Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

# Measurement Error Case Series Models With Application to Infection-Cardiovascular Risk in Older Patients on Dialysis

Sandra M. MOHAMMED, Damla ŞENTÜRK, Lorien S. DALRYMPLE, and Danh V. NGUYEN

Infection and cardiovascular disease are leading causes of hospitalization and death in older patients on dialysis. Our recent work found an increase in the relative incidence of cardiovascular outcomes during the ∼30 days after infection-related hospitalizations using the case series model, which adjusts for measured and unmeasured baseline confounders. However, a major challenge in modeling/assessing the infection-cardiovascular risk hypothesis is that the exact time of infection, or more generally "exposure," onsets cannot be ascertained based on hospitalization data. Only imprecise markers of the timing of infection onsets are available. Although there is a large literature on measurement error in the predictors in regression modeling, to date, there is no work on measurement error on the timing of a time-varying exposure to our knowledge. Thus, we propose a new method, the measurement error case series (MECS) models, to account for measurement error in time-varying exposure onsets. We characterized the general nature of bias resulting from estimation that ignores measurement error and proposed a bias-corrected estimation for the MECS models. We examined in detail the accuracy of the proposed method to estimate the relative incidence. Hospitalization data from the United States Renal Data System, which captures nearly all (>99%) patients with end-stage renal disease in the United States over time, are used to illustrate the proposed method. The results suggest that the estimate of the cardiovascular incidence following the 30 days after infections, a period where acute effects of infection on vascular endothelium may be most pronounced, is substantially attenuated in the presence of infection onset measurement error.

KEY WORDS: Cardiovascular outcomes; Case series models; End-stage renal disease; Infection; Measurement error; Nonhomogeneous Poisson process; Time-varying exposure onset; United States Renal Data System.

## 1. INTRODUCTION

Infection and cardiovascular disease are leading causes of hospitalization and death in patients on dialysis in the United States (United States Renal Data System [USRDS] 2010). Smeeth et al. (2004) found that infection is associated with an increased risk of cardiovascular events, using data from the United Kingdom General Research Practice Database (Walley and Mantgani 1997), the largest source of ongoing data on illness and practice in the United Kingdom. More specifically, they showed that there is a three- to five-fold increase in risk (incidence) of a myocardial infarction or stroke after infection in the general population. Although the precise mechanisms by which infection may affect cardiovascular events are not fully known, infections may affect vascular endothelium (Mahmoudi, Curzen, and Gallagher 2007), create a chronic subclinical inflammatory state that affects atherosclerosis (Zhu et al. 2000), or create a procoagulant state (Macko et al. 1996; Sun 2006).

Only recently has the association between acute infections and cardiovascular events been examined in the (United States) dialysis population, using data from the USRDS that captures nearly all (>99%) patients with end-stage renal disease in the United States. More specifically, Dalrymple et al. (2011) found an increased incidence of cardiovascular events, particularly in the first 30-day risk period after infection-related hospitalizations in older dialysis patients (age $\geq$ 65). Infection, or more generally the "exposure" of interest, was observed over time during the follow-up/observation period for each individual; therefore, the exposure was time varying. The case series model, also called self-controlled case series model (Farrington 1995), was used to estimate the association between the incidence of cardiovascular events and the time-varying exposure, namely infection. The approach uses only cases, that is, individuals with one or more cardiovascular events. There are several important reasons and appealing aspects to this modeling choice. First, the case series method provides consistent estimates of the relative incidence (e.g., cardiovascular events during the 30-day risk period following an infection relative to the control period over the observation time) using only cases. Second, it implicitly controls for all fixed confounders, measured and unmeasured, such as genetics or coexisting illnesses. This latter point is particularly relevant to the dialysis cohort from the USRDS, as dialysis patients who do and do not acquire infections likely differ in important ways not easily measured, therefore making adjustment in (e.g., Poisson) regression modeling infeasible. Third, the case series model is well suited for this type of hypothesis-driven research because (1) it requires specification of the risk period(s) a priori and (2) leads to a specific time window to potentially

Sandra M. Mohammed is Ph.D. Student, Division of Biostatistics, University of California, Davis, CA 95616. Damla Şentürk is Assistant Professor, Department of Biostatistics, University of California, Los Angeles, CA 90095. Lorien S. Dalrymple is Assistant Professor, Division of Nephrology, Department of Medicine, University of California, Sacramento, CA 95691. Danh V. Nguyen (E-mail: *ucdnguyen@ucdavis.edu*) is Professor, Division of Biostatistics, University of California, Davis, CA 95616. This publication was made possible by grant UL1 RR024146 from the National Center for Research Resources and NIH K12 grant through the UC Davis Clinical and Translational Science Center (DVN, LSD) and partially by NIDDK (National Institute of Diabetes and Digestive and Kidney Diseases) grant R01DK092232 (DS, LSD, DVN). We thank Barbara Grimes, Department of Biostatistics, University of California, San Francisco, and Yi Mu at UC Davis Department of Public Health Sciences. We are grateful to two reviewers and an associate editor for thoughtful suggestions, which improved the article. The interpretation and reporting of the data presented here are the responsibility of the authors and in no way should be seen as an official policy or interpretation of the U.S. government. This study was approved by the Institutional Review Board of the University of California Davis Health System.



DOI: 10.1080/01621459.2012.695648

examine more aggressive intervention monitoring, such as the first 30 days after exposure/infection, for cardiovascular risk reduction, in addition to implementation of overall infection control/prevention strategies and early treatment in this population.

The case series model was originally proposed in 1995 to investigate the association between an acute outcome and a time-varying exposure, more specifically, adverse events and vaccination. As mentioned above, only individuals with one or more events are sampled in the case series model. It is derived from an underlying nonhomogeneous Poisson cohort model where events and exposure history are available on the observation period $(a_i, b_i]$ for the $i$th case/individual and with incidence rate $\lambda_{ijk} = \exp(\varphi_i + \alpha_j + \beta_k)$, where $\varphi_i$, $\alpha_j$, and $\beta_k$ are the individual-specific, $j$th age group, and $k$th risk group effects, respectively. The primary effects of interest are the $\beta_k$'s. To take into account the deliberate sampling on cases only, the model likelihood is conditional on an event having occurred and the resulting kernel of the likelihood is multinomial with probabilities depending on the incidence rate $\lambda_{ijk}$. The "self-controlled" aspect of the method refers to the fact that individual effects $\varphi_i$ cancel in the likelihood. The exposure history, that is, the times when the exposures occurred, within the observation period of individual $i$, namely $(a_i, b_i]$, is assumed to be known precisely. For example, in the original application of the case series model, there is little doubt as to when a vaccination occurs as those can be ascertained fairly accurately.

However, a major challenge associated with using the USRDS hospitalization data to address the infection-cardiovascular risk hypothesis is that the exact date or time of infection/exposure onset cannot be ascertained based on hospital claims data, although the discharge date is a surrogate marker for the time of infection as it reasonably assures that the infection has occurred by this date. Thus, our previous work used the date of infection-related hospitalization discharge as the observed time of infection (Dalrymple et al. 2011). Clearly, this is a conservative approximation to the true unknown date of infection, which most likely occurred sometime during hospitalization or prior to the start of hospitalization. From a more general perspective, this can be viewed as a problem of exposure onset measurement error, where one only observes a marker of the unknown time of exposure. The lack of an existing method that can handle this exposure onset measurement error leads us to propose the measurement error case series (MECS) model, to more thoroughly assess the infection-cardiovascular risk hypothesis in the dialysis population. Thus, the proposed MECS model in this work aims to target the true underlying relative incidences using imprecise exposure onset times and still retain the advantages associated with the original case series model.

We note that exposure onset measurement error is distinct from traditional measurement error in the form of mismeasured continuous or misclassified categorical variables (e.g., Carroll et al. 2006). There is indeed an extensive literature on methods for dealing with measurement error in the *covariates*, including time-varying covariates, in general regression modeling. For example, modeling measurement error in time-varying covariates, such as longitudinal dietary intakes (e.g., from food frequency questionnaire [FFQ]), in linear mixed models and survival analysis were considered by Tosteson et al. (1999) and Liao et al. (2011), respectively. Measurement error in time-dependent covariates, such as growth hormone and binding protein levels, in pharmacokinetics nonlinear mixed-effects models were considered by Higgins, Davidian, and Giltinan (1997). Other works include Huang and Wang (2000) and Tsiatis and Davidian (2004), both in the context of time-to-event data. In the literature on time-varying covariate measurement error, including the above referenced works, the main issue is that the longitudinal covariate measurements themselves, for example, dietary/nutrient intakes or protein concentrations, are measured with error. The *timing* of the covariate measurements, such as when the FFQs were administered or hourly measurements of serum protein concentration, is not in doubt. Our work here focuses on measurement error in the timing of when the exposures occur over time, and to date, there has been no work to handle measurement error in the *timing* of exposure onset in case series modeling. There are several works, unrelated to case series modeling and exposure onset measurement error, that involve the "timing error" of covariates, whose meaning differs completely from the current work. For example, in Higgins, Davidian, and Giltinan (1997), this simply refers to covariates measured at different time points than the response measurement times. Li and Ryan (2004) considered "mistiming error," which refers to using a covariate measured at a later *known* time, such as some known time after birth or baseline, in place of an intended covariate at birth, which is not available in a Cox regression model. See Keiding (1992), who also considered mistimed covariates. We note that in mistimed covariates, the timing of the covariate is not in doubt (error). One simply uses the covariate measured at a different known time; hence, the timing of the available covariate itself is not subject to error.

Some interesting case series applications involve antidepressant use and hip fracture (Hubbard et al. 2003) and prescription medications and motor vehicle crashes (Gibson et al. 2009), although the original novel proposal by Farrington (1995) was for investigating associations between time-varying transient exposures, specifically vaccinations, and adverse events. It is increasingly recognized as an important method in the analysis of data from biomedical and epidemiological studies. Excellent expository articles on the practice and implementation of case series modeling are provided by Whitaker et al. (2006; Whitaker, Hocine, and Farrington 2009).

The remainder of this article is organized as follows. In Section 2, we introduce the MECS model for exposure onset measurement error and investigate the bias when ignoring the measurement error. We also propose a bias-correction method in Section 2.3 that requires only fitting the ordinary case series model, ignoring measurement error. In Section 2.4, we examine the accuracy of the proposed bias-correction method in estimating the true relative incidence of interest. The efficacy of the proposed approach is examined in extensive simulation studies in Section 3 and these studies also reveal the nature of bias under different patterns of true exposure effects. We present, for the first time, an assessment of the infection-cardiovascular risk association in the dialysis population taking into account the imprecise exposure/infection onset data in Section 4. Finally, Section 5 provides a discussion of our proposed MECS models and findings.

## 2. MEASUREMENT ERROR CASE SERIES MODELS FOR IMPRECISE EXPOSURE ONSET TIMES

### 2.1 The Model

The self-controlled case series model was proposed by Farrington (1995) and was originally designed to estimate the relative incidence of acute events following transient exposures. It is a retrospective cohort method based on a conditional Poisson model requiring cases only and is self-controlled, since all time-invariant/age-independent confounders are implicitly controlled. The case series method is derived by conditioning on the occurrence of an event and the individual's exposure history during a fixed observation period, where event counts arise from a nonhomogeneous Poisson process. For a cohort of $N$ individuals ($i = 1, \ldots, N$) with one or more events, let $(a_i, b_i]$ denote the observation period over time for subject $i$ (e.g., age in days) which is partitioned into age intervals (groups) $j = 0, \ldots, J$ and exposure risk periods $k = 0, \ldots, K$. The baseline or control risk period corresponds to $k = 0$, which consists of all time periods outside of the exposure risk periods. Similarly, the reference age group is $j = 0$. For example, in the infection-cardiovascular example introduced earlier (Dalrymple et al. 2011), the risk periods are 1–30, 31–60, and 61–90 days ($K = 3$) after an infection (and the baseline period consists of observation times outside of the three risk periods). Figure 1(a) illustrates the follow-up data for a subject.

The case series method compares the incidence within a risk period relative to the incidence in the baseline period, within each individual. Let the length of time individual $i$ spends in age group $j$ and risk period $k$ be $e_{ijk}$. Given the exposure history over the observation period for individual $i$, the number of events in each interval, denoted as $n_{ijk}$, is assumed to follow a nonhomogeneous Poisson process with rate $\lambda_{ijk} = \exp(\varphi_i + \alpha_j + \beta_k)$,

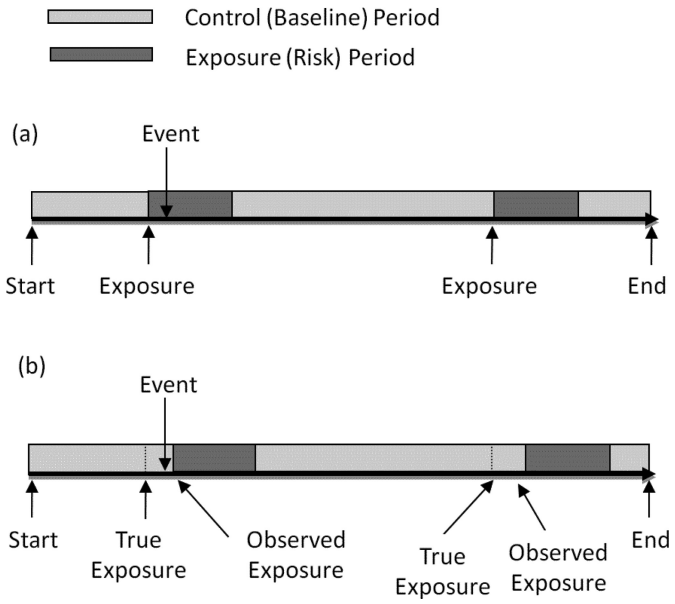


Figure 1. Example of follow-up data for one subject (a) without and (b) with exposure onset measurement error. Note in (b) that when exposure onset measurement error is present, the event is now observed in the baseline period when in truth the event occurs in the exposure period, attenuating the relative incidence in this case.

that is, $n_{ijk} \sim \text{Poisson}(e_{ijk}\lambda_{ijk})$. Here the parameters $\varphi_i$, $\alpha_j$, and $\beta_k$ are, respectively, the individual-specific, $j$th age group relative to age group $j = 0$, and $k$th risk group relative to baseline period $k = 0$ effects, with $\alpha_0 = \beta_0 = 0$. The parameters of primary interest are $\beta_k$, $k = 1, \ldots, K$, the log relative incidences for the exposure risk periods. The case series model is obtained by conditioning on the event $n_{i\cdot\cdot} = \sum_{jk} n_{ijk} \geq 1$, where $n_{i\cdot\cdot}$ is the total number of events for individual $i$. As shown in Farrington (1995), the kernel of the case series likelihood is product multinomial, with the contribution from subject $i$ given as

$$L_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{j,k} \pi_{ijk}^{n_{ijk}} \tag{1}$$

with probabilities

$$\begin{aligned}\pi_{ijk} &= \frac{e_{ijk}\lambda_{ijk}}{\sum_{rs} e_{irs}\lambda_{irs}} = \frac{e_{ijk}\exp(\varphi_i + \alpha_j + \beta_k)}{\sum_{rs} e_{irs}\exp(\varphi_i + \alpha_r + \beta_s)} \\ &= \frac{e_{ijk}\exp(\alpha_j + \beta_k)}{\sum_{rs} e_{irs}\exp(\alpha_r + \beta_s)},\end{aligned} \tag{2}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$. The individual effects $\varphi_i$ cancel out, thus, self-controlling for all fixed covariates. We refer the reader to Farrington (1995) for details and also to the excellent expository articles by Whitaker et al. (2006; Whitaker, Hocine, and Farrington 2009) on the application of case series models, including study planning guidance and assessment of model assumptions.

Next, we introduce the model for exposure onset measurement error, motivated by imprecise infection onset when using USRDS hospitalization data. Because exposure onset measurement error is largely inconsequential to estimation of age effects, we focus here on the exposure effects of primary interest (i.e., the $\beta's$); therefore, we drop subscript $j$. Also, we consider here, in more detail, a positive measurement error model because it is relevant to our current application, where we know that each exposure/infection must occur prior to the observed discharge date of an infection-related hospitalization. Thus, we consider the following positive additive exposure onset measurement error model,

$$w_{il} = v_{il} + u_{il}, \quad l = 1, \ldots, L_i, \tag{3}$$

where $w_{il}$ is the observed exposure onset time, $v_{il}$ is the true (unobserved) exposure onset time, $u_{il}$ is a positive measurement error ($u_{il} > 0$) with mean $\mu_u = E(u_{il})$ and variance $\sigma_u^2 = \text{var}(u_{il})$, and $L_i$ is the number of exposures observed for individual $i$. For our current application, $w_{il}$ is the infection-related discharge time. Figure 1(b) illustrates the effects of exposure onset measurement error. We assume that the amount of measurement error in the exposure time ($u_{il}$) is less than the risk period length of interest. For example, if the relative incidence of events associated with the 30-day period after an infection is of interest, then the uncertainty in the time when the infection actually occurred should not exceed 30 days. If $u_{il} > 30$ days, then one cannot estimate the relative incidence in the 30-day period after an infection, because $u_{il} > 30$ amounts to not having any reliable data for estimation. Thus, this assumption on the magnitude of the measurement error essentially ensures that there must be some amount of reliable data for estimation.

We refer to (1)–(3) as the MECS model. As in the case of classical measurement error problems, naive estimation

ignoring measurement error will be biased. We characterize the target of the naive estimation, and hence the bias, in Section 2.2. For the MECS models, given $\mathbf{u}_i = (u_{i1}, \ldots, u_{iL_i})$, $\tilde{\mathbf{n}}_i = (\tilde{n}_{i0}, \ldots, \tilde{n}_{iK}) \sim \text{Mult}(n_{i\cdot}, \tilde{\boldsymbol{\pi}}_i(\mathbf{u}_i))$, where $\tilde{n}_{ik}$ is the observed number of (e.g., cardiovascular) events in risk period $k$ $(k = 0, \ldots, K)$ and $\tilde{\boldsymbol{\pi}}_i \equiv \tilde{\boldsymbol{\pi}}_i(\mathbf{u}_i) = (\tilde{\pi}_{i0}(\mathbf{u}_i), \ldots, \tilde{\pi}_{iK}(\mathbf{u}_i))$ are modified probabilities depending on measurement error $\mathbf{u}_i$. More specifically, we show in the Appendix that $\tilde{\pi}_{ik}$ is a function of the length of the $k$th risk period, the true underlying rate $\lambda_{ik}$, and a mixture of true rates corresponding to control and risk periods and measurement error.

We provide the following result (Theorem 1) needed to fully characterize the bias resulting from ignoring exposure onset measurement error in Sections 2.2 and 3. For this purpose, we further introduce the following notation. Among individuals with $L_i \geq 2$ exposures, let $L_i'$ denote the number of (disjoint) risk segments ($L_i' \leq L_i$). For example, suppose that individual $i$ has $L_i = 5$ infections, the risk period of interest is 1–30 days after an infection, infection 3 occurs within the 30-day risk period after infection 2 (i.e., the 30-day risk periods for infections 2 and 3 overlap), and similarly, infection 5 occurs within 30 days of infection 4. Then we have $L_i' = 3$ risk segments defined by infection 1, infections 2 and 3, and infections 4 and 5. We refer to the later two risk segments as "overlapping" risk segments. Thus, we have the following result for the general MECS model and a special case with potential overlapping risk segments. It is assumed that risk periods are adjacent following an exposure, for instance, 1–30, 31–60, and 61–90 days following an exposure. We defer the proof to the Appendix.

*Theorem 1.* Under the general MECS models (1)–(3) with $L_i$ exposures and nonoverlapping risk segments,

$$E(\tilde{n}_{ik}) = \begin{cases} n_{i\cdot}\Delta_i^{-1}\{e_{ik}\lambda_{ik} + L_i\mu_u(\lambda_{i,k+1} - \lambda_{ik})\}, \\ \quad k = 0, 1, \ldots, K-1 \\ n_{i\cdot}\Delta_i^{-1}\{e_{iK}\lambda_{iK} + L_i\mu_u(\lambda_{i0} - \lambda_{iK})\}, \\ \quad k = K \end{cases}, \quad (4)$$

where $\Delta_i = \sum_r e_{ir}\lambda_{ir}$. Furthermore, for MECS models with one risk period, individuals with multiple exposures, and possibly with overlaps, $E(\tilde{n}_{ik})$ in (4) holds with $L_i$ replaced by $L_i'$, the number of disjoint risk segments.

*Remarks.* Our studies in Section 3 show that the MECS model with nonoverlapping risk segments leads to the most severe bias on average; therefore, this case is of particular relevance to understanding the extent and nature of bias when ignoring exposure onset measurement error. Overlapping risk segments are relatively rare, although we also examine their effects on estimation bias as well. We discuss the above result for common MECS models (e.g., with $K = 1$ risk period) when there are possibly overlapping risk segments in the Appendix. Also, although not directly applicable to our specific application in this work, the risk periods need not be adjacent in the case series model generally. Equation (4) can be extended to this case and we discuss this in the Appendix section.

In the special case with one risk group ($K = 1$), we have that (4) reduces to

$$\begin{aligned} E(\tilde{n}_{i1}) &= n_{i\cdot}\frac{e_{i1}\lambda_{i1} + L_i\mu_u(\lambda_{i0} - \lambda_{i1})}{e_{i0} + e_{i1}\lambda_{i1}} \\ &= n_{i\cdot}\frac{e_{i1}\exp(\beta_1) + L_i\mu_u(1 - \exp(\beta_1))}{e_{i0} + e_{i1}\exp(\beta_1)}. \end{aligned}$$

Thus, from the above equality, we note that $\lim_{\mu_u \to 0} E(\tilde{n}_{i1}) = n_{i\cdot}\pi_{i1}$, providing an approach to bias correction in Section 2.3.

## 2.2 Bias When Ignoring Exposure Onset Measurement Error

We consider here the bias that results from naive case series model estimation without accounting for exposure onset measurement error. Together with Theorem 1, the result described in this section is important for two purposes. First, it will be used to characterize the general nature of the bias when ignoring exposure time measurement error in Section 3. Second, it will be used to determine the accuracy of the proposed bias-correction method in estimating the true relative incidence of interest in Section 2.4.

Consider the observed data $\{\tilde{\mathbf{n}}_i, \mathbf{w}_i, \mathbf{e}_i\}$, where $\mathbf{w}_i = (w_1, \ldots, w_{L_i})$ is the vector of observed exposure times for the $i$th subject. The conditional maximum likelihood estimator (MLE) for the case series model, denoted $\widehat{\boldsymbol{\beta}}^*$, is obtained as a solution to the set of likelihood equations

$$N^{-1}\sum_{i=1}^{N}(\tilde{n}_{ik} - n_{i\cdot}\widehat{\pi}_{ik}^*) = 0, \quad k = 1, 2, \ldots, K, \quad (5)$$

where $\widehat{\pi}_{ik}^* = e_{ik}\exp(\widehat{\beta}_k^*)/\sum_r e_{ir}\exp(\widehat{\beta}_r^*)$. Details are provided in the section A.2: Estimation via Newton–Raphson. The MLE is consistent for $\boldsymbol{\beta}^*$, which satisfies the estimating equations in (5) in expectation, that is, $\boldsymbol{\beta}^*$ is a solution to

$$N^{-1}\sum_{i=1}^{N}\left\{E(\tilde{n}_{ik}) - n_{i\cdot}\pi_{ik}^*\right\} = 0, \quad k = 1, 2, \ldots, K, \quad (6)$$

with $\pi_{ik}^* = e_{ik}\exp(\beta_k^*)/\sum_r e_{ir}\exp(\beta_r^*)$. The bias of the naive case series model ignoring measurement error is the difference between $\boldsymbol{\beta}^*$ (the target of $\widehat{\boldsymbol{\beta}}^*$) and $\boldsymbol{\beta}$. To solve (6), we note that substituting the expression for $E(\tilde{n}_{ik})$ from (4) gives

$$\begin{aligned} a_k \equiv N^{-1}\textstyle\sum_{i=1}^{N} n_{i\cdot}\left[\Delta_i^{-1}\{e_{ik}\lambda_{ik} + L_i\mu_u(\lambda_{i,k+1} - \lambda_{ik})\} - \pi_{ik}^*\right] = 0, \\ k = 1, \ldots, K-1 \\ a_K \equiv N^{-1}\textstyle\sum_{i=1}^{N} n_{i\cdot}\left[\Delta_i^{-1}\{e_{iK}\lambda_{iK} + L_i\mu_u(\lambda_{i0} - \lambda_{iK})\} - \pi_{iK}^*\right] = 0, \\ k = K. \end{aligned}$$

This set of equations can be solved numerically for $\boldsymbol{\beta}^*$ by the Newton–Raphson method where the update of $\boldsymbol{\beta}^*$ at iteration $t+1$ is $\boldsymbol{\beta}^{*(t+1)} = \boldsymbol{\beta}^{*(t)} - (\mathbf{J}^{(t)})^{-1}\mathbf{a}^{(t)}$, with $\mathbf{a}^{(t)} = (a_1^{(t)}, \ldots, a_K^{(t)})^{\mathrm{T}}$ and $\mathbf{J}^{(t)}$ is a $K \times K$ matrix of partial derivatives evaluated at $\boldsymbol{\beta}^{(t)}$;

$$\begin{aligned} \frac{\partial a_k}{\partial \beta_k^*} &= -N^{-1}\textstyle\sum_{i=1}^{N} n_{i\cdot}\pi_{ik}^*(1 - \pi_{ik}^*), \quad k = 1, \ldots, K \\ \frac{\partial a_k}{\partial \beta_l^*} &= N^{-1}\textstyle\sum_{i=1}^{N} n_{i\cdot}\pi_{ik}^*\pi_{il}^*, \qquad k \neq l. \end{aligned}$$

We use the above theoretical result to study the nature of bias due to exposure onset measurement error in the simulation studies of Section 3 for various patterns of exposure effects and realistic data scenarios. However, some insights are immediate from the above result with the simplifying assumption that the length of observation periods (and risk periods) and the number of exposures are the same for all subjects. For instance, we have from (6)

$$N^{-1}\sum_{i=1}^{N} n_{i\cdot}\frac{e_{i1}\exp(\beta)+L_i\mu_u(1-\exp(\beta))}{e_{i0}+e_{i1}\exp(\beta)} - N^{-1}\sum_{i=1}^{N} n_{i\cdot}\frac{e_{i1}\exp(\beta^*)}{e_{i0}+e_{i1}\exp(\beta^*)}=0,$$

where a closed form expression for $\beta^*$ is possible under equal observation and risk periods for all subjects. Thus, in the simple, but illustrative, case where $K = 1$ (with $L_i = 1$; dropping subscript $k$), we obtain

$$\beta^* = \log\left\{\frac{e_1\exp(\beta)+\mu_u(1-\exp(\beta))}{e_0-\mu_u(1-\exp(\beta))}\right\} - \log(e_1/e_0). \quad (7)$$

From Equation (7), the bias in this special case is apparent; there is increasing attenuation of the true relative incidence, $\exp(\beta)$, as the mean of the exposure onset measurement error, $\mu_u$, increases. However, the nature of the bias generally is not always attenuation, as we will demonstrate subsequently in Section 3. More precisely, if $K = 1$, the naive estimate will be attenuated, but with $K > 1$ risk periods, the bias can be in either directions, depending on the pattern of the true relative incidences of the risk periods. We note that the target of the naive MLE $\widehat{\beta}^*$, namely $\beta^*$ given in (7), follows from (4), the law of large numbers and Slutsky's theorem, since it is straight-forward to show that $\widehat{\beta}^* = \log(\tilde{n}_{\cdot 1}/\tilde{n}_{\cdot 0}) - \log(e_1/e_0)$. Here $\tilde{n}_{\cdot k} = \sum_i \tilde{n}_{ik}$, $k = 0, 1$.

## 2.3 Bias-Corrected Estimation Procedure

To motivate our proposed bias-corrected estimation procedure, consider the target of the naive (conditional) MLE, namely $\beta^*$ given in (7). Note that $\lim_{\mu_u\to 0}\beta^* = \beta$, the true unknown relative incidence parameter of interest. Thus, we propose a practical case series bias-correction procedure where the pattern of bias as a function of increasing amounts of *exposure onset measurement error* ($\mu_u$) is determined/estimated and then extrapolated to the ideal case of no measurement error in the time of exposure. We assume that an estimate of the average amount of exposure onset measurement error $\mu_u$ is available.

The simple steps of this correction procedure are as follows. First, obtain datasets with increasing exposure onset measurement error, that is, with increasing mean $\mu_j = \mu_u + \tau_j$, by adding a sequence of constants $\tau_j$, $j = 0, 1, \ldots, M$ to the observed exposure times where $0 = \tau_0 < \tau_1 < \cdots < \tau_M$ and $\tau_0$ refers to the observed data. Next, for each $j$, compute the naive MLE, $\widehat{\boldsymbol{\beta}}^{*(j)}$, by applying the standard case series model, ignoring exposure time measurement error. For a given estimate of $\mu_u$, fit a least squares regression of $\widehat{\boldsymbol{\beta}}^*_k = (\widehat{\beta}^{*(0)}_k, \widehat{\beta}^{*(1)}_k, \ldots, \widehat{\beta}^{*(M)}_k)^{\mathrm{T}}$ on $\mu_j$ and the bias-corrected estimator is taken to be the extrapolated value at $\mu_j = 0$, that is, the regression intercept. More precisely, define the $(M+1)\times(\nu+1)$ fixed predictor matrix by $\mathbf{D}_\mu$, then the estimated coefficient of the regression fit is $\widehat{\boldsymbol{\alpha}}_k = (\widehat{\alpha}_{k0}, \widehat{\alpha}_{k1}, \ldots, \widehat{\alpha}_{k\nu})^{\mathrm{T}} = (\mathbf{D}_\mu^{\mathrm{T}}\mathbf{D}_\mu)^{-1}\mathbf{D}_\mu^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^*_k$, $k = 1, \ldots, K$. Thus, for bias correction, we propose to use the extrapolated value at $\mu_j = 0$, that is,

$$\widehat{\beta}_k = \widehat{\alpha}_{k0} = \mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^*_k, \quad k = 1, 2, \ldots, K, \quad (8)$$

where the vector of constants $\mathbf{c} = (c_0, c_1, \ldots, c_M)^{\mathrm{T}}$ is the first row of the $(\nu+1)\times(M+1)$ matrix $(\mathbf{D}_\mu^{\mathrm{T}}\mathbf{D}_\mu)^{-1}\mathbf{D}_\mu^{\mathrm{T}}$. As we will detail in Section 2.4, it is adequate in practice to consider a quadratic regression fit ($\nu = 2$) with $\mathbf{D}_\mu$ containing predictors $\mu_j$ and $\mu_j^2$.

First, note that although on the surface the proposed bias-corrected estimator appears to resemble the simulation and extrapolation (SIMEX; Cook and Stefanski 1994) method for classical measurement error in the *predictors* (Carroll et al. 2006), it is quite distinct. SIMEX simulates additional datasets with increasing measurement error *variance* in the predictors. Such an approach is not appropriate for measurement error in the time of exposure onset. This is clear when we consider a salient feature of the bias analysis described in Section 2.2, which is that the bias does not depend on the exposure time measurement error variance, as evident from (7). This facet generally holds, as illustrated in Figure 2, for the more general MECS model described above. Thus, a SIMEX approach is not appropriate to bias correction for measurement error in the time of exposure. Second, note that there is no simulation involved in our proposed estimation.

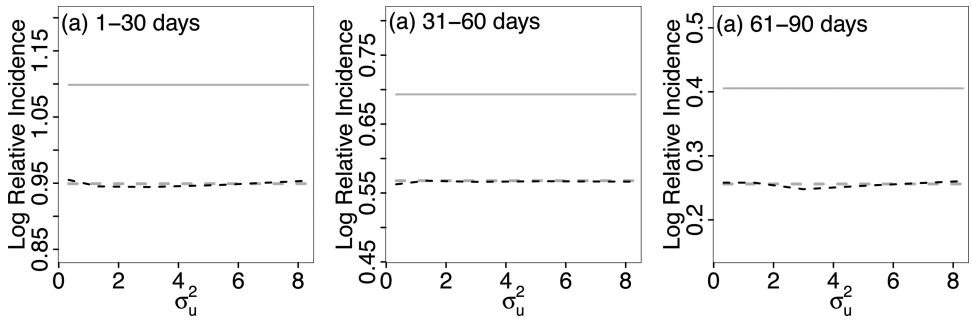


Figure 2. Property of exposure onset time measurement error. Bias does not depend on the variance of the measurement error distribution. Dashed black curves denote naive case series estimates ignoring measurement error, which targets $\boldsymbol{\beta}^*$ (dashed gray curves) instead of $\boldsymbol{\beta}$ (solid gray line).

We also note that given an estimate of $\mu_u$ based on auxiliary data or even knowledge of $\mu_u$ exactly, direct application of the case series model to the data where one subtracts this estimate from each observed exposure onset time will not correctly target the true relative incidences. For this reason, we developed the above estimator where extrapolation is made from a sequence of datasets with additively higher mean measurement error. We show this explicitly in the Appendix and relate this to the more general cases of exposure onset measurement error where $u$ is not strictly positive.

### 2.4 Accuracy of the Bias-Correction Method

In this section, we evaluate the asymptotic accuracy of the bias-corrected estimation procedure to target the relative incidence $\theta_k \equiv \exp(\beta_k)$ of interest. More precisely, we investigate the approximation error in using $\widehat{\theta}_k \equiv \exp(\widehat{\beta}_k) = \exp(\widehat{\alpha}_{k0}) = \exp(\mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k^*)$ as an estimator of $\theta_k$. First, note that since $\widehat{\boldsymbol{\beta}}_k^*$ is consistent for $\boldsymbol{\beta}_k^*$, a solution to (6), we have that $\widehat{\beta}_k = \widehat{\alpha}_{k0} = \mathbf{c}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k^*$ is consistent for $\mathbf{c}^{\mathrm{T}}\boldsymbol{\beta}_k^*$. Because $\boldsymbol{\beta}_k^*$ is a solution to (6), note that it is a function of the true effects $\boldsymbol{\beta}^{\mathrm{T}} = (\beta_1, \ldots, \beta_K)$, the average exposure time measurement error $\mu_u$, and the observation lengths in the risk periods and baseline period $\{e_{ik}\}$ across individuals $i = 1, \ldots, N$. Thus, the accuracy depends on these parameters generally. Therefore, to assess accuracy in approximating the true relative incidence $\theta_k$ with the limit $\theta_k^*(\boldsymbol{\beta}) = \exp(\mathbf{c}^{\mathrm{T}}\boldsymbol{\beta}_k^*)$, we consider the maximum absolute relative approximation error (maxARE), defined as a function of the true relative incidences $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$,

$$\max \mathrm{ARE} = \max_{\boldsymbol{\beta}} \left\{ \left| \frac{\theta_k - \theta_k^*(\boldsymbol{\beta})}{\theta_k} \right|, k = 1, \ldots, K \right\}. \quad (9)$$

Note that we used the notation $\theta_k^*(\boldsymbol{\beta})$ to emphasize the dependence on the true parameters $\boldsymbol{\beta}$. We first consider the simple MECS model, with a single risk period $K = 1$, equal risk period $e_{i1} = e_1$, and equal follow-up times, as it illustrates clearly the factors affecting the maximum approximation error. For this model (dropping subscript $k$), we have $\beta^{*(j)} = \log(a + \mu_j) - \log(b - \mu_j) - \log(e_1/e_0)$, where $a = e_1 \exp(\beta)/(1 - \exp(\beta))$ and $b = e_0/(1 - \exp(\beta))$. Considering a third order Taylor approximation to $\log(\cdot)$ at $\mu_j = 0$ gives

$$\beta^{*(j)} \approx \{\log(a) + \mu_j/a\} + \{\log(b) - \mu_j/b\} - \log(e_1/e_0).$$

It follows that for a quadratic regression fit (see Section 2.3), we have

$$\mathbf{c}^{\mathrm{T}}\boldsymbol{\beta}^* = \sum_{j=0}^{M} c_j \beta^{*(j)} \approx \log(a) - \log(b) - \log(e_1/e_0) = \beta, \quad (10)$$

since $a/b = (e_1/e_0)\exp(\beta)$, $\sum_{j=0}^{M} c_j = 1$ and $\sum_{j=0}^{M} c_j \mu_j = \sum_{j=0}^{M} c_j \mu_j^2 = 0$.

The error in the above approximation involves $\sum_{j=0}^{M} c_j \mu_j^3$, $\sim a^{-3}$, and $\sim b^{-3}$. Thus, the error in (10) will depend on two main factors: the effect size $\beta$ and the average level of exposure onset time measurement error relative to the risk period length $\mu_u/e_1$, as well as the relative risk period length to the baseline period length $e_1/e_0$. Therefore, we examine the error in the approximation with respect to these parameters. We examine the maximum absolute relative error (maxARE) as given in (9). To encompass all reasonable relative incidences of interest in real applications, we evaluate the accuracy for a wide range of true relative incidences, ranging from a tiny effect size of 1% ($\theta = 1.01$) to a very large effect size of 1000% (10-fold; $\theta = 10$). For equal observed risk length $e_1 = 30$ days after an exposure during 700 days of follow-up ($e_1/e_0 \approx 4.5\%$), and with the average level of exposure onset time measurement error of $\mu_u = 4$, 6, and 8 days (i.e., $\mu_u/e_1 \sim 13.3\%$, 20%, and 26.7%, respectively), the maxARE are 0.48%, 1.27%, and 2.84%; thus, even with the higher relative measurement error of nearly 27%, the maximum error of $\theta^*(\beta)$ is less than 3% of the true relative incidence $\theta$. We note that the 2.84% maximum error is associated with the extremely large effect size of $\theta = 10$; therefore, the maximum relative error is quite low.

More generally, when allowing for varying lengths of the observed risk period for each individual and varying number of exposures per person, the maximum relative approximation error is similarly low compared with the simpler case of equal risk observation lengths considered above. For this, we solve Equation (6) for $\beta^*$ using the Newton–Raphson method described in Section 2.2, and then evaluate (9) to determine the maximum error over the range of $\beta$ values. Similar to the simple model considered above, the maxARE over all values of $\beta$ are 0.52%, 1.34%, and 3.00% corresponding to moderate to high relative average exposure onset time measurement error, that is, average $\mu_u/e_1 \sim 13.3\%$, 20%, and 27%. The average time in the risk period across individuals is 48 days and the average percent of time spent in the risk period relative to the baseline period is ~7.5%. These parameters are similar to our data application in Section 4, where the relative incidences of cardiovascular events in the 30 days following infection were estimated to be ~1.5 to 1.8, corresponding to average exposure onset time measurement error of $\mu_u = 4$ and $\mu_u = 8$ days. Under this setting, the approximation error is negligible: 0.02% and 0.20% for $\mu_u = 4$ and 8 days, respectively.

The maximum approximation error is similarly low when extending to models with multiple risk groups, examined in more details in Section 3. For example, with average relative exposure time measurement error of 20%, the maximum relative error is ~5% for tiny effect sizes of $\theta = 1.01$ to 1.05, and it reduces to ~1.5% otherwise. In summary, the limiting value $\theta_k^*(\boldsymbol{\beta})$ is close to the true effect $\theta_k$ for all reasonable range of $\beta_k$ and average relative exposure onset time measurement error. However, $\theta_k^*(\boldsymbol{\beta})$ is not close to $\theta_k$ arbitrarily and particularly under impractical conditions in which we expect it not to perform well. For example, consider a situation in which there is excessive amount of measurement error on the time of exposure. For concreteness, consider the case where we are interested in the relative incidence of events in the 30-day risk period following an exposure with a very high exposure onset measurement error of $\mu_u = 15$ days, so that on average, we are uncertain as to when the true exposure actually occurred by 15 days within a relatively small 30-day risk window of interest (average $100 \times \mu_u/e_1 = 50\%$). The maximum relative error is 25.7%. However, this reduced performance is expected since one cannot expect to be able to estimate the relative incidence of events during a fixed risk period following an exposure when one has excessive uncertainties regarding when the exposure actually occurred.

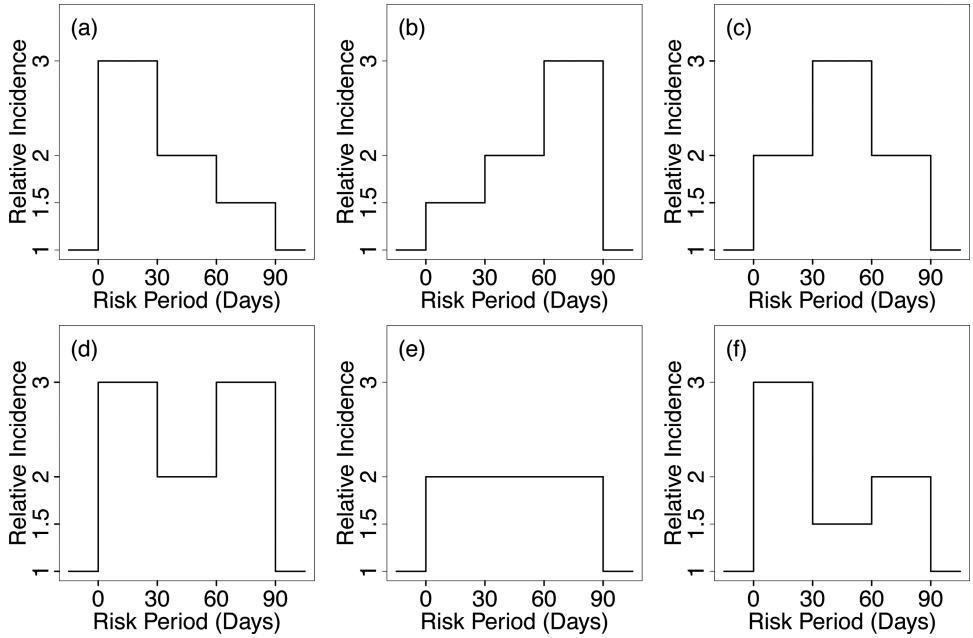


Figure 3. Patterns of true relative incidence ($\boldsymbol{\beta}$) in contiguous risk periods studied in simulations, including (a) decreasing, (b) increasing, (e) constant, and both symmetric and nonsymmetric mixtures of increasing and decreasing $\beta_k$'s (pattern c, d, and f).

## 3. SIMULATION STUDY

### 3.1 Simulation Design

In this section, we implement a set of simulation studies to address two specific objectives: (1) characterize the general nature of bias as a consequence of data with exposure onset measurement error and (2) assess the efficacy of the proposed bias-corrected estimation procedure to target the true relative incidence in MECS models.

We consider various underlying patterns of true relative incidences over multiple risk periods. As illustrated in Figure 3, these patterns include decreasing, increasing, and constant relative incidences over risk periods (cases a, b, and e, respectively), as well as mixtures of increasing and decreasing incidences (cases c, d, and f). The log relative incidences corresponding to these patterns are provided in Table 1, for example, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3) = (1.099, 0.693, 0.0405)$ for pattern (a). For each of these pattern of relative incidences, we generated the data as follows. The observation period $[a_i, b_i]$ is uniformly generated with mean follow-up length of 700 days for $i = 1, \ldots, N = 1000$ individuals and individual effects/baseline rates are set to $\varphi_i = \log(1/10000)$. Next, the marginal number of events, $n_{i\cdot}$, for the $i$th individual is generated from a Poisson distribution, truncated at 1 to obtain cases/individuals with at least one event according to the case series model (Farrington 1995), and these events are distributed within an individual's observation period according to the multinomial distribution with probabilities given in Equation (2). The number of nonoverlapping exposures $L_i$ range from 0 to 3 with probability masses $\{0.2, 0.4, 0.25, 0.15\}$ and true/unobserved exposure times $v_{il}$, $l = 1, \ldots, L_i$, are uniformly distributed over the follow-up period. Next, positive measurement error $u_{il}$, for each exposure, is added to the true exposure times to obtain the observed exposure times, $w_{il} = v_{il} + u_{il}$. We consider uniform, normal, and gamma distributed measurement error distributions with variances 1.33, 1.0, and 2.77, respectively.

To characterize the general nature of bias due to exposure onset measurement error and illustrate how the bias depends on the average amount of measurement error $\mu_u$ as described in Section 2.2, we generated data with increasing $\mu_u =$ 4, 6, 8, 10, 12, or 14 days. All simulation results reported next are based on averages over 200 simulated datasets.

Table 1. Bias-corrected estimates $\widehat{\boldsymbol{\beta}}$ and standard deviations over 200 replications for datasets (each with $N = 1000$) with three 30-day risk periods and for risk/effect patterns (a)–(f) described in Figure 3

| Pattern | True ($\beta_1$) | Est. | SD | True ($\beta_2$) | Est. | SD | True ($\beta_3$) | Est. | SD |
|---|---|---|---|---|---|---|---|---|---|
| (a) | 1.099 | 1.106 | 0.129 | 0.693 | 0.688 | 0.151 | 0.405 | 0.387 | 0.172 |
| (b) | 0.405 | 0.405 | 0.172 | 0.693 | 0.672 | 0.167 | 1.099 | 1.128 | 0.134 |
| (c) | 0.693 | 0.693 | 0.171 | 1.099 | 1.094 | 0.137 | 0.693 | 0.720 | 0.145 |
| (d) | 1.099 | 1.120 | 0.131 | 0.693 | 0.671 | 0.156 | 1.099 | 1.120 | 0.136 |
| (e) | 0.693 | 0.686 | 0.161 | 0.693 | 0.683 | 0.161 | 0.693 | 0.703 | 0.166 |
| (f) | 1.099 | 1.102 | 0.136 | 0.405 | 0.399 | 0.176 | 0.693 | 0.689 | 0.166 |

### 3.2 Results: Bias When Ignoring Measurement Error and Bias-Corrected Estimation

To describe the general nature of the asymptotic bias, we focus in more details on studies of decreasing and increasing effects patterns illustrated in Figure 3(a) and 3(b), respectively. Also, for the description of the bias, we focus on the case with uniform measurement error and sample size $N = 1000$. Bias results for effect patterns (a) and (b) are shown in Figure 4 as a function of $\mu_u$. Generally, and as expected, the bias increases with increasing average measurement error $\mu_u$. As can be seen, the naive case series estimates (black dashed curve) target $\boldsymbol{\beta}^*$ (dashed gray curve), which is obtained as the solution to Equation (6) using the Newton–Raphson algorithm described in Section 2.2.

When the risk is highest in the first 30 days and decreases with time, that is, pattern (a), all estimates ignoring measurement error, $\hat{\beta}_1^*$, $\hat{\beta}_2^*$, and $\hat{\beta}_3^*$, are attenuated. Conversely, when the true risk increases over time, that is, pattern (b), $\hat{\beta}_3^*$ is attenuated but $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ are inflated. This result is expected, since it can seen from (A.1) that the bias of the relative incidence in the $k$th risk period is a function of the incidence in the next contiguous risk period, $(k + 1)$, for $k = 0, 1, \ldots, K - 1$. Thus, if the true relative incidence in risk period $k$ is greater than in risk period $k + 1$, that is, if $\beta_k > \beta_{k+1}$, then $\beta_k^*$ will be attenuated. If $\beta_k < \beta_{k+1}$, then $\beta_k^*$ will be inflated. Note that $\beta_K^*$ will usually be attenuated if exposure onset measurement error is present since the relative incidence in the $K$th risk period is a function of the incidence in the baseline period. Similar patterns of bias hold for different measurement error distributions, for example, normal and gamma measurement errors. We provide details of these cases in the online supplementary materials.

Next, in Figure 5, we compare the relative amount of bias as the average measurement error $\mu_u$ increases for data with no overlapping risk segments, partial overlapping risk segments, and complete overlapping risk segments. The simulation setting for Figure 5 is as described above with a 30-day risk period and uniformly distributed measurement error. As described earlier

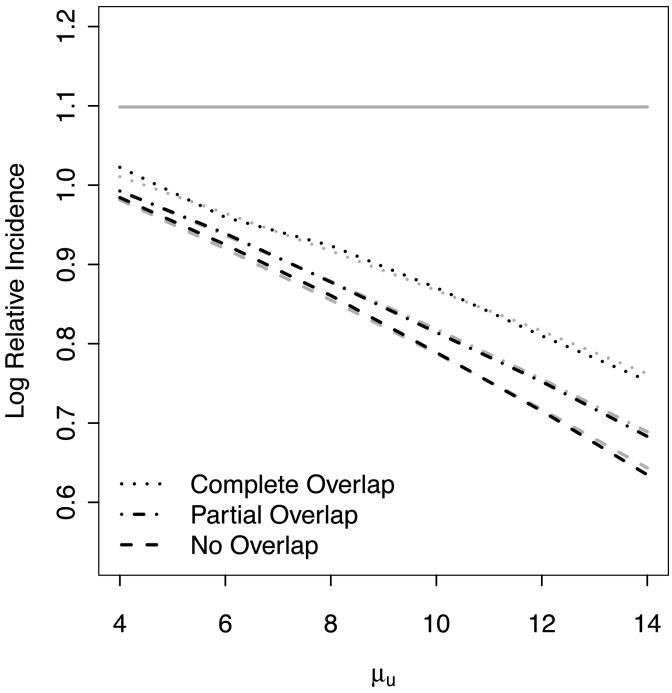


Figure 5. Relative estimation bias when ignoring measurement error for data with (1) complete, (2) partial, or (3) no overlapping risk segments. Bias is greatest when there are no overlapping risk segments, least when all risk segments overlap, and intermediate when there is some/partial overlap. Dashed black curves denote naive case series estimates ignoring measurement error, which targets $\boldsymbol{\beta}^*$ (dashed gray curves) instead of $\boldsymbol{\beta}$ (horizontal line).

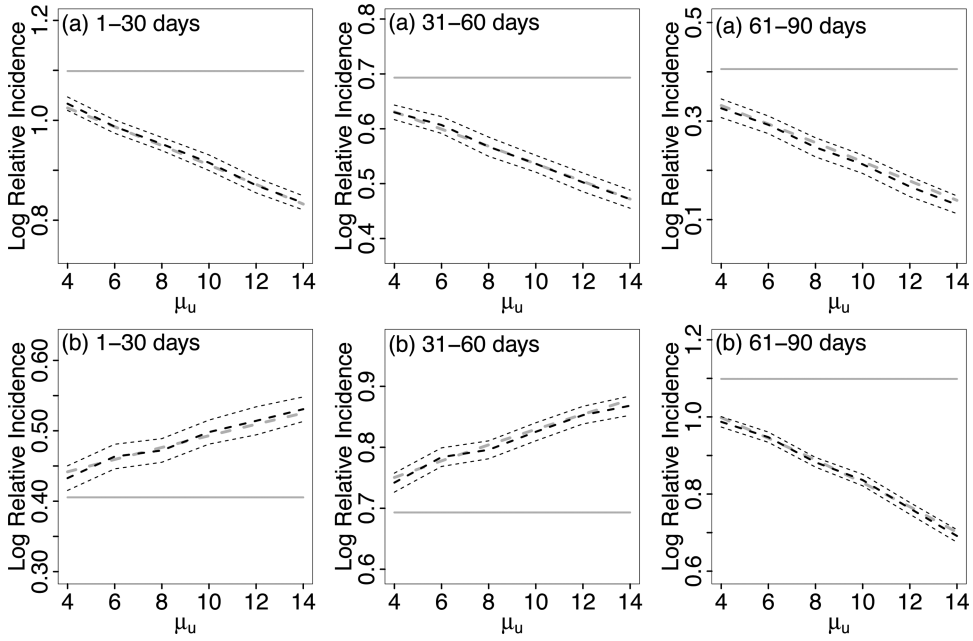


Figure 4. Bias for (a) increasing and (b) decreasing true patterns of log relative incidence $\boldsymbol{\beta}$ (indicated by horizontal lines) over three risk periods. Dashed black curves denote naive case series estimates ignoring measurement error, which targets $\boldsymbol{\beta}^*$ (dashed gray curves) instead of $\boldsymbol{\beta}$. In pattern (a), all three risk period log relative incidence estimates are attenuated. In pattern (b), the first two log relative incidence estimates are inflated and the third is attenuated. Also included are 95% confidence intervals (thin dashed lines). Given are averages over 200 simulated datasets.

Table 2. (A) Absolute bias for the proposed method and the ratio bias (naive/proposed) and (B) mean square error (MSE) of the naive and the proposed bias-corrected estimates corresponding to Table 1. See Table 1 caption for details

| (A) Bias | $\beta_1$ | | $\beta_2$ | | $\beta_3$ | |
|---|---|---|---|---|---|---|
| Pattern | Proposed | Ratio | Proposed | Ratio | Proposed | Ratio |
| (a) | 0.008 | 8.125 | 0.006 | 11.167 | 0.019 | 4.105 |
| (b) | 0.003 | 12.333 | 0.016 | 3.562 | 0.020 | 4.300 |
| (c) | 0.008 | 5.000 | 0.004 | 15.500 | 0.008 | 12.000 |
| (d) | 0.019 | 4.526 | 0.031 | 0.581 | 0.005 | 24.200 |
| (e) | 0.008 | 2.125 | 0.010 | 1.700 | 0.010 | 7.700 |
| (f) | 0.003 | 29.667 | 0.007 | 1.143 | 0.004 | 21.000 |
| (B) MSE | | | | | | |
| Pattern | Naive | Proposed | Naive | Proposed | Naive | Proposed |
| (a) | 0.012 | 0.017 | 0.017 | 0.023 | 0.021 | 0.030 |
| (b) | 0.015 | 0.032 | 0.014 | 0.024 | 0.018 | 0.019 |
| (c) | 0.013 | 0.025 | 0.010 | 0.014 | 0.022 | 0.027 |
| (d) | 0.016 | 0.017 | 0.012 | 0.028 | 0.026 | 0.017 |
| (e) | 0.011 | 0.026 | 0.012 | 0.026 | 0.021 | 0.028 |
| (f) | 0.017 | 0.018 | 0.013 | 0.031 | 0.020 | 0.028 |

in Section 2.1, overlapping risk segments are observed in practice when a recurrent exposure occurs in the risk period of a previous exposure. For example, for a risk period of interest defined as 30 days after an infection and with the first and second infections occurring on days 10 and 35, the length of the risk segment is 55 days instead of 60 days since days 35–40 overlap. Although nonoverlapping exposures, as described above, are vastly more common, we also illustrate here the relative differences in bias when there are overlapping risk segments. As illustrated in Figure 5, the bias is greatest when there are no overlapping risk segments, least when all risk segments overlap, and intermediate when there are some/partial overlaps. Furthermore, the simulation results confirm the theoretical results described in Section 2.2 for realistic data with mixtures of nonoverlapping and overlapping risk segments. Results for different measurement error distributions are similar and are not shown.

We illustrate through extensive simulation studies the efficacy of the proposed bias-corrected estimation procedure for the MECS models described in Section 2.3. Table 1 shows results from the proposed bias-corrected estimation procedure, using the simple quadratic regression intercept estimate, under the more general MECS data scenarios for each true effect pattern in Figure 3(a)–3(f) with $N = 1000$. For each simulated dataset with exposure onset time measurement error, increasing constants ($\tau_m = 0, 2, 4, \ldots, 10$ days) were added to each exposure time and estimates were computed by applying the case series model to each dataset as detailed in Section 2.3. The proposed adjusted estimator performs well for the different patterns of risk (a)–(f) and targets the true parameter $\boldsymbol{\beta}$. Table 2 presents the bias and mean squared error (MSE) for the bias-corrected and the naive estimators. Overall, the MSE for the bias-corrected estimate is of similar order as the naive estimate MSE; however as expected, its bias is drastically reduced. On average, the reduction is ∼9.4 folds across the simulations. We note that the simulation study assumes that $\mu_u$ is known; thus the results may be optimistic.

In addition to uniform exposure onset measurement errors, we also examined gamma and normally distributed measurement errors, as well as the performance for finite sample sizes. The adjusted estimates target the true parameters as expected and these additional results are available in the online supplemental materials. Also, at the suggestion of a reviewer, we provided more details on the performance as the variance of the measurement error distribution ($\sigma_u^2$) increases in the online supplementary materials (Table 7). This result illustrates the robustness of the bias correction to the measurement error variance.

Finally, we remark on the choice of $\mu_M$ (i.e., the $\tau_j$ sequence) in the regression extrapolation. From our experiences, there is flexibility in the choice of the $\tau_j$ sequence and it is not a major factor in the quality of the estimator if chosen reasonably. In practice, the choice of the sequence $\{\tau_j\}_{j=0}^{M}$ will depend on the estimated average exposure onset measurement error ($\mu_u$) relative to the *a priori* specified risk period of interest. For example, for the simulation study (and the data application) where the risk period of interest is the 30-day period(s) after infection/exposure, the choice of $\mu_M = \mu_u + \tau_M$ should not be excessively small relative to the risk period length. Figure 6 displays a regression fit through $\widehat{\boldsymbol{\beta}}^{*(j)}$ versus $\mu_j = \mu_u + \tau_j$ with $\mu_u = 4$ and $\mu_M = 14$, demonstrating a very precise extrapolation (black lines). Here the interval $\mu_M - \mu_u$ relative to the risk period is 1/3. We find that generally $\mu_M$ is selected such that this relative length is about $1/3$ to $1/2$ to be adequate. To illustrate how the extrapolation accuracy declines, we chose $\mu_M = 10$ so that the relative interval is only 20% of the risk period (gray lines). However, even with such a small interval where the regression is fitted, the extrapolation is able to correct for most of the bias relative to the naive estimate (asterisk). Since the computational cost of the naive case series estimator is trivial, we recommend a conservative approach to take $M$ as large as feasible for a given dataset to capture the features of the quadratic regression fit in the extrapolation. In our applications, we find that it is adequate to take $M$ to be 6–10.

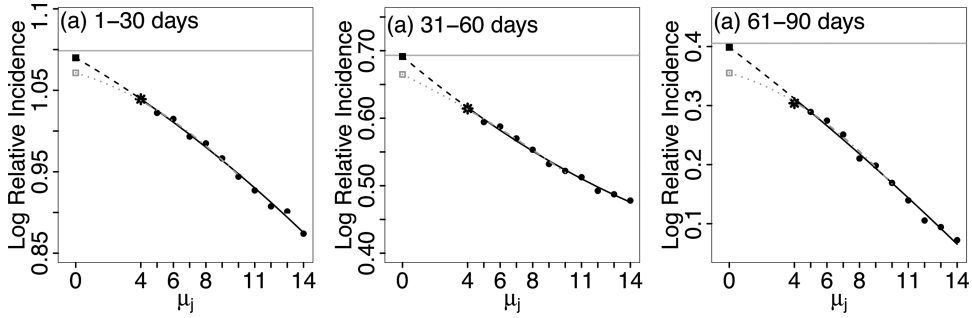


Figure 6. Illustration of the extrapolation from the regression fit to obtain the bias-corrected relative incidence estimates. The black lines are the regression fit using $\mu_M = 14$ and the gray lines are based on $\mu_M = 10$. The horizontal lines are the true $\beta_k$'s; the asterisk marks the naive estimate; the square boxes indicate the bias-corrected estimates for $\mu_M = 14$ (black) and $\mu_M = 10$ (gray).

## 4. APPLICATION: EXAMINING THE INFECTION-CARDIOVASCULAR RISK HYPOTHESIS FOR PATIENTS ON DIALYSIS

### 4.1 MECS Model for Infection-Cardiovascular Risk Using USRDS Data

As introduced in Section 1, a major challenge associated with using the USRDS inpatient hospitalization data to address the infection-cardiovascular risk hypothesis is that the exact date or time of infection onset cannot be ascertained precisely, although the observed discharge date is a good marker for the time of infection since it reasonably assures that the infection has occurred by this date. Thus, using the observed discharge time as a marker of the true infection time is a conservative approximation since the infection most likely occurred sometime during or possibly prior to the hospitalization. This modeling strategy leads to positive infection onset measurement error, and thus we apply the proposed MECS model to determine the relative incidence of cardiovascular events in the risk period following infection. Provided in more detail in Dalrymple et al. (2011), in the short period following an infection, specifically the approximate 30 days after infection, the effects of infection on vascular endothelium are hypothesized to be most pronounced. Thus, we focus on this risk period for illustrating the proposed MECS model. Cardiovascular events were defined as myocardial infarction, unstable angina, stroke, or transient ischemic attack and infections of interest included septicemia, bacteremia, peritonitis, endocarditis, soft-tissue, pulmonary, genitourinary, gastrointestinal, and joint or bone infection. The source population included patients 65–100 years of age with end-stage renal disease who newly initiated dialysis between January 1, 2000, and December 31, 2002. Study follow-up ended on December 31, 2004. We refer the reader to Dalrymple et al. (2011) for further details on the study protocol. The cohort for the analysis reported below includes $N = 16{,}779$ patients with one or more cardiovascular events.

Generally, external or auxiliary data sources are needed to reasonably estimate the measurement error parameter, namely $\mu_u$. However, for our current application, we can derive reasonable bounds on $\mu_u$ by using data on the length of hospitalization stay. If a hospitalization has an infection-related discharge diagnoses, it is likely that the infection occurred some time during the hospitalization stay or, in some cases, possibly shortly before the start of the hospitalization. The median length of hospital stays is 8 days in our cohort. Thus, we used $\mu_u = 8$ days as our intermediate estimate. We also used the lower estimate $\mu_u = 4$, half of the length of a typical hospitalization stay, to illustrate the reduced biased corresponding to a lower average (optimistic) level of measurement error, as expected. Thus, we apply the bias-correction method for the MECS model to obtain adjusted estimates of the true relative incidence using these values of $\mu_u$. Also, we note that due to the data violating the assumption of constant risk within a risk period, for illustration of the bias-correction method, we define the risk period as days 6–30 after an observed infection onset measured with error. The naive and bias-corrected log relative incidences are provided in Table 3. The naive relative incidence is 1.354 ($\widehat{\beta} = 0.303$), that is, the incidence of a cardiovascular event is 35% higher within the risk period after an infection compared with the baseline period. If we estimate that, on average, the observed date of infection is 4 days later than the true date of infection ($\mu_u = 4$), then the relative incidence estimate is increased by 16% to 1.516 ($\widehat{\beta} = 0.416$) after adjusting for infection onset measurement error. Similarly, if we instead consider the intermediate estimate of measurement error $\mu_u = 8$, the median of the hospitalization length distribution, then the bias-corrected relative incidence estimate is 1.768 ($\widehat{\beta} = 0.570$), an increase of 41% above the naive estimate.

Thus, infection onset measurement error led to substantive attenuation of the direct case series estimate of the relative incidence of cardiovascular events following infection in patients on dialysis. However, from the standard error estimates provided in Table 3, it is obvious that all analyses lead to the same scientific conclusion that there is an increased risk of cardiovascular

Table 3. Naive and bias-corrected estimates and standard errors for $\hat{\beta}$

| Estimation method | Log relative incidence | Standard error | 95% CI |
|---|---|---|---|
| Naive | 0.303 | 0.03 | – |
| Bias-correction $\mu_u = 4$ | 0.416 | 0.05 | 0.312–0.513 |
| Bias-correction $\mu_u = 8$ | 0.570 | 0.09 | 0.402–0.734 |
| Bias-correction $\widehat{\mu}_u = 5.5$ | 0.470 | 0.06 | 0.358–0.586 |

NOTE: The naive estimate corresponds to applying the case series method to the observed USRDS data, ignoring the presence of measurement error. The bias-corrected estimates correspond to applying the proposed correction method and approximate standard errors were obtained using the bootstrap method (via resampling of subjects). Bootstrap confidence intervals (CIs) are provided for the bias-corrected estimates

events following infection; only point estimates and precision are affected in this data. Reported standard errors for the bias-corrected estimation are based on 500 bootstrap datasets by resampling subjects, although SE estimates stabilize at ∼100 bootstrap samples. Bootstrap confidence intervals (CIs) are also provided, showing that even under the extremely optimistic assumption of low measurement error ($\mu_u = 4$), the CI is above the naive point estimate.

We note that for the USRDS data analysis above, we took a conservative approach by considering a low and intermediate value for $\mu_u$ because we do not have a direct estimate of the average amount of exposure onset measurement error. Instead, we arrived at this range based on the available data on the duration of hospitalization stay, where the median length of stay is 8 days. Examining both analyses, for $\mu_u = 4$ and 8, allowed us to assess the degrees of attenuation of the true relative incidence for these two levels of measurement error. The analysis corresponding to $\mu_u = 4$ can be interpreted as a reasonable approximation to the situation where the distribution of infection onset is highly skewed to the end of the hospitalization. This is unlikely and too optimistic, but it provides a lower estimate of attenuation in a very optimistic scenario. It is informative that for the USRDS data, even under this optimistic assumption about $\mu_u$, the relative incidence of cardiovascular events is increased by ∼16%, and the increase is more likely to be in the 23%–41% range, corresponding to $\mu_u = 5.5$ to 8, for instance. However, similar to the theory of classical measurement error in the covariates, with a consistent estimate of $\mu_u$, say $\widehat{\mu}_u$, which can be obtained from an internal subsample or external validation data sources generally, the proposed bias-correction method for the MECS model would target the true relative incidence. For our application here, if we assume equal likelihood of infection during a hospitalization stay, then $u_i|l_i \sim U(0, l_i)$, where $l_i$ is the duration/length of hospitalization. Thus, $\mu_u = E(u) = E\{E(u_i|l_i)\}$ and we can use the consistent estimate $\widehat{\mu}_u = 0.5 \sum_i^N l_i/N \sim 11/2 = 5.5$ (which is between the 4 and 8 days range we considered earlier). For comparison, this analysis is also provided in Table 3.

### 4.2 Modeling of USRDS Infection-Cardiovascular Data via Simulation

In this section, we further analyze the infection-cardiovascular hypothesis and the effects of exposure onset measurement error by modeling the USRDS data characteristics relevant to case series modeling by simulating data that matches more precisely the key characteristics of the USRDS cohort. This allows for a more thorough study of the effects of measurement error since the unknown parameters are controlled and can be varied. More precisely, we simulated USRDS case series data by using the simulation approach described in Section 3 and matched key relevant characteristics of the USRDS data, including the distributions of the ages at the start and end of the observation period, the length of follow-up, the length of baseline and risk periods, and the number of exposures per individual. The sample size is kept at $N = 16,779$, as observed. The distribution of the number of exposures for each individual, $L_i$, was based on the distribution of the number of exposures in the observed data. We note that the number of exposures in the observed data ranged from 0 to 19. However, in the simulated data, we restricted the range to no more than 10 exposures as this captured >99% of the original population and did not affect the analyses. We take the bias-corrected estimates of $\beta$ in the previous section to model the USRDS data. Once the number of exposures were generated, the times of exposure were distributed uniformly across each individual's observation history. Figure 7 displays the observed and simulated distribution of the length of baseline period, observed exposure onset ages, number of exposures, and number of events per individual. The characteristics of the simulated USRDS data closely track that of the observed USRDS data.

We apply the case series model directly to both the data with true exposure times and observed exposure times. As done earlier, we assess the effect of measurement error generated from three different distributions: uniform, normal, and gamma. Only results from the uniform distribution are presented here but results from the normal and gamma distributions are similar. Naive and bias-corrected estimates were obtained and averaged over 200 simulated datasets each for $\mu_u = 4$ and $\mu_u = 8$. When relatively modest measurement error was present, that is, when $\mu_u = 4$ days, namely when $u_{il} \sim$ Uniform[1, 7], the average naive and bias-corrected relative incidence estimates were 1.433 and 1.525, respectively. For intermediate level of measurement error, that is, $u_{il} \sim$ Uniform[5, 11] ($\mu_u = 8$ days), then the average relative incidence estimates were 1.518 and 1.784 for the naive and bias-corrected method, respectively. Thus, the results are similar to the percent increases observed in the analysis of the USRDS data of Section 4.1: the bias-corrected relative incidence estimates suggest a 9% (for $\mu_u = 4$) and 27% (for $\mu_u = 8$) increase over the naive relative incidence estimate on average, a significant difference in this high risk population.

## 5. DISCUSSION

Motivated by using the USRDS data to assess the infection-cardiovascular risk association in the dialysis population, where the precise times of infection cannot be ascertained, we proposed the MECS model to take into account the imprecise exposure/infection onset data. We presented, for the first time, a novel analysis of infection-cardiovascular risk association in a national cohort, using the proposed MECS model. The results lend additional support to the hypothesis that the ∼30-day period following infection is associated with a significantly increased risk/incidence of cardiovascular events. Through several different analyses of infection onset measurement error in Section 4, we confirmed the previously reported conclusion of an increased risk of cardiovascular events following infection-related hospitalization, based on the discharge date as a marker of the time of infection (Dalrymple et al. 2011); furthermore, the estimate of relative incidence was conservative when ignoring measurement error.

We also provided the asymptotic bias of the case series model that ignores exposure onset measurement error and proposed a feasible estimation procedure to correct for the bias when using the USRDS data. The method performs well in extensive realistic simulation studies, designed to match key characteristics of the USRDS data for the case series model. Also, an appealing aspect of the proposed MECS models is that

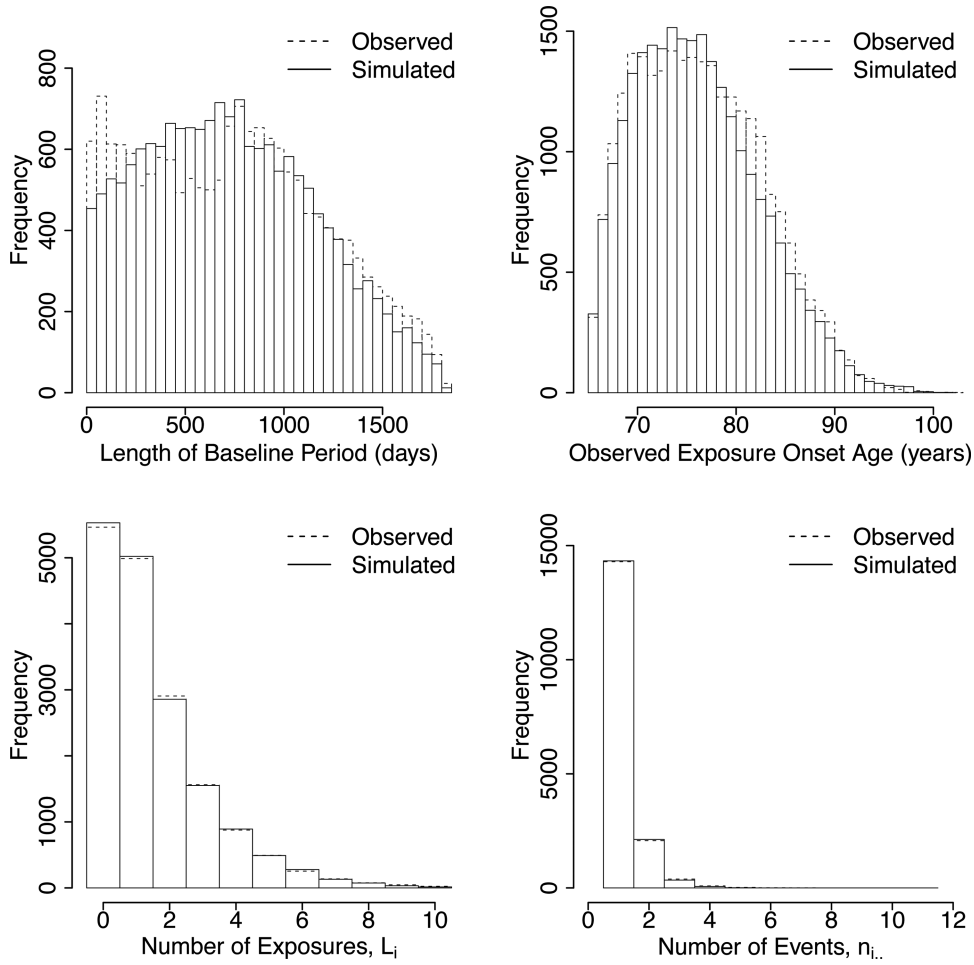


Figure 7. Characteristics of observed USRDS data compared with simulated USRDS data.

no new computational tools are needed for estimation, as they only involve repeated computation of the naive estimator.

We note that for general application of the MECS models, additional/auxiliary data are needed to estimate the mean of the exposure onset measurement error distribution. This is similar to the need for auxiliary data/information in regression modeling with traditional measurement error for mismeasured continuous or misclassified categorical variables. Depending on the specific application, additional data to estimate exposure onset measurement error mean can be obtained through a retrospective review of a subset of subjects to determine true exposure onset times or follow-up data collection if feasible. In the absence of any additional data or if the collection of new data is not practical, the tools developed in this work can be used in sensitivity analyses that assume specific values for the mean of exposure onset measurement error. For our specific application, which uses inpatient hospitalization data, a conservative marker of infection onset time is used and reasonable bounds for $\mu_u$ were determined based on additional data on the length of hospital stay.

Finally, in our application and the proposed MECS model framework, the precise onset of the cardiovascular events that we consider is reasonably measured with accuracy with respect to diagnosis and time. Although more rare, inevitably with the size of the sample analyzed, the timing of some events will not be accurate. However, although less applicable to our current hypothesis on cardiovascular outcomes/events, measurement error on the precise onset of events generally is possible, as pointed out by a reviewer. Case series modeling with measurement errors in both the precise onset of events as well as the timing of exposures presents significant challenges; it is an open problem at this time.

## APPENDIX

### A.1 Proof of Theorem 1

For the general MECS model with $L_i$ exposures for individual $i$ and with $K$ adjacent risk periods, direct calculations yield

$$\tilde{\pi}_{ik}(\mathbf{u}_i) = \begin{cases} \Delta_i^{-1}\{e_{ik}\lambda_{ik} + (\lambda_{i,k+1} - \lambda_{ik})\sum_{l=1}^{L_i} u_{il}\}, \\ \quad k = 0, 1, \ldots, K-1 \\ \Delta_i^{-1}\{e_{iK}\lambda_{iK} + (\lambda_{i0} - \lambda_{iK})\sum_{l=1}^{L_i} u_{il}\}, \\ \quad k = K \end{cases}, \quad \text{(A.1)}$$

where $\Delta_i = \sum_r e_{ir}\lambda_{ir}$. To see that the denominator of $\tilde{\pi}_{ik}(\mathbf{u}_i)$ is $\Delta_i$, denote the numerator in (A.1) by $\delta_{ik}$ for $k = 0, 1, \ldots, K$. Then it can be directly verified that $\sum_k \delta_{ik} = \sum_k e_{ik}\lambda_{ik}$. (Therefore, $\sum_k \tilde{\pi}_{ik} = 1$.) Thus, $E(\tilde{n}_{ik}|\mathbf{u}_i) = n_{i.}\tilde{\pi}_{ik}(\mathbf{u}_i)$ since, conditional on $\mathbf{u}_i$, $\tilde{\mathbf{n}}_i = (\tilde{n}_{i0}, \ldots, \tilde{n}_{iK}) \sim \text{Mult}(n_{i.}, \tilde{\boldsymbol{\pi}}_i(\mathbf{u}_i))$; $E(\tilde{n}_{ik})$ given in (4) follows. Next, we consider the more rare case in applications where two or more exposures may have overlapping risk segments within the risk period (the MECS model with $K = 1$), as previously defined in Section 2.1. Suppose that

there are $L_i \geq 2$ exposures with $L_i'$ disjoint risk segments ($L_i' \leq L_i$) as previously defined prior to Theorem 1. Denote the number of exposures that form the $s$th risk segment by $\zeta(s)$, where $\zeta(s) \geq 1$ ($s = 1, \ldots, L_i'$). Overlapping risk segments correspond to $\zeta(s) > 1$ and note that $\sum_{s=1}^{L_i'} \zeta(s) = L_i$. The risk period length, $e_{i1}$, can be partitioned into disjoint risk segment lengths $e_{i1}^{(s)} > 0$; thus, $e_{i1} = \sum_{s=1}^{L_i'} e_{i1}^{(s)}$. Similarly, the baseline period length, $e_{i0}$, can be partitioned into disjoint baseline segment lengths $e_{i0}^{(s)} > 0$ ($s = 1, \ldots, L_i'$) and $e_{i0}^{(L_i'+1)} \geq 0$, where the boundary is at the end of the observation period, $e_{i0}^{(L_i'+1)}$, may be zero if the end of the observation period occurs within the last exposure risk segment. (Thus, $e_{i0} = \sum_{s=1}^{L_i'+1} e_{i0}^{(s)}$.) Also, denote the measurement errors associated with the $\zeta(s)$ exposures of the $s$th risk segment by $\{u_{is_j}; j = 1, \ldots, \zeta(s)\}$, for $s = 1, \ldots, L_i'$. Then the group probabilities, $\tilde{\pi}_{ik}$, can be shown to be linear functions of the underlying rates $\{\lambda_{ik}\}$, where the coefficients depend on the risk segment lengths $\{e_{i0}^{(s)}\}$ and $\{e_{i1}^{(s)}\}$, and the measurement errors $\{u_{is_j}; j = 1, \ldots, \zeta(s)\}$. Similar to (A.1), the model probabilities for the risk and baseline periods (after simplification) are given by

$$\tilde{\pi}_{ik}(\mathbf{u}_i) = \begin{cases} \Delta_i^{-1} \left\{ e_{i0}\lambda_{i0} + \lambda_{i1} \sum_{s=1}^{L_i'} u_{is_1} - \lambda_{i0} \sum_{s=1}^{L_i'} u_{is_{\zeta(s)}} \right\}, & k = 0 \\ \Delta_i^{-1} \left\{ e_{i1}\lambda_{i1} + \lambda_{i0} \sum_{s=1}^{L_i'} u_{is_{\zeta(s)}} - \lambda_{i1} \sum_{s=1}^{L_i'} u_{is_1} \right\}, & k = 1 \end{cases}.$$

Therefore, $E(n_{ik})$ is as given in (4) with $L_i$ replaced by $L_i'$. This completes the proof of Theorem 1.

For our application of the case series model to infection-cardiovascular association as well as other applications of this model, a heightened risk of adverse events following an exposure justifies the assumption of adjacent risk periods. However, the risk periods need not be adjacent in the case series model generally. In this case (for nonoverlapping exposures), it can be shown that Equation (4) becomes

$$E(\tilde{n}_{ik}) = \begin{cases} n_{i.}\Delta_i^{-1}\{e_{ik}\lambda_{ik} + L_i\mu_u(\lambda_{i,0} - \lambda_{ik})\}, \\ \quad k = 1, 1, \ldots, K \\ n_{i.}\Delta_i^{-1} \left\{ e_{i0}\lambda_{i0} + L_i\mu_u \left( \sum_{k=1}^{K} \lambda_{ik} - K\lambda_{i0} \right) \right\}, \\ \quad k = 0 \end{cases},$$

## A.2 Estimation via Newton–Raphson

Fitting the case series (multinomial) model can be based on standard softwares (including SAS, Stata, and R; see *http://statistics.open.ac.uk/sccs*) that have routines for Poisson models with log link function and allow for an offset term $\log(e_{ijk})$. For the MECS estimation proposed, no new computational tools are required; therefore, existing software can be used. However, it may be more efficient with the repeated applications for different $\mu_j$ to use the Newton–Raphson algorithm directly. We provide here the formulas for straight-forward implementation. Following the notations introduced in Section 2.1, we have that $\log(\pi_{ijk}/\pi_{i00}) = \alpha_j + \beta_k + \log(r_{ijk})$, with $r_{ijk} = e_{ijk}/e_{i00}$. Thus, the log-likelihood for the $i$th subject is

$$\ell_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{j=0}^{J}\sum_{k=0}^{K} n_{ijk}\{\alpha_j + \beta_k + \log(r_{ijk})\} - n_{i..} \log \left\{ \sum_{r=0}^{J}\sum_{s=0}^{K} e_{irs} \exp(\alpha_r + \beta_s)/e_{i00} \right\}$$

and the log-likelihood for $N$ subjects with at least one event is $\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{N} \ell_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$. (Note that the second term above is simply $n_{i..} \log(1/\pi_{i00})$.) Thus, direct calculations yield the $J + K$ likelihood equations:

$$\frac{\partial \ell(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \beta_k} = \sum_{i=1}^{N}\sum_{j=0}^{J}(n_{ijk} - n_{i..}\pi_{ijk}) = 0, \quad k = 1, \ldots, K$$

$$\frac{\partial \ell(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \alpha_j} = \sum_{i=1}^{N}\sum_{k=0}^{K}(n_{ijk} - n_{i..}\pi_{ijk}) = 0, \quad j = 1, \ldots, J.$$

The needed second-order partial derivatives are $\partial^2\ell(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial\alpha_j^2 = -\sum_i n_{i..}\pi_{ij.}(1 - \pi_{ij.})$, $j = 1, \ldots, J$; $\partial^2\ell(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial\alpha_j\alpha_l = \sum_i n_{i..}\pi_{ij.}\pi_{il.}$, $j \neq l$; $\partial^2\ell(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial\beta_k^2 = -\sum_i n_{i..}\pi_{i.k}(1 - \pi_{i.k})$, $k = 1, \ldots, K$; $\partial^2\ell(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial\beta_k\beta_l = \sum_i n_{i..}\pi_{i.k}\pi_{i.l}$, $k \neq l$; $\partial^2\ell(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial\alpha_j\beta_k = -\sum_i n_{i..}(\pi_{ijk} - \pi_{ij.}\pi_{i.k})$. The Newton–Raphson update at iteration $(t + 1)$ is $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - (\mathbf{H}^{(t)})^{-1}\mathbf{q}^{(t)}$, where $\boldsymbol{\theta}^{(t)} = (\alpha_1^{(t)}, \ldots, \alpha_J^{(t)}, \beta_1^{(t)}, \ldots, \beta_K^{(t)})^{\mathrm{T}}$, $\mathbf{q}^{(t)}$ is the vector of first-order partial derivatives, and $\mathbf{H}^{(t)}$ is the $(J + K) \times (J + K)$ Hessian matrix, both evaluated at $\boldsymbol{\theta}^{(t)}$.

## A.3 Notes on Other Cases of Exposure Onset Measurement Error and Bias Correction

In this work, we focus on positive exposure onset measurement error because this model is directly applicable to our primary interest in analyzing infection-related hospitalizations data from the USRDS. However, for the proposed MECS models to be more broadly applicable, we provide here notes on the more general case where the variable $u$ is not strictly positive. First, note that when exposure onset is strictly negative (i.e., $u < 0$), it can be shown that the expression for $E(\tilde{n}_{ik})$ given in (4) holds with $\mu_u$ replaced by $E|u|$. For the more general case, let $p_0 = \Pr(u < 0)$ and $p_1 = \Pr(u \geq 0)$. For simplicity of notations, consider one exposure and risk period ($L_i = 1$, $K = 1$). Then it can be shown that

$$E(\tilde{n}_{i1}) = p_0\{e_{i1}\lambda_{i1} + (\lambda_{i0} - \lambda_{i1})E[|u|I(u < 0)]\}\Delta_i^{-1} + p_1\{e_{i1}\lambda_{i1} + (\lambda_{i0} - \lambda_{i1})E[uI(u \geq 0)]\}\Delta_i^{-1},$$

where $I(E)$ denotes the indicator function for event $E$. Therefore, additional/auxiliary data are needed to estimate $p_0$, $p_1$, $E[uI(u \geq 0)]$ and $E[|u|I(u < 0)]$. Auxiliary data in some applications could be based on retrospective chart review of a subset of subjects to determine true exposure onset, for instance. Otherwise, sensitivity analysis can be performed for assumed values of these characteristics/parameters of the distribution of $u$.

Finally, we point out that the above expression for $E(\tilde{n}_{i1})$ explicitly shows that even when $\mu_u$ is known precisely, case series analysis of data whereby one simply subtracts $\mu_u$ from each observed exposure onset time with positive measurement error will not target $\beta_k$. It is clear that doing so, under positive exposure onset measurement error, will simply lead to the case of general measurement error.

## SUPPLEMENTARY MATERIALS

Additional simulation results referred to in Section 3 are available in mec_sup.pdf.

*[Received September 2011. Revised April 2012.]*

## REFERENCES

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*, Boca Raton, FL: Chapman and Hall/CRC. [2,5]

Cook, J. R., and Stefanski, L. A. (1994), "Simulation-Extrapolation Estimation in Parametric Measurement Error Models," *Journal of the American Statistical Association*, 89, 1314–1328. [5]

Dalrymple, L. S., Mohammed, S. M., Mu, Y., Johansen, K. L., Chertow, G. M., Grimes, B., Kaysen, G. A., and Nguyen, D. V. (2011), "The Risk of Cardiovascular-Related Events Following Infection-Related Hospitalizations in Older Patients on Dialysis," *Clinical Journal of the American Society of Nephrology*, 6, 1708–1713. [1,2,3,10,11]

Farrington, C. P. (1995), "Relative Incidence Estimation From Case Series for Vaccine Safety Evaluation," *Biometrics*, 51, 228–235. [1,2,3,7]

Gibson, J. E., Hubbard, R. B., Smith, C. J. P., Tata, L. J., Britton, J. R., and Fogarty, A. W. (2009), "Use of Self-Controlled Analytical Techniques to Assess the Association Between Use of Prescription Medications and the Risk of Motor Vehicle Crashes," *American Journal of Epidemiology*, 169, 761–768. [2]

Higgins, R. M., Davidian, M., and Giltinan, D. M. (1997), "A Two-Step Approach to Measurement Error in Time-Dependent Covariates in Nonlinear Mixed-Effects Models, With Application to IGF-I Pharmacokinetics," *Journal of the American Statistical Association*, 92, 436–448. [2]

Huang, Y., and Wang, C. Y. (2000), "Cox Regression With Accurate Covariate Unascertainable: A Nonparametric-Correction Approach," *Journal of the American Statistical Association*, 95, 1209–1219. [2]

Hubbard, R., Farrington, P., Smith, C., Smeeth, L., and Tattersfield, A. (2003), "Exposure to Tricyclic and Selective Serotonin Reuptake Inhibitor Antidepressants and the Risk of Hip Fracture," *American Journal of Epidemiology*, 158, 77–84. [2]

Keiding, N. (1992), "Independent Delayed Entry," *Survival Analysis: State of the Art*, 211, 309–326. [2]

Li, Y., and Ryan, L. (2004), "Survival Analysis With Heterogeneous Covariate Measurement Error," *Journal of the American Statistical Association*, 99, 724–735. [2]

Liao, X., Zucker, D. M., Li, Y., and Spiegelman, D. (2011), "Survival Analysis With Error-Prone Time-Varying Covariates: A Risk Set Calibration Approach," *Biometrics*, 67, 50–58. [2]

Macko, R. F., Ameriso, S. F., Gruber, A., Griffin, J. H., Fernandez, J. A., Barndt, R., Quismorio, F. P., Weiner, J. M., and Fisher, M. (1996), "Impairments of the Protein C System and Fibrinolysis in Infection-Associated Stroke," *Stroke*, 27, 2005–2011. [1]

Mahmoudi, M., Curzen, N., and Gallagher, P. J. (2007), "Atherogenesis: The Role of Inflammation and Infection," *Histopathology*, 50, 535–546. [1]

Smeeth, L., Thomas, S. L., Hall, A. J., Hubbard, R., Farrington, P., and Vallance, P. (2004), "Risk of Myocardial Infarction and Stroke After Acute Infection or Vaccination," *New England Journal of Medicine*, 351, 2611–2618. [1]

Sun, H. (2006), "The Interaction Between Pathogens and the Host Coagulation System," *Physiology*, 21, 281–288. [1]

Tsiatis, A., and Davidian, M. (2004), "Joint Modeling of Longitudinal and Time-to-Event Data: An Overview," *Statistica Sinica*, 14, 809–834. [2]

Tosteson, T. D., Buonaccorsi, J. P., and Demidenko, E. (1998), "Covariate Measurement Error and the Estimation of Random Effect Parameters in a Mixed Model for Longitudinal Data," *Statistics in Medicine*, 17, 1959–1971. [xxxx]

United States Renal Data System. (2010), *USRDS 2010 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States*, Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases. [1]

Walley, T., and Mantgani, A. (1997), "The UK General Practice Research Database," *Lancet*, 350, 1097–1099. [1]

Whitaker, H. J., Farrington, C. P., Spiessens, B., and Musonda, P. (2006), "Tutorial in Biostatistics: The Self-Controlled Case Series Method," *Statistics in Medicine*, 25, 1768–1797. [2,3]

Whitaker, H. J., Hocine, M., and Farrington, C. P. (2009), "The Methodology of Self-Controlled Case Series Studies," *Statistical Methods in Medical Research*, 18, 7–26. [2,3]

Zhu, J., Quyyumi, A. A., Norman, J. E., Csako, G., Waclawiw, M. A., Shearer, G. M., and Epstein, S. E. (2000), "Effects of Total Pathogen Burden on Coronary Artery Disease Risk and C-Reactive Protein Levels," *American Journal of Cardiology*, 85, 140–146. [1]