CHAPTER 9

General discussion
The studies presented in this thesis focused on confounding in observational studies. Several designs and analytical methods to quantify confounding were evaluated using single-study and systematic comparisons. We focused on various self-controlled designs and compared our findings with randomized designs as well as more traditional observational designs such as the cohort and case-control design.

In Chapter 2 we evaluated whether the reporting of confounding in observational cohort and case-control studies improved after the publication of the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) statement using a systematic before-after comparison. In total, 174 articles published before and 220 articles published after the publication of the STROBE statement in high-impact general medical and epidemiological journals were included. Out of 8 pre-specified essential items in the reporting of confounding, on average only four were reported in both periods. Results were similar for journals that published and/or endorsed the STROBE statement in their Instruction for Authors and journals that did neither. Hence, research is needed into the development and evaluation of alternative strategies to improve the quality of reporting and adherence to reporting guidelines.

The effect of pravastatin and fosinopril, an angiotensin converting enzyme (ACE) inhibitor, on urinary tract infections was assessed in a post-hoc analysis of a randomized controlled trial (Chapter 3). Intention-to-treat analyses showed that pravastatin was associated with a reduced total number of urinary tract infections, but not with first urinary tract infections. Fosinopril was in contrast associated with an increased occurrence of first urinary tract infections. Pravastatin was in contrast associated with an increased occurrence of first urinary tract infections.

Subsequently, we evaluated the effect of ACE inhibitors on the risk of first urinary tract infections using both a prescription sequence symmetry analysis (Chapter 4) and a case-crossover design (Chapter 5). Using both designs ACEi therapy initiation was associated with an increased risk of developing first urinary tract infections, even after adjustment for several time-varying confounders. Despite the similarities between the case-crossover design and the prescription sequence symmetry analysis, the latter design led to slightly lower effect estimates than the case-crossover design and the results from the randomized design discussed in chapter 3.

In Chapter 6 we re-evaluated a previously published cohort study without adjustments for potential confounders that evaluated the association of combined use of selective serotonin reuptake inhibitors (SSRIs) and nonsteroidal anti-inflammatory drugs (NSAIDs) with the risk of starting peptic ulcer treatment. To evaluate whether the strong synergistically increased risk found in that other study is likely due to confounding we applied a prescription sequence symmetry design (Chapter 6). Using the prescription
sequence symmetry design, the effect estimate of concurrent use of SSRIs and NSAIDs did not exceed the effect estimate of NSAIDs alone, suggesting that at least part of the previously reported association between combined use of SSRIs with NSAIDs and peptic ulcer treatment might be attributed to unmeasured or residual confounding. Furthermore, this study shows that when limited data on potential confounders are available, applying a prescription sequence symmetry design can have more value than applying a traditional cohort design.

Given the ongoing debate about whether antibiotic use during pregnancy increases the risk of asthma development in the offspring, we assessed whether this increased risk is likely causal or due to confounding. We made a comparison between a case-control, case-sibling, a time-trend-adjusted case-sibling design and a maternal-paternal comparison and applied a quantitative bias analysis to evaluate the potential role of confounding (Chapter 7). In both the case-control, case-sibling and time-trend-adjusted case-sibling design exposure to antibiotics in the third trimester of pregnancy was associated with a small increased risk of asthma in preschool children. However, quantitative bias analysis showed that time-varying confounding could not be excluded as an explanation of the statistically significant findings.

Finally, we made a systematic comparison between effect estimates of the two most common case-only designs, the self-controlled case-series and case-crossover design, and the more traditional case-control and cohort design. A predictive model was built to assess whether discrepancies between both types of designs could be predicted by failure to meet assumptions of the case-only designs (Chapter 8). The concordance between effect estimates of case-only and cohort or case-control designs appeared to be moderate, and discrepancies beyond chance were very common. Such discrepancies could be predicted by failure to meet important assumptions of case-only designs. Hence, researchers should be aware that there may be other causes of discrepancies or agreement between both types of designs than the presence or absence of time-invariant unmeasured confounding.

In this final chapter we place the main findings in context and provide future perspectives.

**Selection strategies for measured confounders**

Confounding is mixing or confusing the effect of the exposure of interest with the effects of other variables leading to bias [1]. Understanding and adjusting for confounding in observational intervention research is central to address causality when an association is observed [2]. To be considered a potential confounder, a variable must have an association
with the exposure and a causal relationship with the outcome (or be a surrogate measure of a cause) in the study population. In addition, the variable should not be affected by the exposure [3]. These are traditional requirements for variables to be considered potential confounders, however, there is no general consensus on how to identify and select confounders in research practice.

Some form of confounding selection procedure is required, especially when data are sparse [4]. Further, in any case, inclusion of non-confounders can introduce bias and reduce precision [5-7]. Traditionally, confounders are selected based on different quantitative strategies, including change in effect estimate criterion, univariate associations of the variable with the exposure and the outcome, or stepwise regression modelling [3,8]. More recently, other data-driven selection methods such as high-dimensional propensity score analysis or other exposure models and regularization methods have been proposed as potential alternative methods for confounder selection [4].

A problem with all these purely data-driven variable selection methods is the risk that bias and/or unnecessary large variance is introduced by selecting intermediates, colliders or instrumental variables [9,10-12]. The inclusion of colliders and instrumental variables is especially problematic if it introduces or amplifies uncontrolled unmeasured confounding [9]. A frequently proposed solution is to explicitly define the causal structure and relationships between variables based on subject knowledge [12-14]. A commonly proposed strategy is to summarize and communicate the assumed causal structure using causal diagrams, or directed acyclic graphs (DAG) [15,16]. A causal diagram for a confounder U of the relationship between exposure X and outcome Y would look like Figure 9.1.

![Figure 9.1](image)

**Figure 9.1.** Causal diagram of the relationship between exposure X and outcome Y and confounder U.

Such causal diagrams, often much more complicated, are sometimes used to summarize and communicate the causal structure assumed by the researchers and can be used to identify the minimal sufficient adjustment sets for estimating the total effect of X on Y [15,17]. In theory, causal diagrams are essential for valid selection of confounders for adjustment. Such diagrams can help to identify and remove instrumental variables,
intermediates and any other variable influenced by exposure or outcome, leaving only
the confounders for which one needs to adjust to obtain unconfounded effect estimates
[15]. However, the usefulness of such diagrams has been criticized by researchers because
often it is unknown how all variables interact with each other and causal diagrams quickly
become too complicated to understand when one believes that hundreds of variables can
act as (proxies for) confounders [18]. Hence, causal diagrams are useful for summarizing
and communicating simple causal structures with relatively few variables, but seem to
be less useful when sufficient prior knowledge is lacking as is often the case when many
variables might be involved. In these latter situation one may, dependent on whether
there are concerns about for example bias amplification, want to use confounder selection
methods that are less vulnerable to bias amplification, such as regularization regression.
Regardless of the chosen method, it is important to report which method is used to enable the
reader to identify potential problems that the selection method may have introduced. Hence,
it is worrying that both the reasons why potential confounders are selected for analysis
and the rationale for their inclusion in the final model are reported infrequently as
observed in the study described in Chapter 2.

Controlling unmeasured confounding

In the paragraph above, we evaluated methods that can be very useful when there is a
large number of measured covariates that may act as (proxy for) confounders. However,
often there are several potential confounders that are not measured or captured in, for
example, an administrative health care database. Moreover, even with high-dimensional
propensity scores and shrinkage methods using hundreds of variables, unmeasured
confounders may be uncorrelated with measured covariates. For example, smoking
history is often absent in healthcare claims databases and is not likely to be proxied well
by information available from claims [19]. Self-controlled or case-only designs have lately
been developed to overcome the problem of unmeasured confounding in observational
intervention studies. The general idea behind these designs is that patients can serve as
their own controls, thereby minimizing confounding by factors that are stable over time.
Consequently, these designs are especially useful when applied to evaluate short-term
effects, as the likelihood that most factors will be (approximately) stable over time is in
these situations generally relatively high. Several (variations on) self-controlled designs
have been developed over the years, including case-crossover [20], case-time-control
[21], fixed-effects case-time-control [22], case-case-time-control [23,24], prescription
sequence symmetry [25], exposure-crossover [26], and self-controlled case-series designs
[27]. In this thesis we focused on the prescription sequence symmetry, case-crossover,
case-sibling and self-controlled case-series design.
Prescription sequence symmetry analysis

The prescription sequence symmetry analysis (PSSA) was first proposed by Hallas as a screening tool for adverse drug reactions [25]. This design is often regarded as a simplified version of a case-crossover [20] or self-controlled case-series design [27]. In contrast to those designs, the PSSA does not base the effect estimate on comparisons within patients, but within patient populations (i.e. between different patients using the same drug). Only patients that both use the exposure drug of interest and the outcome drug of interest within a certain period of each other are selected for analysis. Hence, by selecting only patients that initiate the drugs of interest, all included patients have an indication for the exposure drug, thereby limiting confounding by indication.

The PSSA method has several limitations and requires several assumptions. An obvious limitation of the PSSA is that there is a limit to the number of drugs that can serve as a valid proxy for acute clinical outcomes. Further, especially when using short time-windows, which reduces the probability that results are affected by several potential biases as explained below, it is crucial that the timing of drug prescribing is close to the timing of the actual clinical event to prevent misclassification. Hence we examined short-term adverse or beneficial effects which require immediate drug treatment.

In chapter 4 and 5 we used certain antibiotics as a proxy for acute urinary tract infections, which would likely not result in severe misclassification of the timing of the outcome as there will be generally limited time between the onset of symptoms and the prescription of antibiotics. Although it is theoretically possible that increased contact with the general practitioner after angiotensin-converting enzyme inhibitors initiation explains part of the increased risk, the absence of an effect for beta-blocker initiation (sequence ratio of 1.01) suggests that this is not the case. The problem of finding valid drug-proxies can be solved by applying sequence symmetry analysis with outcome events instead of outcome drugs [28,29].

In the PSSA, patients that use the outcome drug both before and after initiation of the exposure drug are simply categorized as patients that use the outcome drug before the exposure drug and information about outcome drug use after exposure drug initiation is ignored. In contrast, a self-controlled case-series design would also incorporate the information after exposure drug initiation of the same patient. Hence, as explained in Chapter 5, the PSSA will result in effect estimates that are lower (closer to the null when evaluating an increased risk, a stronger protective effect when evaluating a protective effect) in comparison with a self-controlled case-series or a case-crossover design that does not have the same limitation. Nevertheless, we found similar results when applying the prescription sequence symmetry analysis to study the effect of ACE-inhibitors on the occurrence of urinary tract infections as in a randomized design (Chapter 3 and 4).
In chapters 4 and 5, we restricted the analysis to patients that initiated the outcome drug within 28 days of exposure drug initiation in an attempt to limit possible confounding by factors that are related to the timing of prescribing (e.g. the closer to the date of exposure initiation the higher the disease severity). The use of relatively short time-windows may have resulted in a relatively small difference between the PSSA and case-crossover design (Chapter 5). When using longer time-windows, e.g. one year as commonly applied [30-34], the influence of ignoring information after exposure initiation in patients that initiate the outcome drug in the before-period may be much larger (due to a higher probability that the outcome also occurs in the after-period). Consequently, when screening for unknown drug-related adverse effects, one needs to take into account that the PSSA method may result in more false negatives than self-controlled case-series designs such as the hierarchical Bayesian self-controlled case-series model [35]. Although the hierarchical Bayesian self-controlled case-series model still has to be empirically evaluated, it is a method that seems to be an interesting screening tool as it was specifically designed for estimating the effects of many drugs on many outcomes [35].

As explained in Chapter 4 and 5, using sequence symmetry analysis to evaluate adverse effects that are already known by prescribers [28,31,36-38] may result in artificially increased risks. This latter phenomenon is also observed with self-controlled case-series when exposure prescribing is outcome dependent. With well-known adverse effects, physicians may postpone prescribing a drug that further increases the risk of a recently experienced event. For the self-controlled case-series, a solution often used in the setting of temporary contraindication for vaccination is to incorporate a ‘pre-vaccination risk’ period that is not used to calculate the baseline outcome incidence [39]. Although a ‘pre-exposure’ period could be excluded from the at risk period in PSSA, one should be aware that this may introduce bias due to confounding by disease severity and potential time-trends by moving the before period further away from exposure initiation than the after period. Therefore, careful consideration of study periods is essential to limit up or downward bias.

Although confounding by indication is limited by including only patients that receive the exposure drug of interest, confounding by factors that are related to both the outcome and the timing of prescribing of the exposure is still possible. When there are concerns that such time-varying confounding may play a role (e.g. disease progression) it is important to use relatively narrow time-windows with all self-controlled methods. This is especially relevant for the PSSA design, as no adjustment for such confounders can be made via regression techniques as in real self-controlled methods (Chapter 5). Therefore, in our own studies we limited the time-windows to 28 days for all prescription sequence symmetry analyses, thereby minimizing confounding by factors.
that are (nearly) stable around the moment of exposure initiation (Chapters 4, 5 and 6). Another reason why the use of relatively narrow time-windows can result in less biased effect estimates is that trends in prescribing of exposure drug and/or outcome drug can influence the results [25,40]. The effects of potential time-trends can be limited by using narrow time-windows (Chapters 4, 5 and 6), because the prescribing practice will generally not change drastically over very short time periods. Given the vulnerability to several biases when using relatively long time-windows, such as the commonly applied 1-year time-windows, an important recommendation would be to restrict the analysis to patients that initiate both drugs of interest within a relatively short-time period. This recommendation is also more in agreement with theory and recommendations of self-controlled designs such as the case-crossover study [20,40].

Although the prescription sequence symmetry design has, as other observational designs, some important limitations as discussed in chapters 4-6 and above, it also has important strengths. Due to the simplicity of the design it can be used to screen Big Data databases without encountering scaling issues. Moreover, when limited data on potential confounders are available, applying a prescription sequence symmetry design can have more value than performing a cohort study and neglecting potential unmeasured confounding. In Chapter 6, the PSSA indicated that at least part of the association of concomitant use of non-steroidal anti-inflammatory drugs (NSAIDs) and selective serotonin reuptake inhibitors (SSRIs) with peptic ulcer drug treatment may be due to unmeasured confounding in a previously published cohort study. Nevertheless, given the ever increasing computer power and recent developments in the self-controlled case-series method [35,41-48] that address much of the limitations encountered by the PSSA and/or standard self-controlled case-series, future research may additionally be focused on the self-controlled case-series methodology.

**Case-crossover design**

The case-crossover design was developed by Maclure to study the effect of transient effect on the risk of acute outcomes [20]. With this self-controlled design, the probability of exposure in the period just before the outcome event (hazard period) is compared with the probability of exposure in control period(s). Because patients (cases) serve as their own control, confounding by time-invariant characteristics, including unmeasured characteristics, are eliminated.

We evaluated the case-crossover design in both a single-study comparison (Chapter 5) and a systematic comparison (Chapter 8). The effect of angiotensin-converting enzyme inhibitors on urinary tract infections was similar using a case-crossover design as in a randomized setup (Chapter 3 and 5), suggesting that, with similar study populations,
exposure and outcome definitions, the case-crossover design appeared to pertain similar results as randomized controlled trials. Inevitably, there are situations in which a case-crossover design will less likely provide valid effect estimates than other designs. In chapter 8, we found that differences between case-only designs, including the case-crossover design, and case-control and/or cohort studies could be predicted by failure to meet important assumptions [49] of the case-only design. Although this finding does not necessarily mean that the case-only study is wrong and the cohort/case-control study right, it implies that researchers should at least be aware that discrepancies between both types of designs may thus be caused by wrong application of the case-only or case-crossover design. Predictors of discrepancies included evaluation of chronic exposures, common events, insidious events and the use of relatively long hazard periods. In chapter 8 we did not include all assumptions of the case-crossover as potential predictors, as for some it was difficult to assess whether the assumption was met without having the original data at hand. Here, we will discuss more in depth some important remaining limitations of the case-crossover design that should be considered when considering whether a case-crossover design would be an appropriate study design for a specific causal research question.

Because the case-crossover design is a unidirectional design, with the control period always preceding the hazard period, the design is vulnerable to time-trends in exposure [20,21,50-52]. Several solutions exist to this potential problem, including an effective and simple solution: using short hazard and control periods without long periods between the hazard and control periods similar to the PSSA design. However, sometimes this is not possible due to possible carry-over effects or when it takes a while before the effect of the drug is noticeable [50]. Carry-over effects occur when the exposure in the control periods is not independent of the outcome event [49]. The PSSA design is not vulnerable to carry-over effects as within population instead of within patients comparisons are made. To solve the problem of potential time-trends in exposure different options have been proposed: the case-time-control design that uses crossover in control patients to assess and adjust for potential time-trends in exposure [21]; the case-case-time-control that uses future cases as present controls to adjust for exposure time-trend bias [23]; and a bidirectional case-crossover design that samples control windows before and/or after the outcome event [53,54]. With this latter method control periods should not be selected as a function of event times in order to get unbiased results using standard conditional logistic regression [49,53-56]. Hence, symmetrically sampling of control periods before and after the hazard period within individuals should preferably be avoided, although the related bias tends to be small [56]. When there is a linear trend in exposure a bidirectional case-crossover design will remove bias due to the time trend. An important assumption of this method is that the outcome does not alter the probability
of future exposure. When this assumption does not hold, alternative strategies need to be adopted to address time-trends in exposure.

The case-time-control design would be a valid alternative in such situations [21]. However, since this method uses crossover in controls to adjust for time-trends, bias may be introduced when the exposure trend is different among controls than among cases [57]. If controls would be different from cases with regard to exposure trends, one could use future cases as present controls using the case-case-time-control design [23]. A problem with this latter approach is that on the one hand future cases should not be too distant in time to eliminate to capture non-linear time-trends, while on the other hand there must be sufficient time between the event and the person-time sampled from the future case to ensure that the exposure is independent of the future event [23]. Moreover, the method may be inefficient as only patients that can be matched to a future case will contribute to the analysis. As those future cases need to be identified, follow-up must be longer than with a conventional case-time-control or case-crossover design. Finally, selection bias may be introduced, because individuals cannot become future cases unless they survive until they experience the outcome event [40].

Although the case-crossover design eliminates confounding by factors that are stable over time [20], the design is still vulnerable to confounding by time-varying factors. When factors that vary between the hazard and control period(s) are measured, a multivariate conditional logistic model can be used to adjust for these time-varying factors as we did in chapter 5. Although a recent review suggested that this is the only correct multivariate model for case-crossover designs [58], a stratified Cox’s proportional hazards model with strata representing the matched sets is equally valid [59]. However, one should mind that applying conditional logistic regression or a stratified Cox’s proportional hazards model to a case-crossover design with multiple control periods can induce bias if exposures at different time points within individuals are not statistically independent [60].

When a selection procedure for inclusion of potential confounders is necessary, earlier mentioned traditional methods (e.g. change in effect estimate criterion) or regularization methods such as Lasso can be used [61,62].

When applied to the right research question, the case-crossover design can be a very useful research design and provide less biased effect estimates than alternative designs, though empirical comparisons with golden standard RCTs are virtually lacking. Especially when examining transient effects of intermittent exposures on acute outcomes in presence of several important time-invariant confounders, the case-crossover design is a very valuable study design. In chapter 5 we found similar effect estimates using a case-crossover design as in a randomized design, despite evaluating the effect of angiotensin-converting
enzyme inhibitors, which are often used more chronically. This suggests that even if some assumptions are not entirely met, still similar results can be obtained as in randomized settings. However, empirical studies and statistical simulations are needed to evaluate how vulnerable the design is to violation of the different assumptions. Such simulations would be especially interesting given the observation that discrepancies between case-only and cohort/case-control studies in empirical studies could be predicted by failure to meet assumptions of the case-only designs (Chapter 8).

**Self-controlled case-series**

The self-controlled case-series was developed by Farrington to evaluate the influence of transient exposures on acute outcomes [27]. However, the method has also been applied to evaluate long-term effects [63] and recently an exposure-adjusted self-controlled case-series method was proposed that was developed to evaluate the long-term effect of accumulated exposure [48]. However, with long follow-up the potential for bias due to time-varying confounding may become large. When large amounts of potential time-varying confounders are available, self-controlled case-series designs can be implemented with regularization methods [48,64].

The original self-controlled case-series design is similar to a bidirectional case-crossover design in the sense that person-time is sampled from both before as after the index date. In addition, in both designs patients serve as their own control, thereby minimizing confounding by factors that are (nearly) stable over time. Moreover, both methods have similar assumptions and limitations. The full-stratum [53] and time-stratified [65] bidirectional case-crossover design are basically versions of the self-controlled case-series designs [66,67]. The most notable difference is that with the bidirectional case-crossover design the index date is the outcome event [53], while the self-controlled case-series is exposure-indexed [27,66].

A necessary condition of the case-crossover design is that the exposure distribution is stationary, i.e. there is no time-trend bias [60]. It would be intuitive to assume that a necessary condition of the self-controlled case-series would be that the outcome distribution is stationary. However, variations in the baseline incidence can be allowed for in the model by incorporating age or time effects [66].

The standard self-controlled case-series has four main assumptions. The most important assumptions, similar to the PSSA method, is that the probability of exposure is not affected by the occurrence of an outcome event [66]. Obviously, this is also an important assumption of the bidirectional case-crossover design [53]. One solution to overcome the problem of outcome dependent exposure is redefining the observation period
as starting from exposure and continuing to the end of the observation period. This approach is only possible when the post-exposure risk period is not indefinite and is not valid for multiple exposure, because bias for only one exposure but not for subsequent exposures is corrected [45,66,68]. Furthermore, selection bias may be introduced by excluding cases for which the event occurred before exposure. In addition, this will result in a loss of power and less ability to adjust for time-varying confounders as unvaccinated cases will also be dropped [45]. As mentioned earlier, with temporary dependencies, a solution often used in the setting of vaccinations is to incorporate a ‘pre-vaccination risk’ period that is not used to calculate the baseline outcome incidence [39]. When the outcome-dependency is long, another solution exist that can also be used to analyse multiple exposures. In this pseudo-likelihood approach the observation period is remained the same as in standard self-controlled case-series, but only exposures experienced prior to the event are used to estimate the relative incidence [42]. The data are subsequently analysed as if there could be no subsequent exposures [42]. When in fact such exposures do occur, those exposure periods are incorporated into the baseline period and event counts during this period are adjusted to the number of events that would have been observed when there was no exposure in this period using Horvitz-Thompson-like estimators [42]. This method can be applied to transient exposures and rare non-recurrent events. The method is only valid when the risk returns to the baseline level at the end of the risk period [42].

Second, because the statistical model is derived from a Poisson cohort model [66,67], analysis of frequent non-recurrent or unique events is not valid. Rare non-recurrent events can be analysed using the self-controlled case-series, because the probability of observing more than one event per individual (given that at least one is observed) tends to be zero [27,66,67]. In addition, clustered recurrent events should also not be analysed using the self-controlled case-series design. However, if events clusters in episodes, but the episodes can be assumed independent, clustered events within a certain time-period can be grouped. When this is not appropriate only first events can be used, provided that the initiating event is rare [66]. More recently, two variations of the self-controlled case-series were developed to allow for dependence of events [43,47].

Third, only time-invariant confounders that act multiplicatively on the baseline incidence are cancelled out [67]. This assumption must also hold for modelled time-varying exposures of interest. For the exposure and measured covariates it can be tested whether these assumptions are not violated using methods similar to those used for the proportional hazards model [67].

Fourth, the observation periods should be independent of event times [67]. Thus, individuals must generally remain observable after an event occurs. This may be
problematic when the outcome of interest is death or strongly related to death [67,69]. When the exposure of interest does not appreciably affect the risk of death besides through the effect on the outcome, the method of Farrington et al. can be applied [69,70]. When death is the outcome of interest, similar methods as described above that address event-dependent exposure can be used [42,66]. Nevertheless, in practice, the self-controlled case-series method is generally robust to failure to meet the assumption that observation periods are independent of event times [67].

In chapter 8, the majority of comparison between case-only and cohort or case-control designs consisted of comparisons between self-controlled case-series and cohort designs. This was largely due to the high number of such comparisons in the study of Madigan et al. [71]. Although the authors also reported comparisons with the case-crossover design in the supplementary files of that paper, we did not include those comparisons as we detected an error in their data and results. Though we communicated with the authors of this study and they were going to look further into this issue, at the time of writing this issue has not been solved yet. Nevertheless that study may have contributed to the relatively high number of discrepancies in chapter 8. For example, several associations were included that potentially suffered from bias due to outcomes that alter the exposure probability, without using a modified version of the self-controlled case-series that could address this limitation [71]. Similarly, a study that concluded that within-person study designs, including the self-controlled case-series, may have greater susceptibility to bias, applied the self-controlled case-series in an incorrect manner [72]. They censored exposure at the outcome which is not valid, when the self-controlled case-series would have been applied in a correct manner it is likely that different conclusion would have been reached [73]. Hence, whenever applying the self-controlled case-series design to a research question, one needs to carefully consider whether the assumptions of the design will be met. When the assumptions are met, the self-controlled design can be a very valuable design, especially since all measured and unmeasured multiplicative time-invariant confounders are adjusted for by design.

Case-sibling design
Although sibling discordance studies were already applied in the 19th century, related designs such as the case-sibling design applied in chapter 7 are increasingly being used in epidemiology [74]. In the context of prenatal exposures, the advantage of the case-sibling design is that both the case and the control sibling share the same mother and household and share genetic and time-invariant environmental factors that may differ when using a conventional case-control or cohort study [74]. Although the design is not self-controlled in the sense that children are compared with themselves, a comparison is made between pregnancies of the same mother. A major advantage over twin studies
is that twins always have the same prenatal exposure, while siblings can be differentially exposed during pregnancy. Although a contrast is often made between a case-sibling and a cohort or case-control design to gain an indication about unmeasured confounding, chapter 7 shows that this is not always as easy as it seems. As described by Frisell et al., agreement between the case-sibling design and cohort or case-control study does not necessarily imply absence of unmeasured time-invariant confounding and disagreement does not necessarily imply presence of unmeasured time-invariant confounding [75]. Confounding by factors that are not perfectly shared by siblings or measurement error of exposure can result in biased estimates and ultimately wrong conclusions [75]. Estimates of within-pair estimates are more susceptible to bias due to non-shared confounder than estimates from unpaired analyses [75]. As only case-sibling pairs that are differently exposed contribute to the analysis, a selection for pairs that differ on non-shared causes of the exposure is made. As non-differential (random) measurement error is not shared by siblings, it is likely that among the discordant pairs more individuals will be misclassified on exposure than in the general population [75]. Consequently, when no other biases than random measurement error play a role, the effect estimate will be closer to the null in a case-sibling design when compared with cases and controls from the general population. Indeed, the greater vulnerability to misclassification applies to all (self)-matched designs [76,77].

In chapter 7, we discussed another potential limitation of the case-sibling design when there is a natural ordering in the occurrence of cases. For example, firstborn children are at higher risk of developing asthma than their later born siblings [78] and it is therefore likely that the birth sequence is unevenly distributed between case and control siblings. This may result in bias when there is a trend in exposure. In chapter 7 we applied a method akin to the case-time-control design that adjusts for such potential time-trends. Although for this application the adjustment did not substantially alter the results, because there was no substantial time-trend in the use of antibiotics during pregnancy, it may remove substantial bias when there is a strong time-trend and a natural ordering in the occurrence of cases. Such ordering of cases may occur in various association studies using case-sibling design [79,80]. Although such studies often adjust for birth order, this may not completely remove potential time trends in exposure. An important limitation of using the trend in controls may be that the trend may be different among control pregnancies, especially if mothers of first-born cases decide to stop with a potential harmful exposure during subsequent pregnancies.

Although a case-sibling design on its own does not proof causality, partly due to the limitations described above, it can provide an important contribution to our understanding of associations. Combined with other designs that may provide more information about potential confounding, such as the maternal-paternal comparison
applied in chapter 7, the increasing use of the case-sibling is an important step forward in the field of pregnancy outcomes, and the evidence base should be enlarged in the coming years.

**Quantitative bias analysis**
The potential impact of unmeasured confounders, misclassification and/or selection bias can also be estimated using quantitative bias analysis [2,81]. Methods of bias analysis have been well known for decades and are endorsed for widespread use [81]. Nevertheless, we found in chapter 2 that quantitative bias analysis is rarely applied. When the exposure is not randomly assigned, as in all observational studies, the comparison between groups provides a probability that the outcome distribution is attributable to chance as opposed to the combined effects of exposure and systematic error [2]. Thus to infer causality, an educated guess about the impact of systematic errors is needed.

Between 2010 and 2012, 85% of published articles on case-control or cohort studies on interventions with a hypothesized beneficial effect added a qualitative comment about the likelihood of unmeasured confounding (chapter 2). However, studies on anchoring and adjustment heuristic suggest that people tend to make adjustments that are insufficient to capture the true impact of the bias [2]. Moreover, an understanding of overconfidence heuristic suggests that people will tend to be overconfident about results [2]. Consequently, qualitative comments on the likelihood and impact of bias are expected to be insufficient. Quantitative bias analysis, informed by internal validation studies or carefully selected external information, can potentially overcome this problem [2,81].

Arguments often posed against quantitative bias analysis are that it is difficult to assign valid values to bias parameters and that the methods are too complex to apply. However, frequentist approaches assume that all bias parameters are set to values that induce no systematic error. It is almost certain that these assigned values are not the true values (no unmeasured confounding, no selection bias and no measurement error) [81]. When different values may seem valid, one can evaluate the sensitivity of the results to variation in the assigned values by assigning a probability distribution or ranging the values using multidimensional bias analysis. The obtained information can also guide the direction of future research. Although Bayesian and probabilistic methods may indeed be complex to apply for the inexperienced researcher, simple bias analysis methods can be easily applied (chapter 7) [2].
In light of the increasing use of Big Data [82-84] leading to effect estimates with very narrow confidence intervals, quantitative bias analysis will become increasingly important to prevent policy actions based on overconfidence in the results. For example, based on conventional meta-analyses that found that residential electromagnetic fields were associated with an increased risk of childhood leukaemia, some stakeholders argued that it was necessary to undertake action and relocate of power lines [81]. However, using probabilistic bias analysis it was shown that there was much more uncertainty in the estimates than captured by the conventional confidence intervals [85]. The uncertainty intervals could easily include the null, indicating that costly policy action was not supported by the available data [85].

**Alternative methods to address unmeasured confounding**

Although the different self-controlled methods as applied in this thesis work have been proven to be useful to address unmeasured confounding, there are some situations in which the assumptions of the methods are likely violated, and other approaches are needed. Two developments are shortly mentioned below: p-value calibration and the instrumental variable design.

**P-value calibration**

Recently, a p-value calibration approach was proposed to take into account that conventional confidence intervals and p-values only take into account random error and ignore selection bias, measurement error and residual or unmeasured confounding [86]. With this method, p-values are calibrated using the observed null distribution for negative controls (drug-outcome pairs for which no causal relationship is assumed). A major problem with this method is that it only works well when negative control drugs really have no causal effect and when the influence of various sources of systematic error is similar for the negative controls as for the association of interest. This is always a potential problem when using negative or positive controls to gain an indication about potential unmeasured confounding. Since several negative controls are needed for calibration, p-value calibration may be difficult to apply in practice without failing to meet crucial assumptions of the method. Therefore, p-value calibration may suffer from a similar problem as conventional methods to calculate confidence intervals and p-values in observational studies: it is likely that the assumptions for the method do not hold. Therefore we would recommend to use quantitative bias analysis to estimate the susceptibility of the results to unmeasured confounding and other forms of systematic error whenever such analysis is productive [81].
**Instrumental variable design**

As mentioned under the heading ‘selection strategies for measured confounders’ (near)-instrumental variables can cause amplification of bias due to unmeasured confounding when wrongly selected as confounders in the final model. However, instrumental variables can in various situations be very helpful tools to address unmeasured confounding. Instrumental variable analysis can be applied when a variable can be identified that 1) is positively correlated with exposure (nonzero average causal effect), 2) is independent of unmeasured confounders conditional on covariates (random assignment), and 3) affects the outcomes only through its effect on exposure (exclusion restriction). The instrumental variable can be used to extract variation in exposure that is free of unmeasured confounding and subsequently use this variation in exposure to estimate the casual effect of the exposure [87]. However, in practice, it is often difficult to identify variables that fulfil all three requirements. The potential instrumental variable should have a relatively strong association with the exposure, because instrumental variable analysis with a weak instrument is sensitive to slight departures from being a valid instrument [87,88]. Hence, when there is some correlation between exposure and an unmeasured confounder, ordinary least squares regression may be less biased then an instrumental variables analysis with a weak instrument that has a weaker correlation with the unmeasured confounder [88]. However, even when a variable can be considered a valid instrument according to the three requirements listed above, two additional assumptions have to be met for valid analysis: the monotonicity assumption, although not necessary for all instrumental variable analyses [89], and the stable unit treatment value assumption (SUTVA). Suppose we have a binary instrumental variable Z and a binary exposure X. The monotonicity assumption requires that there are no subjects who are defiers, i.e. persons that are always exposed opposite of what would be expected on the level of the instrumental variable (no subject i with . Although this assumption will often not be violated, the model only estimates the average treatment effect for compliers . The effect estimate does not inform about the effect of the treatment among always takers or never takers , unless it can be realistically assumed that there is no heterogeneity in the treatment effects [87]. The SUTVA assumes that potential outcomes for each person are unrelated to the treatment status of other individuals (no interference). It is important to note that this assumption is also generally assumed for other methods than instrumental variable analysis and does not necessarily bias the estimation but changes the interpretation of the effect estimates [87]. When this is not taken into account one may misinterpret what is being estimated [90,91].

When all assumptions are met, instrumental variable analysis can be a useful tool to overcome bias due to unmeasured confounding. In practice it will be often difficult to identify a good instrumental variable candidate and to fulfil all assumptions. Hence, whenever performing an instrumental variable analysis, it is important to report about
the likelihood each assumption is met [92,93] and ideally perform sensitivity analysis to estimate the potential impact of violation of the assumptions.

**Conclusion**

In observational research, measuring and adjusting for confounding is crucial to evaluate whether an observed association is possibly causal or biased. Hence, confounding needs to be addressed in the design or analysis phase of each observational study that is intended to study causal associations. In this thesis, we mainly focused on different designs to adjust for measured and unmeasured potential confounding. Various self-controlled methods have been developed to partly overcome the problem of time-invariant unmeasured confounding. When applying such methods it is important to verify whether assumptions are likely met or whether another design might be more applicable to the research question. Quantitative bias analysis with bias parameters based on internal or external validation data can always be applied, regardless of the design and analytical choices. We would like to encourage researchers to further empirically test and modify existing self-controlled methods to partly overcome their main limitations and to develop novel techniques that further reduce or better quantify confounding in observational research.
References


212