Case–Crossover Analyses of Air Pollution Exposure Data

Referent Selection Strategies and Their Implications for Bias

Holly Janes,* Lianne Sheppard,*† and Thomas Lumley*

Abstract: The case–crossover design has been widely used to study the association between short-term air pollution exposure and the risk of an acute adverse health event. The design uses cases only; for each individual case, exposure just before the event is compared with exposure at other control (or "referent") times. Time-invariant confounders are controlled by making within-subject comparisons. Even more important in the air pollution setting is that time-varying confounders can also be controlled by design by matching referents to the index time. The referent selection strategy is important for reasons in addition to control of confounding. The case–crossover design makes the implicit assumption that there is no trend in exposure across the referent times. In addition, the statistical method that is used—conditional logistic regression—is unbiased only with certain referent strategies. We review here the case–crossover literature in the air pollution context, focusing on key issues regarding referent selection. We conclude with a set of recommendations for choosing a referent strategy with air pollution exposure data. Specifically, we advocate the time-stratified approach to referent selection because it ensures unbiased conditional logistic regression estimates, avoids bias resulting from time trend in the exposure series, and can be tailored to match on specific time-varying confounders.

(*Epidemiology* 2005;16: 717–726)

Editors' note: A commentary on this article appears on page 715.

The case–crossover design is an approach for studying acute effects of day-to-day variation in pollution on morbidity or mortality.[1] Various health outcomes have been assessed, including deaths and myocardial infarctions. The exposure is typically ambient air pollution concentration (frequently, particulate matter [PM]) measured at a centrally located monitor in a particular geographic region. The same exposure is used for all subjects in the study and hence is called a "shared" exposure series.

Studies of acute effects face many challenges, including exposure measurement error, the discrepancy between ambient and personal exposures, and confounding. 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 other pollutants. These factors can be strongly associated with day-to-day variation in both exposure and health outcomes. Biases are especially troublesome because the exposure effect is usually very small, whereas the time-dependent confounding can be many times larger.[2]

The case–crossover design represents a novel approach to control for confounding. Given a sample of subjects who experienced the event, exposure just before the event (the "index" time) is compared with exposure at comparable control, or "referent," times.[3] This is similar to a matched case–control study; the index exposures belong to matched sets of exposures for subjects at their individual referent times. The matched sets are called "referent windows." 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 confounding effects are also controlled by design. This is in stark contrast to other approaches used for the estimation of acute pollution effects (eg, time-series studies) in which time-dependent confounding is controlled by modeling.

The selection of referents is a key issue in the case–crossover design. We refer to this choice as the "referent

selection strategy" or "referent scheme." As stated previously, the referent scheme 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, the estimating equations typically used (ie, the conditional logistic regression estimating equations) are unbiased only with certain referent strategies. We call this bias in the conditional logistic regression estimating equations "overlap bias." (The use of the term "overlap" is discussed subsequently.) Although parameter estimates from an unbiased estimating equation are unbiased in large samples, a biased estimating equation produces biased parameter estimates even in large samples. In this article, we focus on bias in estimating equations rather than bias in the estimates themselves, because the latter have small sample bias.

A variety of referent schemes have been used 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 none has provided guidelines for evaluating and choosing a referent scheme. In this article, we review referent selection practices in air pollution case–crossover studies and then give an overview of the case–crossover method and potential biases associated with the design. We relate each of the referent schemes to the various types of bias and provide an illustration of overlap bias. Finally, we explore the relative efficiency of 2 referent schemes and extend our review to unshared exposure series. We conclude with recommendations for choosing a referent scheme for an air pollution case–crossover study.

REFERENT SCHEMES USED IN THE LITERATURE

We reviewed 19 air pollution case–crossover studies published between the introduction of the design in 1991 and June 2004 (see Table 1).[4–6,8–10,23–35] We limit our attention to applied rather than methodologic papers. All of the studies examined ambient, shared air pollution exposures.

By far the most popular referent scheme (used in 63% of the studies) is the symmetric bidirectional design[16] in which referents are at a fixed lag before and after the index time. 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. A less common method is the restricted unidirectional design in which referents are at fixed lags before the index day. Unidirectional referents are usually sampled before the index day, although one study sampled unidirectional referents prospectively.[5] Finally, several studies reported the results of multiple analyses, each using a different referent scheme. We do no recommend this practice, because it makes interpretation difficult and introduces model selection bias.[36–38] Although these referent schemes may appear to be quite similar, we show that the

distinctions among them are important and have implications for bias.

THE CASE–CROSSOVER METHOD

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 assume that the event is rare: for each individual, events occur with a low probability.

Relevant Aspects of Air Pollution Exposure Data

With a shared exposure series, it is appropriate to condition on the exposure when deriving the likelihood of the data and when evaluating the properties of the effect estimate. (The likelihood can be thought of as the probability model for the data.) With only one exposure series, we want to know that the estimate and 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. We restrict our attention primarily to analyses conditional on exposure except in the section on unshared exposures.

Ambient air pollution exposure data is exogenous or generated independently of the individual under study. In contrast, endogenous exposures such as personal air pollution are influenced by the individual. Our review is limited to exogenous exposures.

Attributes of the Design Useful in the Air Pollution Context

The case–crossover design has several strengths: it does not require a control sample (and hence avoids bias associated with improper control selection); it makes effect modification assessment relatively simple; it controls for fixed confounders by design; and it controls for time-dependent confounders by matching.

Time-series regression is also frequently used to assess the short-term health effects of air pollution. The main difference between this method 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. (For further control, confounders could also be included in the regression model.)

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

TABLE 1. Case–Crossover Studies of Air Pollution Exposures*

Authors	Publication 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	Nonaccident 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	Nonaccident 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}	Nonaccident 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	Nonaccident 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	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 yr 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	Nonaccident mortality	Patients with congestive heart failure in Seoul, South Korea	Symmetric bidirectional 7, 14
Levy et al[4]	2001	$PM_{2.5}$;‡ PM_{10}	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	Nonaccident mortality	Adults in Barcelona with chronic obstructive pulmonary disease	Symmetric bidirectional 7
Sunyer et al[8]	2000	Black smoke	Nonaccident mortality	Patients with chronic obstructive pulmonary disease in Barcelona, Spain	Symmetric bidirectional 7
Lee and Schwartz[3]	1999	TSP, SO_2, O_3	Nonaccident mortality	Seoul, South Korea	Unidirectional retrospective and prospective 7 and/or 14; symmetric bidirectional 7
Neas et al[6]	1999	TSP	Nonaccident mortality	Philadelphia, PA	Symmetric bidirectional 7, 14, 21

*Studies are listed in (reverse) chronologic order.
†Stratified by month, year, and one of 4 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.

model assumes that time can be divided into "exposed" and "unexposed" periods. It stipulates that a subject is at high risk for a fixed time after an exposed period and thereafter returns to background risk until the next exposed period. (Note that this model is subject to length bias under certain conditions, as noted by Varachan and Frangakis.[39]) However, the proportional hazards model for a rare disease, using a constant baseline hazard for each individual, is more appropriate for nonbinary air pollution exposures. This was first proposed for case–crossover studies by Navidi[7] and by Marshall and Jackson.[12] We call this the time-varying exposure model.

The time-varying exposure model assumes that there is only one event for each case (which is legitimate with a rare event) and specifies risk as a function of time and exposure. If past exposure lags are included, this model allows the risk to depend on exposure history. Under the time-varying exposure model, the hazard rate of person i at time t given time-varying covariates x_{it} is given by $\lambda_i(t; x_{it}) = \lambda_i e^{x_{it}\beta}$. Over a short time period, the assumption of a constant baseline hazard is often reasonable; this assumption is equivalent to assuming smooth seasonal effects in a time-series analysis.[22] The parameter e^β represents 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 was motivated by the analogy to matched case–control designs in which the conditional logistic regression likelihood is the likelihood of the data. Its use makes sense, because 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 (ie, matched pairs), the 2 estimation procedures are identical.[40] In general, however, conditional logistic regression is a better choice, because it can be used with nonbinary exposures and makes it easier to control for additional confounders (those not used in the matching). These factors can be included in the regression model.

Potential Biases Associated With the Choice of Referent Scheme

The conditional logistic regression estimating equations are unbiased only with certain referent schemes. For most of the commonly used strategies (eg, symmetric bidirectional), the estimating equations have overlap bias.[22,41] In the next section, we identify the referent schemes that are subject to overlap bias.

Greenland[11] and Navidi[7] showed that the case–crossover design requires that there be no time trend in exposure within the referent window. This assumption is necessary because the exposure effect is estimated by contrasting exposures at the index and referent times. If, for example, referents are always before the index time, and there is a decreasing trend in exposure, the effect estimate will be negatively biased. With strong long-term time trends often present in air pollution data, bias resulting from time trend is a concern. We discuss a method for controlling this bias in the next section.

REFERENT SCHEMES

Classes of Referent Schemes

As mentioned, a number of fundamentally different types of referent schemes have been used in air pollution studies (Table 2). We have previously proposed a taxonomy of referent strategies with groups that correspond to the statistical properties of the designs.[41] We classify designs as localizable or nonlocalizable and ignorable or nonignorable. 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 nonlocalizable design, the conditional likelihood is uninformative for β. The symmetric bidirectional design, for example, is nonlocalizable, because 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 because, when estimation is based on making comparisons within the referent windows (which

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	Nonlocalizable	No	No	No
Full-stratum bidirectional	Localizable, ignorable	Yes	No	Yes
Symmetric bidirectional	Nonlocalizable	Yes	Yes	No
Time-stratified	Localizable, ignorable	Yes	Yes	Yes
Semisymmetric bidirectional	Localizable, nonignorable	Yes	Yes	No*

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

are matched on time-dependent confounders), these effects are controlled. Of the designs in Table 2, only the time-stratified, full-stratum bidirectional, and semisymmetric bidirectional designs are localizable.

The localizable designs are classified as either ignorable or nonignorable. Ignorable designs are those for which the referent sampling scheme can be ignored in the analysis; conditional logistic regression can be used to obtain unbiased estimates. With a nonignorable design, the likelihood of the data depends on the referent sampling scheme, and this likelihood must be used for an unbiased analysis. This definition of ignorability is the same as that proposed by Little and Rubin[42] and used in the missing data context. In either case, ignorability implies that the data can be analyzed as if the observed data were the complete data without accounting for how the data were sampled. Of the designs 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 estimates have overlap bias under nonlocalizable or localizable, nonignorable referent selection.[41] With a nonlocalizable 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, nonignorable referent scheme, again, the likelihood of the data must be used for unbiased estimation (although in one case, there is a simple way, as described subsequently, to obtain these estimates).

Unidirectional Referent Selection

With the restricted unidirectional design, confounding by 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 (Fig. 1A). This design is nonlocalizable.[41] Yet, unidirectional sampling has another major disadvantage in air pollution studies: selecting referents only before the index time can lead to time trend bias. For this reason, unidirectional sampling is not common in air pollution studies.

Full-Stratum Bidirectional Referent Selection

Navidi[7] proposed that time trend bias could be eliminated by choosing referents bidirectionally both before and after the index time (sometimes called ambidirectional selection[21]). Technically, bidirectional sampling is valid only

when cases are still at risk after an event, an assumption that is certainly violated 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. Lumley and Levy[22] showed that, with a rare event, the bias resulting from sampling referents after the at-risk period is small, and, more importantly, smaller than the time trend bias.

Navidi[7] proposed full-stratum bidirectional referent selection (Fig. 1B), in which the referents are all days in the exposure series other than the index day. Interestingly, a case–crossover analysis with full-stratum bidirectional referents and a shared exposure series is equivalent to a Poisson regression analysis.[22] However, although time trend bias is controlled by design, time-dependent confounding must be controlled by modeling, because the referent window is so large that confounding is not necessarily controlled. The full-stratum bidirectional design is a localizable, ignorable design.[41]

Symmetric Bidirectional Referent Selection

If referents are within the same season and on the same day of the week as the index time (Fig. 1C),[16] the symmetric bidirectional design controls for bias resulting from time trend and confounding by both season and day of the week. 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 nonlocalizable.[41]

Time-Stratified Referent Selection

The time-stratified design is not subject to bias resulting from time trend because there is no pattern in the placement of referents relative to the index time. In addition, the design can control for season and day of the week by restricting referents to the same day of the week, month, and year as the index day. The time-stratified design is a localizable, ignorable design.[41]

The time-stratified design has some interesting relationships with other designs. The full-stratum bidirectional design is a time-stratified design in which there is one large stratum (although for this design, 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
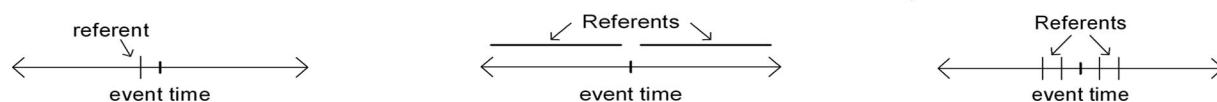

FIGURE 1. Restricted unidirectional, full-stratum bidirectional, and symmetric bidirectional referent selection strategies.

721

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

Semisymmetric Bidirectional Referent Selection

With the semisymmetric bidirectional design,[14] one referent is randomly chosen from days at a fixed lag pre- and postevent; if only one of these days is available (as a result of the case being at either end of the exposure series), it serves as the referent. If the lag is small and a multiple of 7, confounding by season and day of the week can be controlled by design. There is no bias resulting from time trend, because referents are bidirectionally sampled.

The semisymmetric bidirectional design is a localizable, nonignorable design. It turns out that in this case, estimates based on the true likelihood can be obtained using standard conditional logistic regression software when an offset is specified (with a value of $\log_e 2$ for cases that have only one possible referent and zero otherwise; this is equivalent to an offset of $-\log_e 2$ for cases with 2 referents).[41]

OVERLAP BIAS

The term "overlap bias" was first used by Lumley and Levy,[22] who observed that this bias is similar to friend control bias in matched case–control studies.[43,44] Yet, the term "overlap" is somewhat misleading; it suggests that a design in which an individual's set of possible referent windows is overlapping is subject to overlap bias and that a design with disjoint windows is not subject to such bias. This is not the case; in both the full-stratum bidirectional and symmetric bidirectional designs, the referent windows overlap, but only the latter design is subject to overlap bias. Similarly, the phenomenon of overlap bias is not defined by whether or not the index time is fixed within the referent window. Despite the fact that all of the existing referent strategies with index times fixed within the referent window are subject to overlap bias, and those with random index times are not (eg, the time-stratified design), it would be possible to configure a referent strategy with a random index time that is subject to overlap bias. Hence, we have not found a satisfactory heuristic explanation of overlap bias. The bias is purely mathematical; whether or not bias exists for a particular referent strategy depends on the form of the likelihood. In this sense, overlap bias is similar to the well-known bias associated with unconditional logistic regression in a case–control study with finely matched data.[40]

Numerical Example

In this example, we calculate the overlap bias, as a function of the exposure series, for 2 referent schemes. We show that, for a fixed referent strategy, determining which exposure series are prone to overlap bias is virtually impossible. For a given exposure series, there is no intuitive reason

that the symmetric bidirectional design[16] (with 1-day lag) is subject to overlap bias and the time-stratified design (with strata of length 2 and 3) is not.[41] We consider the simple case of a shared binary exposure series of length 10.

We first calculate the overlap bias under symmetric bidirectional referent selection. Let the binary exposure series, z, have length T, with K "positive" exposures. This series can be reduced to a set of 6 parameters (z_{010}, z_{001}, z_{011}, z_{101}, z_{01}, and z_{10}), which refer 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 or 01 at the end. We show in Appendix A (available with the online version of this article) that the overlap bias can be expressed as X/Y where

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

(1)

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

(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 way to characterize the exposure series that have overlap bias. (Table 3) gives 4 different length-10 binary exposure series along with their parameters and the value of X at various values of β. Figure 2 displays the overlap bias associated with each of these series (X/Y). We observe a large amount of variation in bias across the exposure series. For example, although series D differs on just one day from series C, these 2 series have dramatically different amounts of overlap bias. We also find that, in several instances, the bias is larger than β itself and that it can exist for very small β.

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 this design. This exercise does not provide any intuition as to why the time-stratified design is free from bias and the symmetric bidirectional design is not; this difference is a consequence of the differences in the likelihoods.

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						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

A.

B.
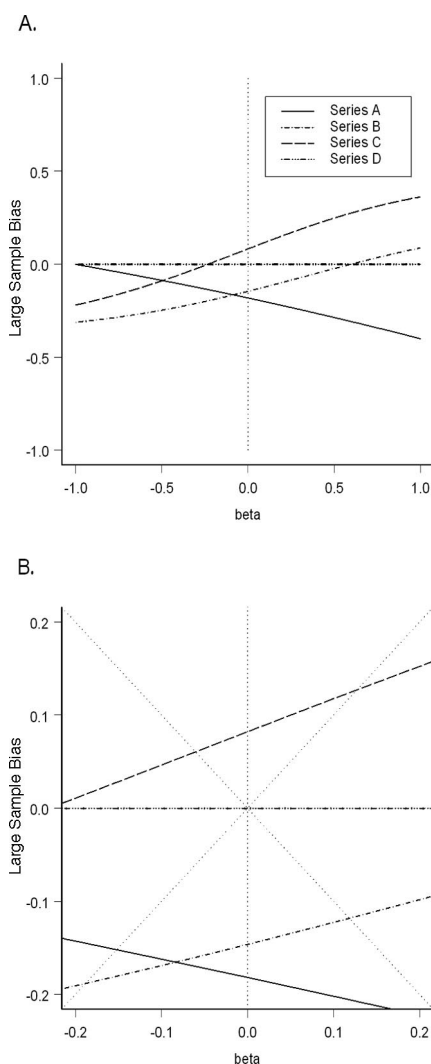
FIGURE 2. The large sample bias as a function of β for the exposure series shown in Table 3. (**A**) β between −1 and 1. (**B**) β between −0.2 and 0.2. In part B, the line y = x is superimposed so that we can observe when the bias is larger than β itself. Note that the large sample bias scale is different on the 2 plots.

The Magnitude of the Overlap Bias

Given an exposure series and a referent scheme, the magnitude of the overlap bias can be calculated, as was shown in the previous section as well as by our previous paper.[41] The bias is generally small, but it is highly unpredictable. Simulation studies examining continuous exposures have shown that bias 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 $\beta = 0$, producing apparent effects when none 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 others. For example, with the symmetric bidirectional design, cases at the beginning of the series will not have preevent referents, and cases at the end will not have postevent referents. Bateson and Schwartz[17] called this "selection bias" and suggested subtracting it off. Although this correction is sufficient when $\beta = 0$, for other β, it 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 pattern is not found for other values of β. In general, the bias will decrease as the length of the exposure series increases, but the rate of this convergence is much slower for $\beta \neq 0$.

REFERENT SELECTION EFFICIENCY CONCERNS

Like with any design, the case–crossover design involves a tradeoff between bias and efficiency. (Efficiency is usually quantified using the variance of the effect estimate.) Increasing the number of referents will improve efficiency but decrease control over confounding. Yet, in the air pollu-

tion setting, bias is generally the dominant concern as a result of the small effect sizes. Hence, we study efficiency alone, assuming confounding has already been controlled by matching. We note, however, that after controlling for confounding, there may be little choice as to the number of referents.

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 valid only for localizable, ignorable designs and not for the nonlocalizable designs studied by these authors. This is another manifestation of overlap bias: for nonlocalizable designs, both the conditional logistic regression estimating equations and variances are biased. Hence, the conclusions of these authors are not necessarily correct.

Here, we compare the efficiency of the time-stratified and full-stratum bidirectional designs (2 localizable, ignorable designs) as a function of the stratum size in the time-stratified design. The variances are given in Appendix B (available with the online version of this article). After controlling for confounding, the exposure will have no seasonality and no 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 2 designs depends on the exposure series, the stratum size (denoted by M), and β. We show in Figure 3A the relative efficiency for 3 different exposure series as a function of M, when $\beta = 0$. The relative efficiency varies across exposure series, but for these 3 exposures, it is approximately 70% when $M = 10$, 40% to 50% when $M = 5$, and 10% to 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 autocorrelated with the index exposure, and, hence, there will be less power to detect an exposure effect. This pattern is illustrated in Figure 3B. For the same 3 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 (eg, if $M = 4$, the first stratum consists of days 1, 2, 3, and 4 in Fig. 3A and days 1, 26, 51, and 76 in Fig. 3B). Indeed, we see that the efficiency of the time-stratified design is higher in Figure 3B than in Figure
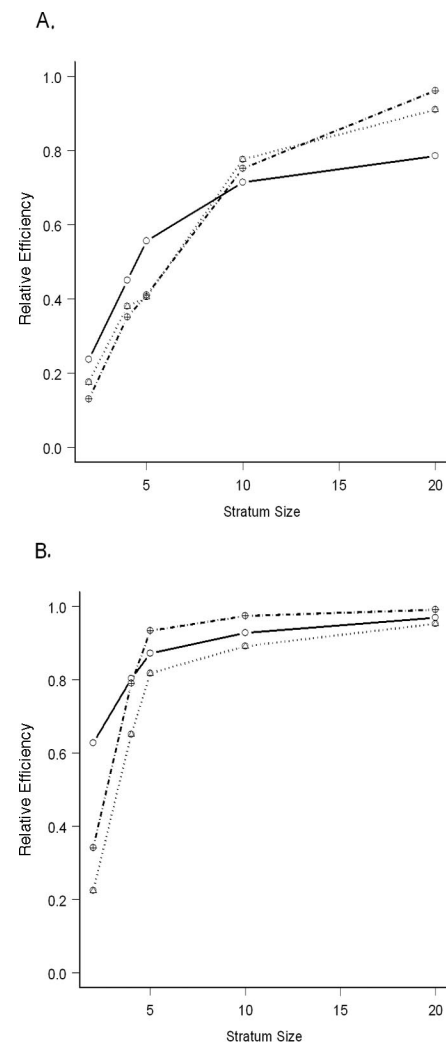

FIGURE 3. Efficiency of the time-stratified design relative to the full-stratum bidirectional design as a function of the stratum size in the time-stratified design. Relative efficiency is shown for 3 randomly chosen simulated exposure series with autocorrelation when $\beta = 0$. In part **A**, the time-stratified design uses strata as sequences of adjacent days (ie, if the stratum size is 4, the first stratum consists of days 1 to 4). In part **B**, the referents are spaced in the series (ie, if the stratum size is 4, the first stratum consists of days 1, 26, 51, 76).

3A. Thus, it is advisable to choose referents that are not adjacent to the index time (eg, 6 days apart) to maximize the information contributed by each referent.

CASE–CROSSOVER METHODS FOR UNSHARED EXPOSURE SERIES

This article has focused on the case of a fixed, shared exposure series. Hence, throughout our discussion, we have conditioned on exposure within the referent windows. For unshared exposures, it may be more appropriate not to condition on

exposure. In particular, if exposures are independent across subjects, the observed exposures can be viewed as a random sample from a given distribution. We call such exposures "random." With random exposures, it is appropriate to examine the properties of the effect estimate and estimating equations averaged over all possible exposure series (rather than conditional on exposure). Several 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 geographic regions, these series may be spatially correlated. With this type of exposure, the question of whether to condition on exposure is debatable.

With a random exposure under nonlocalizable 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, we know of no unbiased estimator with a random exposure and nonlocalizable design.

In contrast, with a localizable, ignorable referent scheme, there is no bias in the conditional logistic regression effect estimate when averaging over exposures. This logically follows because unbiasedness for a given exposure series implies unbiasedness averaged across exposure series (ie, 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 estimation procedure. The most efficient estimate would come from assuming a model for the exposure distribution and then 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 classes. With a nonlocalizable referent scheme and a shared exposure series, there is no existing 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, we know of no method for calculating unbiased effect estimates. In contrast, with a localizable, ignorable referent scheme, the conditional logistic regression estimates are unbiased regardless of whether or not we condition on exposure.

DISCUSSION AND CONCLUSIONS

The case–crossover design is well suited to the study of short-term health effects of air pollution. Time-dependent confounding, time trends, and autocorrelation in air pollution exposure series make proper referent selection 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 as a result of autocorrelation in the exposure series. If there still remains a choice as to the number of referents at this point, larger number of referents will increase efficiency. Finally, the analysis should condition on the fixed exposure series.

If a nonlocalizable (or localizable, nonignorable) referent scheme is used, conditional logistic regression yields biased estimates. This bias is usually small, although it can exist for the small effect sizes seen in air pollution studies and even when there is no exposure effect. The magnitude of the overlap bias varies with the referent scheme and the exposure series; hence, model shopping on referent strategies will tend to exacerbate the bias. (Model shopping occurs when a referent scheme is chosen because it results in larger effect estimates than other candidate referent schemes.) 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 scheme in an air pollution exposure study assume that the exposure is exogenous and the outcome rare. We advocate time-stratified referent selection, because this design avoids overlap bias and time trend bias. 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 the following variables: day of the week, temperature, measurement time, and copollutants) is adequate for most studies. Although the semisymmetric bidirectional design can also achieve these goals, this design requires modification of the traditional conditional logistic regression analysis and will be less efficient than a time-stratified design because fewer referents are used.

REFERENCES

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*. 2004;99:938–948.
3. Maclure M. The case–crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol*. 1991;133:144–153.
4. Kwon HJ, Cho SH, Nyberg F, et al. 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, 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, et al. Increased particulate air pollution and the triggering of myocardial infarction. *Circulation*. 2001; 103:2810–2815.
10. Sunyer J, Schwartz J, Tobias A, et al. 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. *Am J Epidemiol*. 1995;142:91–98.

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

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. *Int J Epidemiol*. 2003;32:1064–1070.

19. Jaakkola JJK. Case–crossover design in air pollution epidemiology. *Eur Respir J*. 2003;21(suppl):40:85s.

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

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: 528–535.

24. Kan H, Chen B. A case–crossover analysis of air pollution and daily mortality in Shanghai. *J Occup Health*. 2003;45:119–124.

25. Lin M, Chen Y, Burnett RT, et al. The influence of ambient coarse particulate matter on asthma hospitalisations in children: case–crossover and time-series analyses. *Environ Health Perspect*. 2002;110:575–581.

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. *J Epidemiol Community Health*. 2003;57:50–55.

27. Schwartz J. Is the association of airborne particles with daily deaths confounded by gaseous air pollutants? An approach to control by matching. *Environ Health Perspect*. 2004;112:557–561.

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. *Am J Epidemiol*. 2003;157: 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. *Int J Epidemiol*. 2001;30:1138–1140.

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;57:687–693.

31. Tsai SS, Huang CH, Goggins WB, et al. Relationship between air pollution and daily mortality in a tropical city: Kaohsiung, Taiwan. *J Toxicol Environ Health A*. 2003;66:1341–1349.

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;34:2612–2616.

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. *J Toxicol Environ Health 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. *Environ Int*. 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. *Inhal Toxicol*. 2003;15:1297–1308.

36. Clyde M. Bayesian model averaging and model search strategies with discussion. In: Bernado JM, Berger JO, Dawid AP, Smith AFM, 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 analyses. *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;159:596–602.

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. *Stat Med*. 2005;24: 285–300.

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. *Int J Epidemiol*. 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.