Biometrics 69, 925–936
December 2013
DOI: 10.1111/biom.12102

Bayesian Analysis of Time-Series Data under Case-Crossover Designs: Posterior Equivalence and Inference

Shi Li,[1,*] Bhramar Mukherjee,[1] Stuart Batterman,[2] Malay Ghosh[3]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, U.S.A.
[2]Department of Environmental Health Sciences, University of Michigan, Ann Arbor, Michigan, U.S.A.
[3]Department of Statistics, University of Florida, Gainesville, Florida, U.S.A.
email: shili@umich.edu

Summary. Case-crossover designs are widely used to study short-term exposure effects on the risk of acute adverse health events. While the frequentist literature on this topic is vast, there is no Bayesian work in this general area. The contribution of this paper is twofold. First, the paper establishes Bayesian equivalence results that require characterization of the set of priors under which the posterior distributions of the risk ratio parameters based on a case-crossover and time-series analysis are identical. Second, the paper studies inferential issues under case-crossover designs in a Bayesian framework. Traditionally, a conditional logistic regression is used for inference on risk-ratio parameters in case-crossover studies. We consider instead a more general full likelihood-based approach which makes less restrictive assumptions on the risk functions. Formulation of a full likelihood leads to growth in the number of parameters proportional to the sample size. We propose a semi-parametric Bayesian approach using a Dirichlet process prior to handle the random nuisance parameters that appear in a full likelihood formulation. We carry out a simulation study to compare the Bayesian methods based on full and conditional likelihood with the standard frequentist approaches for case-crossover and time-series analysis. The proposed methods are illustrated through the Detroit Asthma Morbidity, Air Quality and Traffic study, which examines the association between acute asthma risk and ambient air pollutant concentrations.

Key words: Case-crossover; Conditional likelihood; Dirichlet process; Estimating equation; Markov chain Monte Carlo; Matched case-control; Posterior inference; Time-series.

1. Introduction

Case-crossover design, originally proposed by Maclure (1991), has been widely used to study the effect of short-term exposure on the risk of acute adverse health events, such as temperature on mortality (Basu, Dominici, and Samet, 2005) and ambient air pollution on asthma (Li et al., 2011). Under this design, exposure at the event time of each case is compared to exposure at some referent times (times within a certain period where the same case did not experience any event). For each case, a "referent window" is defined as a set of time points consisting of the event time and all referent times for the same case. It can be viewed as a hybrid of case-control (comparing exposure distribution of cases and controls through a retrospective design) and crossover (the case serves as its own control) studies. The case-crossover design inherently controls for time-invariant confounders (e.g., gender and race) by making within-person comparisons, and controls for potential time-varying confounders (e.g., seasonal trends) by proper choice of the referent times. The time-stratified case-crossover design divides time a priori into disjoint strata, uses the event time to determine which stratum is selected, and selects all or a sub-sample of the remaining times in the stratum as referent times for a given case (Janes, Sheppard, and Lumley, 2005a). For example, time stratum based on the same day of the week in the same calender month that controls for confounding due to day of the week, seasonal and long-term effects is often recommended (Janes, Sheppard, and Lumley, 2005a).

The design and analytic issues related to the referent time selection have been comprehensively discussed (Lumley and Levy, 2000; Levy et al., 2001; Janes, Sheppard, and Lumley, 2005a; Janes, Sheppard, and Lumley, 2005b; Mittleman, 2005). The traditional approach for analyzing case-crossover data is to treat them as coming from a matched case-control structure, where each stratum consists of exposures at event and referent times of a given case. A conditional logistic regression (CLR) is routinely used to obtain estimates of the underlying risk ratio parameters. In terms of referent time selection, a "non-localizable" design (Janes, Sheppard, and Lumley, 2005a) is a case-crossover design for which the CLR estimating equation under the choices of referent times is biased, such as unidirectional (Maclure, 1991), bidirectional (Navidi, 1998) and symmetric bidirectional designs (SBD) (Bateson and Schwartz, 1999). The bias has been termed "overlap bias" (Lumley and Levy, 2000). In contrast, a "localizable" design (Janes, Sheppard, and Lumley, 2005a) is a case-crossover design for which there exists an unbiased CLR estimating equation, such as the time-stratified design (TSD) (Janes, Sheppard, and Lumley, 2005a) and semi-symmetric bidirectional design (Navidi and Weinhandl, 2002). Web Appendix C Figure 1 shows several illustrations of common referent time selection strategies. The TSD is generally preferred compared to any of the alternatives thus far proposed (Janes et al., 2005b; Mittleman, 2005). Based on a 2010 review article (Carracedo-Martínez et al., 2010), though 42% of case-crossover studies during

1999–2008 used SBD, the TSD has become the most popular design since 2005.

An alternative analysis of such exposure and event series data is to use a standard time-series analysis. Lu and Zeger (2007) have shown that the traditional CLR approach to analyze case-crossover data can be viewed as a time-series analysis with an underlying log-linear model of a specific form. This equivalence has also been noted in special cases by Levy et al. (2001) and (Janes, Sheppard, and Lumley, 2005a).

Bayesian data analysis under case-crossover designs appear to be non-existent in the literature though there is substantial work on Bayesian modeling of matched case-control data (Ghosh and Chen, 2002; Sinha, Mukherjee, and Ghosh, 2004). It is true that the use of CLR remains identical in the two contexts for certain "localizable" designs. However, the assumptions and the data structure make the statistical points of discussion distinct in a case-crossover study compared to a matched case-control study under a Bayesian paradigm. In this article we consider a comprehensive treatment of the problem starting with some posterior equivalence results, followed by alternative Bayesian proposals beyond using CLR as the basis for inference in case-crossover studies.

The article is structured as follows. In Section 2, we describe the disease-exposure association model, underlying assumptions and two potential likelihood formulations, the conditional and the full likelihood, under the case-crossover design. In Section 3, we then consider equivalence results analogous to Lu and Zeger (2007) in a Bayesian framework under both formulations. Bayesian equivalence results are intended to characterize the priors that ensure identical posterior inference regarding the risk ratio parameters as derived under case-crossover designs and from time-series analysis. Bayesian equivalence results for case-control studies, relating prospective and retrospective likelihoods appear in several recent papers (Seaman and Richardson, 2004; Staicu, 2010; Ghosh et al., 2012). The full likelihood formulation requires less restrictive assumptions than the conditional one, however, it involves a set of nuisance parameters corresponding to each individual or day that grows with sample size. Thus maximum likelihood estimators (MLE) of the risk ratio parameters can be potentially inconsistent. In Section 4, we present a semi-parametric Bayesian approach using a Dirichlet process prior (Ferguson, 1973; Antoniak, 1974; Müller and Quintana, 2004) to handle the random nuisance parameters in the full likelihood formulation. Section 5 presents a simulation study where we evaluate the performance of both conditional and full likelihood approaches under two common referent time selection strategies: TSD ("localizable") and SBD ("non-localizable"). We study frequentist properties such as bias and mean-squared error (MSE) of the proposed methods. Our numerical results indicate that Bayesian analysis based on the full likelihood has advantages in relaxing certain model assumptions and reducing bias, but both Bayesian and frequentist inference based on conditional likelihood are fairly robust with respect to design choices and model assumptions. Section 6 demonstrates the proposed methods through a study examining the association between acute asthma risk and ambient air pollutant concentrations. We discussed how to use information from published studies through formulation of an informative prior in the context of the example.

We would like to highlight the two fundamentally novel aspects of our study before concluding this section. The present study is the first to consider Bayesian equivalence results between case-crossover and time-series analysis. The proposal to use a full likelihood and use semi-parametric Bayes technique to make the estimation of the nuisance parameters feasible is also completely new in the case-crossover context. The numerical comparison of all proposed Bayesian methods with frequentist alternatives is an added asset of the article.

2. Case-Crossover and Time-Series: Disease Risk Model and Likelihood

2.1. *The Risk Model*

Suppose that a population of $N+M$ initially disease-free individuals are being followed forward in time. Let time point t (follow-up day, say) stand for the time interval $[t, t+1)$ throughout this article. Let Y_{it} be the binary indicator whether subject i has the disease occurring at time t ($Y_{it}=1$ if yes; $Y_{it}=0$ if no). Let N and T be the total number of cases at the end of the follow-up period and the total number of the discrete set of time points t, respectively, and let, without loss of generality, the first N of the $N+M$ individuals denote the cases. We start with the risk ratio model using similar notation as in Lu and Zeger (2007), where the risk of an event for individual i at time t is assumed as

$$P\left(Y_{it}=1 \mid \boldsymbol{X}_{it}\right)=\frac{\lambda_{0it}\exp\left(\boldsymbol{\beta}^{\top}\boldsymbol{X}_{it}\right)}{1+\lambda_{0it}\exp\left(\boldsymbol{\beta}^{\top}\boldsymbol{X}_{it}\right)}. \tag{1}$$

$\boldsymbol{X}_{it}=(X_{it1},\ldots,X_{itp})^{\top}$ is the p-dimensional exposure variable for individual i at time t, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)^{\top}$ is the common set of log risk ratio parameters. Each individual i is assumed to have his/her own baseline risk λ_{0it}.

If the risk of the disease for individual i at time t is small, it will imply the following:

ASSUMPTION 1. $P\left(Y_{it}=1 \mid \boldsymbol{X}_{it}\right)=\lambda_{0it}\exp\left(\boldsymbol{\beta}^{\top}\boldsymbol{X}_{it}\right)/\{1+\lambda_{0it}\exp\left(\boldsymbol{\beta}^{\top}\boldsymbol{X}_{it}\right)\}\approx\lambda_{0it}\exp\left(\boldsymbol{\beta}^{\top}\boldsymbol{X}_{it}\right)$.

For the case-crossover analysis, we do not require Assumption 1 and can proceed with the likelihood governed by model (1); while for the time-series analysis, Assumption 1 is required.

2.2. *Case-Crossover Design*

2.2.1. A traditional conditional likelihood approach. The log risk ratio parameter $\boldsymbol{\beta}$ in model (1) can be ascertained under a case-crossover design. We consider the situation where the diseased individuals can have multiple occurrences of events during the whole follow-up period, but the referent windows corresponding to these multiple events from the same individual must not overlap. We ignore the within-individual correlation among the multiple occurrences of events, and treat these multiple events as independent "cases." Without loss of generality, we still denote N as the number of "cases."

Let t_i and W_i be the event time and referent window for case i. The following assumption on the baseline risk is often made under a case-crossover design, which is natural if the length of W_i is short (typically a month, as in Janes, Sheppard, and Lumley, 2005a).

ASSUMPTION 2. *For each case i with event time t_i, the baseline risk λ_{0it} is constant within the referent window W_i, that is, $\lambda_{0it} = \lambda_{0it_i}$, for any $t \in W_i$, $i = 1, \ldots, N$.*

Given the exposure $\boldsymbol{X} = (\boldsymbol{X}_{it})_{N\times T}$, the referent windows W_i's and that $\sum_{t\in W_i} Y_{it} = 1$, the conditional likelihood of a case-crossover design can be derived as

$$\begin{aligned} L_{cc}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \prod_{i=1}^{N} P\left(Y_{it_i} = 1, Y_{is} = 0, \forall s \neq t_i \mid \boldsymbol{X}, W_i, \sum_{t\in W_i} Y_{it} = 1 \right) \\ &= \prod_{i=1}^{N} \frac{P\left(Y_{it_i} = 1, Y_{is} = 0, \forall s \neq t_i \mid \boldsymbol{X}, W_i\right)}{\sum_{t\in W_i} P\left(Y_{it} = 1, Y_{is} = 0, \forall s \neq t \mid \boldsymbol{X}, W_i\right)} \\ &= \prod_{i=1}^{N} \frac{\lambda_{0it_i} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{it_i}\right)}{\sum_{t\in W_i} \lambda_{0it} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{it}\right)} \\ &= \prod_{i=1}^{N} \frac{\exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{it_i}\right)}{\sum_{t\in W_i} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{it}\right)} = L_{cc}(\boldsymbol{\beta}). \end{aligned} \tag{2}$$

In (2), the second equality holds under a "localizable" design (Lumley and Levy, 2000; Janes et al., 2005b), under which an unbiased estimate of $\boldsymbol{\beta}$ can be obtained using a CLR. Under Assumption 2, the third equality holds. The nuisance parameter λ_{0it_i} was eliminated by conditioning on the sufficient statistic $\sum_{t\in W_i} Y_{it}$.

The case-crossover design is similar to a matched case-control design in the sense that the exposure at the event time of each case is compared to exposures at all referent times for the same case, that is, a matched set of exposures corresponding to each W_i. Lumley and Levy (2000) discussed the differences between the two designs, such as the dependency of exposures between and within stratum. Due to these dependencies, they showed that (2) can be treated as a conditional likelihood of a matched case-control study only under "localizable" referent window. They also showed that, with a "non-localizable" referent window such as a SBD, t_i and W_i are simple functions of each other (t_i is the mid-point of W_i), and $P\left(Y_{it} = 1, Y_{is} = 0, \forall s \neq t | \boldsymbol{X}, W_i\right) = I(t = t_i)$ is deterministic. Thus, $L_{cc}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = 1$, that is, uninformative. The estimating equation for $\boldsymbol{\beta}$ corresponding to $L_{cc}(\boldsymbol{\beta})$ is biased under "non-localizable" designs, and "overlap bias" is incurred if $\boldsymbol{\beta}$ is naively estimated using $L_{cc}(\boldsymbol{\beta})$ (Lumley and Levy, 2000).

The case-crossover design is commonly used in ecological studies concerning issues such as effect of climate change and air pollution on human health, where personal exposure is often not assessed at an individual level. For example, ambient air pollutant concentrations are usually measured from monitoring sites representing the exposure of the nearby population.

ASSUMPTION 3. *The study population has experienced shared exposure at each time t such that $\boldsymbol{X}_{it} = \boldsymbol{X}_t$, for $i = 1, ..., N+M$.*

Under Assumption 3, the conditional likelihood in (2) can be rearranged in terms of the number of events (e.g., daily mortality) at each follow-up time t (e.g., day) and expressed as

$$\begin{aligned} L_{cc}(\boldsymbol{\beta}) &= \prod_{t=1}^{T} \left\{ \prod_{i:\, Y_{it}=1} \frac{\exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{it}\right)}{\sum_{s\in W_i} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{is}\right)} \right\} \\ &= \prod_{t=1}^{T} \left\{ \frac{\exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{t}\right)}{\sum_{s\in W(t)} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{s}\right)} \right\}^{Y_t}, \end{aligned} \tag{3}$$

where $W(t)$ is the referent window containing t as the event time, $Y_t = \sum_{i=1}^{N} Y_{it}$ is the count of events at time t. Assumption 3 can be relaxed and individual exposure values can be accommodated as in our earlier formulation (1) or (2).

2.2.2. A full likelihood approach. We propose an alternative full likelihood formulation of model (1). From (1), before enforcing either Assumption 2 or 3, the full likelihood of a case-crossover design is expressed in terms of individual level exposure and baseline risk as

$$\begin{aligned} L^{NT}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \prod_{i=1}^{N} P\left(Y_{it_i} = 1, Y_{is} = 0, \forall s \neq t_i,\ s \in W_i \mid \boldsymbol{X}\right) \\ &= \prod_{i=1}^{N} \frac{\lambda_{0it_i} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{it_i}\right)}{\prod_{s\in W_i} \left\{1 + \lambda_{0is} \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_{is}\right)\right\}\}}, \end{aligned} \tag{4}$$

which allows for a completely general form of λ_{0it}. We refer to $L^{NT}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ in (4) as full likelihood with individual and day level intercepts. If the baseline risk for individual i does not change in its referent window and depends only on the event time t_i as in Assumption 2, then we can write (4) as $L^{N}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \prod_{i=1}^{N} [\lambda_{0it_i} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it_i}) / \prod_{s\in W_i} \{1 + \lambda_{0it_i} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{is})\}]$. We refer to $L^{N}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ as full likelihood with individual level intercepts. In this case, the full likelihood $L^{N}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ under a case-crossover design is exactly analogous to deriving the full likelihood of a matched case-control study under a stratified logistic regression model logit $P(Y_{is} = 1 \mid \boldsymbol{X}_{is}) = \log(\lambda_{0it_i}) + \boldsymbol{\beta}^\top \boldsymbol{X}_{is}$, $s \in W_i$, $i = 1, \ldots, N$. Under both Assumptions 2 and 3, one can alternatively translate the likelihood in terms of common nuisance parameters $\log(\lambda_{0it_i}) = \nu_{t_i}$ for all cases i that have event time t_i. Then $L^{N}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ can be aggregated as

$$L^{T}_{full}(\boldsymbol{\beta}, \boldsymbol{\nu}) = \prod_{t=1}^{T} \left[\frac{\exp\left(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_{t}\right)}{\prod_{s\in W(t)} \{1 + \exp\left(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_{s}\right)\}} \right]^{Y_t}, \tag{5}$$

where $\boldsymbol{\nu} = (\nu_1, .., \nu_T)$. We refer to $L^{T}_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ as full likelihood with day level intercepts.

2.3. *Time-series Analysis*

The log risk ratio parameter $\boldsymbol{\beta}$ in model (1) can also be estimated using an alternative time-series analysis. The expected number of events at time t can be expressed as the sum of the individual level probabilities (1) over the population. Under both Assumptions 1 and 3, $\log(E(Y_t)) = \log(\sum_{i=1}^{N+M} E(Y_{it})) = \boldsymbol{\beta}^\top \boldsymbol{X}_t + \log(\sum_{i=1}^{N+M} \lambda_{0it})$. For exposures varying at an individual level $\boldsymbol{X}_{it}$, generally one cannot aggregate the risk

$\lambda_{0it}\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it})$ as above to obtain a standard time-series structure involving the day level counts and exposures. Consider $\log(\sum_{i=1}^{N+M}\lambda_{0it})$ as a function of time only, say S_t. Then $\boldsymbol{\beta}$ can be estimated through the log-linear Poisson model

$$\log(E(Y_t)) = \boldsymbol{\beta}^\top \boldsymbol{X}_t + S_t, \tag{6}$$

where S_t is typically modeled as parametric (e.g., season) and/or non-parametric (e.g., natural spline of time) terms. The likelihood corresponding to the log-linear model in (6) is

$$L_{ll}(\boldsymbol{\beta}, S_t) \propto \prod_{t=1}^{T} \left\{ \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_t + S_t\right)\right\}^{Y_t} \exp\left\{ -\exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_t + S_t\right)\right\}. \tag{7}$$

2.3.1. Frequentist equivalence between time-series analysis and case-crossover design using conditional likelihood. The two estimating equations for $\boldsymbol{\beta}$ corresponding to (3) and (6) are $U_{cc}(\beta) = \sum_{t=1}^{T} X_t\{Y_t - \exp(\beta^\top X_t)\sum_{s\in W(t)} Y_s / \sum_{r\in W(s)} \exp(\beta^\top X_r)\}$ and $U_{ll}(\beta) = \sum_{t=1}^{T} X_t\{Y_t - \exp(\beta^\top X_t + S_t)\}$, respectively. By comparing $U_{cc}(\beta)$ and $U_{ll}(\beta)$, Lu and Zeger (2007) showed that, for a certain choice of window $W(t)$ in (3) of a "localizable" design, there exists a choice of S_t in log-linear model (6) such that the two estimating equations provide the same estimate of $\boldsymbol{\beta}$. For example, for a TSD with $W(t)$ representing the time stratum containing time t, if $\widehat{S}_t(\boldsymbol{\beta}) = \log[\{\sum_{s\in W(t)} Y_s\}/\{\sum_{s\in W(t)} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_s)\}]$, then log-linear model (6) will provide the same estimate of $\boldsymbol{\beta}$ as (3). Note that $\widehat{S}_{t'}(\boldsymbol{\beta}) = \widehat{S}_t(\boldsymbol{\beta})$ for any $t' \in W(t)$, implying S_t is a step function of t with a separate value at each time stratum. As the conditional likelihood is uninformative under a "non-localizable" design, there is no equivalence between time-series analysis using a log-linear model and case-crossover design.

2.3.2. Frequentist equivalence using full likelihood. Similarly, by comparing the two estimating equations corresponding to $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ and $L_{ll}(\boldsymbol{\beta}, S_t)$, we showed (in Web Appendix A) that, for a certain choice of window $W(t)$ of a "localizable" design, there exists a choice of S_t in the log-linear model such that the two estimating equations provide the same estimate of $\boldsymbol{\beta}$. Under a TSD, while the conditional likelihood approach or an equivalent log-linear model would only allow the baseline risk to change discontinuously among different time strata, the full likelihood approach does not require such constraints. However, both the full likelihood method and its equivalent log-linear model encounter difficulty in estimating $\boldsymbol{\beta}$ in the presence of T day level nuisance parameters using maximum likelihood. As an alternative estimation strategy, a random effects Bayesian approach could be used to handle these random nuisance parameters under the full likelihood formulations, which are described in Section 4.

3. Bayesian Equivalence Between Case-Crossover and Time-Series Analysis

3.1. *Bayesian Equivalence Using Conditional Likelihood*

We focus on the posterior distributions of the log risk ratio parameter $\boldsymbol{\beta}$ derived under case-crossover and time-series analysis. The Bayesian equivalence result for $\boldsymbol{\beta}$ requires that the posterior distribution of $\boldsymbol{\beta}$ derived from $L_{cc}(\boldsymbol{\beta})$ in (3) and from $L_{ll}(\boldsymbol{\beta}, S_t)$ in (7) are identical under certain forms of S_t and certain prior distributions on $\boldsymbol{\beta}$ and S_t. The validity of using a conditional likelihood as the basis for Bayesian inference has been discussed in previous studies. For example, Rice (2004, 2008) discussed the equivalence between the use of conditional and marginal likelihoods for matched case-control study. Since the conditional likelihood approach is only valid for a "localizable" design, we restrict our attention specifically to a TSD in this section.

THEOREM 1. *Suppose the follow-up time points $t = 1, \ldots, T$ are divided a priori into K disjoint time strata $ts(k)$ under a TSD, $k = 1, \ldots, K$. If S_t in log-linear model (6) is defined as a step function with distinct values of S_t be S'_k on $ts(k)$, $k = 1, \ldots, K$, and if independent improper priors $\pi(S'_k) \propto 1$ for S'_k and a proper prior $\pi(\boldsymbol{\beta})$ for $\boldsymbol{\beta}$ are used where S'_k and $\boldsymbol{\beta}$ are mutually independent, then the posterior distribution of $\boldsymbol{\beta}$ derived from $L_{cc}(\boldsymbol{\beta})$ in (3) is identical to the marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L_{ll}(\boldsymbol{\beta}, S_t)$ in (7).*

Proof of Theorem 1 is given in Web Appendix A. We showed that, given the choice of S_t and prior distribution on S_t and $\boldsymbol{\beta}$ as in Theorem 1, the marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L_{ll}(\boldsymbol{\beta}, S_t)$ is $\pi(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) \propto \int\cdots\int \pi(\boldsymbol{\beta})\pi(S'_1, \ldots, S'_K)L_{ll}(\boldsymbol{\beta}, S_t)dS'_1\cdots dS'_K \propto \pi(\boldsymbol{\beta})L_{cc}(\boldsymbol{\beta})$, that is, the posterior distribution of $\boldsymbol{\beta}$ derived from $L_{cc}(\boldsymbol{\beta})$.

With shared exposure data across all individuals, the time-series model in (6) is more flexible than a case-crossover design using conditional likelihood, in the sense that model (6) allows various smoothing functions of time for S_t where one special choice is equivalent to the analysis of data under case-crossover design. Log-linear models can also account for over-dispersion of the Poisson variance that is typically present in air pollution studies, while case-crossover studies cannot. In contrast, case-crossover design has the advantage of controlling for personal level confounders, and modeling individual level exposures over time-series models.

3.2. *Bayesian Equivalence With Full Likelihood*

We aim to show the marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L^T_{full}(\boldsymbol{\nu}, \boldsymbol{\beta})$ under the shared exposure assumption is identical to that derived from a Poisson likelihood. Let y_{s1t} and y_{s0t} be the numbers of potential event and referent times that equals to t in the sth time stratum. y_{sdt} is assumed to follow a Poisson distribution $Poisson(\mu_{sdt})$ with mean $\mu_{sdt} = \exp(\phi_{st} + d(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t))$, $d = 0, 1$; $s = 1, \ldots, T$; $t = 1, \ldots, T$. The Poisson likelihood with ancillary parameters ϕ_{st} is given by,

$$L_p(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta}) = \prod_{s=1}^{T}\prod_{t=1}^{T}\prod_{d=0}^{1} \left[\exp\{\phi_{st} + d\left(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t\right)\}\right]^{y_{sdt}} \exp\left[-\exp\{\phi_{st} + d\left(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t\right)\}\right]$$

With independent improper priors $\pi(\phi_{st}) \propto 1$ and proper prior on $\boldsymbol{\nu}$ and $\boldsymbol{\beta}$, the joint posterior distribution of $(\boldsymbol{\nu}, \boldsymbol{\beta})$ derived from $L_p(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta})$ is $\pi(\boldsymbol{\nu}, \boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) \propto$

$\int \pi(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta}) L_p(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta}) d\boldsymbol{\phi} \propto \pi(\boldsymbol{\nu}, \boldsymbol{\beta}) L^T_{full}(\boldsymbol{\nu}, \boldsymbol{\beta})$ (shown in Web Appendix A). So the marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L_p(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta})$ and from $L^T_{full}(\boldsymbol{\nu}, \boldsymbol{\beta})$ are the same. This method is inspired by the Multinomial-Poisson transformation (Baker, 1994) and its Bayesian counterpart (Seaman and Richardson, 2004; Ghosh, Zhang, and Mukherjee, 2006). Though this proves theoretical Bayesian equivalence between using a case-crossover full likelihood and a Poisson likelihood, the Poisson model has a large number of nuisance parameters $\boldsymbol{\phi}$. Moreover, the interpretation of the artificially constructed Poisson model is practically not very meaningful. Thus, we focus on full likelihood based methods with more flexible semi-parametric prior distributions on the intercepts in the following section, instead of an equivalent time-series formulation.

4. Bayesian Inference

Traditionally, CLR models were routinely used for frequentist inference on $\boldsymbol{\beta}$ under case-crossover designs. A naive approach would be to also use the conditional likelihood $L_{cc}(\boldsymbol{\beta})$ as the basis for Bayesian inference, where prior specification on only $\boldsymbol{\beta}$ is needed. The posterior distribution of $\boldsymbol{\beta}$ is not a standard distribution, but posterior draws could be generated by using a Gibbs sampler (Web Appendix B).

For the full likelihood approach, though the number of nuisance parameters grows with the sample size, a random effects model can be used to reduce the problem to estimating the parameters corresponding to the random effects distribution. For example, the stratified logistic regression model with likelihood $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ in (5) can be readily fitted through a generalized linear mixed model with $\nu_t \overset{iid}{\sim} N(\mu_\nu, \sigma^2_\nu)$. Methods such as penalized pseudo-likelihood (Breslow and Clayton, 1993) can be used for inference in such models, which is available in standard statistical software.

The misspecification of the random effects distribution could lead to potential bias in the estimation of $\boldsymbol{\beta}$ (Mukherjee et al., 2009). To avoid assuming a parametric normal distribution on the nuisance parameters, we consider a more robust semi-parametric Bayesian approach that allows the random effects to have a nonparametric distribution. To this end, we use the Dirichlet process prior (Ferguson, 1973; Antoniak, 1974; Müller and Quintana, 2004) to handle the random intercepts. In particular, for example with $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ in (5), we assume $\nu_t | G \overset{iid}{\sim} G$, where G is a random distribution generated from a Dirichlet process with concentration parameter α and base distribution G_0, that is, $G|\alpha, G_0 \sim DP(\alpha, G_0)$. Let $\boldsymbol{\nu}_{-t} = (\nu_1, \ldots, \nu_{t-1}, \nu_{t+1}, \ldots, \nu_T)$, for $t = 1, \ldots, T$. The joint prior distribution $\pi(\boldsymbol{\nu})$ can be represented in terms of leave-one-out conditional distributions as $\nu_t | \boldsymbol{\nu}_{-t} \sim \alpha(T - 1 + \alpha)^{-1} G_0 + (T - 1 + \alpha)^{-1} \sum_{s=1, s\neq t}^{T} I_{\nu_s}(\cdot)$ (Blackwell and MacQueen, 1973). Thus $(\nu_1, \ldots, \nu_T)$ will be adaptively reduced to fewer distinct clusters with positive probability. As $\alpha \to \infty$, the Dirichlet process model reduces to specifying a parametric model $\nu_t \overset{iid}{\sim} G_0$; whereas $\alpha \to 0$ implies a parametric model with a common stratum effect, namely $\nu_t = \nu^*$ for $t = 1, \ldots, T$, where $\nu^* \sim G_0$.

To complete the hierarchy, independent hyperpriors are considered as follows: $\alpha | a_0, b_0 \sim Gamma(a_0, b_0)$, $G_0 \sim N(\mu, \sigma^2)$, $\mu | \mu_0, \sigma_0 \sim N(\mu_0, \sigma^2_0)$, $\sigma^{-2} | c_0, d_0 \sim Gamma(c_0, d_0)$. We consider mutually independent normal priors $\boldsymbol{\beta} \sim N(\mu_\beta, \sigma^2_\beta \boldsymbol{I}_p)$. The posterior distributions of $\boldsymbol{\nu}$ and $\boldsymbol{\beta}$ can be obtained by using a Metropolis–Hastings within Gibbs algorithm as described in Neal (2000) and Sinha et al. (2004). The details are presented in Web Appendix B. Similarly, for the individual specific stratum effects in $L^N_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$, we assume $\log(\lambda_{0it_i}) | G \overset{iid}{\sim} G$, for $i = 1, 2, \ldots, N$. For $L^{NT}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$, we only consider a special case by assuming a multiplicative structure on the nuisance parameters: $\lambda_{0it} = \lambda_{0i} \exp(\omega_t)$, where λ_{0i} is a constant frailty for person i and ω_t is the time varying effect on the risk. We model $\log(\lambda_{0i})$ and ω_t through random distributions generated from two independent Dirichlet processes.

5. A Simulation Study

5.1. *Simulation Settings*

In our simulation, we used $\lambda_{0it} = \lambda_{0i} \exp(\omega_t)$ as the form of the true baseline risk. Under both Assumptions 1 and 3, we have $\log(E(Y_t)) = \boldsymbol{\beta}^\top \boldsymbol{X}_t + \log(\sum_{i=1}^{N+M} \lambda_{0it}) = \boldsymbol{\beta}^\top \boldsymbol{X}_t + \omega_t + \eta$, where $\eta = \log(\sum_{i=1}^{N+M} \lambda_{0i})$. We generated the number of events per day from a Poisson model

$$Y_t \sim Poisson(\mu_t), \quad \text{where } \mu_t = \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X}_t + \omega_t + \eta\right). \quad (8)$$

We considered various simulation scenarios with different choices of time effects ω_t on the baseline risk, true effect sizes β^*, and exposure series X_t. Without loss of generality, we considered $\boldsymbol{\beta}$ and $\boldsymbol{X}_t$ to be univariate in our simulation study. We convert the time-series data in the form of individual event referent times according to a given case-crossover design.

Temporal trends on the baseline risk: In order to examine whether the full likelihood method under a case-crossover design is more robust to various baseline risk specifications than the conditional likelihood method, we considered three forms of time-varying effect ω_t involved in the baseline risk. In particular, B1: $\omega_t = \omega$ that satisfies Assumption 2; B2: $\omega_t = c(1 - 0.001t)[1 + 0.5cos(2\pi t/365)]$ that combines seasonal and long-term decreasing trends, where c is a positive scaling factor; B3: B3 is a mixture of B2 (with probability 0.9) and random spikes (with probability 0.1), where the spikes follow a uniform distribution $U(2c, 4c)$. Note that B2 and B3 both violate Assumption 2.

Effect sizes: We considered two typical true effect sizes $\beta^* = 0.1$ (a risk ratio of 1.1, e.g., reflecting the effect of $10\,\mu g m^{-3}$ increase of $PM_{2.5}$ (particulate matter < 2.5 μ m in diameter) on the risk of acute asthma (Li et al., 2011) or 10 °F change of temperature on mortality risk (Basu et al., 2005)) and $\beta^* = 1$ (a risk ratio of 2.7, e.g., reflecting the effect of medication use on preventing elderly falls in case-crossover intervention trials (Luo and Sorock, 2008)).

Exposure series: We simulated exposure series X_t over a 3-year ($T = 1096$) period under two settings. E1: X_t has auto-correlation structure $AR(1)$ ($\rho = 0.6$); E2: X_t has long-term decreasing trend plus seasonal and day of week effects, with the same auto-correlation structure as in E1. We generated E1 and E2 to have the same marginal distributions.

Table 1
Summary of the simulation results in terms of relative bias and MSE, under different settings of temporal terms ω_t on baseline risk and exposure series X_t, with true effect size $\beta^* = 0.1$

	Exposure E1[a]						Exposure E2[a]					
	Baseline B1[a]		Baseline B2[a]		Baseline B3[a]		Baseline B1		Baseline B2		Baseline B3	
	RB[b] (%)	MSE[b]	RB	MSE	RB	MSE	RB	MSE	RB	MSE	RB	MSE
Log-linear model (true[c])	−0.29	2.53	−0.07	2.40	−0.18	2.65	−0.43	3.15	0.37	3.43	−0.93	3.21
Log-linear model (spline[c])	−0.50	2.68	−0.11	2.59	−0.39	3.09	−0.46	3.53	0.40	3.67	−1.09	3.50
Time-stratified design												
Conditional likelihood (CLR[b])	−0.08	3.66	−0.33	3.39	−0.65	3.77	−0.44	4.29	0.92	4.33	−1.44	4.28
Bayesian (non-informative prior[d])												
Conditional likelihood	−0.10	3.67	−0.42	3.39	−0.74	3.77	−0.45	4.33	0.77	4.34	−1.49	4.32
Full likelihood with DP (T)[d]	−0.55	4.87	0.46	4.77	0.57	5.10	−1.54	6.00	−0.44	5.71	−1.25	5.60
Bayesian (informative prior[d])												
Conditional likelihood	−5.67	**2.86**[e]	−5.04	**2.71**	−4.89	2.97	−6.87	**3.27**	−5.97	3.35	−7.88	3.32
Full likelihood with DP (T)	−5.76	2.91	−5.05	2.75	−4.84	**2.87**	−7.03	3.28	−6.35	**3.15**	−7.62	**3.22**
Symmetric bi-directional design												
Conditional likelihood (CLR)	2.59	7.01	2.03	6.80	3.03	7.42	−0.75	10.87	−0.87	9.40	−2.24	9.87
Bayesian (non-informative prior)												
Conditional likelihood	2.44	7.03	1.89	6.80	2.93	7.41	−0.93	10.87	−1.05	9.39	−2.36	9.84
Full likelihood with DP (T)	2.59	7.38	2.01	7.14	2.78	7.80	−3.67	10.55	−2.93	9.26	−4.08	10.05
Bayesian (informative prior)												
Conditional likelihood	−3.64	3.72	−3.76	3.83	−3.24	3.95	−7.90	5.88	−7.91	4.39	−9.35	5.28
Full likelihood with DP (T)	−3.68	3.84	−3.93	3.72	−3.34	4.04	−8.76	5.46	−8.30	4.77	−9.16	5.31

[a] E1, auto-correlation only; E2, auto-correlation plus seasonal trend; B1, $\omega_t = \omega$; B2, $\omega_t = 0.05(1 - 0.001t)[1 + 0.5cos(2\pi t/365)]$; B3, a mixture of B2 and point mass at random spikes.
[b] RB, relative bias; MSE, mean squared error (multiplied by 10^5); CLR, conditional logistic regression.
[c] True: adjusted for true temporal trend (B1–B3); spline: adjusted for a natural cubic spline with 7 df per year.
[d] Informative prior on β, $\beta \sim N(0.08, 0.03^2)$; non-informative prior on β, $\beta \sim N(0, 10^2)$; DP (T), Dirichlet process prior $DP(\alpha, G_0)$ on $\boldsymbol{\nu}$ in $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$.
[e] The best performing method for a case-crossover design in terms of MSE is marked in **bold** under each baseline×exposure setting.

Likelihoods/methods: We would like to compare time-series analysis using log-linear models with the analysis under case-crossover designs. The methods can broadly be divided into three classes in terms of likelihoods we considered. In particular, M1: log-linear models, adjusted for the true temporal trend ω_t as offset (this is the closest to the true generating model) or adjusted for a natural spline term on time t; M2: conditional likelihood approach under a case-crossover design, using both frequentist and Bayesian treatment; M3: full likelihood approach under a case-crossover design, with the random intercepts handled by a Dirichlet process. Within case-crossover analysis, we also compared the two commonly used referent time selection strategies, TSD with SBD, to possibly quantify the "overlap bias."

Prior choices: Within the Bayesian alternatives we implemented two choices of prior distributions on β: non-informative and informative. For non-informative prior, we used a vague prior $\beta \sim N(0, 10^2)$; for informative prior, we considered $\beta \sim N(\mu_\beta, \sigma^2_\beta)$ with μ_β and σ_β potentially elicited from historical data. We described the details of incorporating historical data to construct informative priors under a concrete data example framework in Section 6. We used $\beta \sim N(0.08, 0.03^2)$ when $\beta^* = 0.1$, and $\beta \sim N(0.8, 0.2^2)$ when $\beta^* = 1$ in our simulations. Justification for these prior choices on μ_β and σ_β as well as on the full set of parameters are provided in Web Appendix B.

5.2. *Simulation Findings*

The simulation was repeated 1000 times under each scenario. We summarized the results in terms of relative bias ($RB = (1000^{-1}\sum_{i=1}^{1000}\hat{\beta}_i - \beta^*)/\beta^* \times 100\%$) and MSE ($MSE = 1000^{-1}\sum_{i=1}^{1000}(\hat{\beta}_i - \beta^*)^2$) corresponding to the log risk ratio parameter β. Tables 1 and 2 present results corresponding to 6 baseline×exposure (3×2) settings, for $\beta^* = 0.1$ and 1 respectively. As the individual level model $L^{NT}_{full}(\beta, \boldsymbol{\lambda})$ and $L^{N}_{full}(\beta, \boldsymbol{\lambda})$ are computationally intensive with large N ($N \approx 20{,}000$ for $\beta^* = 0.1$ in our simulations), we only considered $L^T_{full}(\beta, \boldsymbol{\nu})$ under $\beta^* = 0.1$. We considered all three versions of full likelihoods under $\beta^* = 1$ with $N \approx 1000$.

Likelihoods/methods: We present the estimates from the log-linear model using offset terms to be the true values of ω_t as a reference benchmark, against which each of our methods is compared. In practice, while carrying out a time-series analysis, one will not know the true time effect terms and will use a flexible nonparametric spline term (Dominici et al., 2002, 2003). The log-linear model adjusted for a natural cubic spline term of time with 7 degrees of freedom per year approximates the true model quite accurately. The two log-

Table 2

Summary of the simulation results in terms of relative bias and MSE, under different settings of temporal terms ω_t on baseline risk and exposure series X_t, with true effect size $\beta^* = 1$

	Exposure E1[a]						Exposure E2[a]					
	Baseline B1[a]		Baseline B2[a]		Baseline B3[a]		Baseline B1		Baseline B2		Baseline B3	
	RB[b] (%)	MSE[b]	RB	MSE	RB	MSE	RB	MSE	RB	MSE	RB	MSE
Log-linear model (true[c])	−0.58	0.75	−0.65	0.81	−0.80	0.96	−0.38	0.77	1.00	0.86	0.92	0.91
Log-linear model (spline[c])	−0.85	0.93	−0.90	1.06	−0.98	1.11	−0.84	1.01	1.79	1.03	1.44	1.16
Time-stratified design												
Conditional likelihood (CLR[b])	−1.92	1.01	−1.99	1.07	−2.60	1.13	−1.00	1.03	1.89	1.10	1.96	1.15
Bayesian (non-informative prior[d])												
Conditional likelihood	−1.97	1.00	−2.05	1.07	−2.63	1.14	−1.04	1.05	1.82	1.08	1.96	1.15
Full likelihood with DP[d] (T)	−2.17	1.06	−1.18	1.10	−2.50	1.17	−1.25	1.03	2.35	1.15	1.57	1.18
Full likelihood with DP (N)	−2.04	1.06	−1.26	1.11	−2.43	1.17	−1.19	1.03	2.32	1.14	1.71	1.19
Full likelihood with DP (T and N)	−0.67	1.21	−0.65	1.23	−0.75	1.21	−1.00	1.09	1.46	1.29	1.10	1.36
Bayesian (informative prior[d])												
Conditional likelihood	−14.67	**0.30**[e]	−14.60	**0.31**	−15.96	0.40	−13.12	**0.36**	−14.78	**0.29**	−14.51	**0.28**
Full likelihood with DP (T)	−14.70	0.31	−14.28	**0.31**	−16.39	0.38	−13.53	0.37	−15.12	0.31	−14.80	0.30
Full likelihood with DP (N)	−14.72	0.31	−14.33	**0.31**	−16.22	**0.37**	−13.38	**0.36**	−15.13	0.31	−14.82	0.30
Full likelihood with DP (T and N)	−12.32	0.43	−11.93	0.39	−14.99	0.45	−11.63	0.43	−12.48	0.36	−12.24	0.45
Symmetric bi-directional design												
Conditional likelihood (CLR)	−5.21	1.76	−4.89	2.12	−8.34	2.43	−4.23	2.68	−3.54	2.21	−8.40	2.73
Bayesian (non-informative prior)												
Conditional likelihood	−5.71	1.78	−4.41	2.15	−8.84	2.45	−4.92	2.72	−4.09	2.25	−9.04	2.75
Full likelihood with DP (T)	−6.54	1.69	−4.79	2.04	−7.78	2.43	−6.32	2.52	−5.32	2.26	−8.25	2.83
Full likelihood with DP (N)	−6.66	1.67	−4.89	2.03	−7.67	2.41	−5.70	2.53	−5.41	2.29	−8.59	2.80
Full likelihood with DP (T and N)	−1.45	1.83	−2.43	2.10	−3.02	2.61	−1.73	2.77	−1.92	2.30	−5.27	2.84
Bayesian (informative prior)												
Conditional likelihood	−14.70	0.84	−15.60	0.87	−15.79	0.85	−17.17	0.87	−18.25	0.95	−16.12	0.96
Full likelihood with DP (T)	−14.99	0.84	−16.05	0.89	−15.42	0.92	−18.79	0.87	−18.76	0.93	−16.53	0.97
Full likelihood with DP (N)	−14.99	0.85	−16.16	0.90	−15.43	0.93	−18.77	0.87	−18.61	0.92	−16.38	0.96
Full likelihood with DP (T and N)	−12.36	0.93	−11.74	0.93	−12.90	0.99	−15.63	0.92	−17.08	0.97	−12.80	1.08

[a] E1, auto-correlation only; E2, auto-correlation plus seasonal trend; B1, $\omega_t = \omega$; B2, $\omega_t = 0.05(1 - 0.001t)[1 + 0.5cos(2\pi t/365)]$; B3, a mixture of B2 and point mass at random spikes.

[b] RB, relative bias; MSE, mean squared error (multiplied by 10^5); CLR, conditional logistic regression.

[c] True: adjusted for true temporal trend (B1–B3); spline: adjusted for a natural cubic spline with 7 df per year.

[d] Informative prior on β: $\beta \sim N(0.8, 0.2^2)$; non-informative prior on β: $\beta \sim N(0, 10^2)$; DP, Dirichlet process prior on the random intercepts in full likelihood formulation (T: $L^T_{full}(\beta, \boldsymbol{\nu})$; N: $L^N_{full}(\beta, \boldsymbol{\lambda})$; T and N: $L^{NT}_{full}(\beta, \boldsymbol{\lambda})$).

[e] The best performing method for a case-crossover design in terms of MSE is marked in **bold** under each baseline×exposure setting.

linear models both have smaller bias and MSE than the case-crossover designs, especially under B2 or B3. Note that the log-linear models fitted here were not chosen to be the equivalent models to case-crossover designs.

(a) *Design effect*: In comparing the two designs, we found that the TSD generally has smaller bias and MSE than the SBD. Under B1, the only source of bias is the "overlap bias" of a SBD. We observed up to 5% difference in bias between TSD and SBD (Tables 1 and 2). However, the magnitude and direction of the "overlap bias" depend on the particular exposure series and effect size, as previously noted in (Janes, Sheppard, and Lumley, 2005a).

(b) *Conditional versus full likelihood formulation (with non-informative priors)*: The conditional likelihood $L_{cc}(\beta)$ as well as the full likelihood $L^N_{full}(\beta, \boldsymbol{\lambda})$ and $L^T_{full}(\beta, \boldsymbol{\nu})$ require Assumption 2 of constancy of ω_t in each referent window and only allows the risk to change discontinuously across referent windows. The most general form of the full likelihood $L^{NT}_{full}(\beta, \boldsymbol{\lambda})$ does not require this assumption. As for bias, we note that the bias due to violation of Assumption 2 (under B2 and B3) is typically very small (< 1%) when $\beta^* = 0.1$ (Table 1), and up to 3% when $\beta^* = 1$ (Table 2). However, full likelihood methods have greater MSE than conditional counterparts under non-informative priors. This is expected as another level of hierarchy was added to model the uncertainty in the random nuisance parameters.

Prior sensitivity: Both conditional and full likelihood methods show substantial reduction in MSE when informative priors on β are used as shown in Tables 1 and 2, which also lead to substantial shrinkage towards the prior mean. Thus, given the context of the study and prior information, Bayesian methods that utilize informative priors can potentially have advantage over their frequentist counterparts in terms of MSE.

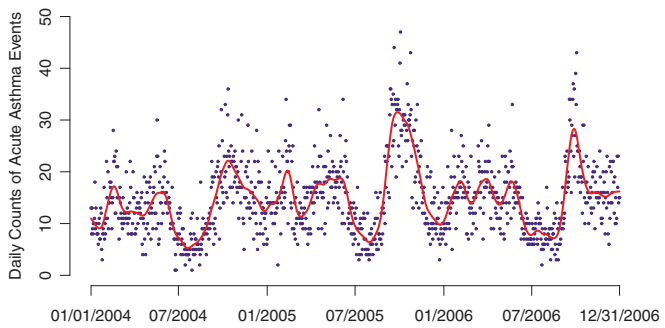


Figure 1. Temporal trend of daily counts of acute asthma events (shown as points) for the pediatric Medicaid population in Detroit, Michigan, 2004–2006, as obtained in the DAMAT study. The overlaying smooth curve is created by using locally estimated scatter-plot smoother. [This figure appears in color in the electronic version of this article.]

Remark: Note that the full likelihood has additional analytic flexibility to handle individual level data and incorporate interaction terms that account for subject level covariates. We generated time-series data with shared exposure for the present simulation study for illustration purposes, leading to best performance by the log-linear models. If we had generated individual level data with personal factors, the time-series analysis would have been more susceptible to residual bias from ignoring personal level confounders.

6. Data Example

We illustrate the proposed methods through the "Detroit Asthma Morbidity, Air Quality and Traffic" (DAMAT) study originally analyzed by Li et al. (2011). One primary goal of the study was to examine the association between acute asthma risk and ambient air pollutant levels, especially $PM_{2.5}$, for the pediatric (children 2–18 years) Medicaid population in Detroit, Michigan, during the 2004–2006 study period ($T = 1096$ days). Daily counts of asthma events, including emergency department visits and hospitalizations, were used as the outcome series Y_t, for $t = 1, \ldots, 1096$. Figure 1 shows the smoothed trend of Y_t indicating a strong seasonal pattern, with the highest frequency during fall, and the lowest during summer. A total of 12,933 asthma events were observed during the 1096 days, representing an average rate of 11.8 events per day. Daily $PM_{2.5}$ concentration was computed as the average concentration across the air quality monitoring sites in the Detroit area. Daily $PM_{2.5}$ data also show a strong seasonal pattern with a mean level of 15.0 μgm^{-3}. Daily meteorological variables, including temperature (TP) and relative humidity (RH), were obtained similarly. To account for other temporal trends that were not controlled by the case-crossover design, a natural quadratic spline term (denoted by $ns(\cdot)$) was used on the TP variable.

Likelihoods: Under the case-crossover design, the conditional likelihood $L_{cc}(\boldsymbol{\beta})$ is given by

$$\prod_{t=1}^{1096}\left[\frac{\exp\{\beta_{PM_{2.5}}\mathrm{PM}_{2.5,t}+\beta_{RH}\mathrm{RH}_t+ns(\mathrm{TP}_t)\}}{\sum_{s\in W(t)}\exp\{\beta_{PM_{2.5}}\mathrm{PM}_{2.5,s}+\beta_{RH}\mathrm{RH}_s+ns(\mathrm{TP}_s)\}}\right]^{Y_t}, \tag{9}$$

where we used the 5-day moving average of $PM_{2.5}$ concentration prior to the asthma events on day t as $\mathrm{PM}_{2.5,t}$. The Medicaid data was also analyzed using the equivalent log-linear model of (9). In particular, for a TSD with time stratum as the same day of the week in the same calender month, we compare it with the following equivalent log-linear model

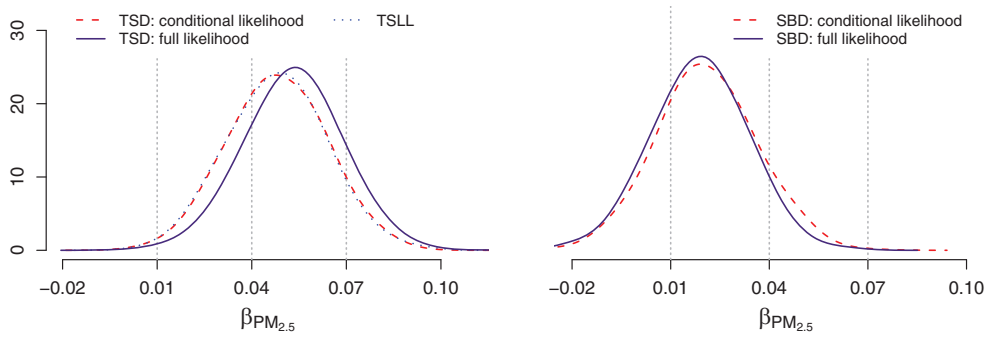
$$\log(E(Y_t)) = \beta_0 + \beta_{PM_{2.5}}\mathrm{PM}_{2.5,t} + \beta_{RH}\mathrm{RH}_t + ns(\mathrm{TP}_t) + S_t, \tag{10}$$

where S_t represents all possible combinations among the three factors of day of the week, month and year, having a total of $7 \times 12 \times 3 = 252$ levels. We refer to (10) as the time-stratified log-linear (TSLL) model. For inference on $\boldsymbol{\beta}$, models (9) and (10) are equivalent under both the frequentist and Bayesian framework (with prior specification as described in Theorem 1). Similarly, a case-crossover SBD using referent times as 7 and 14 days before and after the event day was compared to the corresponding symmetric bidirectional log-linear (SBLL) model with $S_t = \log[\sum_{s=t,\, t\pm7,\, t\pm14}\{Y_s/\sum_{r=s,\, s\pm7,\, s\pm14}\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_r)\}]$. Joint estimation of $\boldsymbol{\beta}$ and S_t was performed iteratively as S_t potentially depends on $\boldsymbol{\beta}$. The comparison between SBD and SBLL is pertinent only under the frequentist framework.

Prior choices: We used $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ under the full likelihood approach. We first considered the random effects model assuming $\nu_t \overset{iid}{\sim} N(0, 10^2)$ without further prior specification on $\boldsymbol{\beta}$, and proceeded with the marginal likelihood to estimate $\boldsymbol{\beta}$. Then we considered the full Bayesian treatment using Dirichlet process prior $\nu_t|G \overset{iid}{\sim} G$ where $G|\alpha, G_0 \sim DP(\alpha, G_0)$. $\alpha \sim Gamma(0.5, 0.1)$, $G_0 \sim N(\mu, \sigma^2)$, $\mu \sim N(0, 10)$ and $\sigma^{-2} \sim Gamma(4, 1)$ were used as the base prior setting in our data example. As part of our sensitivity analysis, we varied the priors on α across four *Gamma* distributions, and two extreme cases when $\alpha \to 0$ (corresponding to $\nu_t = \nu^*$ for $t = 1, \ldots, T$, where $\nu^* \sim G_0$) and $\alpha \to \infty$ (corresponding to $\nu_t \overset{iid}{\sim} G_0$). More details were provided in Figure 2c.

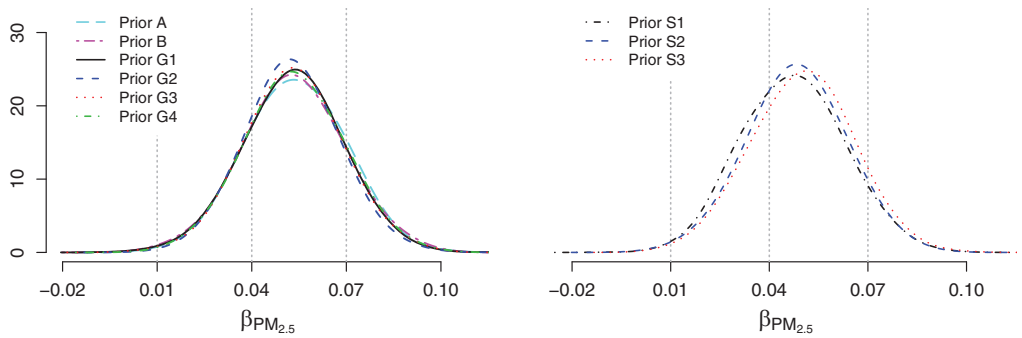
We considered both informative and non-informative priors on $\beta_{PM_{2.5}}$. For non-informative prior, we used a vague prior $\beta_{PM_{2.5}} \sim N(0, 10^2)$. For informative prior, we considered an ad-hoc way of eliciting prior information from published results. From a recent review (Li et al., 2011), we *a priori* postulated that the asthma-$PM_{2.5}$ association is in general modest with a risk ratio ranging in (1.01–1.09) for 10 μgm^{-3} increase in $PM_{2.5}$. Assuming $\beta_{PM_{2.5}} \sim N(\mu_\beta, \sigma^2_\beta)$, if we believed that the 95% confidence interval (CI) for $\exp(\beta_{PM_{2.5}})$ is (1.01, 1.09), the approximate values for μ_β and σ_β can be obtained as $\mu_\beta = [\log(1.09) + \log(1.01)]/2 = 0.05$ and $\sigma_\beta = [\log(1.09) - \log(1.01)]/4 = 0.02$. Then our informative prior was chosen as $N(0.05, 0.02^2)$. There is no general consensus on the best way to elicit a subjective prior though this topic has been studied vastly (Dey and Liu, 2007).

While non-informative or non-subjective priors are often quite adequate as default priors for many Bayesian analyses, prior elicitation when possible can lead to more meaningful results from the data. This is especially true in the presence of historical data, available from series of past studies, for example, past studies relating asthma-$PM_{2.5}$ associations in our



(a) Posterior densities using likelihood under TSD

(b) Posterior densities using likelihood under SBD



(c) TSD using $L^T_{full}(\beta, \nu)$ and 6 different DP priors

(d) TSLL using 3 different priors on S_t

Figure 2. Posterior density plots for the log risk ratio parameter $\beta_{PM_{2.5}}$ corresponding to acute asthma events for a $10\,\mu gm^{-3}$ increase in $PM_{2.5}$ based on data from the DAMAT study, where vague prior $\beta_{PM_{2.5}} \sim N(0, 10^2)$ was used. Panels (a) and (b) used the base prior setting as described in Section 6. Panels (c) and (d) varied prior choices on $\boldsymbol{\nu}$ and S_t as a sensitivity analysis. [This figure appears in color in the electronic version of this article.]

Priors on $\boldsymbol{\nu}$ (panel (c)): A: $\nu_t = \nu^*$ for $t = 1, \ldots, 1096$, where $\nu^* \sim N(0, 10^2)$; B: $\nu_t \overset{iid}{\sim} N(0, 10^2)$;
G1: $\alpha \sim Gamma(0.5, 0.1)$; G2: $\alpha \sim Gamma(2, 0.2)$; G3: $\alpha \sim Gamma(10, 0.5)$; G4: $\alpha \sim Gamma(20, 1)$.

Priors on S_t (panel (d)): S1: $\pi(S'_k) \propto 1$; S2: $S'_k \overset{iid}{\sim} N(0, 10^2)$; S3: $S'_k | G \overset{iid}{\sim} G$, where $G|\alpha, G_0 \sim DP(\alpha, G_0)$.

TSD, time-stratified design; SBD, symmetric bidirectional design; TSLL, time-stratified log-linear; DP, Dirichlet process.

context. These elicited priors based on historical data, when proper, have an operational advantage over non-subjective improper priors. Proper priors are required for computing Bayes factors and posterior model probabilities.

Ibrahim and Chen (2000) proposed a particular approach towards the development of priors based on historical data. They referred to these priors as "power priors." Specifically, if D_0 denotes the historical data from a previous article, the power prior for β is defined as $\pi(\beta|D_0, a_0, c_0) \propto L^{a_0}(\beta|D_0)\pi_0(\beta|c_0)$. Here $\pi_0(\beta|c_0)$ is the initial prior before the data D_0 were observed and c_0 is a specified hyperparameter. The parameter $a_0 \in [0, 1]$ is a scalar parameter which controls the influence of the historical data on the current data. In particular, $a_0 = 1$ corresponds to the past posterior which has become the present prior. On the other hand, $a_0 = 0$ corresponds to a prior specification which ignores completely the historical data. While one can do the analysis by simply assigning also a prior on a_0, we will pursue our analysis both when a_0 is fixed or random, assigning *Beta* priors to a_0 in the latter case. If $\pi_0(\beta|c_0)$ is proper, then the power prior $\pi(\beta|D_0, a_0, c_0)$ is guaranteed to be proper; further, $\pi(\beta|D_0, a_0, c_0)$ can be proper under certain regression settings even if $\pi_0(\beta|c_0)$ is an improper uniform prior (Ibrahim and Chen, 2000). We used a vague initial prior $N(0, 10^2)$ for $\pi_0(\beta|c_0)$, such that the prior $\pi(\beta|D_0, a_0, c_0)$ is a handy proper prior. $L(\beta|D_0)$ was constructed based on several published studies of asthma-$PM_{2.5}$ associations reviewed in (Li et al., 2011). Further details regarding the construction of $L(\beta|D_0)$ are given in Web Appendix B.

Results: The results are shown in Table 3 and Figure 2. In general, evidence of significant increases in acute asthma risk was found with $10\,\mu gm^{-3}$ increase in $PM_{2.5}$ concentrations leading to a risk ratio ranging from 1.02 to 1.06 across different methods.

(a) *Design effect*: Comparing TSD with SBD, $\hat{\beta}_{PM_{2.5}}$ estimated under a TSD (ranging from 1.05 to 1.06 in Table 3) are larger than those estimated under a SBD (from 1.02 to 1.04). The overall pattern of the attenuated effects under the SBD is probably due to the choice of the window and potential "overlap bias", though this

Table 3

Risk ratios of acute asthma events corresponding to a $10\,\mu g\,m^{-3}$ increase in $PM_{2.5}$ in the DAMAT study. The model was adjusted for temperature and relative humidity

	TSD[a]			SBD[a]	
Frequentist	MLE[a]	95% CI[a]		MLE	95% CI
Conditional likelihood	1.049	(1.019, 1.080)	Conditional likelihood	1.022	(0.992, 1.052)
TSLL[a]	1.049	(1.019, 1.080)	SBLL[a]	1.022	(0.992, 1.052)
Full likelihood REM[a] (T)	1.055	(1.026, 1.085)	Full likelihood REM (T)	1.020	(0.991, 1.048)
Bayesian (prior 1[b])	Bayes[a]	95% HPD[a]		Bayes	95% HPD
Conditional likelihood	1.049	(1.021, 1.081)	Conditional likelihood	1.023	(0.993, 1.053)
TSLL	1.049	(1.021, 1.081)			
Full likelihood DP[b] (T)	1.055	(1.026, 1.086)	Full likelihood DP (T)	1.020	(0.992, 1.049)
Bayesian (prior 2[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.052	(1.027, 1.075)	Conditional likelihood	1.035	(1.012, 1.058)
TSLL	1.052	(1.028, 1.076)			
Full likelihood DP (T)	1.055	(1.030, 1.076)	Full likelihood DP (T)	1.034	(1.011, 1.056)
Bayesian (power prior 1[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.045	(1.024, 1.068)	Conditional likelihood	1.033	(1.010, 1.059)
TSLL	1.045	(1.024, 1.068)			
Full likelihood DP (T)	1.049	(1.025, 1.069)	Full likelihood DP (T)	1.031	(1.008, 1.058)
Bayesian (power prior 2[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.054	(1.027, 1.080)	Conditional likelihood	1.040	(1.015, 1.068)
TSLL	1.054	(1.027, 1.080)			
Full likelihood DP (T)	1.059	(1.036, 1.085)	Full likelihood DP (T)	1.041	(1.014, 1.070)
Bayesian (power prior 3[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.046	(1.018, 1.075)	Conditional likelihood	1.031	(1.007, 1.063)
TSLL	1.046	(1.017, 1.074)			
Full likelihood DP (T)	1.050	(1.022, 1.076)	Full likelihood DP (T)	1.030	(1.005, 1.063)
Bayesian (power prior 4[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.055	(1.027, 1.084)	Conditional likelihood	1.041	(1.012, 1.070)
TSLL	1.055	(1.027, 1.083)			
Full likelihood DP (T)	1.060	(1.030, 1.087)	Full likelihood DP (T)	1.040	(1.011, 1.069)

[a] TSD: time-stratified design; SBD: symmetric bidirectional design; TSLL: time-stratified log-linear; SBLL: symmetric bidirectional log-linear; REM: random effects model; MLE: maximum likelihood estimate (penalized pseudo-likelihood for REM); CI: confidence interval; Bayes: Bayes estimates in terms of posterior mean; HPD: highest posterior density.

[b] DP (T): Dirichlet process prior $DP(\alpha, G_0)$ on $\boldsymbol{\nu}$ in $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ under the base prior setting described in Section 6; Prior 1: non-informative prior $\beta_{PM_{2.5}} \sim N(0, 10^2)$; Prior 2: informative prior $\beta_{PM_{2.5}} \sim N(0.05, 0.02^2)$; Power prior 1: $a_0 = 0.5$; Power prior 2: $a_0 = 1.0$; Power prior 3: $a_0 \sim Beta(20, 20)$ with mean 0.50 and variance 0.08; Power prior 4: $a_0 \sim Beta(50, 1)$ with mean 0.98 and variance 0.02.

direction does not hold in general as noted in our simulation study. Comparing the case-crossover TSD with corresponding TSLL model (in Table 3), we noted that they provided identical numerical results for $\hat{\beta}_{PM_{2.5}}$ under both frequentist and Bayesian framework (except possible Monte Carlo errors), indicating the numerical validity of our equivalence results. Frequentist equivalence results also appear to hold numerically for the case-crossover SBD and the corresponding time-series SBLL.

(b) *Conditional versus full likelihood*: Under a case-crossover TSD, full likelihood methods provided slightly stronger effects (risk ratio ranging from 1.05 to 1.06 in Table 3) than those derived using conditional likelihood (from 1.04 to 1.05). As noted in our simulation study, the violation of the constant baseline risk assumption within each window probably led to this difference. We also fit a log-linear model adjusted for a natural cubic spline term of time with 7 degrees of freedom per year, which shows an estimate of 1.055 (95% CI: (1.027, 1.084)) that is similar to the results using $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$. Under SBD, there is no substantial difference of using full versus conditional likelihood.

Prior sensitivity:

(a) *Priors on $\beta_{PM_{2.5}}$*: When a vague prior $\beta_{PM_{2.5}} \sim N(0, 10^2)$ is used, the Bayesian approaches yielded results that are quite similar to maximum likelihood-based inferences (Table 3). It is reassuring that with modest to large sample sizes, we observed similar results from Bayesian and frequentist methods. With a smaller sample size using only 1-year data ($T = 365$), even the use of non-informative prior increased the precision as compared to the frequentist methods (Web Appendix C Table 1). In Table 3, the use of informative priors (including the ad-hoc prior 2 and the four power priors) increased the precision as compared to the results under

the vague prior. Given one historical study having relatively larger effect (Web Appendix B), the use of power priors 2 ($a_0 = 1$) and 4 ($a_0 \sim Beta(50, 1)$ with prior mean ≈ 1) provided stronger effects than those under power priors 1 ($a_0 = 0.5$) and 3 ($a_0 \sim Beta(20, 20)$ with prior mean 0.5), because power priors 2 and 4 put more weight on $L(\beta|D_0)$. The use of another layer of uncertainty on a_0 (power priors 3 and 4) creates a heavier tail for the marginal power prior distribution of $\beta_{PM_{2.5}}$ than that using a fixed a_0 (power priors 1 and 2), and thus provides wider HPD credible intervals.

(b) *Priors on* $\boldsymbol{\nu}$: Figure 2c shows the posterior distributions of $\beta_{PM_{2.5}}$ derived under the six different prior settings on α, where the posterior distributions of $\beta_{PM_{2.5}}$ remain robust. We observed that the random effects model and the Dirichlet process prior model provided very similar results in Table 3. Although the prior support allows the number of clusters ranging from 1 to 252, we observed only one cluster under four out of the six prior settings (shown in Web Appendix C Table 2). The results suggested that a parametric constant random intercept model was adequate for this data set.

(c) *Priors on* S_t: For TSLL (10), we considered a sensitivity analysis of prior on S_t instead of the flat prior ($\propto 1$) indicated in Theorem 1. In particular, a Dirichlet process prior as well as an i.i.d. normal prior $N(0, 10^2)$ on S'_k was also used. Figure 2d shows that the posterior distribution of $\beta_{PM_{2.5}}$ remained very similar for all these priors on S'_k. So there is evidence that the results in Theorem 1 are robust with respect to prior specification on S_t.

7. Discussion

The article presents two novel ideas in the context of case-crossover studies, and it is the first treatment of the problem in a Bayesian domain. The first contribution is to study equivalence properties in terms of obtaining identical posterior inference under case-crossover and time-series analysis. The second and more important contribution is to propose different forms of full likelihood and strategies for flexible semiparametric Bayesian estimation and inference under such likelihoods. Our numerical example and simulation studies illustrate that the Bayesian specification has advantages in terms of robustness to model misspecification on the baseline risks and efficiency advantages if an informative prior is used on the risk ratio parameter. A major potential advantage for using the full likelihood could be to include individual level data. This formulation makes it possible to test for evidence of effect modification of exposure effect by individual level factors, an analysis that is not feasible under a conditional likelihood formulation.

This work leads to many other potential extensions where a Bayesian analysis may have attractive features under a case-crossover design. For example, extensions to distributed lag linear/non-linear models (Welty et al., 2009; Gasparrini, Armstrong, and Kenward, 2010), hierarchical models for meta-analysis (e.g., Dominici, Samet, and Zeger, 2000; Dominici et al., 2002), and recurrent events (Luo and Sorock, 2008) are natural directions to pursue.

8. Supplementary Materials

Web Appendix referenced in Sections 1–6, the Detroit Medicaid data analyzed in Section 6, and the R codes implementing the proposed methods in Sections 5 and 6, are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

The research of Dr. Bhramar Mukherjee and Dr. Malay Ghosh was partially supported by NSF DMS Grants 1007494 and 1007417. The research of Dr. Mukherjee was supported by NIH Grant ES 020811 and CA 156608. Dr. Stuart Batterman acknowledges support by grant P30ES017885 from the NIEHS, NIH entitled "Lifestage Exposure and Adult Disease." DAMAT study was funded in part by the USEPA grant EPA-G2007-STAR-A1 and Science to Achieve Results (STAR) Program.

References

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics* **2**, 1152–1174.

Baker, S. G. (1994). The multinomial-Poisson transformation. *The Statistician* **1**, 495–504.

Basu, R., Dominici, F., and Samet, J. M. (2005). Temperature and mortality among the elderly in the United States: A comparison of epidemiologic methods. *Epidemiology* **16**, 58–66.

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**, 539–544.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* **1**, 353–355.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **97**, 9–25.

Carracedo-Martínez, E., Taracido, M., Tobias, A., Saez, M., and Figueiras, A. (2010). Case-crossover analysis of air pollution health effects: A systematic review of methodology and application. *Environmental Health Perspectives* **118**, 1173–1182.

Dey, D. K. and Liu, J. (2007). A quantitative study of quantile based direct prior elicitation from expert opinion. *Bayesian Analysis* **2**, 137–166.

Dominici, F., Daniels, M., Zeger, S. L., and Samet, J. M. (2002). Air pollution and mortality: Estimating regional and national dose–response relationships. *Journal of the American Statistical Association* **97**, 100–111.

Dominici, F., McDermott, A., Zeger, S. L., and Samet, J. M. (2003). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology* **156**, 193–203.

Dominici, F., Samet, J. M., and Zeger, S. L. (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: A hierarchical modelling strategy. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **163**, 263–302.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.

Gasparrini, A., Armstrong, B., and Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine* **29**, 2224–2234.

Ghosh, M. and Chen, M. H. (2002). Bayesian inference for matched case-control studies. *Sankhyā: The Indian Journal of Statistics, Series B* **64**, 107–127.

Ghosh, M., Song, J., Forster, J. J., Mitra, R., and Mukherjee, B. (2012). On the equivalence of posterior inference based on retrospective and prospective likelihoods: Application to a case-control study of colorectal cancer. *Statistics in Medicine* **31**, 2196–2208.

Ghosh, M., Zhang, L., and Mukherjee, B. (2006). Equivalence of posteriors in the Bayesian analysis of the multinomial-Poisson transformation. *Metron-International Journal of Statistics* **64**, 19–28.

Ibrahim, J. G. and Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science* **15**, 46–60.

Janes, H., Sheppard, L., and Lumley, T. (2005a). Case-crossover analyses of air pollution exposure data: Referent selection strategies and their implications for bias. *Epidemiology* **16**, 717–726.

Janes, H., Sheppard, L., and Lumley, T. (2005b). Overlap bias in the case-crossover design, with application to air pollution exposures. *Statistics in Medicine* **24**, 285–300.

Levy, D., Lumley, T., Sheppard, L., Kaufman, J., and Checkoway, H. (2001). Referent selection in case-crossover analyses of acute health effects of air pollution. *Epidemiology* **12**, 186–192.

Li, S., Batterman, S., Wasilevich, E., Wahl, R., Wirth, J., Su, F. C., and Mukherjee, B. (2011). Association of daily asthma emergency department visits and hospital admissions with ambient air pollutants among the pediatric Medicaid population in Detroit: Time-series and time-stratified case-crossover analyses with threshold effects. *Environmental Research* **111**, 1137–1147.

Lu, Y. and Zeger, S. L. (2007). On the equivalence of case-crossover and time-series methods in environmental epidemiology. *Biostatistics* **8**, 337–344.

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

Luo, X. and Sorock, G. S. (2008). Analysis of recurrent event data under the case-crossover design with applications to elderly falls. *Statistics in Medicine* **27**, 2890–2901.

Maclure, M. (1991). The case-crossover design: A method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* **133**, 144–153.

Mittleman, M. A. (2005). Optimal referent selection strategies in case-crossover studies: A settled issue. *Epidemiology* **16**, 715–716.

Mukherjee, B., Ahn, J., Liu, I., and Sánchez, B. N. (2009). On elimination of nuisance parameters in stratified proportional odds model by amalgamating conditional likelihoods. *Statistics in Medicine* **27**, 4950–4971.

Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19**, 95–110.

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

Navidi, W. and Weinhandl, E. (2002). Risk set sampling for case-crossover designs. *Epidemiology* **13**, 100–105.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9**, 249–265.

Rice, K. M. (2004). Equivalence between conditional and mixture approaches to the Rasch model and matched case-control studies, with applications. *Journal of the American Statistical Association* **99**, 510–522.

Rice, K. M. (2008). Equivalence between conditional and random-effects likelihoods for pair-matched case-control studies. *Journal of the American Statistical Association* **103**, 385–396.

Seaman, S. R. and Richardson, S. (2004). Equivalence of prospective and retrospective models in the Bayesian analysis of case-control studies. *Biometrika* **91**, 15–25.

Sinha, S., Mukherjee, B., and Ghosh, M. (2004). Bayesian semiparametric modeling for matched case–control studies with multiple disease states. *Biometrics* **60**, 41–49.

Staicu, A. M. (2010). On the equivalence of prospective and retrospective likelihood methods in case-control studies. *Biometrika* **97**, 990–996.

Welty, L. J., Peng, R. D., Zeger, S. L., and Dominici, F. (2009). Bayesian distributed lag models: Estimating effects of particulate matter air pollution on daily mortality. *Biometrics* **65**, 282–291.

Received November 2012. *Revised July* 2013.
Accepted August 2013.