Published in final edited form as:

Biometrics. 2013 June ; 69(2): 520–529. doi:10.1111/biom.12033.

Naive Hypothesis Testing for Case Series Analysis with Time-Varying Exposure Onset Measurement Error: Inference for Infection-Cardiovascular Risk in Patients on Dialysis

Sandra M. Mohammed¹, Lorien S. Dalrymple², Damla entürk³, and Danh V. Nguyen^{1,4}

Danh V. Nguyen: ucdnguyen@ucdavis.edu

¹Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, California 95616, U.S.A.

²Division of Nephrology, Department of Medicine, University of California, Sacramento, California 95817, U.S.A.

³Department of Biostatistics, University of California, Los Angeles, California 90095, U.S.A.

⁴Clinical and Translational Science Center, University of California, Sacramento, California 95817, U.S.A.

Summary

The case series method is useful in studying the relationship between time-varying exposures, such as infections, and acute events observed during the observation periods of individuals. It provides estimates of the relative incidences of events in risk periods (e.g., 30-day period after infections) relative to the baseline periods. When the times of exposure onsets are not known precisely, application of the case series model ignoring exposure onset measurement error leads to biased estimates. Bias-correction is necessary in order to understand the true directions and effect sizes associated with exposure risk periods, although uncorrected estimators have smaller variance. Thus, inference via hypothesis testing based on uncorrected test statistics, if valid, is potentially more powerful. Furthermore, the tests can be implemented in standard software and do not require additional auxiliary data. In this work, we examine the validity and power of naive hypothesis testing, based on applying the case series analysis to the imprecise data without correcting for the error. Based on simulation studies and theoretical calculations, we determine the validity and relative power of common hypothesis tests of interest in case series analysis. In particular, we illustrate that the tests for the global null hypothesis, the overall null hypotheses associated with all risk periods or all age effects are valid. However, tests of individual risk period parameters are not generally valid. Practical guidelines are provided and illustrated with data from patients on dialysis.

Correspondence to: Danh V. Nguyen, ucdnguyen@ucdavis.edu.

Supplemental Materials: Web Appendices in Sections 3 and 6 are available with this paper at the *Biometrics* website on Wiley Online Library.

Keywords

case series models; exposure timing measurement error; hypothesis testing; inference; longitudinal observational database; non-homogeneous Poisson process

1 Introduction

The self-controlled case series method, or case series (CS) method, was developed by Farrington (1995) to assess the relationship between time-varying exposures and acute events. The CS method was originally designed for studies of associations between vaccines and adverse events (Farrington, Nash and Miller, 1996). For example, it can be used to examine whether the incidence of adverse events is increased in specific time (risk) periods after vaccination, such as one month after vaccination/exposure, relative to baseline periods. It has since been used in a variety of non-vaccine epidemiologic studies, including the assessment of the risk of hospitalization for stroke after initiation of antipsychotics in the elderly (Pratt et al., 2010). The method is also useful in active surveillance for drug safety using longitudinal observational databases (Madigan et al., 2011). In contrast to cohort or case-control approaches, it uses data on only cases, individuals with at least one event, and is self-matched. It differs from the self-matched case-crossover method (Maclure, 1991), a case-control method based on cases only, which compares the time period immediately preceding the event time to a referent period selected from the case's own history. The CS method allows for consistent estimates of relative incidences of adverse events in the risk periods and implicitly controls for all time-invariant confounders. It is obtained by conditioning on each individual having one or more events, where events arise from an underlying time-dependent Poisson cohort model. An excellent paper on the application of the method can be found in Whitaker et al. (2006).

Our work is motivated by the application of the CS method to the study of the relative incidence of cardiovascular events after infections, based on hospitalization data for patients on dialysis (Dalrymple et al., 2011) assembled from the United States Renal Data System (USRDS, 2010). For example, it is of interest to assess whether there is an increase in the incidence of cardiovascular events during the 1-30 day period or 31-60 day period after an infection. We note that the risk periods of interest need not be contiguous generally.

The main issue we consider in this work is inference via naive hypothesis testing for the CS analysis when the timing of infection, or “exposure” more generally, is not known precisely. For example, in the aforementioned USRDS hospitalization data, the exact dates of infections are unknown and available information on admission and discharge dates associated with discharge diagnoses of hospitalizations are used as markers of infection onset times. This issue is relevant to other applications of the CS method to longitudinal observational databases (e.g., hospital claims or administrative databases) such as adverse events due to medications or other types of hospitalizations. Dalrymple et al. (2011) applied the CS analysis to data from the USRDS and used the date of discharge of infection-related hospitalizations to approximate the date of infection onset. Although this approach ensures that the infection occurred by the discharge date, the actual onset time was most likely some time during, or immediately before, the hospitalization. Thus, the application of the CS

analysis to such data has positive exposure onset measurement error. To account for the measurement error in the exposure times, Mohammed et al. (2012) proposed the measurement error case series (MECS) model and bias-corrected estimation. While bias-corrected relative incidence estimation is available, inference procedures, such as hypothesis testing, have not been examined.

Thus, in this work we examine naive hypothesis tests, which are tests simply based on naive estimates, ignoring measurement error. When valid, naive hypothesis tests are relatively more powerful and they can be easily performed without any additional data or information, analogous to naive hypothesis tests in classical measurement error methods (see Carroll et al., 2006, chapter 10). Also, the tests can be implemented in standard software. We note that the MECS models were recently introduced; therefore, to date there has been no work on inferential aspects of the method. Therefore, in this work we examine the validity and power of the naive hypothesis testing for the general MECS models.

This paper is organized as follows. Details of the approach and relevant hypothesis tests are described in Section 2. Theoretical calculations to assess validity of the naive tests are discussed in Section 3. Section 4 describes simulation studies assessing the validity and power of the naive hypothesis tests. An illustration with data from the United States population of older patients on dialysis and a discussion of practical guidelines are provided in Sections 5 and 6, respectively.

2 Models and Hypothesis Tests

For a cohort of N individuals, each of whom has at least one event, let $(a_i, b_i]$ denote the observation period for individual i which is further partitioned into $J + 1$ age groups, $j = 0, \dots, J$, and $K + 1$ exposure risk periods, $k = 0, \dots, K$, where $k = 0$ corresponds to the baseline period and $j = 0$ refers to the reference age group. The number of events, n_{ijk} , for individual i in age group j and risk period k is modeled as a non-homogeneous Poisson process. That is, n_{ijk} is distributed as $\text{Poisson}(e_{ijk}\lambda_{ijk})$, where $\lambda_{ijk} = \exp(\phi_i + \delta_j + \beta_k)$ is the incidence rate and e_{ijk} is the length of time spent in age group j and risk period k for person i . The incidence rate parameters, ϕ_i , δ_j and β_k , are the individual-specific, j^{th} age group and k^{th} risk period effect, respectively.

The CS likelihood is obtained after conditioning on the occurrence of at least one event for each individual. The kernel of the CS likelihood is product multinomial (Farrington, 1995)

with contribution from individual i given by $L_i(\delta, \beta) = \prod_{j,k} \pi_{ijk}^{n_{ijk}}$, where

$$\pi_{ijk} = \frac{e_{ijk}\lambda_{ijk}}{\sum_{r=0}^J \sum_{s=0}^K e_{irs}\lambda_{irs}} = \frac{e_{ijk}\exp(\delta_j + \beta_k)}{\sum_{r=0}^J \sum_{s=0}^K e_{irs}\exp(\delta_r + \beta_s)}, \quad (1)$$

$\delta = (\delta_1, \dots, \delta_J)^T$, $\beta = (\beta_1, \dots, \beta_K)$ and $\delta_0 = \beta_0 = 0$ correspond to the referent age and risk group. Maximum likelihood (ML) estimates can be obtained by maximizing the log-likelihood

$l(\delta, \beta) = \sum_{i=1}^N \log \{L_i(\delta, \beta)\}$. While the log relative incidences of the exposure risk periods, β_k , $k = 1, \dots, K$, are the primary parameters of interest, the log relative incidences of the age

groups, $\delta_j, j = 1, \dots, J$, are sometimes of interest in practice as well. Therefore, we will examine hypothesis testing for both sets of parameters. We note that for our application and the work here, the risk periods are contiguous; however, this is not a requirement for the case series model generally and non-contiguous periods may be of interest in other applications. (The shifted pattern of bias for non-contiguous risk periods is very similar, as previously described in Mohammed et al. (2012).)

Measurement error case series models (Mohammed et al., 2012) are extensions of the above case series models to handle data with imprecise times of exposure onset. More specifically, the positive additive exposure onset measurement error model is given by $w_{il} = v_{il} + u_{il}$ ($l = 1, \dots, L_i$), where w_{il} is the observed exposure onset time (e.g., infection-related hospitalization discharge time), v_{il} is the true *unobserved* exposure (infection) onset time, u_{il} a positive measurement error with mean $\mu_u = E(u_{il})$ and L_i is the number of exposures for individual i . Note that not knowing precisely v_{il} , the time of infection onset, leads to misclassification of the risk periods; i.e., error in exposure timing. This measurement error combined with case series models (1) above defines the MECS model. As detailed in Mohammed et al. (2012), the amount of measurement error in the exposure times cannot be unrestricted. A practical and necessary assumption is that u_{il} is less than the length of the risk period of interest. For instance, with a “30-day risk period after an infection, the uncertainty in the time when the infection actually occurred should not exceed 30 days; otherwise, one could not estimate the relative incidence in the 30-day risk period after an infection because $u_{il} > 30$ amounts to not having any reliable data for estimation.”

Naive hypothesis testing regarding the underlying parameters of interest (δ, β) will not require any new data since it is based on tests from applying the standard CS model (1) to the observed data, ignoring exposure onset measurement error (i.e., assuming $u_{il} = 0$). Let $\tilde{n}_{ijk}$ denote the number of events in age group j and risk group k based on the *observed* exposure times, w_{il} , $i = 1, \dots, N$, $l = 1, \dots, L_i$. Denote the targets of the naive CS age and risk estimates obtained when ignoring measurement error as $\delta^* = (\delta_1^*, \dots, \delta_J^*)^T$ and $\beta^* = (\beta_1^*, \dots, \beta_K^*)^T$. Thus, the naive MLEs of (δ^*, β^*) , denoted $(\hat{\delta}^*, \hat{\beta}^*)$, are obtained by solving the set of $(J + K)$ likelihood equations

$$\begin{aligned} N^{-1} \sum_{i=1}^N \sum_{j=0}^J (\tilde{n}_{ijk} - n_{i..} \hat{\pi}_{ijk}^*) &= 0, \quad k=1, \dots, K, \\ N^{-1} \sum_{i=1}^N \sum_{k=0}^K (\tilde{n}_{ijk} - n_{i..} \hat{\pi}_{ijk}^*) &= 0, \quad j=1, \dots, J, \end{aligned} \quad (2)$$

where $\hat{\pi}_{ijk}^* = e_{ijk} \exp(\hat{\delta}_j^* + \hat{\beta}_k^*) / \sum_{r=0}^J \sum_{s=0}^K e_{irs} \exp(\hat{\delta}_r^* + \hat{\beta}_s^*)$; $\tilde{n}_{ijk}$ is the observed number of events in age group j and risk period k for individual i ; and $n_{i..}$ is the total number of events for individual i . Generally, (δ^*, β^*) are nonlinear functions of the true underlying parameter of interest (δ, β) and the naive MLEs $(\hat{\delta}^*, \hat{\beta}^*)$ are biased. The relationships between (δ, β) and (δ^*, β^*) do not have closed-form expressions generally, except for the simple models with equal risk periods and equal follow-up times for all subjects.

Since the naive estimates are MLEs they are distributed asymptotically as multivariate normal, and hypothesis testing in practice utilizes the likelihood ratio or Wald test statistics. We will focus our study of naive hypothesis testing using the likelihood ratio test (LRT); our preliminary studies also indicate that there is not much difference between the two. The LRT statistic is $T = -2(\ell_R - \ell_F)$ where ℓ_R is the log-likelihood of the reduced model and ℓ_F is the log-likelihood of the full model. It is well-known that the distribution of T is distributed chi-square under the null hypothesis: $T \sim \chi_p^2$, where χ_p^2 denotes the chi-square distribution with p degrees of freedom (which is the difference in parameters between the full and reduced models). We focus on testing the following four types of null hypotheses useful in practice: (1) Global null: $H_0 : (\delta, \beta)^T = \mathbf{0}$; (2) Overall null age or overall null risk effects: $H_0 : \delta = \mathbf{0}$ and $H_0 : \beta = \mathbf{0}$; (3) Specific null age group effect: $H_0 : \delta_j = 0$ (component-wise tests) for $j = 1, \dots, J$; (4) Specific null risk period effect: $H_0 : \beta_k = 0$ (component-wise tests) for $k = 1, \dots, K$. (1.) For our application, the global null corresponds to the hypothesis that the incidences of cardiovascular events in the risk periods following infections are not different from baseline and that, furthermore, the baseline incidence does not depend on age. (2.) Clearly, the null risk or age effects (alone) allows for testing the hypothesis that there may be differential risk associated with periods of exposure, but no overall age effects; and vice versa. (3. & 4.) Finally, the component-wise tests allow for investigation of increased incidence of cardiovascular events in the first 30 days after infections, but not in the 60-90 day period after infections, for instance.

Our study objectives are two-fold. The first question we will address is, “Which of the above tests are valid?” A test is valid if its asymptotic Type I error rate under the null approaches the nominal test level α (Carroll et al., 2006, Chap. 10). Secondly, we determine the power of the naive tests and compare it to the power of the “optimal” test, which is based on the same data *without* exposure onset measurement error. The empirical power will be calculated as the proportion of likelihood ratio tests that reject the null hypothesis at a fixed significance level.

3 Validity of Naive Tests: Theoretical Calculations

In this section we consider theoretical calculations to study the validity of the naive tests. The corresponding simulation experiments are considered in Section 4 below. The naive ML estimates $(\hat{\delta}^*, \hat{\beta}^*)$ are obtained by solving the set of likelihood equations given in (2). Thus, they are consistent for (δ^*, β^*) , which satisfy the estimating equations in expectation:

$$\begin{aligned} a_j &\equiv \sum_{i=1}^N \sum_{k=0}^K \{E(\tilde{n}_{ijk}) - n_{i..} \pi_{ijk}^*\} = 0, \quad j=1, 2, \dots, J, \\ b_k &\equiv \sum_{i=1}^N \sum_{j=0}^J \{E(\tilde{n}_{ijk}) - n_{i..} \pi_{ijk}^*\} = 0, \quad k=1, 2, \dots, K, \end{aligned} \quad (3)$$

$$\text{where } \pi_{ijk}^* = e_{ijk} e^{\delta_j^* + \beta_k^*} / \sum_{r,s} e_{irs} e^{\delta_r^* + \beta_s^*},$$

$$\begin{aligned} E(\tilde{n}_{ijk}) &= n_{i..} \frac{e_{ijk} e^{\delta_j + \beta_k} + L_{ij} \mu_u (e^{\delta_j + \beta_{k+1}} - e^{\delta_j + \beta_k})}{\sum_{r,s} e_{irs} e^{\delta_r + \beta_s}}, \quad k=0, 1, \dots, K-1, \\ E(\tilde{n}_{ijK}) &= n_{i..} \frac{e_{ijK} e^{\delta_j + \beta_K} + L_{ij} \mu_u (e^{\delta_j + \beta_0} - e^{\delta_j + \beta_K})}{\sum_{r,s} e_{irs} e^{\delta_r + \beta_s}}, \quad k=K, \end{aligned} \quad (4)$$

and L_{ij} is the number of exposures for person i in age group j under the general MECS model described in Section 2. We omit the proof of (4) since it is a straightforward generalization of Theorem 1 in Mohammed et al. (2012). The set of equations given in (3) can be solved numerically to obtain (δ^*, β^*) by the Newton-Raphson method. Thus, although there is no closed-form expression for (δ^*, β^*) , it can be determined for any configuration of (δ, β) , the data $\{\tilde{n}_{ijk}, e_{ijk}\}$ and average level of exposure onset measurement error μ_u . More specifically, the Newton-Raphson update of (δ^*, β^*) at iteration $t+1$ is $(\delta^*, \beta^*)^{(t+1)} = (\delta^*, \beta^*)^{(t)} - (\mathbf{J}^{(t)})^{-1} \mathbf{d}^{(t)}$ with $\mathbf{d}^{(t)} = (a_1^{(t)}, \dots, a_J^{(t)}, b_1^{(t)}, \dots, b_K^{(t)})^T$ and $\mathbf{J}^{(t)}$ is a $(J+K) \times (J+K)$ matrix of partial derivatives evaluated at $(\delta^*, \beta^*)^{(t)}$; see Web Appendices in the Supplemental Materials.

Therefore, despite the complex relationship between (δ^*, β^*) and (δ, β) , using the above calculations, one can determine directly whether (a) $(\delta^*, \beta^*) = \mathbf{0}$ when $(\delta, \beta) = \mathbf{0}$, (b) $\delta^* = \mathbf{0}$ when $\delta = \mathbf{0}$, (c) $\beta^* = \mathbf{0}$ when $\beta = \mathbf{0}$, (d) $\delta_j^* = 0$ when $\delta_j = 0$, for $j = 1, \dots, J$, and (e) $\beta_k^* = 0$ when $\beta_k = 0$, for $k = 1, \dots, K$. We will take this approach to determine the validity of naive tests for the general MECS models. More generally, if the true parameters for the model without measurement error, say $\Omega = \mathbf{0}$, implies that parameters of the model with measurement error, say $\Omega^* = \mathbf{0}$, then the naive test is valid (Carroll et al., 2006, Chap. 10).

We first provide here some insights into the validity of the naive tests by considering a simplified, but instructive, MECS model that provides a closed-form solution to (3). It can be shown (see Supplemental Materials) that under the simplifying assumption of equal follow-up time, equal risk period lengths (i.e., $e_{ik} = e_k$, for $k = 0, \dots, K$) and no age effects

$$\beta_k^* = \begin{cases} \log \left\{ \frac{e_k \theta_k - \mu_u (\theta_k - \theta_{k+1})}{e_0 - \mu_u (1 - \theta_1)} \right\} - \log(e_k / e_0), & k=1, \dots, K-1 \\ \log \left\{ \frac{e_K \theta_K - \mu_u (\theta_K - 1)}{e_0 - \mu_u (1 - \theta_1)} \right\} - \log(e_K / e_0), & k=K \end{cases}, \quad (5)$$

where $\theta_k = \exp(\beta_k)$ is the true relative incidence of events in the k th risk period ($k = 1, \dots, K$) relative to the baseline period ($k = 0$). Furthermore, as noted in the Appendix, the ML estimates for β_k^* in the MECS model without age effects is $\hat{\beta}_k^* = \log(\tilde{n}_{.k} / \tilde{n}_{.0}) - \log(e_k / e_0)$, where $\tilde{n}_{.k} = \sum_i \tilde{n}_{ik}$ is the total number of events in the k th period ($k = 0, \dots, K$).

Thus, the global naive test of $H_0 : \beta = \mathbf{0}$ is valid. However, an examination of (5) reveals that with multiple risk periods, the individual component test of $H_0 : \beta_k = 0$ is generally not valid. More specifically, when $\beta_1 = 0$,

$$\beta_1^* = \log\{1 - \text{RME}_1(1 - \theta_2)\}, \quad (6)$$

which increases away from zero as θ_2 increases away from 1 (i.e., β_2 increases away from 0). Here $\text{RME}_1 = \mu_u/e_1$ is a measure of the relative measurement error (relative to length of the risk period 1), so that RME_1 is not negligible. For example, in our data application, $\text{RME}_1 \approx 6/30 = 0.2$ represents infections occurring about 6 days on average prior to infection-related hospitalization discharge and the risk window of interest is 30 days after infections. From the above equation (6), it is clear that since RME_1 is not small, the naive test of $H_0 : \beta_1 = 0$ is valid only if the relative incidence in the second risk period is 1 or close to 1; i.e., the effect size for risk period 2 is small.

Next, consider the test of $H_0 : \beta_k = 0$ for $k = 2, \dots, K - 1$. From (5), we have

$$\beta_k^* = \log \left\{ \frac{1 - \text{RME}_k(1 - \theta_{k+1})}{1 - \text{RME}_0(1 - \theta_1)} \right\}, \quad (7)$$

where $\text{RME}_k = \mu_u/e_k$. Note that the denominator of (7) involves $\text{RME}_0 = \mu_u/e_0$. For our study of infection-cardiovascular risk (more in Section 5), the mean baseline period is over 750 days (for 2 risk periods) with average exposure onset measurement error of 5.5 days; therefore, $\text{RME}_0 < 0.01$ is small. Hence, when $\text{RME}_0 \approx 0$, as is typical in applications, and for $\theta_1 \approx 1$ (e.g., $\theta_1 \approx 1.6$ in our application),

$$\beta_k^* \approx \log\{1 - \text{RME}_k(1 - \theta_{k+1})\},$$

which is of the same form as β_1^* in (6) for the first risk period. Therefore, the test that the incidence of events in k th risk period is not different from baseline, namely $H_0 : \beta_k = 0$, is valid if the relative incidence for the $k + 1$ risk period, $\theta_{k+1} = 1$ or it is approximately valid if $\theta_{k+1} \approx 1$, since RME_k ($k = 1, \dots, K - 1$) is not close to zero typically. Finally, for last risk period K , the test of $H_0 : \beta_K = 0$ can be seen to be approximately valid from

$$\beta_K^* = \log[\{1 - \text{RME}_0(1 - \theta_1)\}^{-1}] \approx 0, \text{ under similar bounds on } \text{RME}_0 \text{ and } \theta_1.$$

Although we have described the above results for an arbitrary number of risk periods K , for the MECS models with multiple risk periods in practice, typically $K = 1, 2$ or 3 . Combined with the above results, this allows for some additional tractability in determining validity of some naive tests in practice. For example, with $K = 2$, the naive test of $H_0 : \beta_2 = 0$ is approximately valid and failing to reject this hypothesis provides support for the validity of the naive test of $H_0 : \beta_1 = 0$ as well. (This approach, of course, assumes that there is adequate power for detecting β_2 .) Finally, we make two remarks regarding the above results. First, the results hold for the more general MECS models with differential follow-up times among individuals as well as models with multiple exposures. Second, because the exposure onset measurement error directly affects the timing of exposures, age-specific incidence parameters are not affected (biased) critically; therefore, the tests associated with the age effects parameters, δ_j 's, are approximately valid. Data supporting these extensions are provided in the supplemental materials using the above Newton-Raphson calculations.

4 Simulation Assessment of Validity and Power of Naive Tests

As shown above, the individual component naive test of the risk period k is not valid generally in models with multiple risk periods and is only valid or approximately valid when the subsequent period's relative incidence (θ_{k+1}) is 1 or close to 1, respectively. This result holds under assumptions on the relative amount of measurement error, follow-up time and risk period lengths that are reasonable in practice. Thus, in this section, we consider simulation experiments where *the effect sizes (relative incidences) ranging from small (e.g., 25%) to moderate (e.g., 80%)* and with model settings (RME's, follow-up length, risk periods) guided by our previous studies of data from the USRDS.

4.1 Simulation Study Design

To study the validity and relative power of naive hypothesis testing, we simulated data with 3 age groups ($J = 2$), each with length 250 days on average. Follow-up times are different for each individual as in real data applications. For the main simulation results reported in the next section, we simulated data with $N = 500$ individuals with an average follow-up time of 750 days. Similar to data from the USRDS, each individual has $L_i = 1, 2$ or 3 exposures with probabilities 0.55, 0.3 and 0.15, respectively. Exposures are randomly assigned throughout an individual's observation history. Risk periods of length 15, 30 or 45 days are then formed for two risk periods ($K = 2$). These risk period lengths were chosen to correspond to averages of risk length relative to total follow-up times of $r \approx 0.02, 0.04$ and 0.06, respectively, and reflect typical ranges in real data. Because the results are similar, we discuss the case for $r \approx 0.04$ and the other studies are presented in the supplemental materials. Marginal totals are generated according to the case series model from the non-homogeneous Poisson model given by $n_{i..} \sim \text{Poisson}(\sum_{jk} e_{ijk} \lambda_{ijk})$ where $\lambda_{ijk} = \exp(\phi_i + \delta_j + \beta_k)$ with $\phi_i = \log(1/10000)$ fixed. Finally, these marginal totals are randomly distributed throughout each individual's observation period based on the multinomial probabilities shown in (1).

Uniformly distributed exposure onset measurement error is added to each individual's true exposure times to create the observed data with error. The naive tests are then applied to the observed data which have exposure onset measurement error. More specifically, for each risk period length, we set the average exposure onset measurement error μ_u to be such that the relative measurement error (RME), relative to the length of the risk period, is about 10%, 20% and 30%. The patterns of the age and risk effects, $\delta = (\delta_1, \delta_2)$ and $\beta = (\beta_1, \beta_2)$, respectively, are varied throughout the simulations. A summary of the simulation parameter settings can be found in Table 1. For each simulation configuration, we generated 2000 datasets. As summarized in Table 1, we also considered normal and gamma distributed exposure onset measurement error; however, the results are similar and we present only the uniformly distributed error case and defer the remaining results to the supplemental materials.

We apply LRT's to each observed dataset and calculate the power as the proportion of 2000 datasets to reject the null hypothesis under the alternative. To determine the optimal power as a benchmark, the LRT is based on the same true dataset without measurement error. To empirically determine the validity of a specific hypothesis test, its Type I error rate is

similarly calculated as the proportion of 2000 datasets to reject the null hypothesis when the null hypothesis is true. Note that with 2000 replicates, we expect with 0.95 simulation probability that, if the true Type I error rate is 5%, the simulated estimate of that rate will be between 4% and 6%.

4.2 Validity of Naive Tests

Empirical Type I errors for naive hypothesis testing at level $\alpha = 0.05$ are presented in Table 2 for risk lengths of 30 days. Since the results are similar, we will refer mainly to results in data with 30-day risk periods in the summary of results below and results for other risk lengths are presented in the supplemental materials.

Testing under $H_0 : (\delta, \beta)^T = \mathbf{0}$ —The naive test of this global null hypothesis is valid. The Type I error is close to the nominal level α for all levels of average exposure onset measurement error (RME of 0.1, 0.2 and 0.3). As expected under the global null, the Type I errors for all naive tests of hypotheses 2, 3 and 4 achieve the nominal level similar to the optimal test based on data without exposure onset measurement error.

Testing under $H_0 : \delta = \mathbf{0}$ and $H_0 : \beta = \mathbf{0}$ —The naive test of the overall null risk or overall null age effects hypotheses are both valid. For example, with the true model $(e^{\delta_1}, e^{\delta_2}, e^{\beta_1}, e^{\beta_2}) = (1.25, 1.5, 1, 1)$ the Type I error for the test of $\beta = \mathbf{0}$ achieves the nominal level similar to the optimal test. This holds for different patterns of age-specific relative incidence e^{δ_j} (i.e., increasing or decreasing; see tables 2). This similarly holds for testing $H_0 : \delta = \mathbf{0}$.

Testing under $H_0 : \delta_j = 0, j = 1, \dots, J$ —The observed Type I error for the test of an individual component of δ also achieved the nominal test level. For instance, when $\delta_j = 0$ and all other parameters are non-null ($j = 1, \dots, J$) then the test of $H_0 : \delta_j = 0$ is valid, such as in the case of $(e^{\delta_1}, e^{\delta_2}, e^{\beta_1}, e^{\beta_2}) = (1, 1.3, 1.8, 1.5)$; see Table 2. This result confirms that exposure onset measurement error has a small impact on age effects for small to moderate effect sizes; however, this holds for even larger effect sizes (as illustrated in the supplemental materials).

Testing under $H_0 : \beta_k = 0, k = 1, \dots, K$ —Not all tests of the individual components of the risk period effects are valid. More specifically, the test for β_k is invalid when the relative incidence corresponding to the next contiguous period is not 1, $\beta_{k+1} \neq 0$ ($k = 1, \dots, K - 1$). That is, the Type I error does not approach the nominal level of the test and, in fact, increases as the average amount of measurement error (μ_u) increases. When $\beta_{k+1} \approx 0$, the test for β_k is approximately valid. The test for β_K is approximately valid since RME_0 is small. Thus, under the overall null risk effects, the component-wise tests are valid. For instance, with $(e^{\delta_1}, e^{\delta_2}, e^{\beta_1}, e^{\beta_2}) = (1.25, 1.5, 1, 1)$ the tests of $H_0 : \beta_1 = 0$ and $H_0 : \beta_2 = 0$ are both valid. (See highlighted entries of Table 2.) To illustrate the case where the test of $H_0 : \beta_k = 0$ is not valid, consider the case with $(e^{\delta_1}, e^{\delta_2}, e^{\beta_1}, e^{\beta_2}) = (1.2, 1.3, 1, 1.5)$, highlighted in Table 2. Here the test of $H_0 : \beta_1 = 0$ is no longer valid since $\beta_2 \neq 0$, where the Type I errors are 5.5%, 8.25% and 12.65% corresponding to increasing RME of 10% (low), 20% (moderate) and 30% (high).

Although for ease of exposition, we have described the results above based on the model with two risk periods, the results hold more generally with more than two risk periods; additional studies are provided as supplementary materials. Taken together, the results provide specific guidance on the use of naive hypothesis testing for the MECS models. In particular, for the model most often used in practice, with unequal follow-up times, multiple age groups and a single risk period (e.g., 30 days after infection), all associated tests of interest are valid. Some caution must be exercised for MECS models with several risk periods when testing for a specific risk period effect. Guidelines are further discussed in Section 6.

4.3 Power of Naive Tests

We next summarize the power of the naive tests using the data while ignoring error in the exposure onset times. We also compare this to the power of the “optimal” test which is based on the corresponding data with precisely measured exposure onset times. Results from the case with 30-day risk periods are given in Table 3 ($r = 0.04$) and are discussed in more detail below. Similar patterns hold for 15-day ($r = 0.02$) and 45-day ($r = 0.06$) risk periods as well as normal and gamma distributed measurement error. (See supplemental materials.)

Power for testing the global null—When the relative incidences in the risk periods are null ($\beta = 0$) the power of the naive tests is similar to the optimal power, regardless of the relative incidences in the age groups. However, the naive test of the global null hypothesis (when $\beta = 0$) has reduced power and the reduction is more with increasing levels of average amount of exposure onset measurement error. For instance, with $(e^{\delta_1}, e^{\delta_2}, e^{\beta_1}, e^{\beta_2}) = (1.2, 1.3, 1.5, 1.3)$ the power of the naive test is 82.2%, 78.2% and 76.1% for RME of 0.1, 0.2 and 0.3, while the optimal test has about 84% power. See highlighted entries in Table 3(A).

Power for testing age effects—Regardless of whether the relative incidences in the risk periods are null or non-null, i.e., under the models with $\{\delta = 0, \beta = 0\}$ or $\{\delta = 0, \beta \neq 0\}$ all naive tests associated with any age effects have similar power as their corresponding optimal tests. For example, consider the case of $(e^{\delta_1}, e^{\delta_2}, e^{\beta_1}, e^{\beta_2}) = (1.25, 1.5, 1.25, 1.5)$ with the relative measurement error of 30%, highlighted in Table 3(B). Evidently, the differences between the power of optimal tests and the power of the corresponding naive tests for $H_0 : \delta = 0$ and $H_0 : \delta_j = 0$ are all negligible. This is expected because the measurement error is on the times of exposures (unrelated to age groups), and therefore does not significantly bias age effects.

Power for testing effects of exposure risk periods—These are the tests of primary interest in the case series models. As expected, exposure onset measurement error has substantial impact on power, which is not always attenuation, with increasing levels of average error under the MECS models. This is analogous to the patterns of bias of the case series naive estimators characterized in Mohammed et al. (2012). The following summarizes the impact:

(A) The naive test of the overall null risk effect across all exposure risk periods, $H_0 : \beta = 0$, has reduced power. There is increasing loss of power with increasing level of average

amount of exposure onset measurement error (RME increasing from 0.1 to 0.3) generally. For example, with $(e^{\delta_1}, e^{\delta_2}, e^{\beta_1}, e^{\beta_2}) = (1.3, 1, 1.8, 1.5)$ the powers of the naive tests are 95.3%, 91.4% and 87.7% for RME of 0.1, 0.2 and 0.3, respectively, while the optimal test has about 97% power. See highlighted entries in Table 3(C). As expected, this loss in power occurs regardless of whether the baseline incidence is age-dependent.

(B) The power of the naive test of a specific exposure risk period, $H_0 : \beta_k = 0$, does not have a simple pattern of attenuation as in the above tests. The power can be attenuated or inflated (beyond the power of the “optimal” test based on data without measurement error), depending on the pattern of the true relative incidences of the risk periods. More precisely, the reduction or increase of power of the naive test for β_k is dependent on the true pattern of relative incidences across risk periods and the targets of the biased naive case series estimates, namely $\{\beta_k^*, k = 1, \dots, K\}$. The patterns of bias of the naive estimates have been characterized by Mohammed et al. (2012), who found that if the true relative incidence in risk period k is greater than the true relative incidence in the next contiguous risk period, i.e., if $\beta_k > \beta_{k+1}$, then β_k^* will be *attenuated*. However, if $\beta_k < \beta_{k+1}$, then β_k^* will be *inflated*. Since the next contiguous risk period to β_K (the last risk period) is the baseline period, β_K^* will always be attenuated. Thus, it is expected that if $\beta_k < \beta_{k+1}$, the power in the naive test is increased and it is decreased when $\beta_k > \beta_{k+1}$, for $k = 1, \dots, K - 1$, relative to the benchmark. The power of the naive test for β_K is always reduced relative to the optimal test. These observations are confirmed in the simulation studies summarized in Tables 3(C). For example, the power of the naive test of $H_0 : \beta_1 = 0$ is increased to 61.5% compared to 49.4% for the optimal test, when the true log relative incidences are $\beta_1 = 1.4$ and $\beta_2 = 1.8$ at RME = 0.2 (see highlighted entries in Table 3(C)). As expected, the power of the naive test of $H_0 : \beta_2 = 0$ is reduced.

5 Application: Inference for Infection-Cardiovascular Risk

We illustrate the naive hypothesis testing for the MECS models using USRDS hospitalization data to investigate infection-cardiovascular (CV) risk in older patients on dialysis. As reported in Dalrymple et al. (2011), in the short period following an infection, specifically the approximate first 30 days after infections, the effects of infection on vascular endothelium is hypothesized to be most pronounced. Infection may also contribute to a state of subclinical inflammation that potentially affects atherogenesis, progression of atherosclerosis, or possibly both. Previous studies have supported the association between infections and cardiovascular events, both in the general population (Smeeth et al., 2004) and in the dialysis population (Dalrymple et al., 2011; Mohammed et al., 2012). We first focus on this single risk period in a MECS model with three age groups (age 65-75, 76-85, and > 85). The USRDS dialysis cohort for this analysis includes $N = 16,779$ patients 65-100 years of age who started dialysis between January 1, 2000 and December 31, 2002. Individuals were followed until death, transplant or study end on December 31, 2004. Cardiovascular events of interest were myocardial infarction, unstable angina, stroke, or transient ischemic attack and infections of interest included septicemia, bacteremia, peritonitis, endocarditis, soft-tissue, pulmonary, genitourinary, gastrointestinal, joint or bone infection. As detailed in Section 1, the precise time of exposure/infection for USRDS

inpatient hospitalization data is not available, but the observed discharge date is a good marker for the time of infection since it reasonably assures that the infection had occurred by this date.

We first apply the CS analysis with 3 age groups and a single risk period. We note that due to a violation of the assumption of constant risk within a risk period in the data, we define the risk period to be days 6-30 after an infection, following our previous work in Mohammed et al. (2012). By applying the CS analysis to the USRDS data while ignoring exposure onset measurement error, we obtain the naive estimates $\hat{\beta}_1^* = 0.2938$, $\hat{\delta}_1^* = 0.6346$ and $\hat{\delta}_2^* = 1.4076$ (with naive standard errors 0.0313, 0.0517 and 0.0908, respectively). Although not needed for naive hypothesis testing, for comparison we also obtained bias-corrected estimates using the method introduced in Section 1 and described in more details in Mohammed et al. (2012): $\hat{\beta}_1 = 0.4597$, $\hat{\delta}_1 = 0.6302$, $\hat{\delta}_2 = 1.3987$. As expected for models with a single risk period, there is attenuation in the relative incidence of CV events. Thus, the approximate 30-day risk period after infection is associated with a 58% increase in the incidence of CV events compared to the baseline period ($\exp(\hat{\beta}_1) = 1.58$). Note also that the incidences of CV events in the older age groups are higher relative to the youngest age group (age 65-75). For this MECS model, all tests from Section 2 are valid. Specifically, these are $H_0 : (\beta_1, \delta) = \mathbf{0}$, $H_0 : \beta_1 = 0$, $H_0 : \delta = 0$, $H_0 : \delta_1 = 0$ and $H_0 : \delta_2 = 0$. The corresponding observed LRT test statistics T are 351.07, 82.53, 262.83, 153.39 and 244.94, respectively, each with p-value < 0.0001 . Thus, of particular interest, the relative incidence of CV events is significantly elevated in the approximate 30-day risk period following infections.

As a second illustration of the naive hypothesis testing approach and its limitations, we consider a second MECS model with two risk periods where the second risk period spans days 31-60. This MECS model allows for potentially different risk effects associated with the first and second months following infections. The naive CS estimates are $\hat{\beta}_1^* = 0.3146$, $\hat{\beta}_2^* = 0.1544$, $\hat{\delta}_1^* = 0.6304$ and $\hat{\delta}_2^* = 1.3999$. The bias-corrected estimates are $\hat{\beta}_1 = 0.4860$, $\hat{\beta}_2 = 0.1919$, $\hat{\delta}_1 = 0.6250$ and $\hat{\delta}_2 = 1.3891$. Focusing on the primary interest, we observe that the relative incidence of CV events is about 63% higher ($\exp(\hat{\beta}_1) = 1.63$) in the first risk period and is 21% higher in the second risk period, relative to the baseline period. Proceeding to naive hypothesis testing, we first reject the global null, $H_0 : (\delta, \beta)^T = \mathbf{0}$, of no age and risk effects (LRT with 4 degrees of freedom (DF) is $T = 369.82$ with p-value < 0.0001). Baseline relative incidences significantly vary with age; $H_0 : \delta = \mathbf{0}$, $H_0 : \delta_1 = 0$ and $H_0 : \delta_2 = 0$ are all rejected (all p-values < 0.0001). Finally, we turn to testing the main parameters of interest involving the risk periods. The hypothesis that the incidences of CV events in both risk periods simultaneously are not different from baseline, i.e., $H_0 : \beta = \mathbf{0}$, is rejected ($T = 101.27$, 2 DF, p-value < 0.0001). All naive tests performed thus far are valid. Furthermore, since the test of $H_0 : \beta_K = 0$ is approximately valid (because $\text{RME}_0 = 5.5/766.7$ is small), we test the significance of the relative incidence of CV events in the 31-60 day risk period ($H_0 : \beta_2 = 0$); the data suggests that the relative incidence of CV events is significantly elevated in this second risk period ($T = 18.74$, 1 DF, p-value < 0.0001). Thus, we conclude that $\beta_2 > 0$. The remaining test regarding β_1 associated with the first risk period is generally not valid

unless $\beta_2 = 0$. In this case the naive test of $H_0 : \beta_1 = 0$ is also rejected ($T = 92.32$, 1 DF, p-value < 0.0001), possibly supporting that the incidence of CV events in the approximate first 30 days after infection is also significantly elevated, but the Type I error is likely inflated (as shown in Section 4.2). Therefore, we also considered the test for β_1 based on the biased-corrected test statistic; although less powerful, this test led to the same conclusion that the first risk period is associated with significantly increased CV incidence ($T_{BC} = 8.3959$). In summary, these MECS model results suggest that the incidence of CV events is highest immediately following infection (risk period 1) and the CV risk persists in the second risk period although with reduced CV incidence, but still significantly greater than baseline.

6 Discussion

In this work, we provided a practical approach to inference for the MECS model based on naive hypothesis testing. The rationale of the approach parallels the standard used in measurement error problems generally, which is that if the naive test is valid then it is more powerful because naive (uncorrected) estimators have smaller variance compared to bias-corrected estimators. However, bias-corrected estimation is necessary to understand the direction and magnitude of effect sizes; in our application this is the relative cardiovascular incidence following infections in patients on dialysis. It is necessary because the bias does not decline with infinite sample size. Thus, an overall practical approach to inference for CS analysis in the presence of exposure onset measurement error is to consider both (a) bias-correction in order to obtain valid estimates of relative incidences and (b) naive hypothesis testing to improve power.

A practical advantage to naive tests is that no new data is needed since they only involve carrying out standard CS analysis while ignoring the exposure onset measurement error. CS analysis can be implemented easily in standard software; details are provided in the Supplemental Materials. The totality of the results in the current study led us to several practical guidelines for hypothesis testing in the presence of exposure onset measurement error in the CS analysis as follows. (1) In the commonly applied model with one risk period, with or without age groups, all relevant naive tests described in Section 2 are valid.

(2) In the common model with two risk periods, with or without age groups, the naive test regarding the second risk period (β_2) is approximately valid when $RME_0 \approx 0$, a condition that is typically satisfied in practice, but more importantly, can be easily checked. The naive test to determine the statistical significance of the first risk period is valid if $\beta_2 = 0$, the incidence of events in the second risk period is not different from baseline. It is approximately valid if $\beta_2 \approx 0$. (All other naive tests in Section 2 are valid or approximately valid.)

(3) In the general model with $K - 3$ risk periods and $J - 0$ age groups, the naive test of the global null and all naive tests regarding age group effects are valid and approximately valid, respectively. The naive test of the last risk period β_K is approximately valid when $RME_0 \approx 0$. However, the naive test regarding β_k has inflated Type I error that increases with increasing average exposure onset measurement error if β_{k+1} is non-null ($k = 1, \dots, K - 1$). Our study suggests that the inflation is small to moderate (e.g., inflation to $< 15\%$ for testing

at the 5% level for the case of relatively high relative measurement error of 30% and the inflation is negligible at 10% RME). Thus, the naive test of β_k should be augmented with other valid approaches, such as bootstrap tests or tests based on the bias-corrected estimates; although these valid alternatives are less powerful, they will nonetheless be informative in these cases.

Finally, we note that although our specific research application is focused on outcomes for patients on dialysis using the the United States Renal Data System, the use of ICD-9-CM (International Classification of Disease, 9th Revision, Clinical Modification) discharge diagnosis is an international standard. Thus, similar case series applications to longitudinal hospitalization or other administrative databases with ICD-9-CM codes have the same timing or exposure onset error. Thus, our application is relevant to other applications of the case series method to longitudinal observational databases, including hospital claims or administrative databases, such as adverse events due to medications or other types of hospitalizations. Although not our expertise, other applications include drug safety surveillance and pharmacoepidemiology. Also, as mentioned by a referee, the use of administrative records of the date of vaccine purchase, which predates its use is another example of exposure onset measurement error. In this case the exposure onset measurement error is negative and $\mu_u = E(u)$ is simply replaced with $E|u|$. See Mohammed et al. (2012). Generally, records of medication prescriptions do not necessarily coincide with their time of uses and case series analysis of such data needs to account for exposure onset measurement error.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute of Health grants #UL1RR024146 and #R01DK092232. 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 United States government. This study was approved by the Institutional Review Board of the University of California, Davis. We thank the Editor, Associated Editor and 2 reviewers.

References

- Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. Measurement error in nonlinear models: A modern perspective. Boca Raton: Chapman and Hall/CRC; 2006.
- Dalrymple LS, Mohammed SM, Mu Y, Johansen KL, Chertow GM, Grimes B, Kaysen GA, Nguyen DV. The risk of cardiovascular-related events following infection-related hospitalizations in older patients on dialysis. *Clinical Journal of the American Society of Nephrology*. 2011; 6:1708–1713. [PubMed: 21566109]
- Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics*. 1995; 51:228–235. [PubMed: 7766778]
- Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. *American Journal of Epidemiology*. 1996; 143:1165–1173. Erratum (1998) **147**, 93. [PubMed: 8633607]
- Madigan D, Ryan P, Simpson S, Zorych I. Bayesian methods in pharmacovigilence. *Bayesian Statistics*. 2011; 8:421–438.

- Maclure M. The case-cross-over design: A method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*. 1991; 21:144–153. [PubMed: 1985444]
- Mohammed SM, Senturk D, Dalrymple DS, Nguyen DV. Measurement error case series models with application to infection-cardiovascular risk in older patients on dialysis. *Journal of the American Statistical Association*. 2012; 107:1310–1323. [PubMed: 23650442]
- Pratt NL, Roughead EE, Ramsay E, Salter A, Ryan P. Risk of hospitalization for stroke associated with antipsychotic use in the elderly: A self-controlled case series. *Drugs and Aging*. 2010; 27:885–889. [PubMed: 20964462]
- Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. *New England Journal of Medicine*. 2004; 351:2611–2618. [PubMed: 15602021]
- U S Renal Data System. *USRDS 2010 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States*. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; Bethesda, MD: 2010.
- Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: The self-controlled case series method. *Statistics in Medicine*. 2006; 25:1768–1797. [PubMed: 16220518]

Table 1

Summary of the simulation study design parameters.

Parameter	Values
Sample size, N	500 (200, 700, 1000 in supplemental materials)
Age group relative incidences, $\exp(\delta_j)$	1.2, 1.25, 1.3, 1.4, 1.5, 1.8
Risk period relative incidences, $\exp(\beta_k)$	1.25, 1.4, 1.4, 1.5, 1.8
Number of exposures, L_i	1, 2 or 3 with probabilities 0.55, 0.3 and 0.15
Length of risk period	15, 30, 45 days (i.e., $r \approx 0.02, 0.04, 0.06$)
Relative measurement error (RME)	10%, 20%, 30%
Measurement error distribution	Uniform, Normal, Gamma
Average measurement error, μ_u	3, 4.5, 6, 9, 12, 13.5, 18

Table 2

Empirical Type I errors (percent) based on 2000 simulated datasets with 2 30-day risk periods and 3 age groups for varying amounts of relative measurement error (RME).

RME	(A) Global						(B) Age Effects						(C) Risk Effects					
	$H_0: (\delta, \beta) = 0$						$H_0: \delta_i = 0$						$H_0: \delta_2 = 0$					
	e^{δ_1}	e^{δ_2}	e^{β_1}	e^{β_2}	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive
0.1	1	1	1	1	4.6	4.8	4.7	4.7	4.7	4.7	4.7	4.7	5.8	3.8	4.0	4.0	4.3	4.2
0.2	1	1	1	1	4.9	5.4	5.5	5.6	4.6	4.6	4.9	5.0	5.2	4.8	4.9	4.5	4.5	5.3
0.3	1	1	1	1	4.7	4.5	5.1	5.1	4.6	4.9	5.3	5.2	4.4	4.4	5.0	4.3	4.1	4.9
0.1	1.25	1.5	1	1									5.1	5.3	5	5.6	5.2	4.95
0.2	1.25	1.5	1	1									5.4	5.8	4.8	4.9	5.75	5.8
0.3	1.25	1.5	1	1									5.9	5.3	5.45	5.15	4.35	4.7
0.1	1.5	1.25	1	1									5.0	5.0	5.8	6.1	4.6	4.5
0.2	1.5	1.25	1	1									4.5	5.2	4.8	4.4	5.4	5.9
0.3	1.5	1.25	1	1									5.6	4.8	6.0	5.5	4.9	4.4
0.1	1.4	1.8	1	1									5.0	5.3	5.0	5.6	4.4	4.7
0.2	1.4	1.8	1	1									4.0	4.9	5.0	4.7	4.5	5.0
0.3	1.4	1.8	1	1									6.8	6.5	5.8	5.9	5.7	6.1
0.1	1.8	1.4	1	1									5.7	5.8	6.4	6.6	4.5	5.0
0.2	1.8	1.4	1	1									4.9	5.0	4.4	4.3	5.4	5.3
0.3	1.8	1.4	1	1									4.6	4.8	5.0	5.1	4.8	4.5
0.1	1	1	1.25	1.5			4.6	4.7	4.4	4.4	4.7	4.8						
0.2	1	1	1.25	1.5			5.7	5.6	5.6	5.5	4.8	4.8						
0.3	1	1	1.25	1.5			5.9	5.8	5.4	5.4	5.3	5.4						
0.1	1	1	1.5	1.25			4.9	4.8	5.0	5.0	4.7	4.5						
0.2	1	1	1.5	1.25			5.9	6.1	5.3	5.4	4.8	4.8						
0.3	1	1	1.5	1.25			4.4	4.3	4.7	4.8	4.1	4.3						
0.1	1	1	1.4	1.8			4.6	4.7	5.5	5.6	4.1	4.2						

RME	(A) Global						(B) Age Effects						(C) Risk Effects					
	$H_0 : (\delta, \beta) = 0$						$H_0 : \delta = 0$						$H_0 : \delta_1 = 0$					
	e^{δ_1}	e^{δ_2}	e^{β_1}	e^{β_2}	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive
0.2	1	1	1.4	1.8	5.0	5.0	5.3	5.2	4.7	4.6								
0.3	1	1	1.4	1.8	5.4	5.2	5.4	5.3	4.9	5.0								
0.1	1	1	1.8	1.4	4.9	4.8	5.0	5.1	5.9	5.9								
0.2	1	1	1.8	1.4	4.6	4.3	5.0	5.0	4.4	4.8								
0.3	1	1	1.8	1.4	4.9	5.1	4.9	4.9	4.8	5.0								
0.1	1	1.3	1.8	1.5			5.7	5.7										
0.2	1	1.3	1.8	1.5			4.8	4.8										
0.3	1	1.3	1.8	1.5			5.4	5.3										
0.1	1.3	1	1.8	1.5					5.1	5.1								
0.2	1.3	1	1.8	1.5					5.5	5.6								
0.3	1.3	1	1.8	1.5					5.4	5.5								
0.1	1.2	1.3	1	1.5							5.0	5.4						
0.2	1.2	1.3	1	1.5							5.5	8.25						
0.3	1.2	1.3	1	1.5							5.0	12.65						
0.1	1.2	1.3	1.5	1									3.6	4.4				
0.2	1.2	1.3	1.5	1									5.3	5.8				
0.3	1.2	1.3	1.5	1									5.4	5.8				

Table 3

Empirical power (percent) based on 2000 simulated datasets with 2 30-day risk periods and 3 age groups for varying amounts of relative measurement error (RME).

RME	(A) Global						(B) Age Effects						(C) Risk Effects					
	$H_0 : (\delta, \beta) = 0$			$H_0 : \delta = 0$			$H_0 : \delta_1 = 0$			$H_0 : \delta_2 = 0$			$H_0 : \beta = 0$			$H_0 : \beta_1 = 0$		
	e^{δ_1}	e^{δ_2}	e^{β_1}	e^{β_2}	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive
0.1	1.25	1.5	1	1	86.9	87.4	93.6	93.7	50.4	50.5	97.0	97.0						
0.2	1.25	1.5	1	1	86.8	86.8	92.6	92.7	49.2	49.3	96.1	96.1						
0.3	1.25	1.5	1	1	86.0	85.8	92.2	92.3	51.3	51.2	96.1	96.2						
0.1	1.5	1.25	1	1	86.9	86.2	92.7	92.7	95.9	96.0	49.4	49.8						
0.2	1.5	1.25	1	1	88.4	88.3	94.2	94.1	97.1	97.2	50.7	51.0						
0.3	1.5	1.25	1	1	88.2	87.8	93.3	93.2	96.6	96.5	49.6	49.5						
0.1	1.4	1.8	1	1	99.8	99.9	100.0	100.0	83.2	83.3	100.0	100.0						
0.2	1.4	1.8	1	1	99.8	99.7	99.9	99.9	82.2	82.0	100.0	100.0						
0.3	1.4	1.8	1	1	99.8	99.9	100.0	100.0	83.5	83.6	100.0	100.0						
0.1	1.8	1.4	1	1	99.8	99.6	99.9	99.9	100.0	100.0	83.5	83.4						
0.2	1.8	1.4	1	1	99.8	99.8	100.0	100.0	100.0	100.0	84.6	84.5						
0.3	1.8	1.4	1	1	99.5	99.7	99.9	99.9	100.0	100.0	80.0	80.2						
0.1	1	1	1.25	1.5	55.0	50.3							66.4	61.3	24.6	27.0	69.1	60.8
0.2	1	1	1.25	1.5	55.5	45.1							65.5	55.4	22.3	29.9	68.8	51.5
0.3	1	1	1.25	1.5	54.5	39.7							65.7	50.2	25.7	37.1	66.6	39.8
0.1	1	1	1.5	1.25	55.0	49.8							65.9	59.9	67.7	64.0	25.3	21.1
0.2	1	1	1.5	1.25	56.8	45.4							66.2	54.2	68.5	59.6	24.7	16.9
0.3	1	1	1.5	1.25	56.8	37.8							66.9	48.2	69.3	53.7	24.7	12.0
0.1	1	1	1.4	1.8	90.5	87.6							94.7	93.2	52.3	58.3	94.6	90.8
0.2	1	1	1.4	1.8	92.6	83.6							95.9	91.0	49.4	61.5	95.2	84.2
0.3	1	1	1.4	1.8	91.7	80.2							95.3	87.1	50.7	68.9	94.8	73.6
0.1	1	1	1.8	1.4	91.9	87.5							95.8	93.4	95.6	93.4	48.9	41.9

(A) Global										(B) Age Effects						(C) Risk Effects					
RME	e^{δ_1}	e^{δ_2}	e^{β_1}	e^{β_2}	$H_0 : (\delta, \beta) = 0$		$H_0 : \delta = 0$		$H_0 : \delta_1 = 0$		$H_0 : \delta_2 = 0$		$H_0 : \beta = 0$		$H_0 : \beta_1 = 0$		$H_0 : \beta_2 = 0$				
					Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	
0.2	1	1	1.8	1.4	90.8	81.3								95.0	88.7	94.7	89.5	50.3	32.2		
0.3	1	1	1.8	1.4	91.1	73.9								95.6	83.9	94.8	86.1	51.1	26.2		
0.1	1.5	1.25	1.25	1.5	98.0	97.5	92.4	92.4	96.2	96.2	50.7	50.4	68.3	63.1	26.8	29.2	69.4	62.6			
0.2	1.5	1.25	1.25	1.5	97.7	96.7	92.4	92.4	96.6	96.5	47.7	46.9	67.2	56.5	24.5	31.5	70.1	51.3			
0.3	1.5	1.25	1.25	1.5	97.65	96.45	93.75	94	96.4	96.5	51.35	50.7	66.6	50.2	25.9	36.7	69.9	40.4			
0.1	1.5	1.25	1.8	1.4	99.9	99.7	93.3	93.4	96.8	96.8	51.1	50.3	95.9	94.0	94.9	93.2	50.8	44.5			
0.2	1.5	1.25	1.8	1.4	100.0	99.8	93.6	93.4	97.0	97.1	49.8	48.1	96.0	89.2	95.3	89.8	51.7	34.3			
0.3	1.5	1.25	1.8	1.4	99.8	99.0	94.0	94.1	97.0	97.0	49.5	47.2	95.9	84.1	96.1	86.7	50.6	24.9			
0.1	1.4	1.8	1.25	1.5	100.0	100.0	100.0	100.0	82.4	82.4	100.0	100.0	68.0	63.0	24.2	27.9	70.5	62.8			
0.2	1.4	1.8	1.25	1.5	100.0	99.8	99.9	99.9	82.5	82.6	99.9	99.9	65.1	52.9	23.8	30.7	67.3	50.4			
0.3	1.4	1.8	1.25	1.5	100.0	100.0	100.0	100.0	83.4	83.5	100.0	100.0	67.7	52.6	24.2	37.2	69.9	41.2			
0.1	1.4	1.8	1.8	1.4	100.0	100.0	100.0	100.0	81.7	81.6	100.0	100.0	94.9	92.6	94.3	92.4	51.5	42.4			
0.2	1.4	1.8	1.8	1.4	100.0	100.0	100.0	100.0	82.9	83.3	100.0	100.0	95.5	88.1	94.8	89.8	52.3	33.0			
0.3	1.4	1.8	1.8	1.4	100.0	100.0	99.9	99.8	81.4	82.2	100.0	100.0	95.2	82.0	94.7	85.4	52.4	23.9			
0.1	1.2	1.3	1.5	1.3	84.3	82.2	59.1	58.8	38.0	37.9	67.0	66.4	69.0	63.3	67.9	62.9	33.1	28.9			
0.2	1.2	1.3	1.5	1.3	83.85	78.2	59.7	58.4	37.5	37.7	68.5	67.8	68.5	55.3	66.2	59.6	31.6	21.4			
0.3	1.2	1.3	1.5	1.3	84.65	76.05	59.4	58.5	38.4	38.8	66.9	65.5	68.6	50.4	68.3	55.8	33.2	16.7			
0.1	1	1.3	1.8	1.5	99.1	98.5	73.0	71.6			70.3	69.5	97.4	95.3	94.7	93.1	66.2	58.1			
0.2	1	1.3	1.8	1.5	98.2	96.4	74.3	73.1			72.4	70.8	96.0	91.2	94.6	90.7	65.7	46.6			
0.3	1	1.3	1.8	1.5	98.7	94.6	74.0	71.2			69.8	67.9	96.7	86.7	95.0	87.1	64.7	36.5			
0.1	1.3	1	1.8	1.5	99.2	98.9	74.2	74.6	71.7	71.8			97.25	95.3	95.6	93.6	68.8	59.4			
0.2	1.3	1	1.8	1.5	99.4	98.1	74.2	75.8	69.9	70.2			96.75	91.4	95.0	91.0	66.9	47.4			
0.3	1.3	1	1.8	1.5	99.3	97.2	74.1	75.9	70.2	70.8			96.85	87.7	96.1	88.2	67.0	36.5			
0.1	1.2	1.3	1	1.5	80.0	76.0	56.5	56.4	37.7	37.6	66.7	66.5	58.7	51.1			68.5	60.9			
0.2	1.2	1.3	1	1.5	82.7	73.4	59.8	60.0	39.2	39.5	68.6	68.5	60.0	43.1			68.4	51.0			

RME	e^{δ_1}	e^{δ_2}	e^{β_1}	e^{β_2}	(A) Global				(B) Age Effects				(C) Risk Effects							
					$H_0 : (\delta, \beta) = 0$				$H_0 : \delta_1 = 0$				$H_0 : \delta_2 = 0$		$H_0 : \beta = 0$		$H_0 : \beta_1 = 0$		$H_0 : \beta_2 = 0$	
					Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive	Opt.	Naive
0.3	1.2	1.3	1	1.5	81.7	70.9	59.6	59.3	38.9	39.0	67.5	67.2	58.7	37.1			68.5	41.3		
0.1	1.2	1.3	1.5	1	79.2	74.8	58.5	58.2	38.6	38.9	66.9	66.6	59.3	50.2	69.8	61.1				
0.2	1.2	1.3	1.5	1	81.2	71.8	61.2	60.4	40.2	40.2	68.4	67.6	58.7	42.1	69.7	49.7				
0.3	1.2	1.3	1.5	1	79.9	66.8	59.8	59.0	40.4	41.1	67.4	66.0	59.6	31.8	68.2	38.6				