Revealing and Addressing Length Bias and Heterogeneous Effects in Frequency Case-Crossover Studies

Ravi Varadhan and Constantine E. Frangakis

From the Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Received for publication October 22, 2002; accepted for publication October 16, 2003.

The case-crossover design is useful for assessing whether a recurrent exposure (e.g., drug) triggers an event (e.g., myocardial infarction), using only cases, when finding good controls is impractical. In the basic frequency design, the observed exposure odds among cases, during a period immediately before the event, are compared with the expected exposure odds, based on their usual frequency of past exposures. This is equivalent to comparing observed gap times between the event and the last exposure with the expected gap times based on the subjects’ exposure experience under the null hypothesis of no exposure-event relation. Such a comparison reveals two problems in the usual-frequency analyses: 1) length bias that exists even under the null hypothesis; and 2) loss of efficiency when exposure effects do exist. The first problem arises because the event will more likely fall on a longer-than-average period between exposures, even under the null hypothesis, resulting in a systematic downward bias of risk ratios. The second problem arises from categorizing cases as exposed or unexposed and from not fully using the data on gap times between events and preceding exposures. A new method of analysis is presented that is free from length bias and that efficiently uses gap time data.

Abbreviation: MH, Mantel-Haenszel.

The case-crossover design of Maclure (1) has been useful for assessing whether a recurrent exposure triggers an acute event. The design is based only on cases, that is, subjects with the event, and is, therefore, useful when finding representative control subjects is not practical. In the case-crossover design, the case and control units being compared are different time periods within the subject, which, as a result, are matched on factors that remain constant within the subject over the study period, such as sex, race, or other genotypic characteristics.

Information in the original version of the design, for each case, consists of the subject’s past usual exposure experience and the gap time between the event and the last exposure before it. The task then is to assess how different the observed gap times are from those that would be expected on the basis of the usual exposure experience and if there were no true event exposure association (null hypothesis). This basic design is mostly recommended (2) and is being frequently used; for more recent applications, see, for example, these reports (3–6). Literature exists also on other versions of the design, such as relying on proportional hazards models for inference (7), using only short periods of potential exposure before the event (8, 9), or including periods of exposure after the event (10). Time trends in exposure have also been discussed by several author groups (11–15).

In this paper, we reveal and address two problems that are more generally present in the various versions of the case-crossover design. These problems exist even without time trends (stationarity of exposure experience); they can invalidate testing even when there is truly no exposure-event relation, and they are not specific to any single particular model such as the proportional hazards assumption. Moreover, to our knowledge, these two problems have not yet been appropriately addressed. To better focus on the main ideas, we present these problems in the original design for stationary exposure experience.

These problems are as follows: first, length bias even when the null hypothesis of no exposure-event relation is
true; and second, low precision when such a relation exists. In particular, usual-frequency analyses discretize time into blocks of exposed and unexposed periods and then calculate the following: 1) the exposed time periods having an event, as a fraction of the total exposed periods; 2) the unexposed periods having an event, as a fraction of the total unexposed periods; and 3) a “risk ratio,” using “1” and “2” across all subjects. This risk is then compared with unity. We show that such analyses are length biased in the sense that, under the null, the event will more likely fall on a longer-than-average period between exposures. As a result, we show that, under the null, the standard risk ratios are generally lower than 1 when some subjects’ average periods between exposures are comparable with the assumed effect period. We obtain a formula for the bias and adjust the usual-frequency analyses. The second problem arises because usual-frequency analyses do not use the full gap times between events and last exposures. We provide and demonstrate a new method of analysis that is valid under the null and also takes account efficiently of the full information on the gap times.

MATERIALS AND METHODS

Data and formulation using stochastic processes

We wish to study if a systematically recurrent exposure (e.g., drug injection) is associated with the timing of an adverse event (e.g., myocardial infarction). For comparability, we consider the type of information and assumption of no time trends analogous to the original frequency design of Maclure (1). To better demonstrate our arguments, however, it is important that we first formulate these data and conditions in a more fundamental framework of stochastic processes.

Consider a sample of cases, \(i = 1, \ldots, n\), who have the event, and assume that they are a random sample from the population of cases to which we wish to generalize. Assume that each case is experiencing the exposure in a systematically recurrent way; the times between successive exposures are called the exposure’s interarrival times and are denoted by \(T\) (figure 1). To reflect the systematic pattern in the interarrivals within a person, when appropriate, assume that the successive interarrivals are independent and identically distributed samples from a cumulative distribution that is specific to the subject, denoted by \(F_i = \Pr(T \leq t \mid \text{subject } i)\).

The data in the simple (frequency) approach include a summary of \(F_i\), \(i = 1, \ldots, n\), that characterizes the periodicity of the exposure and that is assumed known, although the arguments can be extended to the situation where \(F_i\) is estimated from preceding interarrivals within each person (see Discussion). Specifically, the data here are assumed to be 1) the average interarrival time of exposure, for example, \(\mu_i\), where \(\mu_i = \mathbb{E}(T \mid F_i)\), and 2) the gap time, \(G_i\), between the subject’s event and the last exposure before the event (figure 1).

The general analytical task with these data is to compare the gap times \(G_i\) that are observed with those that would be expected under the null hypothesis, \(H_0\), of no association between exposure and timing of the event. Next, we review briefly the usual-frequency analysis for this comparison.

Review of usual-frequency analyses

The usual-frequency analysis tries to quantify the association between exposure and event by estimating a “rate ratio” calculated as follows.

First, a fixed period \(d\), for instance, 1 hour, is chosen to reflect a time window after exposure and during which exposure can affect the event, if such an effect exists. This period is called the “assumed effect period.” Then, a long-time period of length \(L\) (e.g., a year), counting backwards from the event of each subject, is binned into “exposed periods” and “nonexposed” periods, each of length \(d\). Exposed periods are defined to be those that start at an exposure and last for one effect period of length \(d\), and nonexposed periods are the remaining periods of length \(d\) (figure 1).
TABLE 1. Classification of subject i’s periods by exposure and event status as used in the usual-frequency analyses

<table>
<thead>
<tr>
<th></th>
<th>With event</th>
<th>Without event</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>$X_i$</td>
<td>$L/d - L/\mu_i$</td>
<td>$L/\mu_i$</td>
</tr>
<tr>
<td>Unexposed</td>
<td>$1 - X_i$</td>
<td>$L/d - L/\mu_i - 1 + X_i$</td>
<td>$L/d - L/\mu_i$</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>$L/d - 1$</td>
<td>$L/d$</td>
</tr>
</tbody>
</table>

The event for a subject i is classified to occur either during an exposed period or not, which is represented by the variable $X_i$, as equation 1:

$$X_i = \begin{cases} 
1 & \text{if } G_i \leq d \\
0 & \text{otherwise} 
\end{cases} .$$

The two-way classification that results from categorizing a subject’s periods based on exposure and event status and that is used in usual-frequency analyses is given in Table 1.

Finally, the usual-frequency analysis estimates the increased event risk associated with exposure (over all subjects) using the Mantel-Haenszel (MH) estimator according to Maclure (1):

$$MH^{(o)} = \frac{\sum_{i=1}^{n} X_i \left( \frac{L}{d} \frac{L}{\mu_i} \right)}{\sum_{i=1}^{n} (1 - X_i) \left( \frac{L}{d} \frac{L}{\mu_i} \right)} = \frac{n^{-1} \sum_{i=1}^{n} X_i \left( \frac{1}{d} \frac{1}{\mu_i} \right)}{n^{-1} \sum_{i=1}^{n} (1 - X_i) \left( \frac{1}{d} \frac{1}{\mu_i} \right)} .$$

where $L/\mu_i$ is the number of exposed times for subject i, and $L/d$ is the maximum number of times that subject could have been exposed. Use of this estimator implicitly defines the target of estimation to be the ratio, for example, MH, to which MH$^{(o)}$ would converge with an increasing number of subjects from the case population. It is important, therefore, to gain insights into MH by interpreting it as a ratio of two fractions: the exposed time periods having an event (as a fraction of the total exposed periods) divided by the unexposed periods having an event (as a fraction of the total unexposed periods), combined across all the subjects with the usual Mantel-Haenszel weights. MH is then compared with 1 to assess the exposure-event association. For example, consider a hypothetical large study in which all subjects have a common period of $\mu_i = 8$ hours as the average interarrival between injections of a drug, and where the effect period $d$ for myocardial infarction is taken to be 5 hours. Then, the MH ratio will be 1 when and only when, we observe that five of eight subjects have the myocardial infarction event in an exposed period, and values of MH that are larger (smaller) than 1 are interpreted by the usual-frequency analysis to indicate a positive (negative) exposure-event association. More generally, the usual-frequency analyses take the value of 1 for MH to be the reference value indicating no exposure-event association and around which it judges other degrees of association.

Such analyses generally have two related problems. First, the classification of the gap times $G_i$ based on a common effect period cannot address possible heterogeneous effect periods across subjects and so loses power for tests and efficiency for estimation in situations when exposure does affect the event. Consider, for example, subject 2 of Maclure’s data on sexual activity (1). That subject has a frequency of exposure of 2/week, and the gap time between last exposure and the event is 90 minutes. The ratio of that gap time to the period between exposures, therefore, is 0.018, so the data from that person alone suggest that exposure has affected that person’s event. This individual-level degree of information is lost, however, in the usual-frequency analyses where the person’s gap time is simply classified as either “having” or “not having” the event in an exposed period, and this problem exists regardless of how the common bin length $d$ of the effect period is chosen for making that classification. To address this problem, one should use the full gap times $G_i$ in an analysis that compares them with the usual periods of exposure. This task brings up a more fundamental problem, which is to understand the reference distribution of the gap times $G_i$ in the case-crossover design if the null hypothesis, $H_0$, of no exposure-event association is true. For this reason, in the next section we discuss first the more fundamental properties that $H_0$ induces on the gap times $G_i$ and demonstrate them on the MH estimator. The results of that section are then used to better address the heterogeneous effects of exposure on the timing of the event.

**Revealing and addressing the length bias of gap times in the case-crossover design**

In this section, we assume that the null hypothesis $H_0$ is true. By definition, under $H_0$, for a subject, the event is a random occurrence along the time pattern of exposures. It follows that the event is more likely to be found in an interarrival between exposures, for example, $T_e$ that is longer than the average interarrival $\mu_i$ (Figure 1). Therefore, the case-crossover design exhibits a phenomenon that is known in other applications in epidemiology and sampling as “waiting-time bias” or “length bias” (16–18).

In particular, when subject i’s timing of the event is unrelated to the subject’s process of exposures, and using standard results of renewal process theory (19), the distribution of the gap times $G_i$ between the event and the last exposure is related to the interarrival distribution $F_e$:

$$ \text{pr}(G_i \leq t | F_e) = \frac{1}{\mu_i} \int_0^t (1 - F_e(u))du ,$$

for times $t > 0$. The distribution function 3 is the reference distribution against which the gap times should be compared at a subject-specific level under $H_0$. This means that no matter what method is used to model the effects of exposure on the event, it should be such that its properties under $H_0$ reduce to those determined by equation 3.

To demonstrate this argument, note that equation 3 also determines the properties of the MH estimator under the null $H_0$. In particular, from equations 3 and 1, we find that the probability that the subject’s event will fall in an “exposed” period of length $d$ is...
\[ \Pr (X_i = 1 | F_i) = \Pr (G_i \leq d | F_i) = \frac{d}{\mu_{\ell}} (1 - \alpha_i^2) \leq \frac{d}{\mu_{\ell}}, \quad (4) \]

where

\[ \alpha_i^2 = \frac{1}{d} \int_0^d F_i(u) du. \]

Using \( E(X_i | F_i) = \Pr (X_i = 1 | F_i) \) and the weak law of large numbers, we find, after some algebra, that the estimator \( \text{MH}^{(\text{null})} \) will, in large samples, converge to \( \text{MH}^{(\text{null})} \), where

\[ \text{MH}^{(\text{null})} = \frac{E(k_i)}{E(k_i) + E\left( \frac{\alpha_i^2}{\mu_{\ell}} \right)}, \quad (5) \]

where

\[ k_i = \frac{(1 - \alpha_i^2)}{\mu_{\ell}} \left( 1 - \frac{d}{\mu_{\ell}} \right), \]

and where the expectation, \( E(\cdot) \), in equation 5 denotes the average over all subjects in the case population.

Generally, to study transient effects, the effect period \( d \) is chosen to be smaller than all subject-specific average interarrivals, \( \mu_{\ell} \). It is evident, then, from equation 5, that \( \text{MH}^{(\text{null})} \) is smaller that 1 whenever some \( \alpha_i^2 > 0 \), that is, whenever some actual interarrival times are smaller than the assumed effect period \( d \). As noted by a referee, Maclure (1) deals with length bias by originally assuming in table 4 therein that some actual interarrival times are smaller than the assumed effect period \( d \). Therefore, in equation 3, such a subject we indicate by \( A_i = 1 \), and we do not directly observe the indicator \( A_i \), which is 0 in the special case when exposure does not affect the event. For stable estimation in small samples, it is preferable that \( A_i \) be set to a fixed distribution, the variance of which can encompass a reasonably wide range of actual effect periods. The approach proposed here, while requiring increased effort in terms of estimating interarrival distribution, is valid (consistent) and more efficient. Estimation of \( \text{MH}^{(\text{null})} \) requires estimation of \( \alpha_i^2 \) for each subject, and this requires 1) assumptions on the distribution of \( F_i \) and/or 2) estimation of \( F_i \) from actual interarrivals of exposure within the subject. Then, the ratio \( \text{MH}^{(\text{adj})} = \text{MH}^{(\text{null})}/ \text{MH}^{(\text{null})} \) measures the relative risk between exposed and unexposed periods of having the event that is in excess of that relative risk that is induced by length bias. We demonstrate these arguments numerically in a later section.

**Using mixture models to increase efficiency**

We can now use the reference distribution \( \pi_{\text{aff}} \) for the gap times to formulate a model that allows different subjects to be possibly susceptible to different effect periods as follows.

Each subject \( i \)’s event-exposure gap time \( G_i \) may or may not have been affected by the exposure. We assume that \( G_i \) was not affected by exposure if, conditionally on that subject’s usual exposure experience as characterized by \( F_i \), the subject’s gap time \( G_i \) was a random draw from the null cumulative distribution function of \( G_i \),

\[ H_i^{\text{null}}(t) = (\mu_{\ell})^{-1} \int_0^t (1 - F_i(u)) du, \]

from equation 3; such a subject we indicate by \( A_i = 0 \). Otherwise, we assume that the subject’s gap time was affected by exposure, and, in this case, we let \( G_i \) be a draw from some other cumulative distribution, for instance, \( H_{aff} \), such a subject we indicate by \( A_i = 1 \) (table 2).

We do not directly observe the indicator \( A_i \) of whether the subject has or has not been affected, so, when a subject has null cumulative distribution \( H_{i}^{\text{null}} \) for \( G_i \), then the observed gap time \( G_i \) for that subject has cumulative distribution, for example, \( H_{i}^{\text{obs}} \), given by equation 4:

\[ H_{i}^{\text{obs}}(t) = \pi_{\text{aff}} H_{aff}(t) + (1 - \pi_{\text{aff}}) H_{i}^{\text{null}}(t), \quad (6) \]

where \( \pi_{\text{aff}} \) is the case population’s fraction of subjects whose gap time is affected by exposure, which is assumed to be common across \( F_i \).

Estimation in the above model here focuses on the fraction \( \pi_{\text{aff}} \), which is 0 in the special case when exposure does not affect the event. For stable estimation in small samples, it is preferable that \( H_{aff} \) be set to a fixed distribution, the variability of which can encompass a reasonably wide range of actual effect periods. Model 6 then explicitly allows that 1) any affected subjects were affected by possibly different

---

**TABLE 2. Components of mixture model for subjects who are affected and not affected by exposure in the case-crossover design**

<table>
<thead>
<tr>
<th>Subject’s event is</th>
<th>Subject’s gap time ( G_i ) is from</th>
<th>Cumulative distribution of ( G_i )</th>
<th>Proportion of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not affected ( (A_i = 0) )</td>
<td>Null</td>
<td>( H_i^{null} )</td>
<td>1 – ( \pi_{\text{aff}} )</td>
</tr>
<tr>
<td>Affected ( (A_i = 1) )</td>
<td>Nonnull</td>
<td>( H_i^{aff} )</td>
<td>( \pi_{\text{aff}} )</td>
</tr>
</tbody>
</table>
actual effect periods and 2) the null model 3 is true (when πaff = 0), so tests based on model 6 are valid. In larger samples, Haff can sometimes also be estimated and be allowed to depend on covariates (see Discussion). For identifiability of the above model, it is best that the chosen Haff be centered at a different time than the average of the centers of Hnull.

The above model is useful for inference, such as constructing confidence intervals for πaff and testing the null πaff = 0. For testing, when πaff is expected to be relatively small, the score test using the full data (full score test) gives largest power. Analogously, when we keep only the binned gap times based on the assumed effect period d, which we call the “reduced data,” the corresponding reduced score test gives largest power among all tests that use only the reduced data. Generally, a full score test is more efficient than a reduced score test in the sense that it can achieve the same power as that of the reduced score test using only a fraction of the reduced score test’s number of subjects (20). The gain in efficiency, then, can be expressed as the percent saving in sample size, when designing the study, between the use of the full score test and the use of the reduced score test for testing, and it is calculated as 1 – the ratio of the variances of sample size, when designing the study, between the use of the full score test’s number of subjects (20). The gain in efficiency, then, can be expressed as the percent saving in sample size, when designing the study, between the use of the full score test and the use of the reduced score test for testing, and it is calculated as 1 – the ratio of the variances of the reduced score test statistic versus the full score test statistic. For the examples of Maclure (1), we report the gain in efficiency of the full score test versus the reduced score test in the next section.

To obtain 95 percent confidence intervals for πaff, we first find its maximum likelihood estimator, ˆπaff. Because the null value πaff = 0 is at the boundary of the allowed values of πaff, the usual theory for constructing confidence intervals based on the standard error and using a normal approximation to the distribution of ˆπaff, or of a transformation of it, is not applicable here. For this reason, we use the approach described by Frangakis and Varadhan (21). With this approach, we find the small sample distribution of the maximum likelihood estimator, ˆπaff (using simulation), as a function of the different possible true values of πaff. Then, by inverting that relation, we obtain a 95 percent confidence interval for πaff that has, to any desired approximation, 95 percent coverage for the true fraction πaff, whether or not that fraction is on the boundary. For more details of this method, given in terms of a different model and data, see the article by Frangakis and Varadhan (21). Using this method, we report confidence intervals for πaff in examples in the next section.

DEMONSTRATION

We use the original data of Maclure (1) to demonstrate our methods for using the full gap times. As we saw, this requires assumptions about the stationary distribution, Fn, of interarrivals between exposures within each subject i and/or estimation of Fi from actual interarrivals of exposure within the subject. Because the data of Maclure (1) provide only the periods ˆµi, for demonstration we assume here that Fi’s are exponential with means ˆµi. In addition, because the proposed methods require exact data, in the sexual activity example we omit subject 10, because the reports of 0 frequency and of having an exposure are inconsistent; and in the coffee-drinking example we set the gap time of subject 9, originally reported as “less than 1 hour,” to 0.99 hour. Programs for the calculations given here are available from the authors.

Under the exponential model for Fi, the distribution of the gap time is the same exponential, Hnull(τ) = Fi(τ). Moreover, when a bin of length d is used as the assumed effect period, the MHnull, at which the standard risk ratio is actually centered under the null, can be approximated by replacing the expectations in expression 5 with the sample averages:

\[ MH\text{(null)} = \frac{\sum_{i=1}^{n} (1 - d / \hat{\mu}_i) (1 - e^{-d / \hat{\mu}_i})}{\sum_{i=1}^{n} (d / \hat{\mu}_i) e^{-d / \hat{\mu}_i}}. \]  \tag{7}

It can be seen from equation 7 that the null risk ratio depends only on the parameters d/µ, when the interarrivall distribution is exponential. The smaller the d/µ are, the closer the null risk ratio is to 1, and the larger the d/µ are, the closer the null risk ratio is to 0 (note that d < µ, for all subjects).

Table 3 gives MHnull (null risk ratio), along with the original MH(n) estimate (standard risk ratio) and the adjusted risk ratio MH(n)/MH(null). As the table demonstrates, for sexual activity the bin length as a percentage of the median of periods ˆµ, is negligible, and therefore so is the length bias. For coffee drinking, the three smallest average interarrivals ˆµ were 2.4, 3, and 4.8 hours, and the bin length as a percentage of the median of periods ˆµ is 8–16 percent. Then the null value MHnull is estimated by the exponential model to be 20–40 percent smaller than the null value of 1 assumed by the usual-frequency analysis. In practical terms, the bias arises because the model predicts that some actual interarrivals for some subjects are smaller than the effect period if the latter is assumed to be 1 or 2 hours. This prediction can be tested using the data on previous actual interarrivals other than the last ones that have the events. Note, however, that it

<table>
<thead>
<tr>
<th>Example*</th>
<th>Length used if gap times are binned as “exposed” or “not exposed”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bin of 1 hour</td>
<td>Bin of 2 hours</td>
</tr>
<tr>
<td>Sexual activity</td>
<td></td>
</tr>
<tr>
<td>Bin/median, ˆµi (%)</td>
<td>0</td>
</tr>
<tr>
<td>Standard risk ratio, MHnull</td>
<td>22.2</td>
</tr>
<tr>
<td>Null risk ratio, MHnull</td>
<td>1.0</td>
</tr>
<tr>
<td>Adjusted risk ratio, MHnull/MHnull</td>
<td>22.2</td>
</tr>
<tr>
<td>Coffee drinking</td>
<td></td>
</tr>
<tr>
<td>Bin/median, ˆµi (%)</td>
<td>8</td>
</tr>
<tr>
<td>Standard risk ratio, MHnull</td>
<td>1.8</td>
</tr>
<tr>
<td>Null risk ratio, MHnull</td>
<td>0.8</td>
</tr>
<tr>
<td>Adjusted risk ratio, MHnull/MHnull</td>
<td>2.2</td>
</tr>
</tbody>
</table>

* The examples use data from Maclure (1) as described in the text.
† MH, Mantel-Haenszel.
TABLE 4. Percent gain in efficiency (columns 3 and 4) when using the full gap times with model 6 versus when using binned gap times, as a function of the average, $\mu^\text{aff}$, of the true effect periods and of the bin length used when binning the gap times

<table>
<thead>
<tr>
<th>Example*</th>
<th>$\mu^\text{aff}$ (hours)</th>
<th>Length used to bin gap times as &quot;exposed&quot; or &quot;not exposed&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bin of 1 hour (%)</td>
</tr>
<tr>
<td>Sexual activity</td>
<td>0.5</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>38</td>
</tr>
<tr>
<td>Coffee drinking</td>
<td>0.5</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>39</td>
</tr>
</tbody>
</table>

* The examples use data from Maclure (1) as described in the text.

is generally not appropriate to remove from the analyses such subjects for which that prediction is actually true, because efficiency would be decreased and because the effect can be different between those subjects and the remaining subjects.

For these examples, we also used model 6 to demonstrate calculations of the 95 percent confidence interval for the proportion affected and for the efficiency gained when testing the null hypothesis when using the full versus binned gap times. Assuming $H^\text{aff}$ to be exponentially distributed with an average effect period (if affected) $\mu^\text{aff} = 1$ hour, and using the procedure outlined in Materials and Methods, we estimate that in the case population the myocardial infarction event is associated with coffee drinking for $\pi^\text{aff} = 20$ percent of the subjects (95 percent confidence interval: 0, 59; $p = 0.36$) and with sexual activity also for $\pi^\text{aff} = 20$ percent of the subjects (95 percent CI: 3, 53; $p = 0.01$).

In table 4, we calculate the efficiency gained when using the full gap times with model 6 versus when using binned gap times. We take $H^\text{aff}$ to be the exponential distribution with mean $\mu^\text{aff}$ set to three different effect periods, 0.5, 1, and 2 hours. The results show gains of 20–52 percent in efficiency. Moreover, the gains in both examples, coffee and sexual activity, are approximately the same function of the ratio of the true average effect period $\mu^\text{aff}$ to the length $d$ used when binning the gap times. This suggests that the relative efficiency between the two methods does not depend much on the interarrival distribution. Note, however, that even when the true average effect period $\mu^\text{aff}$ equals the length used for binning the gap times $G_t$, there is an efficiency gain of 20 percent when using the full $G_t$ versus the binned $G_t$ into $X_r$. This is because the method that fully uses data on gap times better capitalizes on the available subject-specific information about possible effects than the usual-frequency analysis.

**DISCUSSION**

We demonstrated that the event-exposure gap times in the frequency case-crossover design are length biased if interar-
allow for cofactor effects by simultaneously modeling other exposures, as done in existing approaches (8, 10). Second, as stated earlier, model 6 can be used to estimate the distribution of the gap time conditionally on affected individuals, even though affected status \( \pi_{\text{aff}} \) is not directly observed. Third, the model can allow for covariates in both the proportion of affected subjects \( \pi_{\text{aff}} \), for example, using a logistic regression, and the distribution of affected subjects \( H_{\text{aff}} \). Such joint modeling, not explored by the usual-frequency analyses, can clarify the effect process, because it distinguishes between a relation of the covariate in \( \pi_{\text{aff}} \), which informs about the likelihood of being affected when exposed, and the relation of the covariate in \( H_{\text{aff}} \), which informs about the duration of the effect when affected. In such mixture modeling, the expectation-maximization algorithm (23) can facilitate estimation by maximum likelihood.

We recommend our approach when there are accurate data on gap times and on past exposure experience, but not when such data are prone to large measurement error. We hope that, with increasing capability to accurately record processes in real time, for example, with modern medical monitoring devices, our approach will contribute to both addressing length bias when needed and improving the understanding of mechanisms in transient effects when using the case-crossover design.

ACKNOWLEDGMENTS

This study is supported in part by grant EY-014314-01 from the National Institutes of Health and in part by the Military Office of the Presidency of the Hellenic Republic. The authors thank Drs. Eleni Petridou, Dimitrios Trichopoulos, and Mei-Cheng Wang for valuable comments.

REFERENCES