12-13-2004

Referent Selection Strategies in Case-Crossover Analyses of Air Pollution Exposure Data: Implications for Bias

Holly Janes
University of Washington, hjanes@u.washington.edu

Lianne Sheppard
University of Washington, sheppard@u.washington.edu

Thomas Lumley
University of Washington, tlumley@u.washington.edu

1. Introduction

The case-crossover design belongs to a class of designs in air pollution epidemiology that is aimed at assessing the short term health effects of air pollution.[1] Referred to as acute effects studies, these designs enable estimation of the effect of day-to-day variation in pollution on morbidity and mortality. Various adverse health events have been studied, including deaths and myocardial infarctions. Ambient air pollution concentrations, measured at centrally located monitors in particular geographic regions, typically serve as the exposure of interest. Hence, the same exposure is used for all subjects in the study; we refer to this as a "shared" exposure series. Particulate matter (PM) has been the most commonly studied pollutant.

There are many challenges involved in acute effects studies, including error in the measurement of exposure and the discrepancy between ambient and personal exposures. Confounding is also a major concern. Time-independent confounding can occur, even with a shared exposure series, if subjects are observed at different points in time. Even more problematic are time-dependent confounders, such as season and concentrations of other pollutants. These factors are strongly associated with day-to-day variation in both exposure and many adverse health outcomes. Biases are especially troublesome because the effect of exposure is usually very small, and the effects of some time-dependent confounders many times larger.[2]

The case-crossover design represents a novel approach to controlling for confounding. Given a sample of subjects who experienced the event of interest, exposure just prior to the event, the "index" time, is compared with exposure at comparable control, or "referent" times.[3] The idea is

similar to a matched case-control study; the exposure at each index time is part of a matched set of exposures consisting of exposures for the subject at his/her referent times. This matched set is called a "referent window". By making within-subject comparisons, time-independent confounders are controlled by design. More importantly, if the referent times are matched to the index time with respect to time-dependent confounders (for example, if the referents are restricted to the same season as the index time), these effects are also controlled by design. This is in stark contrast to other approaches to the estimation of acute pollution effects (e.g. time series studies), in which time-dependent confounding is controlled by modeling.

The selection of referent times is a key issue in the case-crossover design. We refer to this choice as the "referent selection strategy" or "referent scheme". As stated above, the referent selection strategy is important in terms of controlling for time-dependent confounding. In addition, the case-crossover design makes the implicit assumption that there is no time trend in exposure within the referent window. Moreover, only with certain referent strategies are the estimating equations typically used, the conditional logistic regression estimating equations, unbiased. We call this bias in the conditional logistic regression estimating equations "overlap bias". (The use of the term "overlap" will be discussed in Section 5.) Bias in any estimating equation is problematic, since it implies that the associated parameter estimates are biased, even in large samples. In contrast, an unbiased estimating equation guarantees unbiased estimates in large samples. In this paper, we focus on bias in the estimating equations, rather than in the estimates themselves, since the latter have small sample bias.

A variety of referent schemes have been utilized in air pollution studies.[4-10] General [7;11-15] and air pollution specific [16-22] case-crossover methods papers have addressed referent selection, but so far none have presented a cohesive set of guidelines for evaluating and choosing a referent selection strategy. The purpose of this paper is to review referent selection practices in air pollution case-crossover studies, and to clarify key referent selection issues. In Section 2, we review the referent selection strategies that have been used in air pollution case-crossover studies. Section 3 provides an overview of the case-crossover method. We present the statistical model appropriate for air pollution exposures, the estimation method that is used, aspects of the design that are useful with air pollution exposures, and potential biases associated with the design. In Section 4, we describe the different types of referent selection strategies and discuss how each deals with the various types of bias. Section 5 illustrates the phenomenon of overlap bias using a numeric example, and describes the magnitude of the bias. In Section 6, we discuss the relative efficiencies of two referent selection strategies. Section 7 extends our review to situations in which the exposure is not shared (in contrast to all previous sections). Finally, we conclude with recommendations for choosing a referent selection strategy for a case-crossover analysis of air pollution exposure data.

2. Review of Referent Selection Strategies Used in Case-Crossover Studies of Air Pollution Exposures

We reviewed 19 case-crossover studies of air pollution exposure data, published between the introduction of the design in 1991 and June of 2004. We limit our attention to applied, rather than methodological papers (see Table 1). [4-6;8-10;23-35] All of the studies that we reviewed

examined ambient, shared air pollution exposures. Outcomes such as non-accidental mortality and hospitalization for myocardial infarction and asthma were studied.

By far, the most popular referent selection strategy is the symmetric bidirectional design,[16] in which referents are at a fixed lag before and after the index time. Sixty-three percent of the studies we reviewed used this design. A distant second in popularity is the time stratified design, used in 26% of studies. With this design, [22] time is stratified *a priori,* and all other days in the stratum in which the index day falls serve as referents. The restricted unidirectional design was also used. Unidirectional sampling usually means that referents are selected only prior to the index day, although one study sampled unidirectional referents prospectively.[5] Restricted unidirectional referents are at fixed lags relative to the index day. Finally, several studies reported the results of multiple analyses, each using a different referent selection strategy. This practice is not recommended, since it makes interpretation difficult and induces model selection bias.[36-38] While these referent schemes may appear to be quite similar, we will show that the distinctions between them are important, and have implications for bias.

3. The Case-Crossover Method

3.1 Appropriate Exposures and Outcomes

The case-crossover design is appropriate for assessing the association between a short-term exposure and the risk of an acute event. It is most suited to exposures that are transient and do not have carryover effects. In this review, we focus on the case of a rare event. In other words, we assume that, for each individual, events occur with a low probability.

3.2 Relevant Aspects of Air Pollution Exposure Data

In most air pollution case-crossover studies, the exposure is shared among all individuals in the study. Hence, when deriving the likelihood of the data, it is appropriate to condition on this fixed exposure series. (The likelihood can be thought of as the probability model for the data.) In addition, when evaluating the properties of the effect estimate, we should condition on the exposure. Because we have only one exposure series, we want to know that the estimate and the corresponding estimating equations are unbiased for this specific exposure series; knowing that they are unbiased when averaged over all possible exposure series is of little value. We note that most standard regression analyses, such as linear and logistic regression, also condition on exposure. With the exception of a review of case-crossover methods appropriate for unshared exposures in Section 7, we restrict our attention to analyses conditional on exposure.

Another important property of ambient air pollution exposure data is that it is exogenous. Exogenous exposures are generated independently of the individual under study. In contrast, endogenous exposures are influenced by the individual; they can only be observed while the individual is observed, and may change as a result of having experienced the event. Personal air pollution is an example of an endogenous exposure. Our review is limited to exogenous exposures.

3.3 Attributes of the Design Useful in the Air Pollution Context

The main strengths of the case-crossover design are that it does not require a control sample (and hence avoids bias associated with improper control selection), makes effect modification assessment relatively simple, and controls for fixed confounders by design. Also very appealing

is the fact that time-dependent confounders can be controlled by design, by matching referents to the index time with respect to these factors.

Another method commonly used to assess the association between short-term air pollution exposure and adverse health events is time series regression. The main difference between time series regression and the analysis of a case-crossover design using conditional logistic regression is that the former requires modeling the confounders. With a conditional logistic regression analysis, the confounding effects of all matching variables are controlled by design (through matching). Further control could be accomplished by also modeling these factors in the conditional logistic regression model.

3.4 The Time-Varying Exposure Model and Estimation Method

The statistical model first postulated for the case-crossover design was a precipitating event model.[3] This model assumes that time can be discretized into "exposed" and "unexposed" periods. It stipulates that a subject is at high risk for a fixed time following an exposed period, and thereafter returns to background risk, until the next exposed period. (Note that, under certain conditions, this model is subject to length bias, as noted by Varachan and Frangakis. [39]) However, the proportional hazards model for a rare disease with a constant baseline hazard for each individual, first proposed for case-crossover studies by Navidi [7] and Marshall and Jackson[12], is more appropriate for air pollution exposures, which are not binary. We call this the time-varying exposure model.

The time-varying exposure model states that there is only one event for each case (a legitimate assumption for a rare event), and specifies risk as a function of time and exposure. If past exposure lags are included, this model also allows the risk to increase due to past exposures. Under the time-varying exposure model, the hazard rate of person *i* at time *t* given time-varying covariates x_{it} is given by $\lambda_i(t; x_{it}) = \lambda_i \exp(x_{it}\beta)$. Over a short time period, the assumption of a constant baseline hazard is often reasonable, and is equivalent to assuming smooth seasonal effects in a time series analysis.[22] The parameter e^{β} can be interpreted as the change in the risk of an event associated with a short-term unit increase in exposure.

Conditional logistic regression is typically used to estimate β in the time-varying exposure model. The use of this method has been motivated by the analogy to matched case-control designs, where the conditional logistic regression likelihood is exactly the likelihood of the data. Its use makes sense, since the idea is to control for confounding by making comparisons within referent windows, and hence we want to condition on the referent windows in the analysis.

The Mantel-Haenszel estimator was used in some early case-crossover studies with binary exposures.[3] With only one referent for each case, (i.e., matched pairs), the two estimation procedures are identical.[40] However, in general, conditional logistic regression is a better choice, since it can be used with non-binary exposures, and makes control of additional confounders (those not used in the matching) easier. These factors can simply be included in the regression model.

3.5 Potential Biases Associated With the Choice of Referent Selection Strategy

The likelihood of the data for the case-crossover design depends on the choice of referent selection strategy. The conditional logistic regression estimating equations are only unbiased with certain referent selection strategies. For most of the commonly used referent selection strategies (e.g. symmetric bidirectional referents), the conditional logistic regression estimating equations are *not* unbiased [22;41]; this is called overlap bias. In Section 4, we identify the referent schemes that are subject to overlap bias, and in Section 5, we describe overlap bias in more detail.

Greenland [11] and Navidi [7] showed that the case-crossover design relies on the assumption that there is no trend in exposure within the referent window. This assumption is required since the effect of exposure is estimated by contrasting exposures at the index and referent times. If, for example, referents are always prior to the index time, and there is a decreasing trend in exposure over time, the effect estimate will be negatively biased. With the strong long-term time trends often present in air pollution data, bias due to time trend is a concern. We discuss a method for controlling for this bias in Section 4.

Finally, the case-crossover design assumes that the referent exposures are representative of the usual distribution of exposure, and that the index exposure represents the exposure that generated the event.

4. Referent Selection Strategies

4.1 Classes of Referent Selection Strategies

As mentioned, a number of fundamentally different types of referent selection strategies have been used in air pollution studies, including the unidirectional, full stratum bidirectional, symmetric bidirectional, time stratified, and semi-symmetric bidirectional designs, listed in Table 2. We have proposed a taxonomy of referent selection strategies, with groups that correspond to the statistical properties of these designs.[41] We classify designs as localizable or non-localizable, and, within the localizable designs, ignorable or non-ignorable. Localizable designs are those for which the likelihood of the index times conditional on the referent windows contains information about β. In contrast, with a non-localizable design, the conditional likelihood is uninformative for β. An example of a non-localizable design is the symmetric bidirectional design, in which the index time is fixed in the center of the referent window, and hence the location of the index time within the referent window yields no information about β. Localizability is desirable since, when estimation can be based on making comparisons within the referent windows, which are presumably matched on time-dependent confounders, these confounders are controlled. Of the designs in Table 2, only the time stratified, full stratum bidirectional, and semi-symmetric bidirectional designs are localizable.

The localizable designs are classified as either ignorable or non-ignorable. Ignorable designs are those which have a referent sampling scheme that can be ignored when analyzing the data; conditional logistic regression can be used to obtain unbiased estimates. With a non-ignorable design, the likelihood of the data depends on the referent sampling scheme, and this likelihood must be used for an unbiased analysis. We note that this definition of ignorability is the same as that proposed by Little and Rubin, and used in the missing data context.[42] In either case, ignorability implies that the data can be analyzed as if the observed data were the complete data,

without having to account for how the data were sampled. Of the referent strategies listed in Table 2, only the time stratified and full stratum bidirectional designs are localizable and ignorable.

These referent class distinctions are relevant when choosing a statistical analysis. Conditional logistic regression yields unbiased estimates for localizable, ignorable designs. However, the conditional logistic regression estimates have overlap bias under non-localizable or localizable, non-ignorable referent selection. [41] With a non-localizable design, the likelihood of the data must be used to obtain unbiased effect estimates, but the use of this likelihood in applications is impractical. [41] With a localizable, non-ignorable referent scheme, again, the likelihood of the data must be used for unbiased estimation, but in this case, there is a simple way to obtain these estimates (see Section 4.6).

4.2 Unidirectional Referent Sampling

With a restricted unidirectional design, confounding due to season and day of the week are controlled by selecting referents close to, and on the same day of the week as the index day (see Figure 1A). Yet, this design is non-localizable, and thus estimates from conditional logistic regression have overlap bias.[41]

Unidirectional sampling has a major disadvantage in air pollution studies. Selecting referents only prior to the index time can lead to time trend bias. The bias will be larger the further the referents are from the index time. For this reason, unidirectional sampling is not commonly done in air pollution studies.

4.3 Full stratum Bidirectional Referent Selection

Greenland [11] was the first to recognize the pitfall with unidirectional referent selection in the presence of an exposure trend. Navidi [7] proposed that the time trend bias could be eliminated by choosing referents both before and after the index time, a strategy called *bidirectional* referent selection (sometimes called *ambidirectional* selection[21]). Technically, bidirectional sampling is only valid when cases are still at risk after an event, an assumption that is certainly not valid when the event is death. Navidi justified bidirectional sampling by noting that, with air pollution data, the exposure is exogenous and is available for all cases both before and after the event. A more rigorous justification was given by Lumley and Levy [22] who showed that, with a rare event, the bias due to sampling referents after the at-risk period is small. More importantly, the bias associated with sampling referents after the time at risk is smaller than the bias that would be incurred with unidirectional referent selection in the presence of a time trend.

Navidi proposed full stratum bidirectional referent selection (Figure 1B), in which the referents are all days in the exposure series other than the index day. [7] Interestingly, Lumley and Levy showed that a case-crossover analysis with full stratum bidirectional referents and a shared exposure series is equivalent to a Poisson regression analysis. [22] However, while bias due to time trend is controlled, time-dependent confounding (e.g. season) is not. Time-dependent confounding must be controlled by modeling, since the referent window is so large that confounding is not controlled by design. The full stratum bidirectional design is a localizable, ignorable design, so overlap bias is not a problem.[41]

4.4 Symmetric Bidirectional Referent Selection

The symmetric bidirectional design [16] is a popular alternative to the full stratum bidirectional design (Figure 1C). This design controls for bias due to time trend and confounding by both season and day of the week if referents are within the same season and on the same day of the week as the index time.[16] Simulation studies have shown that shorter lags ensure less confounding bias, and that confounding is not as well controlled if the seasonal pattern of exposure is not symmetric.[16;20]

The fundamental problem with the symmetric bidirectional design is that it is non-localizable. Hence, conditional logistic regression estimates are subject to overlap bias.[41]

4.5 Time Stratified Referent Selection

The time stratified design is not subject to bias due to time trend since there is no pattern in the placement of referents relative to the index time. In addition, the design controls for time-dependent confounding by matching. For instance, restricting referents to the same day of the week, month, and year as the index day controls for season and day of the week. The time stratified design is a localizable, ignorable design;[41] hence, conditional logistic regression can be used to obtain unbiased effect estimates.

The time stratified design has some interesting relationships with other designs. The full stratum bidirectional design is a special case of a time stratified design in which there is one large stratum (although in this case confounding must be controlled by modeling).[22] In addition, a conditional logistic regression analysis of a shared exposure series with time stratified by year,

month, and day of the week is the same as a Poisson regression analysis with dummy variables to adjust for day of the week within each month, and month within each year.[21]

4.6 Semi-Symmetric Bidirectional Referent Selection

With the semi-symmetric bidirectional design,[14] one referent is randomly chosen from days at a fixed lag pre- and post-event; if only one of these days is available (due to the case being at either end of the exposure series), it serves as the referent. If the lag is small and a multiple of seven, confounding by season and day of the week can be controlled by design. There is no bias due to time trend, since referents are bidirectionally sampled.

The semi-symmetric bidirectional design is a localizable, non-ignorable design. The likelihood of the data must be used in order to obtain an unbiased effect estimate, and standard conditional logistic regression estimates have overlap bias. It turns out, however, that in this case estimates based on this likelihood can be obtained using standard conditional logistic regression software when an offset is used (that takes a value of log2 for cases with only one possible referent, and zero otherwise). [41]

5. Overlap Bias

The term "overlap bias" was first used by Lumley and Levy,[22] who observed that this bias is similar to the bias due to using friend controls in matched case-control studies.[43;44] Yet, the term overlap is somewhat misleading in that it suggests that a design in which an individual's set of possible referent windows is overlapping is subject to overlap bias, and a design with disjoint windows is not. This is not the case: in both the full stratum bidirectional and symmetric

bidirectional designs, the referent windows overlap. Yet, the full stratum bidirectional design is free from overlap bias, and the symmetric bidirectional design is not. Alternatively, we considered identifying designs subject to overlap bias according to whether or not the index time is fixed within the referent window. While all of the existing referent strategies for which the index time is fixed within the referent window (e.g. the symmetric bidirectional design) are subject to overlap bias, and those with random index times within the referent window (e.g. the time stratified design) are not, it would be possible to configure a referent strategy with a random index time which is subject to overlap bias. Hence, we have not found a satisfactory heuristic explanation of overlap bias. The bias is a purely mathematical phenomenon. Whether or not bias exists for a particular referent strategy depends on the form of the likelihood under that strategy. In this sense, overlap bias is similar to the well-known bias associated with using unconditional logistic regression in a case-control study with finely matched data.[40] We seek to illustrate, with the following numerical example, the mathematical basis of overlap bias, and its non-intuitive nature.

5.1 Numerical Example

In this example, we calculate the overlap bias, as a function of the exposure series, for two different referent selection strategies. We will demonstrate that, for a fixed referent strategy, determining what types of exposure series are prone to overlap bias is virtually impossible. In addition, for a given exposure series, there is no intuitive explanation as to why certain referent strategies induce overlap bias, and others do not. We consider the simple case of a shared binary exposure series of length 10. We contrast the symmetric bidirectional referent strategy[16] (with referents one day before and after the index day), a design which is subject to overlap bias, with

the time stratified referent strategy (with strata of length 2 and 3), which does not have overlap bias.[22;41]

We first calculate the overlap bias under symmetric bidirectional referent selection, as a function of the exposure series. Let T be the length of the series, and K the number of days with "positive" exposure. The binary exposure series, z, can be reduced to a set of six parameters which refer to the number of instances of particular arrangements of exposures. Let $z_{010}, z_{001}, z_{011}, z_{101}, z_{01},$ and z_{10} denote to the number of instances of the following exposure arrangements: `010`; `001` or `100`; `011` or `110`; `101`; `01` at the beginning of the series or `10` at the end; and `10` at the beginning of the series or `01` at the end. We show in Appendix A that the overlap bias can be expressed as X / Y where

$$X = \frac{e^{\beta}}{(T-K)+Ke^{\beta}}\left(\frac{2z_{010}-z_{001}}{2+e^{\beta}}+\frac{z_{011}-2z_{101}}{1+2e^{\beta}}+\frac{z_{10}-z_{01}}{1+e^{\beta}}\right) \quad (1)$$

$$Y = \frac{e^{\beta}}{(T-K)+Ke^{\beta}}\left(\frac{2e^{\beta}z_{010}+2z_{001}}{(2+e^{\beta})^2}+\frac{2e^{\beta}z_{011}+2z_{101}}{(1+2e^{\beta})^2}+\frac{e^{\beta}z_{10}+z_{01}}{(1+e^{\beta})^2}\right). \quad (2)$$

(Here, X is the expected value of the conditional logistic regression estimating equations.) Hence, overlap bias occurs whenever X is nonzero.

This expression for overlap bias is a complex function of the parameters of the exposure series. Its form reveals that there is no simple intuitive way to characterize the exposure series which have overlap bias. Table 3 gives four different length-10 binary exposure series, along with their parameters and the value of X at various values of β. Figure 2 shows the overlap bias associated with each of these series (X / Y). We see that there is a large amount of variation in the amount of bias across the exposure series. For some series, the bias is substantial, while for others, there

is virtually none. For example, although Series D differs on just one day from Series C, these two series have dramatically different amounts of overlap bias. Series D has essentially no bias, while for Series C, there is a large amount of bias for all β. We also observe that, in several instances, the bias is larger than β itself. Finally, the bias can exist for very small β, typical in air pollution exposure studies.

We show in Appendix A a similar expression for overlap bias for the time stratified design. Algebraically, terms cancel, and the expression reduces to zero for all β. Hence, there is never overlap bias with the time stratified design. Once again, this exercise does not reveal any intuition as to why the time stratified design is free from bias and the symmetric bidirectional design is not. This is merely a consequence of the differences in the likelihoods under the two referent strategies.

5.2 The Magnitude of the Overlap Bias

Given an exposure series and a referent selection strategy, the magnitude of the overlap bias can be calculated, as was done in Section 5.1 (see also [41]). The bias is generally small, but it is highly unpredictable. Simulation studies examining continuous exposures have shown that it depends on the particular exposure series. [41] Bias can exist even for small β, which is particularly worrying for air pollution studies. In addition, for a given exposure series, there may be bias with some referent strategies and not with others. Moreover, there is no existing method for predicting in advance the magnitude or direction of the overlap bias, thereby making it impossible to know if the effect estimate is being dampened or magnified. Therefore, it is prudent to choose a referent strategy that avoids this bias entirely.

Overlap bias can exist even when $\beta = 0$, thus making it possible to erroneously detect effects that do not in fact exist. However, when there is no exposure effect, the nature of the overlap bias is somewhat different. When $\beta = 0$, the bias occurs only if cases at the ends of the exposure series have a different referent strategy than other cases. For example, with the symmetric bidirectional design, cases at the beginning of the series will not have pre-event referents, and cases at the end will not have post-event referents. Bateson and Schwartz[17] called this "selection bias", and suggested subtracting it off. While this correction is sufficient when $\beta = 0$, for other β, this correction does not in fact subtract off all of the overlap bias. Another unique aspect of overlap bias when $\beta = 0$ is that it decreases rapidly with the length of the exposure series.[22] This is not true for other values of β. In general, the bias will decrease as the length of the exposure series increases, but the speed of this convergence is much slower for $\beta \neq 0$.

6. Referent Selection: Efficiency Concerns

As is the case with any design, there is a clear bias-efficiency tradeoff in the case-crossover design. (Efficiency is usually quantified using the variance of the effect estimate.) Increasing the number of referents will lead to gains in efficiency, but decreased control over confounders. Yet, in the air pollution setting, bias is generally the dominant concern, due to the small effect sizes. For this reason, we study efficiency alone, assuming confounding has already been controlled by matching the referents to the index times. We note, however, that the necessity of controlling for confounding may leave little choice as to the number of referents. For example, with the time stratified design, the desire to match with respect to year, season, and day of the week (or another confounder) restricts the stratum size to four or five.

Mittleman et al. [13]and Bateson and Schwartz [16] investigated the statistical efficiency of a variety of referent strategies. However, both sets of investigators assumed the conditional logistic regression model to be true when calculating the variances. This model is only valid for localizable, ignorable designs, and not for the non-localizable designs studied by these authors. This is another manifestation of overlap bias: for non-localizable designs, not only are the conditional logistic regression estimating equations biased, but the conditional logistic regression variances are as well. Hence, the conclusions of these authors are not necessarily correct.

Here, we compare the efficiency of the time stratified and full stratum bidirectional designs, two localizable, ignorable designs, as a function of the stratum size in the time stratified design. The variances for the two designs for a given exposure series are given in Appendix B. After controlling for confounding by matching, the exposure will have no seasonality or time trend within-stratum. Hence, we simulated exposures to mimic Seattle PM_{10} data, without seasonality, day-of-week effects, or long-term time trend. The lognormal exposure series are 100 days long, have serial correlation on adjacent days ($\rho = 0.6$), a mean of 3.6, and a variance of 0.2.

The relative efficiency (the ratio of the variances) of the two designs depends on the exposure series, the stratum size in the time stratified design, denoted by M, and β. We show in Figure 3A the relative efficiency for three different exposure series as a function of M, when $\beta = 0$. There is variation in relative efficiency across the exposure series, but for these three exposures, we find relative efficiency to be approximately 70% when $M = 10$, 40-50% when $M = 5$, and 10-25% when $M = 2$. More extensive simulations revealed that the relative efficiency tends to

decrease as β increases, but for the small β observed in air pollution studies, the plots of relative efficiency look almost identical to Figure 3A.

Positive autocorrelation in the exposure series also decreases efficiency. If referents are close to the index time, they will be auto-correlated with the index exposure, and, hence, there will be less power to detect an exposure effect. This is illustrated in Figure 3B. For the same three exposures shown in Figure 3A, we show the relative efficiency of the time stratified and full stratum bidirectional designs. In contrast to Figure 3A, in which the strata in the time stratified design are sets of adjacent days, the referents in Figure 3B are interspersed throughout the exposure series (e.g., if $M = 4$, the first stratum consists of days 1, 2, 3, and 4 in Figure 3A, and days 1, 26, 51, and 76 in Figure 3B). Indeed, we see that the efficiency of the time stratified design is higher in Figure 3B than in Figure 3A. Thus, it is advisable to choose referents that are not adjacent to the index time (e.g. six days apart) in order to maximize the information contributed by each referent.

7. Case-Crossover Methods for Unshared Exposure Series

This paper has focused on the case of a fixed, shared exposure series. Hence, throughout our discussion we have conditioned on exposure within the referent windows. If, however, exposure is not shared, it may be more appropriate not to condition on exposure. In particular, if exposures are independent across subjects, the set of observed exposures can be thought of as a random sample from a given distribution. We call such exposures "random". With random exposures, it is appropriate not to condition on exposure, but to examine the properties of the

effect estimate and estimating equations averaged over all possible exposure series. Several of the case-crossover methods papers have considered random exposures.[12;15]

We note, however, that not all unshared exposures are random. If, for example, ambient exposure series are available for several different geographic regions, these series may be spatially correlated. With this type of exposure, whether or not to condition on exposure is debatable.

With a random exposure, under non-localizable referent selection, Vines and Farrington [15] showed that the conditional logistic regression estimate may be biased when averaging over exposures, except under very strict conditions. In fact, with a random exposure and non-localizable design, we know of no unbiased estimator.

In contrast, with a localizable, ignorable referent scheme, and any exposure distribution, there is no bias in the conditional logistic regression effect estimate when averaging over exposures. This follows since, for localizable, ignorable referent schemes, there can be no overlap bias for a given exposure series, and hence there is no bias averaged across exposure series (i.e. for exposures $X, E(\hat{\beta} - \beta) = E_X(E(\hat{\beta} - \beta \mid X)) = E_X(0) = 0$). However, conditional logistic regression may not be the most efficient way to estimate the exposure effect. The most efficient estimate would come from assuming a model for the exposure distribution and basing estimation on the likelihood of the exposure conditional on the index times and referent windows.

This discussion emphasizes, once again, the different properties of the referent selection classes. With a non-localizable referent scheme, and a shared exposure series, there does not exist a method for obtaining unbiased effect estimates conditional on exposure.[41] Specifically, the conditional logistic regression estimates have overlap bias. In addition, with a random exposure series, there is no existing method for calculating unbiased effect estimates. In contrast, with a localizable, ignorable referent scheme, the conditional logistic regression estimates are unbiased regardless of whether we do or don't condition on exposure.

8. Discussion and Conclusions

The case-crossover design is well suited to the study of the association between short-term air pollution exposure and the risk of an acute adverse health event. By making within-person comparisons, confounding by time-independent confounders is eliminated. Moreover, if referents are matched on important time-dependent confounders, these effects are also controlled. Effect modification can easily be assessed, and standard conditional logistic regression methods can be used for analysis.

With air pollution exposures, confounding is of particular concern, since confounders tend to dominate the exposure effects. Due to time-dependent confounding and time trends and autocorrelation in air pollution exposure series, proper referent selection is particularly important. Referents should be matched on the most dominant time varying confounders, and should be sampled bidirectionally. Sampling referents too close to the index day will result in a loss of power due to autocorrelation in the exposure series. If there remains a choice as to the number of referents after these concerns are taken into account, using more referents will

increase efficiency. Finally, the analysis should condition on the fixed and known exposure series.

If a non-localizable or localizable, non-ignorable referent scheme is used, conditional logistic regression yields biased estimates. This bias is usually small, though it can exist for effect estimates in the range of those typically seen in air pollution studies, and even when there is no exposure effect. The magnitude of the overlap bias varies according to the referent scheme and depends on the particular exposure series; hence, model shopping on referent strategies will tend to exacerbate the bias. (Model shopping occurs when a referent selection strategy is chosen because it results in larger estimates of effect than other candidate referent selection strategies). Therefore, it is wise to avoid overlap bias entirely. Localizable, ignorable referent schemes allow for unbiased estimation using a standard conditional logistic regression analysis.

Our recommendations for the choice of referent selection strategy in an air pollution exposure study assume that the exposure is exogenous and the outcome rare. The time stratified design should be used for referent selection, since it avoids overlap bias and bias due to time trend. The stratification can be tailored to match on the most important time-dependent confounders. Stratifying on year and month, as well as one or more of day of the week, temperature, measurement time, and co-pollutants should be adequate for most studies. While the semi-symmetric bidirectional design can also achieve these goals, it requires modification of the traditional conditional logistic regression analysis, and will be less efficient than a time stratified design because fewer referents are used.

Appendix A

Standard estimating equation theory tells us that we can approximate the bias in $\hat{\beta}$ by

$$\hat{\beta} - \beta \approx \frac{E_{t_i}(U_i(\beta))}{-E_{t_i}(\frac{d}{d\beta}U_i(\beta))}, \tag{1a}$$

where $U_i(\beta)$ is the conditional logistic regression estimating equation for subject *i*, and the expectations are taken with respect to the event time, t_i. At this point, we drop the *i* subscript, and assume that we have a shared exposure series. It can be shown that

$$E_t(U(\beta)) = \sum_{t=1}^{T} \frac{e^{z_t\beta}}{\sum_{s=1}^{T} e^{z_s\beta}} \left(z_t - \sum_{u \in W_t} z_u \frac{e^{z_u\beta}}{\sum_{v \in W_t} e^{z_v\beta}} \right), \tag{2a}$$

where z_t is the exposure on day *t* and W_t is the referent window for day *t*.[22;41] Also,

$$-E_t(\frac{d}{d\beta}U(\beta)) = \sum_{t=1}^{T} \frac{e^{z_t\beta}}{\sum_{s=1}^{T} e^{z_s\beta}} \left[\frac{\sum_{u \in W_t} {z_u}^2 e^{z_u\beta}}{\sum_{v \in W_t} e^{z_v\beta}} - \left(\frac{\sum_{u \in W_t} z_u e^{z_u\beta}}{\sum_{v \in W_t} e^{z_v\beta}} \right)^2 \right]. \tag{3a}$$

Notice that both (2a) and (3a) are sums over all days in the exposure series. The summands are functions of the exposures within the referent windows surrounding each day. They are independent of the outcome, and depend only on the exposures.

We consider the case of a single binary exposure series of length *T*, where there are *K* "positive" exposure days. Using the symmetric bidirectional design, we sample referents one day before and after the index day. With a binary exposure series, the overlap bias (1a) reduces to a very simple form. The form of the bias depends on the number of each of the possible exposure arrangements within a referent window. Since the referent window includes the index day and

the previous and subsequent days, there are 2^3 possible exposure arrangements within the window. Yet index days on the first or last day of the series have only one referent. Hence there are 2^2 possible values for the 2 exposures within these referent windows. Some of the exposure arrangements contribute the same amount to the bias, and hence are equivalent in terms of our calculations. Exposure arrangements with no variation in exposure contribute nothing to the bias. The unique arrangements of exposures that concern us are: `010`, `001` (equivalent to `100`), `011` (equivalent to `110`), `101`, `01` at the beginning of the series (equivalent to `10` at the end of the series), and `10` at the beginning of the series (equivalent to `01` at the end of the series). The numbers of each of these arrangements are represented by z_{010}, z_{100}, z_{110}, z_{101}, z_{01}, and z_{10}, respectively. The arrangements `000` and `111` do not contribute to the summations in (2a) and (3a). Using this notation and some simple algebra, we find that (2a) and (3a) reduce to equations (1) and (2) shown in the text.

In contrast, with time-stratified referent selection, and strata of length two and three, X becomes

$$X = \frac{e^{\beta}}{(T-K)+Ke^{\beta}}\left((z_{010}+z_{001})\left(\frac{-2e^{\beta}}{2+e^{\beta}}+e^{\beta}\left(1-\frac{e^{\beta}}{2+e^{\beta}}\right)\right)+(z_{011}+z_{101})\left(\frac{-2e^{\beta}}{1+2e^{\beta}}+2e^{\beta}\left(1-\frac{2e^{\beta}}{1+2e^{\beta}}\right)\right)\right.$$
$$\left.+(z_{01}+z_{10})\left(\frac{-e^{\beta}}{1+e^{\beta}}+e^{\beta}\left(1-\frac{e^{\beta}}{1+e^{\beta}}\right)\right)\right)$$

which simplifies to zero.

Appendix B

We use the inverse of expression (3a) in Appendix A as the variance of $\hat{\beta}$ for a given exposure series. Note that this expectation is taken over all event times (not conditional on the referent

windows). Hence, it is not the same as the conditional logistic regression variance of $\hat{\beta}$,

$$\left[\frac{\sum_{s \in W_t} z_s^2 e^{z_s \beta}}{\sum_{u \in W_t} e^{z_u \beta}} - \left(\frac{\sum_{s \in W_t} z_s e^{z_s \beta}}{\sum_{u \in W_t} e^{z_u \beta}} \right)^2 \right]^{-1}$$

, which is conditional on the observed referent windows, and thus depends on the observed outcome. We use the inverse of (3a) because it does not depend on the outcome, and thus ensures that our efficiency results are more generalizable.

.

Reference List

(1) Dominici F, Sheppard L. Health effects of air pollution: a statistical review. *International Statistical Review* 2003; 71:243-276.

(2) Dominici F, McDermott A, Hastie T. Improved semi-parametric time series models of air pollution and mortality. *Journal of the American Statistical Association*, in press.

(3) Maclure M. The case-crossover design: A method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991; 133(2):144-153.

(4) Kwon HJ, Cho SH, Nyberg F, Pershagen G. Effects of Ambient Air Pollution on Daily Mortality in a Cohort of Patients with Congestive Heart Failure. *Epidemiology* 2001; 12:413-419.

(5) Lee JT, Schwartz J. Reanalysis of the effects of air pollution on daily mortality in Seoul, Korea: A case-crossover design. *Environ Health Perspect* 1999; 107:633-636.

(6) Levy D, Sheppard L, Checkoway H, Kaufman J, Lumley T, Koenig J et al. A case-crossover analysis of particulate matter air pollution and out-of-hospital primary cardiac arrest. *Epidemiology* 2001; 12:193-199.

(7) Navidi W. Bidirectional case-crossover designs for exposures with time trends. *Biometrics* 1998; 54:596-605.

(8) Neas LM, Schwartz J, Dockery D. A case-crossover analysis of air pollution and mortality in Philadelphia. *Environ Health Perspect* 1999; 107:629-631.

(9) Peters A, Dockery DW, Muller JE, Mittleman MA. Increased particulate air pollution and the triggering of myocardial infarction. *Circulation* 2001; 103:2810-2815.

(10) Sunyer J, Schwartz J, Tobias A, Macfarlane D, Garcia J, Anto JM. Patients with chronic obstructive pulmonary disease are at increased risk of death associated with urban particle air pollution: A case-crossover analysis. *Am J Epidemiol* 2000; 151:50-56.

(11) Greenland S. Confounding and exposure trends in case-crossover and case time-control designs. *Epidemiology* 1996; 7:231-239.

(12) Marshall RJ, Jackson RT. Analysis of case-crossover designs. *Stat Med* 1993; 12:2333-2341.

(13) Mittleman MA, Maclure M, Robins JM. Control sampling strategies for case-crossover studies: An assessment of relative efficiency. *American Journal of Epidemiology* 1995; 142(1):91-98.

(14) Navidi W, Weinhandl E. Risk set sampling for case-crossover designs. *Epidemiology* 2002; 13(1):100-105.

(15) Vines SK, Farrington CP. Within-subject exposure dependency in case-crossover studies. *Stat Med* 2001; 20:3039-3049.

(16) Bateson TF, Schwartz J. Control for seasonal variation and time trend in case crossover studies of acute effects of environmental exposures. *Epidemiology* 1999; 10:539-544.

(17) Bateson TF, Schwartz J. Selection bias and confounding in case-crossover analyses of environmental time-series data. *Epidemiology* 2001; 12:654-661.

(18) Fung KY, Krewski D, Chen Y, et al. Comparison of time series and case-crossover analyses of air pollution and hospital admission data. *International Journal of Epidemiology* 2003; 32(6):1064-1070.

(19) Jaakkola JJK. Case-crossover design in air pollution epidemiology. *European Respiratory Journal* 2003; 21: Suppl. 40(81s):85s.

(20) Lee JT, Kim H, Schwartz J. Bidirectional case-crossover studies of air pollution: Bias from skewed and incomplete waves. *Environ Health Perspect* 2000; 108(12):1107-1111.

(21) Levy D, Lumley T, Sheppard L, et al. Referent selection in case-crossover analyses of acute health effects of air pollution. *Epidemiology* 2001; 12:186-192.

(22) Lumley T, Levy D. Bias in the case-crossover design: Implications for studies of air pollution. *Environmetrics* 2000; 11:689-704.

(23) D'Ippoliti D, Forastiere F, Ancona C, et al. Air pollution and myocardial infarction in Rome: A case-crossover analysis. *Epidemiology* 2003; 14(5):528-535.

(24) Kan H, Chen B. A case-crossover analysis of air pollution and daily mortality in Shanghai. *Journal of Occupational Health* 2003; 45(2):119-124.

(25) Lin M, Chen Y, Burnett RT, et al. The influence of ambient coarse particulate matter on asthma hospitalisations in children: Case-crossoover and time-series analyses. *Environmental Health Perspectives* 2002; 110(6):575-581.

(26) Lin M, Chen Y, Burnett RT, et al. Effect of short-term exposure to gaseous pollution on asthma hospitalisation in children: A bidirectional case-crossover analysis (Research report). *Journal of Epidemiology and Community Health* 2003; 57(1):50-55.

(27) Schwartz J. Is the association of airbone particles with daily deaths confounded by gaseous air pollutants? An approach to control by matching. *Environmental Health Perspectives* 2004; 112(5):557-561.

(28) Sullivan J, Ishikawa N, Sheppard L, et al. Exposure to ambient fine particulate matter and primary cardiac arrest among persons with and without clinically recognized heart disease. *American Journal of Epidemiology* 2003; 157(6):501-509.

(29) Sunyer J, Basagana X. Particles, and not gases, are associated with the risk of death in patients with chronic obstructive pulmonary disease. *International Journal of Epidemiology* 2001; 30(5):1138-1140.

(30) Sunyer J, Basagana X, Belmonte J, et al. Effect of nitrogen dioxide and ozone on the risk of dying in patients with severe asthma. *Thorax* 2002; 57(8):687-693.

(31) Tsai SS, Huang CH, Goggins WB, et al. Relationship between air pollution and daily mortality in a tropical city: Kaohsiung, Taiwan. *Journal of Toxicology and Environmental Health, Part A* 2003; 66(14):1341-1349.

(32) Tsai SS, Goggins WB, Chiu HF, et al. Evidence for an association between air pollution and daily stroke admissions in Kaohsiung, Taiwan. *Stroke* 2003; 34(11):2612-2616.

(33) Yang CY, Yong-Shing C, Chiang-Hsing Y, et al. Relationship between ambient air pollution and hospital admissions for cardiovascular diseases in Kaohsiung, Taiwan. *Journal of Toxicology and Environmental Health, Part A* 2004; 67:483-493.

(34) Yang CY, Chang CC, Hung-Yi C, et al. Relationship between air pollution and daily mortality in a subtropical city: Taipei, Taiwan. *Environment International* 2004; 30:519-523.

(35) Yang Q, Chen Y, Shi Y, et al. Association between ozone and respiratory admissions among children and the elderly in Vancouver, Canada. *Inhalation Toxicology* 2003; 15(13):1297-1308.

(36) Clyde M. Bayesian model averaging and model search strategies (with discussion). In: JM Bernado, JO Berger, AP Dawid, AFM Smith, eds. *Bayesian Statistics 6*. Oxford: Oxford University Press; 1999: 157-185.

(37) Clyde M. Model uncertainty and health effect studies for particulate matter. *Environmetrics* 2000; 11:745-763.

(38) Lumley T, Sheppard L. Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control anayses. *Environmetrics* 2000; 11:705-718.

(39) Varachan R, Frangakis CE. Revealing and addressing length bias and heterogeneous effects in frequency case-crossover studies. *Am J Epidemiol* 2004; 159(6):596-602.

(40) Breslow NE, Day NE. Statistical Methods in Cancer Research. Lyon: International Agency for Research on Cancer, 1980.

(41) Janes H, Sheppard L, Lumley T. Overlap bias in the case-crossover design, with application to air pollution exposures. *Statistics in Medicine*, in press. Published online: www.interscience.wiley.com.

(42) Little RJA, Rubin DB. Statistical analysis with missing data. New York: John Wiley, 1987.

(43) Austin H, Flanders WD, Rothman KJ. Bias arising in case-control studies from selection of controls from overlapping groups. *International Journal of Epidemiology* 1989; 18:713-716.

(44) Robins J, Pike M. The validity of case-control studies with nonrandom selection of controls. *Epidemiology* 1990; 1:273-284.

Table 1 Case-crossover studies of air pollution exposures. Studies are listed in (reverse) chronological order.

Authors	Pub. Date	Exposure	Outcome	Study Population	Referent Strategy
Yang et al.[31]	2004	PM_{10}, CO, NO_2, SO_2, O_3	cardiovascular disease hospital admissions	Kaohsiung, Taiwan	symmetric bidirectional 7
Yang et al.[32]	2004	PM_{10}, CO, NO_2, SO_2, O_3	non-accident mortality	Taipei, Taiwan	symmetric bidirectional 7
D'Ippoliti et al.[21]	2003	TSP, CO, NO_2, SO_2	acute MI hospital admission	Rome, Italy	time stratified by DOW, month, year
Kan et al.[22]	2003	PM_{10}, NO_2, SO_2	non-accident mortality	Shanghai, China	unidirectional 7,14,21 symmetric bidirectional 7,14,21
Lin et al.[24]	2003	CO, NO_2, SO_2, O_3	asthma hospital admission	children in Toronto, Canada	symmetric bidirectional 14
Schwartz[25]	2004	PM_{10}	non-accident mortality	14 US cities	time stratified by month, year, gaseous pollutant*
Sullivan et al.[26]	2003	PM_{10}, CO, SO_2	out of hospital primary cardiac arrest	Washington State	time stratified by DOW, month, year
Tsai et al.[29]	2003	PM_{10}, CO, NO_2, SO_2, O_3	non-accident mortality	Kaohsiung, Taiwan	symmetric bidirectional 7
Tsai et al.[30]	2003	PM_{10}, CO, NO_2, SO_2, O_3	stroke hospital admission	Kaohsiung, Taiwan	symmetric bidirectional 7
Yang et al.[33]	2003	COH, CO, NO_2, SO_2, O_3	ozone and respiratory hospital admission	children and elderly in Vancouver, Canada	symmetric bidirectional 7
Lin et al.[23]	2002	$PM_{10-2.5}$, $PM_{2.5}$, PM_{10}	asthma hospital admission	children in Toronto, Canada	unidirectional 14 symmetric bidirectional 14
Sunyer et al.[28]	2002	PM_{10}, black smoke, CO, NO_2, SO_2, O_3	asthma visits to the emergency room	asthma patients 14 years and older in Barcelona, Spain	time stratified by DOW, month, year
Kwon et al.[2]	2001	PM_{10}, CO, NO_2, SO_2, O_3	non-accident mortality	patients with congestive heart failure in Seoul, South Korea	symmetric bidirectional 7,14
Levy et al.[4]	2001	$PM_{2.5}$**, PM10	out-of hospital primary cardiac arrest	Seattle, WA	time stratified by DOW, month, year
Peters et al.[7]	2001	$PM_{2.5}$	acute MI	Boston, MA	unidirectional 2,3,4
Sunyer et al.[27]	2001	PM_{10}, CO, NO_2, O_3	non-accident mortality	adults in Barcelona with chronic obstructive pulmonary disease	symmetric bidirectional 7
Sunyer et al.[8]	2000	black smoke	non-accident mortality	patients with chronic obstructive pulmonary disease in Barcelona, Spain	symmetric bidirectional 7
Lee and Schwartz[3]	1999	TSP, SO_2, O_3	non-accident mortality	Seoul, South Korea	unidirectional retrospective and prospective 7 and/or 14; symmetric bidirectional 7
Neas et al.[6]	1999	TSP	non-accident mortality	Philadelphia, PA	symmetric bidirectional 7,14,21

* stratified by month, year, and one of four gaseous pollutants: CO (within 0.03 ppm), NO_2 (within 1 ppb), SO_2 (within 1 ppb), O_3 (within 2 ppb).

** measured by light scattering

TSP = total suspended particulates; MI = myocardial infarction; PM_x = particulate matter less than x μm in diameter; DOW = day of week

Table 2 Characteristics of the referent selection strategies commonly used in air pollution studies.

Referent Selection Strategy	Referent Class	Controls for time trend bias?	Controls for confounding by design?	CLR estimates unbiased?*
Restricted unidirectional	Non-localizable	No	No	No
Full stratum bidirectional	Localizable, ignorable	Yes	No	Yes
Symmetric bidirectional	Non-localizable	Yes	Yes	No
Time stratified	Localizable, ignorable	Yes	Yes	Yes
Semi-symmetric bidirectional	Localizable, non-ignorable	Yes	Yes	No**

* CLR: conditional logistic regression.

** Conditional logistic regression with an offset of log(2) for cases with only one referent, and zero otherwise, will produce unbiased estimates.

Table 3 Four different length 10 binary exposure series, their parameters, and the values of the expected conditional logistic regression estimating equation (X) for different values of β.

		Parameters						X = expected estimating equation		
	Exposure Series	z_{010}	z_{001}	z_{101}	z_{011}	z_{01}	z_{10}	$\beta = 0$	$\beta = 0.05$	$\beta = 0.15$
A	0111100000	0	1	0	2	1	0	-0.0167	-0.0176	-0.0193
B	0101101010	3	0	3	2	2	0	-0.0331	-0.0305	-0.0248
C	1101010111	2	0	3	2	0	0	0.0169	0.0202	0.0269
D	1101110111	0	0	2	4	0	0	<0.0001	<0.0001	<0.0001

Figure 1 Restricted unidirectional, full stratum bidirectional, and symmetric bidirectional referent selection strategies.

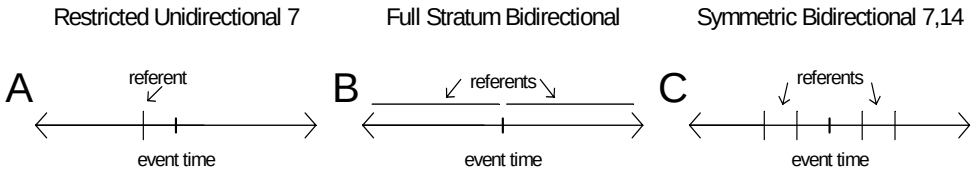

Figure 2 The large sample bias as a function of β for the exposure series shown in Table 3. We show bias on two scales: for β between -1 and 1, and β between -.2 and .2. In the second plot, the line y = x is superimposed so that we can observe when the bias is larger than β itself. Note that the large sample bias scale is different on the two plots.

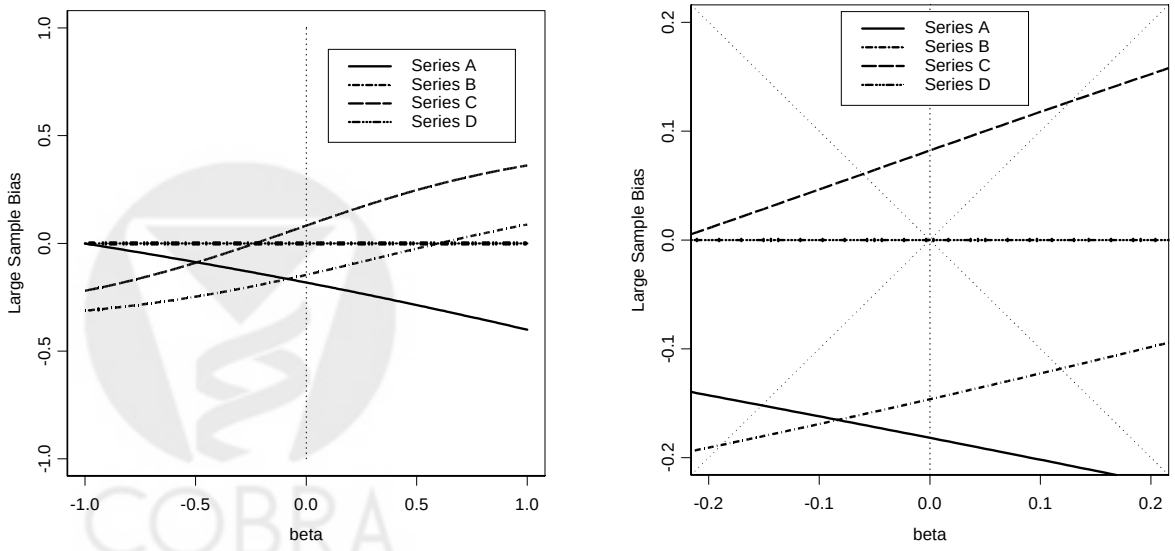

Figure 3 Efficiency of the time stratified design relative to the full stratum bidirectional design, as a function of the stratum size in the time stratified design. Relative efficiency is shown for three randomly chosen simulated exposure series with autocorrelation, when $\beta = 0$. On the left, the time stratified design uses strata as sequences of adjacent days (i.e., if the stratum size is 4, the first stratum consists of days 1 to 4), and in the right panel, the referents are spaced in the series (i.e., if the stratum size is 4, the first stratum consists of days 1, 26, 51, 76).