

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

Ravi Varadhan¹ and Constantine E. Frangakis¹.

¹ Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205.

Address for correspondence: rvaradha@jhsph.edu.

Keywords: bias (epidemiology); cross-over studies; epidemiologic methods; epidemiologic research design; length-biased distribution; mixture models; recurrent events.

Abstract. 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, is 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 to the expected gap times based on the subjects' exposure experience under the null hypothesis of no exposure-event relationship. Such a comparison reveals two problems in the usual-frequency analyses: (i) length-bias that exists even under the null hypothesis; and (ii) 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 due to categorizing cases as exposed or unexposed, and 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.

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, i.e., 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 subject, which, as a result, are matched on factors that remain constant within 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 based on 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, (3)- (6). Literature exists also on other versions of the design, such as relying on proportional hazards models for inference (e.g., (7)), or using only short periods of potential exposure before the event (e.g., (8), (9)), or including periods of exposure after the event (e.g., (10)). Time trends in exposure have also been discussed, e.g., by (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: (i) length-bias even when the null hypothesis of no exposure-event relation is true, and (ii) low precision when such a relation exists. In particular, usual-frequency

analyses discretize time into blocks of exposed and unexposed periods, and then calculate: (a) the exposed time periods having an event, as a fraction of total exposed periods; (b) the unexposed periods having an event, as a fraction of total unexposed periods; and (c) a “risk ratio”, using (a) and (b) across all subjects. This risk is then compared to 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 exposure are comparable to the 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 Maclure’s (1) original frequency design. In order 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 case subjects, $i = 1, \dots, n$, who have the event, and assume they are a random sample from the population of cases to which we wish to generalize. Assume each subject 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 $F_i = \text{pr}(T \leq t \mid \text{subject } i)$. Periods during which the systematic pattern is interrupted, e.g., by sleep, do not count (2). For extensions, see the discussion.

Figure 1 here

The data in the simple (frequency) approach include a summary of $F_i, i = 1, \dots, n$, 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: (i) the average interarrival time of exposure, say μ_i , where $\mu_i = E(T \mid F_i)$; and (ii) the gap time, G_i , between the subject's event and the last exposure before the event (Figure 1).

The general analytic task with these data is to compare the gap times G_i that are observed to 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 , say of 1 hour, is chosen to reflect a time window following 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 “non-exposed”

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 non-exposed periods are the remaining periods of length d (see Fig. 1).

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:

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

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.

Table 1 here.

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

$$\text{MH}^{(n)} = \frac{\sum_{i=1}^n X_i \left(\frac{L}{d} - \frac{L}{\mu_i} \right)}{\sum_{i=1}^n (1 - X_i) \frac{L}{\mu_i}} = \frac{n^{-1} \sum_{i=1}^n X_i \left(\frac{1}{d} - \frac{1}{\mu_i} \right)}{n^{-1} \sum_{i=1}^n (1 - X_i) \frac{1}{\mu_i}}, \quad (2)$$

where L/μ_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, say, MH, to which $\text{MH}^{(n)}$ would converge with increasing number of subjects from the case population. It is important, therefore, to gain insight into MH by interpreting it as a ratio of two fractions: the exposed time periods having an event (as a fraction of total exposed periods), over the unexposed periods having an event (as a fraction of total unexposed periods), combined across all subjects with the usual Mantel-Haenszel weights. MH is then compared to 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 MI is taken to be 5 hours. Then, the MH ratio will be 1 when, and only when, we observe that 5 out of 8 subjects have the MI 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 no. 2 of Maclure's (1) data on sexual activity. 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, the full gap times G_i should be used in an analysis that compares them to 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 heterogeneous effects of exposure on the timing of the event.

Revealing and addressing 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, say T_e , that is longer than the average interarrival μ_i . 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” (e.g., (16)-(18)).

In particular, when subject i 's timing of 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_i as:

$$\text{pr}(G_i \leq t \mid F_i) = \frac{1}{\mu_i} \int_0^t (1 - F_i(u)) du, \quad (3)$$

for all 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 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 expressions 3 and 1, we find that the probability that the subject's event will fall in an “exposed” period of length d is

$$\text{pr}(X_i = 1 \mid F_i) = \text{pr}(G_i \leq d \mid F_i) = \frac{d}{\mu_i} (1 - \alpha_i^2) \leq \frac{d}{\mu_i}, \quad (4)$$

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

Using $E(X_i \mid F_i) = \text{pr}(X_i = 1 \mid F_i)$ and the weak law of large numbers, we obtain, after some

algebra, that the estimator $\text{MH}^{(n)}$ will, in large samples, converge to $\text{MH}^{(\text{null})}$ where

$$\text{MH}^{(\text{null})} = \frac{E(k_i)}{E(k_i) + E\left(\frac{\alpha_i^2}{\mu_i}\right)}, \quad \text{where} \quad k_i = \frac{(1 - \alpha_i^2)}{\mu_i} \left(1 - \frac{d}{\mu_i}\right), \quad (5)$$

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, μ_i . It is evident, then, from equation 5, that $\text{MH}^{(\text{null})}$ is smaller than 1 whenever some $\alpha_i^2 > 0$, i.e., whenever some actual interarrival times are smaller than the assumed effect period d . As noted by a referee, Maclure deals with length bias by originally assuming that such cases are not possible (Maclure, 1991, caption of Table 4). However, when, more generally, such cases occur then, if there is truly no exposure-event relation, the usual-frequency analyses will incorrectly indicate an inverse association, i.e., a protective effect of exposure. By the same argument, whenever there is an exposure-event relation, the usual-frequency analysis will tend to pull that relation towards the negative association present due to length-bias. In short, the usual-frequency analysis will tend to underestimate positive risk ratios and exaggerate true protective effects. The degree of bias, i.e., how different $\text{MH}^{(\text{null})}$ is from 1, depends on the relative magnitude of the quantities $E(\alpha_i^2/\mu_i)$ and $E(k_i)$, and thus on the relative mass of the distribution F_i of interarrivals that lies in $(0, d)$. Therefore, in practice, the bias is expected to be large for exposures with relatively short interarrivals compared to the assumed effect periods, such as with injection of illegal drugs, and expected to be small for exposures with longer interarrivals, such as sexual activity.

In such cases where length-bias is a concern, equation 5 also provides a way to adjust the standard MH ratio for that bias, by estimating $\text{MH}^{(\text{null})}$ from the data. As suggested by an anonymous referee, there can also be other approaches to adjust for the length-bias in the MH

estimator, based on accounting for the overlap of exposure interarrivals (overlap occurs when the interarrival time is less than the postulated effect period). However, such approaches do not address the more important issue of loss of efficiency resulting from ignoring the variability in effect periods between subjects. 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 α_i^2 for each subject, and this requires: (a) assumptions on the distribution of F_i , and/or (b) estimation of F_i from actual interarrivals of exposure within the subject. Then, the ratio $\text{MH}^{(\text{adj})} = \text{MH}^{(n)} / \text{MH}^{(\text{null})}$ measures the relative risk between exposed and unexposed periods 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 3 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 say 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 , $H_i^{\text{null}}(t) = (\mu_i)^{-1} \int_0^t (1 - F_i(u)) du$, from equation 3; such a subject we indicate by $A_i = 0$. Otherwise, we say 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, say H^{aff} ; such a subject we indicate by $A_i = 1$ (Table 2).

Table 2 here

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^{null} for G_i , then the observed

gap time G_i for that subject has cumulative distribution, say H_i^{obs} , given by:

$$H_i^{\text{obs}}(t) = \pi^{\text{aff}} H^{\text{aff}}(t) + (1 - \pi^{\text{aff}}) H_i^{\text{null}}(t), \quad (6)$$

where π^{aff} is the case population's fraction of subjects whose gap time is affected by exposure.

Estimation in the above model here focuses on the fraction π^{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 whose variability can encompass a reasonably wide range of actual effect periods. Model 6 then explicitly allows (a) that any affected subjects were affected by possibly *different* actual effect periods, and (b) the null model 3 (when $\pi^{\text{aff}} = 0$), so tests based on model 6 are valid. In larger samples, H^{aff} can sometimes also be estimated, and allowed to depend on covariates (see discussion). For identifiability of the above model, it is best that the chosen H^{aff} be centered at a different time than the average of the centers of H_i^{null} .

The above model is useful for inference such as constructing confidence intervals (CIs) for π^{aff} and testing the null $\pi^{\text{aff}} = 0$. For testing, when π^{aff} is expected to be relatively small, the score test using the full data (full score test, FST) 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” (RST) gives largest power among all those that use only the reduced data. Generally, an FST is more efficient than an RST in the sense that it can achieve the same power as that of RST using only a fraction of RST's number of subjects (20). The gain in efficiency, then, is the percent saving in sample size, when designing the study, between using FST versus RST for testing, and is calculated as one minus the ratio of the variances of the RST statistic versus the FST statistic. For the examples of Maclure (1), we report the gain in efficiency of FST versus RST in the next section.

To obtain 95% CIs for π^{aff} , first we find its maximum likelihood estimator (MLE), say $\hat{\pi}^{\text{aff}}$. Because the null value $\pi^{\text{aff}} = 0$ is at the boundary of the allowed values of π^{aff} , the usual theory for constructing CIs based on the standard error and using a normal approximation to the distribution of $\hat{\pi}^{\text{aff}}$, or of a transformation of it, is not applicable here. For this reason, we use the approach described in (21). With this approach, we find the small sample distribution of the MLE, $\hat{\pi}^{\text{aff}}$ (using simulation), as a function of different possible true values of π^{aff} . Then, by inverting that relation, we obtain a 95% CI for π^{aff} that has, to any desired approximation, 95% coverage for the true fraction π^{aff} , 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 (21). Using this method we report CIs for π^{aff} in examples in the next section.

DEMONSTRATION

In this section, we use Maclure’s (1) original data to demonstrate our methods for using the full gap times. As we saw, this requires assumptions on the stationary distribution, F_i , of interarrivals between exposures within each subject i and/or estimation of F_i from actual interarrivals of exposure within the subject. Because Maclure’s (1) data provide only the periods μ_i , for demonstration we assume here that F_i are exponential with means μ_i . In addition, because the proposed methods require exact data, in the sexual activity example we omit subject no. 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 no. 9, originally reported as “less than 1 hour”, to 0.99 hours. Programs for the calculations given here are available from the authors.

Under the exponential model for F_i , the distribution of the gap time is the same exponential, $H_i^{\text{null}}(t) = F_i(t)$. Moreover, when a bin of length d is used as the effect period, the $\text{MH}^{(\text{null})}$, at which the standard risk ratio is actually centered under the null, can be obtained by replacing

the expectations in expression 5 with the sample averages:

$$\text{MH}^{(\text{null})} = \frac{\sum_{i=1}^n (1 - d/\mu_i) (1 - e^{-d/\mu_i})}{\sum_{i=1}^n (d/\mu_i) e^{-d/\mu_i}} \quad (1)$$

It can be seen from equation 1 that the null risk ratio depends only on the parameters d/μ_i , when the interarrival distribution is exponential. The smaller the d/μ_i are, the closer the null risk ratio is to 1, and the larger the d/μ_i are, the closer the null risk ratio is to 0 (note that $d < \mu_i$, for all subjects). Table 3 gives $\text{MH}^{(\text{null})}$ (null risk ratio), along with the original $\text{MH}^{(n)}$ estimate (standard risk ratio) and the adjusted risk ratio $\text{MH}^{(n)}/\text{MH}^{(\text{null})}$. As the table demonstrates, for sexual activity the bin length as a percent of the median of periods μ_i is negligible, and therefore so is the length-bias. For coffee drinking, the three smallest average interarrivals μ_i were 2.4, 3, and 4.8 hours, and the bin length as a percent of the median of periods μ_i is 8-16 %. Then the null value $\text{MH}^{(\text{null})}$ is estimated by the exponential model to be 20-40% 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 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 the model 6 to demonstrate calculations of the 95% CI for the proportion affected, and for the efficiency gained when testing the null hypothesis when using the full versus binned gap times. Assuming an average effect period (if affected) $\mu^{\text{aff}} = 1$ hour, and using the procedure outlined in the methods section, we estimate that in the case population the MI event is associated with coffee drinking for $\hat{\pi}^{\text{aff}} = 20\%$ of the subjects

($p = 0.36$, 95% CI: 0%, 59%), and with sexual activity also for $\hat{\pi}^{\text{aff}} = 20\%$ of the subjects ($p = 0.01$, 95% CI: 3%, 53%).

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^{aff} to be the exponential distribution with mean μ^{aff} set to three different effect periods, 1/2, 1, and 2 hours. The results show gains of 20% to 52% in efficiency. Moreover, the gains in both examples, coffee and sexual activity, are approximately the same function of the ratio of true average effect period μ^{aff} to 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 μ^{aff} equals the length used for binning the gap times G_i , there is an efficiency gain of 20% when using the full G_i versus the binned G_i into X_i . 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.

Tables 3, 4 here

DISCUSSION

We demonstrated that the event-exposure gap times in the frequency case-crossover design are length-biased if interarrivals vary in length, and that the usual-frequency analyses that ignore this can be biased and have low precision. We showed how to use the properties of the length-biased gap times in an approach that addresses these problems. To make arguments clearer, we discussed them in terms of the simple frequency version of the case-crossover design, although these same issues also arise in more general versions of the design.

For example, a modified version of the frequency design is to replace using the usual past frequency with using discrete periods in the past (or future) as control periods to compare with the case period that has the event (e.g., (8), (9)). The usual frequency design is then

a special case where the window of the control periods becomes contiguous and increases to a long period of, say, one year. Versions of this design with small control window-widths close to the case period can be more appropriate for the stationarity assumption when there are time trends in exposure for longer periods, and less prone to reporting biases, compared to the frequency elicited by the subjects (1). As also noted by a referee, when only one control period is used, length bias will not be a concern because it will be equally present in both the case and the control period. As the control window-width increases, however, length bias decreases for the control period, whereas it remains a problem for the case period and therefore still needs to be addressed. Addressing length bias in such intermediate designs is a topic for future work. Also, in such versions of the design, the distribution F_i would have to be estimated from the few reference interarrivals. Such estimation can also allow that different interarrivals be dependent within a subject, through an unobserved frailty, e.g., as in (22). To increase efficiency when using a mixture model 6, F_i could alternatively be estimated jointly with the components representing the affected subjects.

With regards to modeling exposure interarrival times using a probability distribution, it is useful to distinguish two cases, (1) assuming stationary interarrivals (i.e. absence of a temporal trend), and (2) not assuming stationary interarrivals. In case 1, the assumption of stationarity expedites the modeling of exposure interarrival times. For example, we can model the interarrival distribution using a semi-parametric approach, where we model the null interarrival distribution, $H_i^{\text{null}}(t)$, nonparametrically, and the interarrival distribution under the alternative, $H_i^{\text{aff}}(t)$, using a parametric model. This semiparametric approach, when used for testing, will have the correct α -level (Type-I error rate) for any $H_i^{\text{null}}(t)$, since any null distribution can be obtained from (6) by letting $\pi^{\text{aff}} = 0$. More importantly, this yields a more powerful procedure for detecting departures from the null, with heterogeneous effect periods. In case 2, time trends or nonstationarity in the exposure interarrival affects not only our approach, but the very prin-

principle of the case-crossover design in that the basic question of “did something *unusual* happen before the event?” needs to be more clearly formulated. Additional assumptions are needed to model such time trends in exposure and to explicitly incorporate time in the null distribution, $H_i^{\text{null}}(t)$, in expression (6).

The mixture modeling approach discussed here can incorporate a number of improvements in larger data sets, where more stable estimation is possible. For example, we can allow for co-factor effects by modeling other exposures, as done in existing approaches (e.g., (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 A_i is not directly observed. Third, the model can allow for covariates in both the proportion of affected subjects π^{aff} , for example using a logistic regression, as well as in the distribution of affected subjects H^{aff} . Such joint modeling, not explored by the usual-frequency analyses, can help better understand the effect process, because it distinguishes between a relation of the covariate in π^{aff} , which informs about the likelihood of being affected when exposed, and the relation of the covariate in H^{aff} , which informs about the duration of the effect when affected. In such mixture modeling, the EM algorithm (23) can facilitate estimation by maximum likelihood.

A more general note is that 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, e.g., 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

The authors thank Drs. Eleni Petridou, Dimitrios Trichopoulos, Mei-Cheng Wang, the Editor and three anonymous reviewers for valuable comments, and the NIH grant EY-014314-

01 and the Military Office of the Presidency of the Hellenic Republic for partial support.

References

- [1] Maclure, M. (1991). “The case-crossover design: a method for studying transient effects on the risk of acute events” *American Journal of Epidemiology* 133, 144–53.
- [2] Maclure, M, and Mittleman, M. A. (2000). “Should we use a case-crossover design ?” *Annu. Rev. Public Health* 21, 193–221.
- [3] Mittleman, M. A., Maclure, M., Nachnani, M., Sherwood, J. B., Muller, J. E. (1997). “Educational attainment, anger, and the risk of triggering myocardial infarction onset. The Determinants of Myocardial Infarction Onset Study Investigators” *Arch Intern Med* 157, 769–75.
- [4] Moller, J., Hallqvist, J., Diderichsen, F., Theorell, T., Reuterwall, C., Ahlbom, A. (1999). “Do episodes of anger trigger myocardial infarction? A case-crossover analysis in the Stockholm Heart Epidemiology Program (SHEEP)” *Psychosom Med* 66, 842–9.
- [5] Mittleman, M. A., Lewis, R. A., Maclure, M., Sherwood, J. B., Muller, J. E. (2001). “Triggering myocardial infarction by marijuana” *Circulation* 103, 2805–09.
- [6] Brugal, M. T., Barrio, G., De L. F., Regidor, E., Royuela, L., Suelves, J. M. (2002). “Factors associated with non-fatal heroin overdose: assessing the effect of frequency and route of heroin administration” *Addiction* 97, 319–27.
- [7] Marshall, R. J., and Jackson R. T. (1993). “Analysis of case-crossover designs” *Stat Med* 12, 2333–41.

- [8] Mittleman, M. A., Maclure M., Robins J. M. (1995). "Control sampling strategies for case-crossover studies: an assessment of relative efficiency" *Am J Epidemiol* 142, 91–8.
- [9] Fagot, J. P., Mockenhaupt, M., Bouwes-Bavinck, J. N., Naldi, L., Viboud, C., Roujeau, J. C.; EuroSCAR Study Group (2002). "Nevirapine and the risk of Stevens-Johnson syndrome or toxic epidermal necrolysis" *AIDS* 15, 1843–8.
- [10] Navidi, W. (1998). "Bidirectional case-crossover designs for exposures with time trends." *Biometrics* 54, 596–605.
- [11] Greenland, S. (1996). "Confounding and exposure trends in case-crossover and case-time-control designs" *Epidemiology* 7, 231–9.
- [12] Bateson, T. F. and Schwartz, J. (1999). "Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures." *Epidemiology* 10, 539–44.
- [13] Lumley, T., and Levy, D. (2000). Bias in the case-crossover design: implications for studies of air pollution. "Environmetrics" 11, 705–17.
- [14] Levy, D., Lumley, T., Sheppard, L., Kaufman, J., Checkoway, H. (2001). "Referent selection in case-crossover analyses of acute health effects of air pollution" *Epidemiology*, 12, 186–92.
- [15] Navidi, W., and Weinhandl, E. (1998). "Risk set sampling for case-crossover designs." *Epidemiology* 13, 100–5.
- [16] Daniels, H. E. (1942). "A new technique for the analysis of fibre length distribution in wool" *J. Text. Inst* 33, 137–50.

- [17] Cox, D. R. (1969). "Some sampling problems in technology" In *New Development in Survey Sampling*, Ed. N. L. Johnson and H. Smith, Jr, pp. 506–27. New York: Wiley-Interscience.
- [18] Zelen, M., and Feinleib, M. (1969). "On the theory of screening for chronic diseases" *Biometrika*, 56, 601–14.
- [19] Grimmett, G. R., and Stirzaker, D. R. (1991). "Probability and random processes" Oxford: Clarendon press.
- [20] Cox, D. R. and Hinkley, D. V. (2000). *Theoretical Statistics*, (8th ed.) Chapman and Hall/CRC: New York.
- [21] Frangakis, C. E., and Varadhan, R. (2002). "On Confidence Intervals for Seasonal Risk of Suicides, with Null Values on the Boundary", *Epidemiology*, 13, 734-737.
- [22] Wang, M. C., and Chang, S. H. (1999). "Nonparametric estimation of a recurrent survival function", *Journal of the American Statistical Association*, 94, 146–53.
- [23] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm" *Journal of the Royal Statistical Society, Ser B* 39, 1–38.

Table 1: Classification of subject i 's periods by exposure and event status as used in the usual-frequency analyses.

	With event	Without event	Total
Exposed	X_i	$\frac{L}{\mu_i} - X_i$	$\frac{L}{\mu_i}$
Unexposed	$1 - X_i$	$\frac{L}{d} - \frac{L}{\mu_i} - 1 + X_i$	$\frac{L}{d} - \frac{L}{\mu_i}$
Total	1	$\frac{L}{d} - 1$	$\frac{L}{d}$

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

subject's event is	subject's gap time G_i is from	cumulative distribution of G_i	proportion of subjects
not affected ($A_i = 0$)	null	H_i^{null}	$1 - \pi^{\text{aff}}$
affected ($A_i = 1$)	non-null	H^{aff}	π^{aff}

Table 3: Length-bias as function of relative magnitude between median of periods μ_i and bin length used as effect period. The examples use Maclure's (1) data as described in the text.

		Length used if gap times are binned as "exposed" or "not exposed"	
		bin is 1hr	bin is 2hr
Example			
sexual activity	bin/median $\{\mu_i\}$	0%	0%
	standard risk ratio, $MH^{(n)}$	22.2	29.8
	null risk ratio, $MH^{(null)}$	1.0	1.0
	adjusted risk ratio, $\frac{MH^{(n)}}{MH^{(null)}}$	22.2	29.8
coffee drinking	bin/median $\{\mu_i\}$	8%	16%
	standard risk ratio, $MH^{(n)}$	1.8	0.6
	null risk ratio, $MH^{(null)}$	0.8	0.6
	adjusted risk ratio, $\frac{MH^{(n)}}{MH^{(null)}}$	2.2	1.0

Table 4: Percent gain in efficiency when using the full gap times with model 6 versus when using binned gap times, as function of the average, μ^{aff} , of the true effect periods, and of the bin length used when binning the gap times. The examples use Maclure's (1) data as described in the text.

Example	mix. μ^{aff}	Length used if gap times are binned as "exposed" or "not exposed"	
		bin is 1hr	bin is 2hr
sexual activity	$\frac{1}{2}$ hr	25%	52%
	1hr	20%	25%
	2hr	38%	20%
coffee drinking	$\frac{1}{2}$ hr	25%	52%
	1hr	20%	25%
	2hr	39%	21%

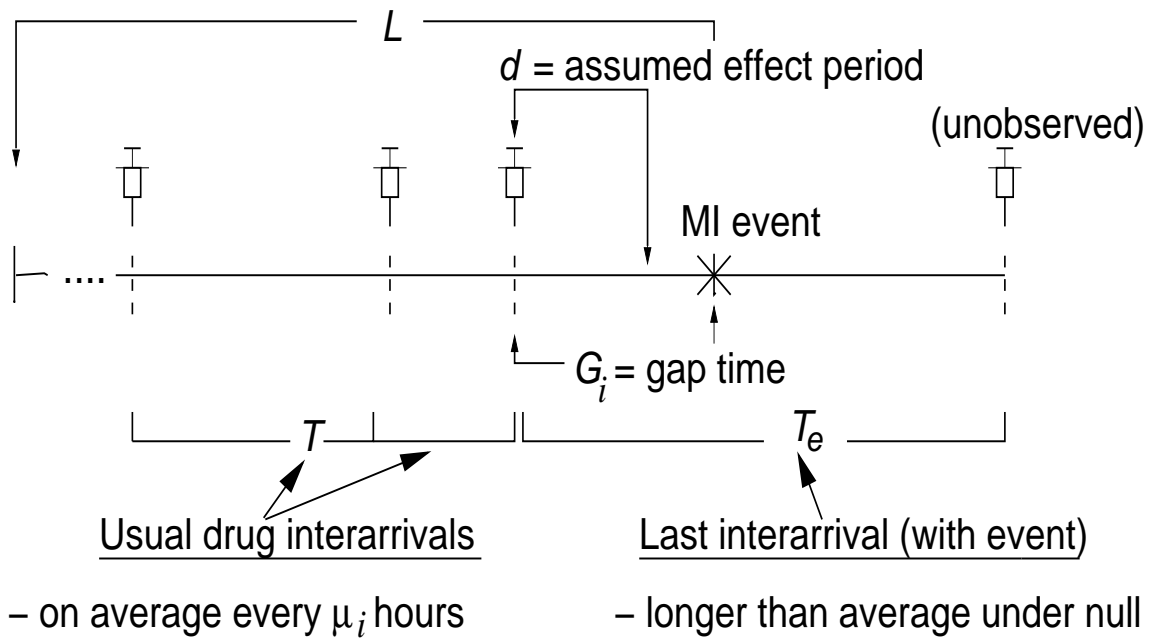


FIGURE 1: Usual interarrivals versus last (with event) interarrival of exposures (shown as syringes) for patient i in the case-crossover design. The display shows the patient having the Myocardial Infarction (MI) event in a non-exposed period (see text).