at start of follow-up (baseline)	<ul style="list-style-type: none"> - Exclusion and inclusion criteria should be defined at the start of follow-up (baseline) and solely based on information available at this point in time (i.e., ignoring potentially known events after baseline). - The definition should include the time window for assessment (usually the same for all cohort members) 	<input type="checkbox"/>
Define exposure (treatments) of interest at start of follow-up		<input type="checkbox"/>
Define outcome(s) of interest	<ul style="list-style-type: none"> - Provide information on measures of accuracy if possible 	<input type="checkbox"/>
Define potential confounders	<ul style="list-style-type: none"> - Potential confounders known to be associated with treatment and outcome should be pre-specified when possible - Confounders should be assessed prior to exposure or treatment initiation to ensure they are not affected by the exposure - Approaches to empirical identification of confounders should be described if planned 	<input type="checkbox"/>

References

- ¹ Schneeweiss S, Maclure M, Walker AM, Grootendorst P, Soumerai SB. On the evaluation of drug benefits policy changes with longitudinal claims data: the policy maker's versus the clinician's perspective. *Health Policy*. 2001 Feb;55(2):97-109.
- ² Maclure M, Carleton B, Schneeweiss S. Designed delays versus rigorous pragmatic trials: lower carat gold standards can produce relevant drug evaluations. *Med Care*. 2007 Oct;45(10 Supl 2):S44-9.
- ³ Rücker G. A two-stage trial design for testing treatment, self-selection and treatment preference effects. *Stat Med*. 1989 Apr;8(4):477-85.
- ⁴ Fava M, Rush AJ, Trivedi MH, et al. Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatr Clin North Am*. Jun 2003;26(2):457-494.
- ⁵ Rothman KJ, Greenland S, Lash T. (Eds.). (2008). *Modern Epidemiology*. (3rd ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.
- ⁶ Miettinen OS, Caro JJ. Principles of nonexperimental assessment of excess risk, with special reference to adverse drug reactions. *J Clin Epidemiol*. 1989;42(4):325-31.
- ⁷ Yusuf, S., Collins, R. & Peto, R. (1984). Why do we need some large, simple randomized trials? *Statistics in Medicine* 3, 409–420.
- ⁸ Glynn RJ, Knight EL, Levin R, Avorn J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology*. 2001 Nov;12(6):682-9.
- ⁹ Stürmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. *American Journal of Epidemiology* 2005; 161: 891-898.
- ¹⁰ Stürmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment effects in the presence of unmeasured confounding: Dealing with observations in the tails of the propensity score distribution – a simulation study. *American Journal of Epidemiology* 2010;172:843-854.
- ¹¹ 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 Apr;35(2):337-44. Epub 2005 Dec 20.
- ¹² Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*. 1986 Sep;15(3):413-9.
- ¹³ Greenland S, Robins JM. Identifiability, exchangeability and confounding revisited. *Epidemiol Perspect Innov*. 2009 Sep 4;6:4.
- ¹⁴ Hill, Austin Bradford (1965). "The environment and disease: association or causation?". *Proceedings of the Royal Society of Medicine* 58: 295–300.
- ¹⁵ Cadarette SM, Katz JN, Brookhart MA, Stürmer T, Stedman MR, Solomon DH. Relative effectiveness of osteoporosis drugs for nonvertebral fracture prevention: a cohort study. *Annals of Internal Medicine* 2008;148:637-46.
- ¹⁶ Stürmer T, Poole C. Matching in cohort studies: return of a long lost family member. [Symposium] *American Journal of Epidemiology* 2009;169(Suppl):S128.
- ¹⁷ Stürmer T, Brenner H. Degree of matching and gain in power and efficiency in case-control studies. *Epidemiology* 2001; 12: 101-108.

- ¹⁸ Stürmer T, Brenner H. Flexible matching strategies to increase power and efficiency to detect and estimate gene-environment interactions in case-control studies. *American Journal of Epidemiology* 2002; 155: 593-602.
- ¹⁹ Martinez C, Assimes TL, Mines D, Dell'aniello S, Suissa S. Use of venlafaxine compared with other antidepressants and the risk of sudden cardiac death or near death: a nested case-control study. *BMJ*. 2010 Feb 5;340:c249. doi: 10.1136/bmj.c249.
- ²⁰ Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; 73: 1–11.
- ²¹ Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol*. 1991 Jan 15;133(2):144-53.
- ²² Maclure M. 'Why me?' versus 'why now?'--differences between operational hypotheses in case-control versus case-crossover studies. *Pharmacoepidemiol Drug Saf*. 2007 Aug;16(8):850-3.
- ²³ Suissa S. The case-time-control design. *Epidemiology*. 1995 May;6(3):248-53.
- ²⁴ Greenland S. Confounding and exposure trends in case-crossover and case-time-control designs. *Epidemiology*. 1996 May;7(3):231-9.
- ²⁵ Farrington CP. Control without separate controls: evaluation of vaccine safety using case-only methods. *Vaccine* 2004;22(15-16):2064-70.
- ²⁶ Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006; May 30;25(10):1768-97.
- ²⁷ Gibson JE, Hubbard RB, Smith CJP, Tata LJ, Britton JR. The Use of Self-Controlled Analytical Techniques to Assess the Temporal Association Between use of Prescription Medications and the Risk of Motor Vehicle Crashes. *Am J Epidemiol* 2009;169(6):761-68.
- ²⁸ Glynn RJ, Schneeweiss S, Stürmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. *Basic and Clinical Pharmacology and Toxicology* 2006;98:253-9.
- ²⁹ Patrick AR, Shrank WH, Glynn RJ, Solomon DH, Dormuth CR, Avorn J, Cadarette SM, Mogun H, Brookhart MA. The Association between Statin use and Outcomes Potentially Attributable to an Unhealthy Lifestyle in Older Adults. *Value Health*. 2011 Jun;14(4):513-20. Epub 2011 Apr 22.
- ³⁰ Kramer MS, Lane DA, Hutchinson TA. Analgesic use, blood dyscrasias, and case-control pharmacoepidemiology. A critique of the international agranulocytosis and aplastic anemia study. *J Chron Dis* 1987;40:1073-1081.
- ³¹ Guess HA. Behavior of the exposure odds ratio in a case-control study when the hazard function is not constant over time. *J Clin Epidemiol* 1989;42:1179-1184.
- ³² Dormuth CR, Patrick AR, Shrank WH, Wright JM, Glynn RJ, Sutherland J, Brookhart MA. Statin adherence and risk of accidents: a cautionary tale. *Circulation*. 2009 Apr 21;119(15):2051-7. Epub 2009 Apr 6.
- ³³ Simpson SH, Eurich DT, Majumdar SR, Padwal RS, Tsuyuki RT, Varney J, Johnson JA. A meta-analysis of the association between adherence to drug therapy and mortality. *BMJ* 2006;333:15.
- ³⁴ Andersen M, Brookhart MA, Glynn RJ, Stovring H, Stürmer T. Practical issues in measuring cessation and re-initiation of drug use in databases. [abstract] *Pharmacoepidemiol Drug Saf* 2008;17(suppl 1):S27
- ³⁵ Moride Y, Abenheim L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol*. 1994 Jul;47(7):731-7.
- ³⁶ Wayne A. Ray. Evaluating Medication Effects Outside of Clinical Trials: New-User Designs. *Am J Epidemiol* 2003;158:915-20.

- ³⁷ Valkhoff VE, Romio SA, Schade R, Schuemie M, Garbe E, Herings R, Lucchi S, Picelli G, Schink T, Straatman H, Villa M, Zambon A, Kuipers EJ, Sturkenboom MCJM. Influence of run-in period on incidence of NSAID use in European population in the SOS project. *Pharmacoepidemiol Drug Saf* 2011;20 (supl 1):S250.
- ³⁸ Suissa S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: immortal time bias in observational studies. *Am J Respir Crit Care Med* 2003;168(1):49-53.
- ³⁹ Gail MH. Does cardiac transplantation prolong life? A reassessment. *Ann Intern Med* 1972;76(5):815-7.
- ⁴⁰ Pocock SJ, Smeeth L. Insulin glargine and malignancy: an unwarranted alarm. *Lancet* 2009;374(9689):511-3.
- ⁴¹ 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 May;9(3):309-15.
- ⁴² 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 Supl 2):S158-65.