Title
Bias and estimation under misspecification of the risk period in self-controlled case series studies

Journal
Stat, 6(1)

ISSN
2049-1573

Authors
Campos, Luis Fernando
Şentürk, Damla
Chen, Yanjun
et al.

Publication Date
2017

DOI
10.1002/sta4.166

Peer reviewed

HHS Public Access

Author manuscript

Stat (Int Stat Inst). Author manuscript; available in PMC 2018 November 21.

Published in final edited form as:

Stat (Int Stat Inst). 2017 ; 6(1): 373–389. doi:10.1002/sta4.166.

Bias and estimation under misspecification of the risk period in self-controlled case series studies

Luis Fernando Campos[a], **Damla Şentürk**[b], **Yanjun Chen**[c], and **Danh V. Nguyen**[d]

[a]Department of Statistics, Harvard University, Cambridge, MA 02138, USA

[b]Department of Biostatistics, University of California, Los Angeles, CA 90095, USA

[c]Institute for Clinical and Translational Science, Irvine, CA 92617, USA

[d]Department of Medicine, University of California Irvine, Orange, CA 92868, USA

Abstract

The self-controlled case series (SCCS) method is useful for estimating the relative incidence (RI) of acute events, such as adverse events (AEs) during a specified risk period following an exposure (e.g., 6-week period after vaccinations or 30-day period after infection-related hospitalizations). In practice, the "optimal" risk period is unknown and must be specified. To date, two approaches are available to guide the specification of the risk period. Both methods do not fully utilize the nature of the bias due to misspecification, which to date has not been characterized. Thus, we elucidate the bias of SCCS estimate of the RI when the risk period is misspecified. We then propose a novel method that more effectively estimates the optimal risk period and the associated RI of AEs. The new method incorporates information on the functional form of the bias. Efficacy of the proposed approach is illustrated with substantial reduction in bias and variance in simulation studies. The proposed method is illustrated with two SCCS studies to determine the (1) risk of idiopathic thrombocytopenic purpura after measles-mumps-rubella vaccination in children and (2) risk of cardiovascular events after infection-related hospitalizations in older patients on dialysis.

Keywords

case series analysis; dialysis; idiopathic thrombocytopenic purpura; infection; maximum likelihood; measles-mumps-rubella; risk length; non-homogeneous Poisson process; vaccination

1. Introduction

The self-controlled case series (SCCS) method was proposed by Farrington (1995) as an approach to study the relationship between time-varying exposures and adverse events (AEs), such as AEs following vaccination (Farrington, 1995; Farrington et al., 1996). The SCCS method has been used to examine the relationship between rare vaccine reactions,

Correspondence to: Danh V. Nguyen.

Supporting Information

Additional supporting information (proof of the bias function (2), Jacobian matrix for equation (8), and Table S1) may be found in the online version of this article at the publisher's web site.

such as idiopathic thrombocytopenic purpura (ITP) after measles-mumps-rubella (MMR) vaccination (Miller et al., 2001; Black et al., 2003). For example, the incidence of AEs in the pre-specified 6-week post-MMR vaccination *risk* period is compared to the incidence in the *control* period, defined as time periods outside of the 6-week risk period. The length of the risk period (e.g., the 6-week risk period) is typically pre-specified in practice based on prior studies.

The SCCS method has also emerged as a useful method in diverse areas of applications, such as the study of the relative incidence (RI) of cardiovascular events during the 30-day period after infection-related hospitalization in patients on dialysis (Dalrymple et al., 2011; Mohammed et al., 2012). In SCCS studies, the primary quantity of interest is the RI of AEs in the risk period relative to the control period. Although of secondary interest, estimation of the optimal risk period is of interest as well because it provides the time frame for monitoring/surveillance of AEs.

An advantage of the SCCS method is that it controls for all measured and unmeasured baseline confounders and is self-matched. Thus, the SCCS estimate of the RI of AEs is not confounded by differences in individual baseline factors, such as socioeconomic status, underlying genetics, and baseline health status, which are difficult to accurately ascertain in exposed and unexposed individuals (e.g., in vaccinated and unvaccinated populations or in dialysis patients who did and did not acquire infections). Traditional cohort and case-control methods are susceptible to these baseline confounding biases. See reviews in Fine and Chen (1992).

As described in the aforementioned studies, the risk period (e.g., the 6-week period after MMR vaccination or 30-day period after an infection-related hospitalization) must be specified *a priori* in SCCS studies. However, the true risk period, which is referred to as the "optimal" risk period in the literature (Xu et al., 2011; 2013), is not known and must be hypothesized or explored in practice. If the goal is to unbiasedly estimate the RI of events within the optimal risk period relative to the control period, then misspecification of the risk period will lead to biased estimates. More precisely, misspecification of the optimal risk period in SCCS studies will result in either (a) over-specification where a portion of the *true control period* is included in the specified risk period or (b) under-specification where a portion of the *true risk time period* is included as part of the control period. The consequence of this misspecification is biased estimate of the RI of adverse events in the optimal risk period relative to the control period.

Previous works offer two approaches to the estimation of the RI and associated true risk period. The first is a graphical approach based on the observed linearity pattern of the estimated RIs and the maximum RI (Xu et al., 2011). Although it can be informative, the subjectivity of the graphical approach poses challenges in practice and a second approach using a hypothesis testing framework based on a scan likelihood ratio test statistic was proposed (Xu et al., 2013).

An impediment to a more effective approach to estimation of the optimal risk period length and the corresponding RI is that the (asymptotic) bias due to misspecification of the risk

period is not fully understood. Thus, in this work, our first objective is to fill this knowledge gap by fully characterizing this bias. We will show that for practical purposes, the bias can be well-approximated by a simple parametric linear-quadratic spline (or a linear-nonlinear spline function generally). Second, after elucidating this pattern of bias, we then incorporate this information to propose a novel estimation procedure (Section 2). Through simulation studies, we demonstrate the overall superior performance of the new method compared to previous methods (Section 3). The proposed method is illustrated with two SCCS studies to examine the (1) risk of ITP after MMR vaccination in children and (2) risk of cardiovascular (CV) events after infection-related hospitalizations in older patients on dialysis (Section 4).

2. Methods

In this section we describe the SCCS method commonly used in the evaluation of vaccine safety, bias under misspecification of optimal risk period, previous approaches to estimate the optimal risk period length and the associated relative incidence, and our novel estimation method.

2.1. SCCS model and formulation of the risk period misspecification problem

Consider a cohort of N individuals, where each individual has at least one event during the observation or follow-up period. More formally, denote the observation period by $(a_i, b_i]$ for individual i, where the start and and end of the observation period are a_i and b_i, respectively. In SCCS studies, adjustment for the confounding effect of the (time-varying) age distribution of the events during the observation period may be needed. Thus, to account for age effects, the observation period is partitioned into $J+1$ age groups, $j = 0,\ldots J$, along with the control period ($k = 0$) and the risk period ($k = 1$). Given the exposure history (e.g., infection or vaccination) over the observation period for individual i, the number of events in each age-risk (or age-control) interval, denoted n_{ijk}, is modeled as a non-homogeneous Poisson process with rate $\lambda_{ijk} = \exp(\varphi_i + \alpha_j + \beta \cdot k)$, for $k = 0, 1$. That is, $n_{ijk} \sim \text{Poisson}(e_{ijk}\lambda_{ijk})$, where e_{ijk} is the length of time in the j^{th} age group and k^{th} risk period for individual i, where $k = 0$ corresponds to the control period. Here the parameters φ_i, α_j and β are, respectively, the individual-specific, j^{th} age group (relative to age group $j = 0$) and risk group (relative to control period $k = 0$) effects, with $\alpha_0 = 0$. The main parameter of interest is β, the log RI of events in the risk period relative to the control period.

Farrington (1995) formulated the SCCS method and showed that when conditioned on $n_{i\cdot\cdot} = \sum_{jk} n_{ijk} \geq 1$, where $n_{i\cdot\cdot}$ is the total number of events for individual i, the kernel of the SCCS (conditional) likelihood is product multinomial. That is, the contribution to the SCCS likelihood from subject i is $L_i(\boldsymbol{\alpha}, \beta) = \prod_{j,k} \pi_{ijk}^{n_{ijk}}$, with probabilities

$$\pi_{ijk} = \frac{e_{ijk}\lambda_{ijk}}{\sum_{r=0}^{J}\sum_{s=0}^{1} e_{irs}\lambda_{irs}} = \frac{e_{ijk}\exp(\alpha_j + \beta \cdot k)}{\sum_{r=0}^{J}\sum_{s=0}^{1} e_{irs}\exp(\alpha_r + \beta \cdot s)}, \quad (1)$$

where $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_J)^T$. The term "self-controlled" refers to the fact that the individual effects φ_i cancel out from (1), thus, self-controlling for all fixed covariates. Estimation can

be based on maximum likelihood (ML) by using the log-likelihood $\ell(\boldsymbol{\alpha}, \beta) = \sum_{i=1}^{N} \log\{L_i(\boldsymbol{\alpha}, \beta)\}$.

We note that in the more general SCCS model, especially in non-vaccine applications, several risk periods are allowed (e.g., the first, second and third month after an infection) as well as non-contiguous risk periods. Our work here focuses on the specialized SCCS model (1), tailored to the applications considered in Section 4.

In practice the risk period length must be specified in SCCS studies. Not surprisingly, when the specified risk length, denoted $\tilde{\tau}$, is not equal to the optimal risk length, denoted τ, the estimate of the RI is biased. To formulate this problem, denote the specified risk length as $\tilde{\tau} = \tau + u$, where u is the amount of risk length error ($u \neq 0$). Also, let $\tilde{n}_{ijk}$ and $\tilde{e}_{ijk}$ denote the number of events and the amount of time spent in the risk period ($k = 1$) or the control period ($k = 0$) in age group j under the misspecified model. Although the bias will be characterized under this more general SCCS model in Section 2.3, we first consider here a simple, but illustrative, special case to track the bias.

In this special case, assume equal follow-up for all individuals, i.e., $(a_i, b_i] = (a, b]$, and without age effects. Also, consider one exposure per individual. Dropping the subject index i and age index j in $\tilde{e}_0$ and $\tilde{e}_1$ since the follow-up is the same for each individual, the ML estimate of the RI is

$$\hat{R}^* = \frac{\sum_{i=1}^{N} \tilde{n}_{i1} / \sum_{i=1}^{N} \tilde{n}_{i0}}{\tilde{e}_1 / \tilde{e}_0}, \quad (2)$$

which targets $R^* \equiv \exp(\beta^*)$, a quantity that will not be equal to the correct relative incidence $R = \exp(\beta)$, unless $\tilde{\tau} = \tau$ (i.e., when $u = 0$). Note that in this case, the average observed time in the risk period and the control period are $\tilde{e}_1 = \tilde{\tau}$ and $\tilde{e}_0 = T - \tilde{\tau}$, respectively, where $T = b - a$ is the follow-up time. Also, note that these are functions of the specified risk length $\tilde{\tau}$. For over-specification ($\tilde{\tau} > \tau$), applying the law of large numbers and Slutsky's theorem, $\hat{R}^*$ is consistent for $R^* = \gamma_0 + \gamma_1 (\tilde{\tau}^{-1} - \tau^{-1})$, which is a linear function of $1/\tilde{\tau}$, with $\gamma_0 = R$ and $\gamma_1 = \tau(R - 1)$. A similar analysis for the case when the optimal risk period is under-specified, $\tilde{\tau} < \tau$ (i.e., $u < 0$), shows that $R^* = \gamma_0 + \{-\gamma_1 (\tilde{\tau}^{-1} - \tau^{-1})\}/\{\gamma_2(\tilde{\tau}^{-1} - \tau^{-1}) + \gamma_3\}$, which is a nonlinear function of $1/\tilde{\tau}$, with $\gamma_2 = \{T + \tau(R - 1)\}/R$ and $\gamma_3 = (T - \tau)/(R\tau)$. The optimal risk period (when $u = 0$) and the associated unbiased RI estimate are obtained where the linear and nonlinear parts of the function R^* joins at a single knot point (at $\tilde{\tau} = \tau$). Figure 1 illustrates this pattern of bias for both $R > 1$ and $R < 1$. Thus, the bias as a function of the inverse of the specified risk length, $\tilde{\tau}^{-1}$, can be expressed as

$$R^* = \gamma_0 + \gamma_1(\tilde{\tau}^{-1} - \tau^{-1})_- + \frac{-\gamma_1\left(\tilde{\tau}^{-1} - \tau^{-1}\right)_+}{\gamma_2\left(\tilde{\tau}^{-1} - \tau^{-1}\right)_+ + \gamma_3}, \quad (3)$$

where $(x)_+ = x$ if $x > 0$ and 0 otherwise and, similarly, $(x)_- = x$ if $x < 0$ and 0 otherwise. A proof of equation (3) is deferred to the Supplemental Information. The first main result of this paper, presented in Section 2.3, will be a systematic approach to elucidate the bias function due to misspecification of the optimal risk more generally.

2.2. Previous approaches under risk period misspecification

Recognizing that R^* is linear in $1/\tilde{\tau}$ when $\tilde{\tau} > \tau$, Xu et al. (2011) proposed a graphical approach to determine the optimal risk period and RI based on the maximum estimated RI, $\hat{R}_{\max} = \max_m \hat{R}^*_m$, where $\{\hat{R}^*_m\}_{m=1}^M$ are the RI estimates obtained from the standard SCCS model (1) for a given sequence of specified risk lengths, $\tilde{\tau}_m$, $m = 1,\ldots,M$ *and* where pattern of relationship between R^*_m and $1/\tilde{\tau}_m$ is "visually" linear. The estimated optimal risk length is taken to be the $\tilde{\tau}_m$ corresponding to $\hat{R}_{\max}$, which we denote as $\hat{\tau}_{\max}$. There are several drawbacks with this approach. First, it assumes that $\tilde{\tau} > \tau$ (i.e., $u > 0$) and $R > 1$ is known. This is impossible to determine in practice. (Also, when $R < 1$, then the minimum RI should be taken instead, i.e., $\hat{R}_{\min} = \min_m \hat{R}^*_m$; see Figure 1, second row.) Second, as we will demonstrate in Section 3, the high bias of the estimated maximum RI, $\hat{R}_{\max}$, is problematic. Third, the transition between linearity and nonlinearity of $\hat{R}^*_m$ as a function of $1/\tilde{\tau}_m$ is difficult to assess from visual inspection which introduces much subjectivity, particularly for small to moderate effect sizes.

A second method to estimate the optimal risk period and RI was subsequently proposed to circumvent the subjectivity of the $\hat{R}_{\max}$/graphical approach (Xu et al., 2013). It is based on a scan likelihood ratio test (LRT) statistic. The idea is to examine the difference between the log-likelihood for a specified risk length under the alternative hypothesis where RI ≠ 1 ($\beta \neq 0$) and the null log-likelihood under RI = 1 ($\beta = 0$), where the incidence of AEs is constant throughout the follow-up period (i.e., no time period of elevated risk). That is, the LRT statistic for a specified $\tilde{\tau}$ is

$$T(\nu, \tilde{\tau}) = \ell_1(\nu, \tilde{\tau}, \hat{\beta}, \hat{\boldsymbol{\alpha}}) - \ell_0(\nu, \tilde{\tau}, \hat{\boldsymbol{\alpha}}),$$

where $\ell_1(\nu, \tilde{\tau}, \hat{\beta}, \hat{\boldsymbol{\alpha}})$ is the maximized SCCS log-likelihood under the alternative hypothesis ($\beta \neq 0$) and $\ell_0(s, \tilde{\tau}, \hat{\boldsymbol{\alpha}})$ is the corresponding maximized log-likelihood under the null hypothesis that $\beta = 0$. Here ν denotes the fixed time when the risk period starts (e.g., $\nu = 1$ for the first day after vaccination). Similar to the graphical approach, this method estimates the optimal risk period length to be $\hat{\tau}_{\text{lrt}} = \text{argmax}_{\tilde{\tau}_m} T(\nu, \tilde{\tau}_m)$ for a sequence of specified risk lengths $\{\tilde{\tau}_m\}_{m=1}^M$ (and for ν fixed). The optimal RI estimate is obtained from the SCCS model fitted with the estimated optimal risk period $\hat{\tau}_{\text{lrt}}$, which we denote as $\hat{R}_{\text{lrt}}$.

2.3. Bias due to non-optimal risk period specification

In this section, we elucidate the bias functional form due to misspecification of the risk period. We then use this theoretical result to propose a new estimation method.

For the SCCS model (1) with $J+1$ age groups, denote the misspecified risk period length by $\tilde{\tau} = \tau + u$. Let $\tilde{n}_{ijk}$ and $\tilde{e}_{ijk}$ denote the number of events and the amount of time spent in the misspecified risk period ($k = 1$) or misspecified control period ($k = 0$) in age group j, $j = 0, \ldots, J$, respectively. Furthermore, denote the targets of the MLEs under the misspecified SCCS model for the risk period and age effect as β^* and $\boldsymbol{\alpha}^* = (\alpha_1^*, \ldots, \alpha_J^*)^{\mathrm{T}}$, respectively. Under this misspecified model, the MLEs of (β^*, $\boldsymbol{\alpha}^*$), denoted ($\hat{\beta}^*$, $\hat{\boldsymbol{\alpha}}^*$), are obtained by solving the set of ($J+1$) estimating equations:

$$N^{-1} \sum_{i=1}^{N} \sum_{j=0}^{J} (\tilde{n}_{ij1} - n_{i..} \hat{\pi}_{ij1}^*) = 0$$

$$N^{-1} \sum_{i=1}^{N} \sum_{k=0}^{1} (\tilde{n}_{ijk} - n_{i..} \hat{\pi}_{ijk}^*) = 0, \qquad j = 1, \ldots, J, \quad (4)$$

where $\hat{\pi}_{ijk}^* = \tilde{e}_{ijk} \exp(\hat{\alpha}_j^* + \hat{\beta}^* \cdot k) / \sum_{r=0}^{J} \sum_{s=0}^{1} \{\tilde{e}_{irs} \exp(\hat{\alpha}_r^* + \hat{\beta}^* \cdot s)\}$. The MLEs, ($\hat{\beta}^*$, $\hat{\boldsymbol{\alpha}}^*$), are consistent for ($\beta^*$,$\boldsymbol{\alpha}^*$), which satisfy the estimating equations (4) in expectation. Thus, for a given design and specified risk length $\tilde{\tau}$, the solution (β^*, $\boldsymbol{\alpha}^*$) can be obtained by solving the following ($J+1$) equations:

$$h_0(\beta^*, \boldsymbol{\alpha}^*) \equiv N^{-1} \sum_{i=1}^{N} \sum_{j=0}^{J} \{E(\tilde{n}_{ij1}) - n_{i..} \pi_{ij1}^*\} = 0$$

$$h_j(\beta^*, \boldsymbol{\alpha}^*) \equiv N^{-1} \sum_{i=1}^{N} \sum_{k=0}^{1} \{E(\tilde{n}_{ijk}) - n_{i..} \pi_{ijk}^*\} = 0, \qquad j = 1, \ldots, J, \quad (5)$$

where $\pi_{ijk}^* = \tilde{e}_{ijk} \exp(\alpha_j^* + \beta^* \cdot k) / \Delta_i^*$ and $\Delta_i^* = \sum_{r=0}^{J} \sum_{s=0}^{1} \tilde{e}_{irs} \exp(\alpha_r^* + \beta^* \cdot s)$.

Because the distribution of the events in the misspecified periods for subject i, namely $\tilde{n}_{ijk}$, conditioned on the total number of events, exposure history and misspecification error ($n_{i.}$, e_{ijk}, u) is multinomial, the probabilities of seeing an event in the specified risk and control time periods in the j^{th} age group are

$$\tilde{\pi}_{ij0} \equiv \tilde{\pi}_{ij0}(u) = \frac{\{e_{ij0} - (e_{ij1} - \tilde{e}_{ij1})\} \exp(\alpha_j)}{\Delta_i},$$

$$\tilde{\pi}_{ij1} \equiv \tilde{\pi}_{ij1}(u) = \frac{e_{ij1}\exp(\alpha_j + \beta) + (e_{ij1} - \tilde{e}_{ij1})\exp(\alpha_j)}{\Delta_i}, \quad (6)$$

when $u < 0$ and $\Delta_i = \sum_{j=0}^{J}\sum_{k=0}^{1} e_{ijk}\exp(\alpha_j + \beta k)$. Hence, $E(\tilde{n}_{ijk}) = n_{i..}\tilde{\pi}_{ijk}$. Similarly, for $u > 0$ these probabilities are

$$\tilde{\pi}_{ij0} \equiv \tilde{\pi}_{ij0}(u) = \frac{e_{ij0}\exp(\alpha_j) - (e_{ij1} - \tilde{e}_{ij1})\exp(\alpha_j + \beta)}{\Delta_i},$$

$$\tilde{\pi}_{ij1} \equiv \tilde{\pi}_{ij1}(u) = \frac{\{e_{ij1} + (e_{ij1} - \tilde{e}_{ij1})\}\exp(\alpha_j + \beta)}{\Delta_i}. \quad (7)$$

The bias as a consequence of optimal risk period misspecification error is the difference between β^* (the target of $\hat{\beta}^*$) and β, the true parameter corresponding to the unknown optimal risk period τ. Similarly, for the age effects, the difference between $\boldsymbol{\alpha}^*$ and $\boldsymbol{\alpha}$ can be evaluated.

To determine the bias, the set of equations (5) can be solved numerically for $(\beta^*, \boldsymbol{\alpha}^*)$ using a Newton-Raphson (NR) method. Thus, $(\beta^*, \boldsymbol{\alpha}^*)$ can be determined for any set of parameters $(\beta, \boldsymbol{\alpha})$ and $\{e_{ijk}, \tilde{e}_{ijk}, \tilde{n}_{ijk}\}$. More precisely, let $(\beta^*, \boldsymbol{\alpha}^*)^{(t)}$ be the NR update at iteration t and $\mathbf{h}^{(t)} = (h_0(\beta^*, \boldsymbol{\alpha}^*)^{(t)}, \ldots, h_J(\beta^*, \boldsymbol{\alpha}^*)^{(t)})$. Then the next NR update is

$$(\beta^*, \boldsymbol{\alpha}^*)^{(t+1)} = (\beta^*, \boldsymbol{\alpha}^*)^{(t)} - (\mathbf{J}^{(t)})^{-1}\mathbf{h}^{(t)}, \quad (8)$$

where $\mathbf{J}^{(t)}$ is the $(J+1) \times (J+1)$ Jacobian matrix of partial derivatives evaluated at $(\beta^*, \boldsymbol{\alpha}^*)^{(t)}$. The entries of $\mathbf{J}^{(t)}$ are provided in the Supplemental Information. See also Mohammed et al. (2013) for a similar application.

2.4. Proposed estimation of the optimal risk period and relative incidence

The motivation for our proposed estimation procedure is based on the general patterns of bias due to the optimal risk misspecification in the SCCS model. These patterns of bias are illustrated in Figure 1 for three cases: (i) a special case where all individuals have equal follow-up times, 1 exposure and 1 event, where a closed-form solution to (8) is available (see Section 2.1, equation 3); (ii) a general case with unequal follow-up times among individuals, one or more events per person, and one exposure, analogous to the MMR-ITP data of Section 4; and (iii) a general case as in case (ii), but with multiple exposures. As Figure 1 illustrates, for the SCCS model without age effects (left column) and with age effects (right column), the bias is linear in $1/\tilde{\tau}_m$ when the misspecified length is greater than the optimal risk length ($\tilde{\tau}_m > \tau$) and it is nonlinear in $1/\tilde{\tau}_m$ when the misspecified risk length is less than the optimal risk length ($\tilde{\tau}_m < \tau$). Furthermore, the single knot point where the

linear bias region transitions to the nonlinear region of the bias function corresponds to the optimal risk period length (at $\tilde{\tau}_m = \tau$, or equivalently at $u = 0$). This analysis of the form of the bias function suggests a direct approach to estimate the optimal risk period and RI. More specifically, the proposed estimation approach is to fit a sequence of linear-quadratic spline models, one for each knot point. For a given knot point, a linear-quadratic spline model is fitted through the scatterplot of the estimated RIs and the (inverse of) a sequence of risk period lengths. Selection of the optimal knot is carried out via minimization of a square-error loss criterion. Our choice to fit a linear-quadratic spline model is illustrated in Figure 2, which shows how closely the chosen model (dotted line) fits the theoretical bias function (gray solid line).

More precisely, let $\{\tilde{\tau}_m\}_{m=1}^{M}$ be a sequence of M specified risk lengths covering τ for the SCCS model (1) and $\{\hat{\beta}_m^*\}_{m=1}^{M}$ be the corresponding estimated log RIs from fitting a standard SCCS model. Next, fit the linear-quadratic spline model,

$$g(t, \tau_m, \boldsymbol{\delta}) = \delta_0 + \delta_1 \tilde{\tau}_m^{-1} + \delta_2 (\tilde{\tau}_m^{-1} - t^{-1})_+ + \delta_3 (\tilde{\tau}_m^{-1} - t^{-1})_+^2, \quad (9)$$

through the scatterplot of $\{\tilde{\tau}_m^{-1}, \exp(\hat{\beta}_m^*)\}_{m=1}^{M}$. This will capture the bias relationship with t as a given single knot with coefficients $\boldsymbol{\delta} \equiv (\delta_0, \delta_1, \delta_2, \delta_3)$. Let $\hat{\boldsymbol{\delta}}$ denote the estimated coefficients from the model fit. We fit model (9) for a sequence of equally spaced knots $\tilde{\tau}_M < t < \tilde{\tau}_1$ and choose the knot with minimum sum of square error as our estimate of the optimal risk period:

$$\hat{\tau} = \arg\min_t \sum_{m=1}^{M} \left\{ \exp(\hat{\beta}_m^*) - g(t, \tau_m, \hat{\boldsymbol{\delta}}) \right\}^2. \quad (10)$$

Corresponding to this proposed optimal risk estimate, $\hat{\tau}$, the proposed estimate of the relative incidence is defined as the ML estimate from fitting the standard SCCS using $\hat{\tau}$ as the risk length. We denote this proposed estimator as $\hat{R}$. We examine the properties and performance of the proposed estimators ($\hat{\tau}$, $\hat{R}$) and compare them to the two previous estimators, namely ($\hat{\tau}_{\max}$, $\hat{R}_{\max}$) and ($\hat{\tau}_{\text{lrt}}$, $\hat{R}_{\text{lrt}}$).

Finally, we note that there are several reasons for the proposed choice of the linear-quadratic spline model (9). The linear-quadratic spline form provides a simple, but accurate, parametric approximation to the true bias function and can be fitted with a standard linear regression model (Figure 2). Although the 4-parameter nonlinear spline form given by (3) provides a slightly better approximation to the theoretical bias for very large effect size and large window length (compare dashed and solid gray lines in Figure 2 for the case of $R = 4$ and $\tau = 45$), it would require an iterative nonlinear fitting routine with specification of initial values. Also, unlike standard fitting of nonlinear models through a scatterplot, stable fitting of the nonlinear spline is difficult in the current problem because there is only a single data

point at each specified risk period length, $\tilde{\tau}$. These difficulties are avoided with the proposed linear-quadratic spline.

3. Simulation studies and results

3.1. Study objectives and design

In this section, we implement studies to address several specific objectives. First, using the theoretical calculation described in Section 2.3 we characterize the bias due to misspecification of the true risk for the SCCS model (1). These bias patterns, determined by the set of Newton-Raphson equations (5), are then verified with extensive simulation studies. Second, we implement simulation studies to evaluate the performance (bias, variance and mean square error).

For our studies, we consider the following general scenarios to assess our proposed methods: (a) a general case with unequal follow-up times among individuals, one or more events per person, and one exposure, analogous to the MMR-ITP data of Section 4 where a single MMR immunization is administered typically around 1 year of age; and (b) a general case, as in case (a), but with multiple exposures. Under both cases (a and b), we simulate data with 2 and 3 age groups ($J = 1, 2$) as well as a single age group (no age effects). Follow-up times are different for each individual as in real data applications and with an average of 365 days of follow-up. The true risk lengths, τ, are 15, 30 and 45 days after an exposure. Exposures are randomly assigned throughout an individual's observation period. For the estimation procedures described in Section 2.4 the sequence of specified risk lengths $\{\tilde{\tau}_m\}_{m=1}^{M}$ are {8,…,**15**,…,24}, {15,…,**30**,…,45} and {20,…,**45**,…,70} corresponding to $\tau =$ 15, 30 and 45 days, respectively. The choice of each of these sequences reflects several considerations to guide implementation in practice: 1) It should be as broad (wide) as possible, which is applicable to all methods (not just the proposed method). 2) The first guideline must be balanced with the fact that for a very small risk length, events will be extremely sparse.

For case (b) with multiple exposures, individuals can have 1, 2 or 3 exposures with probabilities 0.7, 0.2 and 0.1, respectively. The marginal totals of the number of AEs for individuals $i = 1,\ldots,N$ are generated according to the SCCS model. That is, AEs are generated from a non-homogeneous Poisson model given by $n_{i\cdot} \sim \text{Poisson}(\Sigma_{jk}\, e_{ijk}\lambda_{ijk})$, where $\lambda_{ijk} = \exp(\varphi_i + \alpha_j + \beta \cdot k)$ with $\varphi_i = \log(1/10000)$ fixed. These marginal totals are randomly distributed throughout each individual's observation period based on the multinomial probabilities shown in (1). Supplemental Table S1 summarizes the 32 different experiments: (i) single and multiple exposures models, (ii) 1, 2 and 3 age groups, (iii) 3 risk lengths, and (iv) 2 distribution of exposure settings. Each experiment was replicated at 6 different RIs, $\exp(\beta) \in \{0.7, 0.9, 1.2, 1.5, 2, 4\}$, and at 4 sample sizes, $N = 100, 200, 400,$ and 800 individuals.

3.2. Patterns of bias: Theoretical calculations and simulation

As summarized in Section 2.3, the general patterns of bias due to non-optimal risk period specification are characterized by the equations (5–7). We first check the validity of

equations (5–7) via simulation. Figure 3 presents Monte Carlo simulation results, averaged over 200 simulated data sets. It shows that the theoretical quantity R^* (solid gray line) determined from equation (5–7) tracks closely the simulation results (dashed line). Also, given are 95% confidence intervals (CIs; dotted lines) along with the benchmark case where the *optimal* risk length is specified in the model fitting (i.e., $\tilde{\tau} = \tau$); this is indicated with the vertical blue line at $1/\tau = 1/45$. As expected, the CI width decreases as N increases and, also not surprisingly, substantial under-specification ($1/\tilde{\tau} \uparrow$) of optimal risk period length is associated with increased variance. We present in Figure 3 the cases where $R = 2$ and $R = 0.7$ for the general SCCS model with unequal follow-up times among individuals, one or more events per person, multiple exposures, with age groups and true risk length of 45 days. The results are similar for the other cases (results not shown). We also note that the age effects α^*, which is of secondary interest, have small to negligible bias since they are not substantially affected by the misspecified risk period. These results are not shown here but are available upon request.

As described earlier, when $\beta > 0$ (RI > 1), the target of the RI due to misspecification, $R^* = \exp(\beta^*)$, is linearly and nonlinearly attenuated towards the null for over- and under-specification of the optimal risk period (τ) respectively (e.g., see Figure 1, top row). The attenuation bias systematically increases as $\tilde{\tau}$ moves away from τ. When $\beta < 0$, the bias in R^* is linearly and nonlinearly increasing towards the null for over- and under-specification, respectively.

3.3. Efficacy of the proposed estimator and comparison to current methods

We report here extensive simulation studies to assess the efficacy of the proposed estimation of the relative incidence, R, and the corresponding optimal risk period, τ. We also compare the performance of the proposed method ($\hat{\tau}$, $\hat{R}$) to the two existing methods based on the maximum RI ($\hat{\tau}_{\max}$, $\hat{R}_{\max}$) and the scan likelihood ratio test statistics ($\hat{\tau}_{\text{lrt}}$, $\hat{R}_{\text{lrt}}$). Relative efficacy of the methods were assessed under 32 different simulation studies summarized in supplemental Table S1 (exposure models × age groups × risk lengths × distribution of exposures) and each experiment was replicated at 6 different effect sizes ($R = 0.7$, 0.9, 1.2, 1.5, 2, and 4) and 4 sample sizes ($N = 100$, 200, 400, and 800 individuals). For each of the 320 (32 × 10) experimental combinations, 200 datasets were simulated.

Table 1 presents the results for estimation of the relative incidence, $R = 1.5$, based on the (a) SCCS benchmark where the true risk period τ is used; (b) proposed linear-quadratic spline fit; (c) $\hat{R}_{\max}$ approach; and (d) scan LRT statistic approach. Given are mean RI estimates over 200 simulated datasets, along with bias, variance and mean square error (MSE) for true risk period lengths of $\tau = 15$, 30, and 45 and for sample sizes of $N = 100$ to 800. Several salient finite sample performance characteristics emerge. First, the bias of the proposed linear-quadratic spline approach is closest to the benchmark (for all τ), while the (absolute) biases for the LRT and $\hat{R}_{\max}$ methods remain substantial, even for the largest sample size of 800. For example, for $\tau = 15$ and $N = 200$, the mean estimate of the RI of 1.5 are 1.52, 1.55, 1.74 and 1.68 for the benchmark, proposed method, $\hat{R}_{\max}$ and LRT method, respectively. Thus, relative to the benchmark, for this case the bias is about 2%, 14.5%, and 10.5% for the proposed method, $\hat{R}_{\max}$ and LRT method, respectively. Second, although the recently

proposed LRT method improves on bias reduction relative to $\hat{R}_{\max}$, it still performs relatively poorly with respect to bias compared to the proposed linear-quadratic spline method. However, LRT has higher variance; the proposed method and the $\hat{R}_{\max}$ method have similar lower variances compared to the LRT method. Third, the MSE of the proposed method is lowest and closest to the benchmark and the MSEs of the $\hat{R}_{\max}$ and LRT methods are similarly higher. Overall, the proposed linear-quadratic spline method significantly improves both bias and variance and the gains are substantial for small sample sizes (N = 100, 200). These characteristics are easily seen in Figure 4 (rows 1–3). We note that although Table 1 presents the results for the case of multiple exposures uniformly distributed during the follow-up of each individual and two age groups, these patterns of results are similar in other settings (which are available upon request).

We next turn to estimation of the optimal risk period length, τ, that corresponds to the RI estimation presented in Table 1. Table 2 summarizes the results from estimation of the true τ = 15, 30, and 45 days. There are several striking observations that can be made. First, the estimation of τ is highly variable and the variance dominates bias; for example, with $\tau = 30$ and $N = 200$, the mean estimates of $\tau = 30$ are (24.6, 29.30, 29.0) with variance (59.6, 71.23, 35.54) for the $\hat{R}_{\max}$ approach, the LRT method, and the proposed linear-quadratic spline fit, respectively. For this case, the reduction in variance of the proposed method is substantial, 40.4% and 50.1% relative to the $\hat{R}_{\max}$ approach and LRT method, respectively. We note that the LRT and the proposed method have similarly low bias and target the true τ well, even at small to moderate sample sizes. However, the LRT method has higher variance in the estimation of τ and this adversely affects the estimation of the relative incidence, R, as described above. Not surprisingly, the $\hat{R}_{\max}$ approach has both high bias and variance in the estimation of τ and, therefore, negatively impacts the estimation of R. These patterns of results for τ estimation can be seen in Figure 4 (rows 4–6).

4. Applications

4.1. CV event risk following infection-related hospitalization

Infection and cardiovascular (CV) disease are leading causes of hospitalization and death in older (age ≥ 65) patients on dialysis. For this unique population where infection rates are high, infections may affect vascular endothelium, induce a chronic sub-clinical inflammatory state, or may create a procoagulant state, all factors potentially contributing to increased CV risk. Previous SCCS studies found that the RI of CV events (myocardial infarction, unstable angina, stroke, or transient ischemic attack) was elevated during the 30-day period following infection-related hospitalizations (Dalrymple et al., 2011; Mohammed et al., 2012), although knowledge of the precise risk period is lacking. These studies suggest that the “short-term” risk of CV events may extend to 90 days following infection. We use the proposed method to explore the optimal risk period and associated RI of CV events.

The study is based on data from the United States Renal Data System (USRDS) which captures nearly all patients with end-stage renal disease (ESRD) in the U.S. More specifically, the source population included patients 65–100 years of age with ESRD who newly initiated dialysis between January 1, 2000 and December 31, 2002. Study follow-up ended December 31, 2004. We refer the reader to (Dalrymple et al., 2011) for further details

on the study protocol. The analysis cohort consisted of N = 16,779 patients with one or more CV events. The large size of the data allows for examination of the risk period, incremented daily, from day 15 to day 90; i.e., $\tilde{\tau}_m \in \{15, 16, \ldots, 90\}$. Thus, the SCCS model was fitted for each specified risk length $\tilde{\tau}_m$, $m = 1, \ldots, 76 = M$ resulting in the corresponding RI estimates $\hat{R}^*_m = \exp(\hat{\beta}^*_m)$, for $m = 1, \ldots, M$. Following previous works, age was divided into groups of age 65–75, 76–85, and > 85 to account for potential age effects.

The results are displayed in Figure 5 (top). Note that the sequence of estimates in Figure 5 exhibit the linear-nonlinear spline pattern theoretically predicted by equation (3) and also illustrated earlier in Figure 1. For this large data where the linear-nonlinear pattern is clear, all three approaches produced similar results. The graphical/$\hat{R}_{\max}$ estimate of the log relative incidence was $\hat{\beta}_{R_{\max}} = 0.2839$; thus, $\hat{R}_{\max} = 1.33$ (95% CI: 1.26–1.41) corresponding to an estimated optimal risk length of $\hat{\tau}_{\max} = 36$ days after infection. For the LRT approach, we searched window lengths from 15 to 90 days and the maximum $T(\nu = 1, \tilde{\tau})$ occurs at $\tilde{\tau} = \hat{\tau}_{\text{lrt}} = 36$ days. Therefore, the corresponding RI estimation result was identical to the graphical/$\hat{R}_{\max}$ result (i.e., $\hat{R}_{\text{lrt}} = \hat{R}_{\max}$). The optimal risk period length estimate using our proposed method was similar: $\hat{\tau} = \arg\min_t^{-1} \sum_{m=1}^{M} \{\exp(\hat{\beta}^*_m) - g(t, \hat{\tau}_m, \hat{\boldsymbol{\theta}})\}^2 = 34$ days and the corresponding log relative incidence estimate was $\hat{\beta} = 0.2899$ so the RI estimate was $\hat{R} = 1.34$ (95% CI: 1.26–1.40). Finally, we note that the results here provided an “automatic” exploration of the risk period and associated RI estimate. Incidentally, the results support previous works that identified the 30-day period after infection-related hospitalization as a period of high risk for CV events in patients on dialysis (Dalrymple et al., 2011; Mohammed et al., 2012).

4.2. ITP after MMR vaccination

To further illustrate the proposed methodology, we consider a second SCCS study to examine the relationship between the occurrence of idiopathic thrombocytopenic purpura (ITP), a blood disorder characterized by abnormal decreased platelets count in the blood, after MMR vaccination in children aged 12–23 months.

As described in the Introduction section, several studies have examined the MMR-ITP relationship using various risk periods ranging widely from about 2 to 6 weeks (e.g., Miller et al., 2001; Farrington et al., 1995; O’Leary et al., 2012). Using the MMR-ITP data from Miller et al. (2001) we illustrate the proposed method and compare it to the previous approaches, namely the graphical/ $\hat{R}_{\max}$ and the scan LRT methods. Briefly, the data consist of ascertained records of hospitalization discharges for primary thrombocytopenia (ICD9 code 287.3) and linked to immunization data. A total of 35 children were admitted to the hospital for ITPs (events) at least once: 29 children with 1 event; 5 children had 2 events; and 1 child had 5 events. The start of the follow-up for 33 children started on day 365 of age, 1 child on day 438 and another on day 453 of age. The end of the follow-up period was day 730 for 31 children, day 723 for 1 child, day 677 for 2 children, and day 674 for 1 child. To account for age effects, the follow-up period was divided into 60-day intervals (similar to Xu et al. (2011) and Whitaker et al. (2006)) as follows: 366–426, 427–487, 488–548, 549–609, 610–670, and 671–730 days.

To directly compare with the two previous approaches, we also use a 7-day increment starting at 21 days after MMR vaccination, so that $\tilde{\tau}_m \in \{21, 28, \ldots, 147\}$ as in Xu et al. (2011). Thus, the SCCS model was fitted for each specified risk length $\tilde{\tau}_m$, $m = 1, \ldots, 19 = M$ resulting in the corresponding RI estimates $\hat{R}^*_m = \exp(\hat{\beta}^*_m)$, for $m = 1, \ldots, M$, displayed in Figure 5 (bottom). The graphical/ $\hat{R}_{\max}$ and LRT yielded the same estimate of the log relative incidence of $\hat{\beta}_{R_{\max}} = 1.2268$; thus, $\hat{R}_{\max} = 3.4$ (95% CI: 1.6–7.3) corresponding to an estimated optimal risk length of $\hat{\tau}_{\max} = 77$ days. Next, for the proposed method, we fitted a sequence of linear-quadratic spline models, $g(t, \tilde{\tau}_m, \hat{\boldsymbol{\theta}})$, one for each knot $t = 21, 22, \ldots, 147$. The optimal risk period length estimate using our proposed method was $\hat{\tau} = 84$ days and the corresponding log relative incidence estimate was $\hat{\beta} = 1.1497$; $\hat{R} = 3.2$ (95% CI: 1.5–6.8).

Note from Figure 5 that the estimated RI of ITP at the optimal estimated risk period length of 84 (RI = 3.2) remains fairly stable for risk lengths down to about 30 days (i.e., $1/\tilde{\tau}_m = 0.03$ with RI = 3.5). This has important potential clinical implication because the RI of ITP remains similarly high in the estimated optimal risk period of 84 days compared to previous studies with *a priori* selected shorter risk period of 42 days. Thus, the estimated optimal risk period suggests that surveillance of ITP may need to extend to more than 6 weeks after MMR vaccination.

5. Discussion

Our work here focuses on the problem of bias and estimation for the SCCS method when the optimal risk period is unknown, as is typically the case in practice. It provides an "automatic" tool, useful for exploring risk periods when such knowledge is lacking. We provide the first comprehensive characterization of the bias in the estimation of the relative incidence of AEs for the SCCS method commonly used in vaccine safety studies and also illustrated with an example on the infection-CV risk in the dialysis population. Our linear-quadratic spline estimation method incorporates information on the specific patterns of bias when the optimal risk period length is misspecified. The new method yielded substantial improvement in performance in simulation studies over previous methods. The finite sample performance for small and moderate sample sizes is clearly superior with respect to both bias and variance reduction. Implementation of our proposed method is straight-forward because it only requires fitting a linear regression model after fitting a sequence of standard SCCS models, which can be done with readily available software (e.g., in R, SAS, Stata).

Finally, we note that we have chosen to focus on the standard SCCS model (1) where the risk is constant (step function) during the exposure period because that is the model that is widely used in practice. One can extend this model to allow for non-constant risk function as well as non-contiguous risk periods. However, the bias due to incorrectly specified risk period considered in this work would still apply; one can simply visualize this as a shift in the risk period (whether it be shifting of a constant risk function or a non-constant risk function).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the reviewer and editor for their reviews and comments on this work. This work was supported by grant R01DK092232. The interpretation/reporting of the data 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.

References

Black C, Kaye J, Jick H. MMR vaccine and idiopathic thrombocytopenic purpura. Journal of Clinical Pharmacology. 2003; 55:107–111.

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]

Fine PE, Chen RT. Confounding in studies of adverse reactions to vaccines. American Journal Epidemiology. 1992; 136:121–135.

Miller E, Waight P, Farrington P, Andrews N, Stowe J, Taylor B. Idiopathic thrombocytopenic purpura & MMR vaccine. Archives of Diseases in Childhood. 2001; 84:227–229. [PubMed: 11207170]

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]

Mohammed SM, Dalrymple DS, Senturk D, Nguyen DV. Naive hypothesis testing for case series models with time-varying exposure onset measurement error: Inference for infection-cardiovascular risk in patients on dialysis. Biometrics. 2013; 69:520–529. [PubMed: 23731166]

O'Leary ST, Glanz JM, McClure DL, Akhtar A, Daley MF, Nakasato C, Baxter R, Davis RL, Izurieta HS, Lieu TA, Ball R. The risk of immune thrombocytopenic purpura after vaccination in children and adolescents. Pediatrics. 2012; 129:248–255. [PubMed: 22232308]

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]

Xu S, Hambidge SJ, McClure DL, Daley MF, Glanz JM. A scan statistic for identifying optimal risk windows in vaccine safety studies using self-controlled case series design. Statistics in Medicine. 2013; 22:3290–3299.

Xu S, Zhang L, Nelson JC, Zeng C, Mullooly J, McClure DL, Glanz JM. Identifying optimal risk windows for self-controlled case series studies of vaccine safety. Statistics in Medicine. 2011; 30:742–752. [PubMed: 21394750]

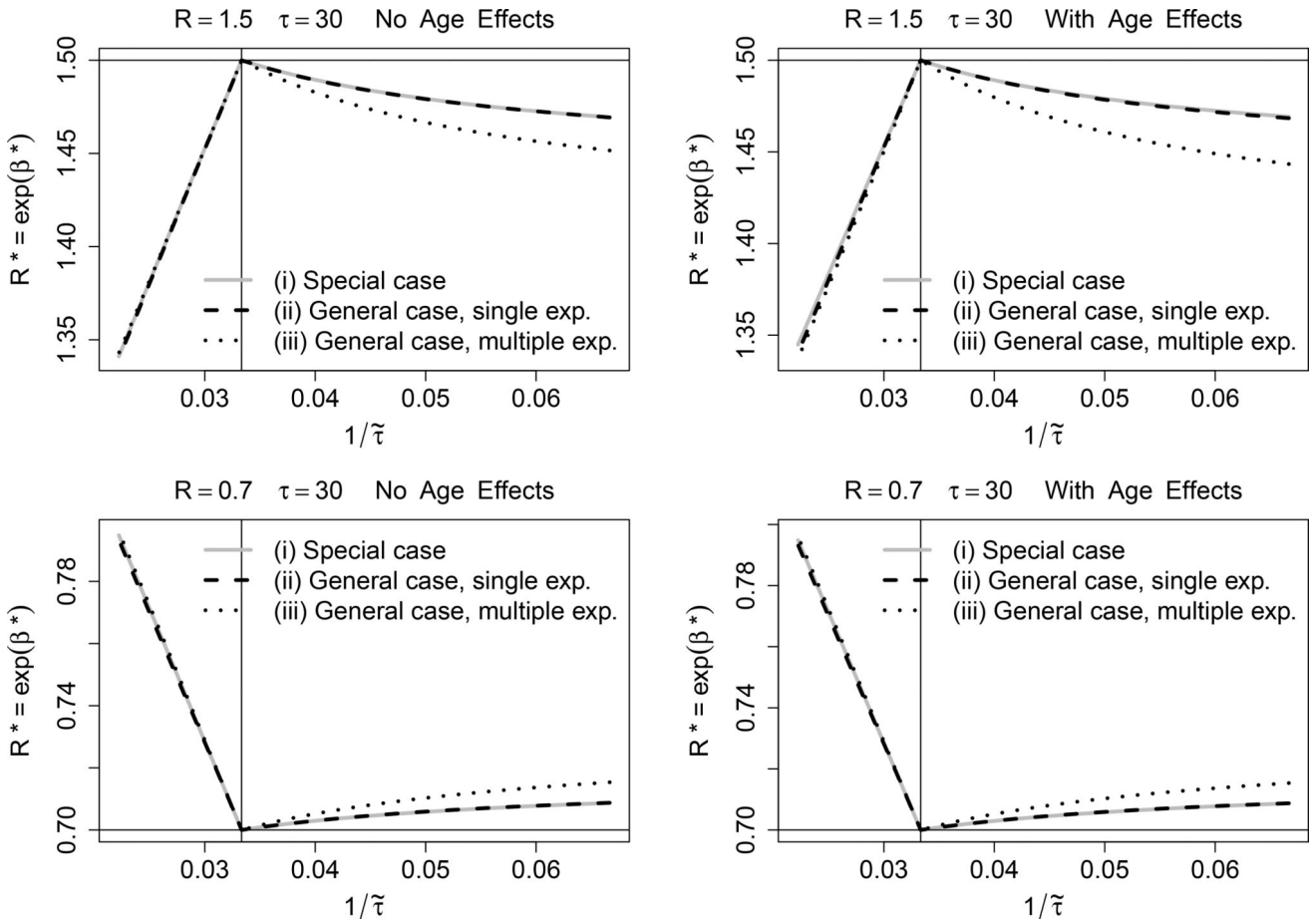

Figure 1.

Relative incidence pattern of bias as a function of the specified risk period length $\tilde{\tau}$ for three models/cases: (i) a special case where all individuals have equal follow-up time, 1 exposure and 1 event; (ii) a general case with unequal follow-up times among individuals, 1 or more events per person, and 1 exposure, analogous to the MMR-ITP data; (iii) a general case as in case (ii), but with multiple exposures. Given are both models with and without age effects as well as $R > 1$ (first row) and $R < 1$ (second row).

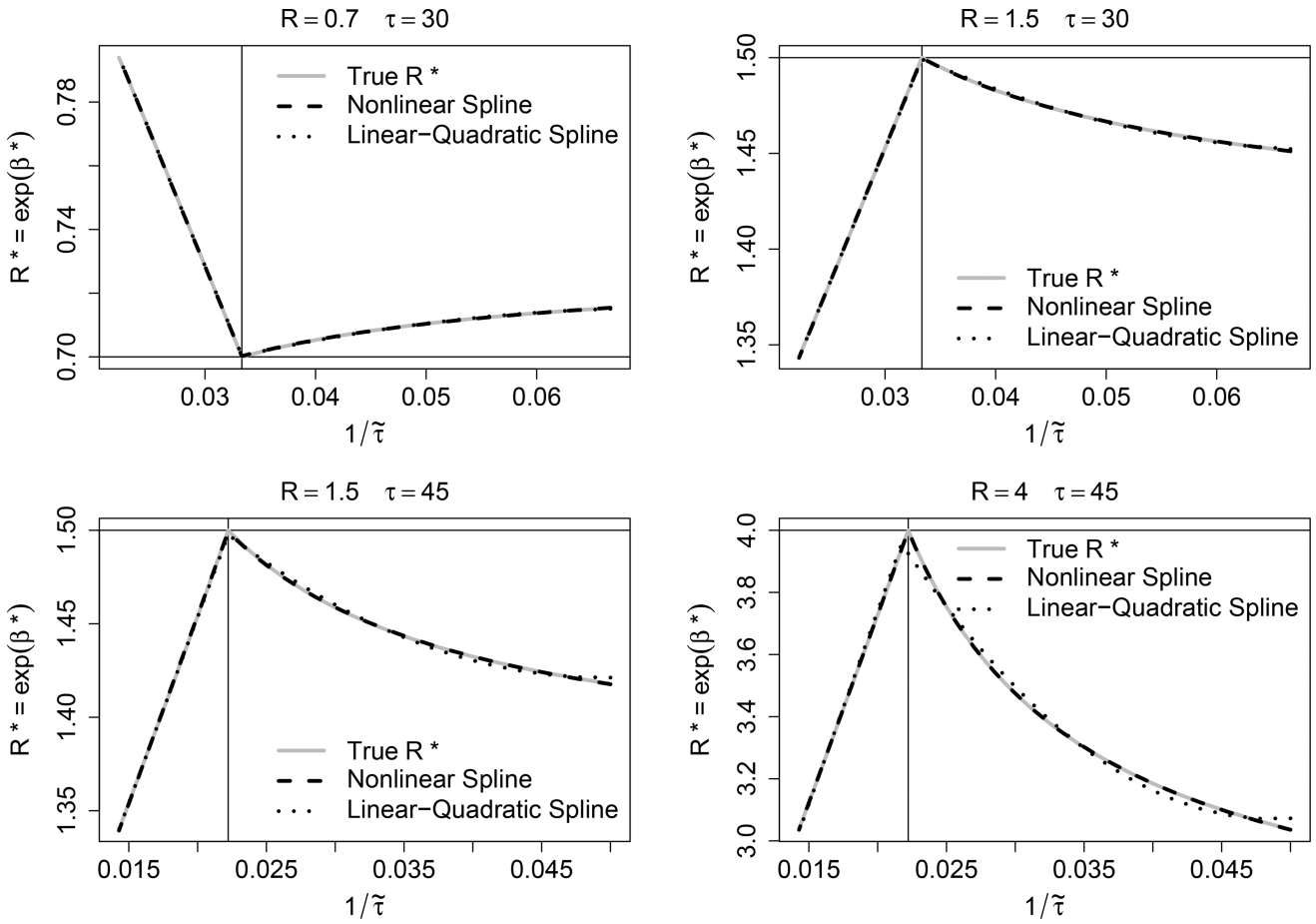

Figure 2.

Linear-quadratic spline (dotted line) as well as nonlinear spline fits to the theoretical (true) bias, $R^* = \exp(\beta^*)$, as a function of the inverse of the specified risk period length, $1/\tilde{\tau}$. The specific examples given are for combinations of the true relative incidence of 0.7, 1.5, 4 and risk length of $\tau = 30$ and 45. The simple linear-quadratic spline tracks the true bias function R^* well, although for very large effect size ($R = 4$, $\tau = 45$) the nonlinear spline function improves slightly.

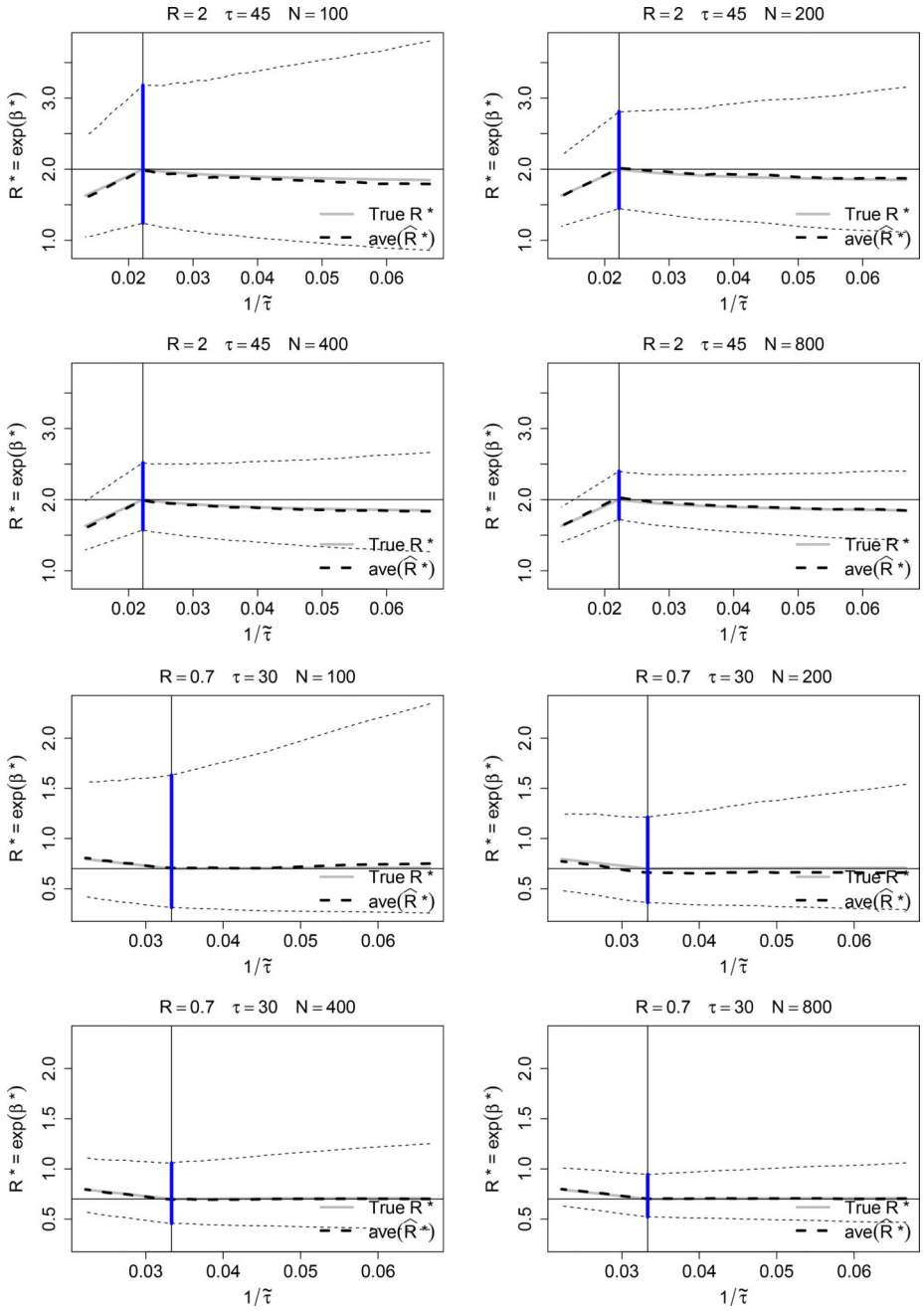

Figure 3.

Theoretical characterization of bias (solid gray) in the relative incidence estimate for varying specified risk period length, $1/\tilde{\tau}$. Dashed black curve denotes the naive SCCS estimate for a given risk period length along with 95% confidence interval; given are averages (ave($\hat{R}$*)) over 200 simulated datasets. Given are for $R = 2$ (top 4 plots) and 0.7 (bottom 4 plots) for sample size $N = 100$ to 800. The blue vertical bar indicates the SCCS model using the true τ (benchmark).

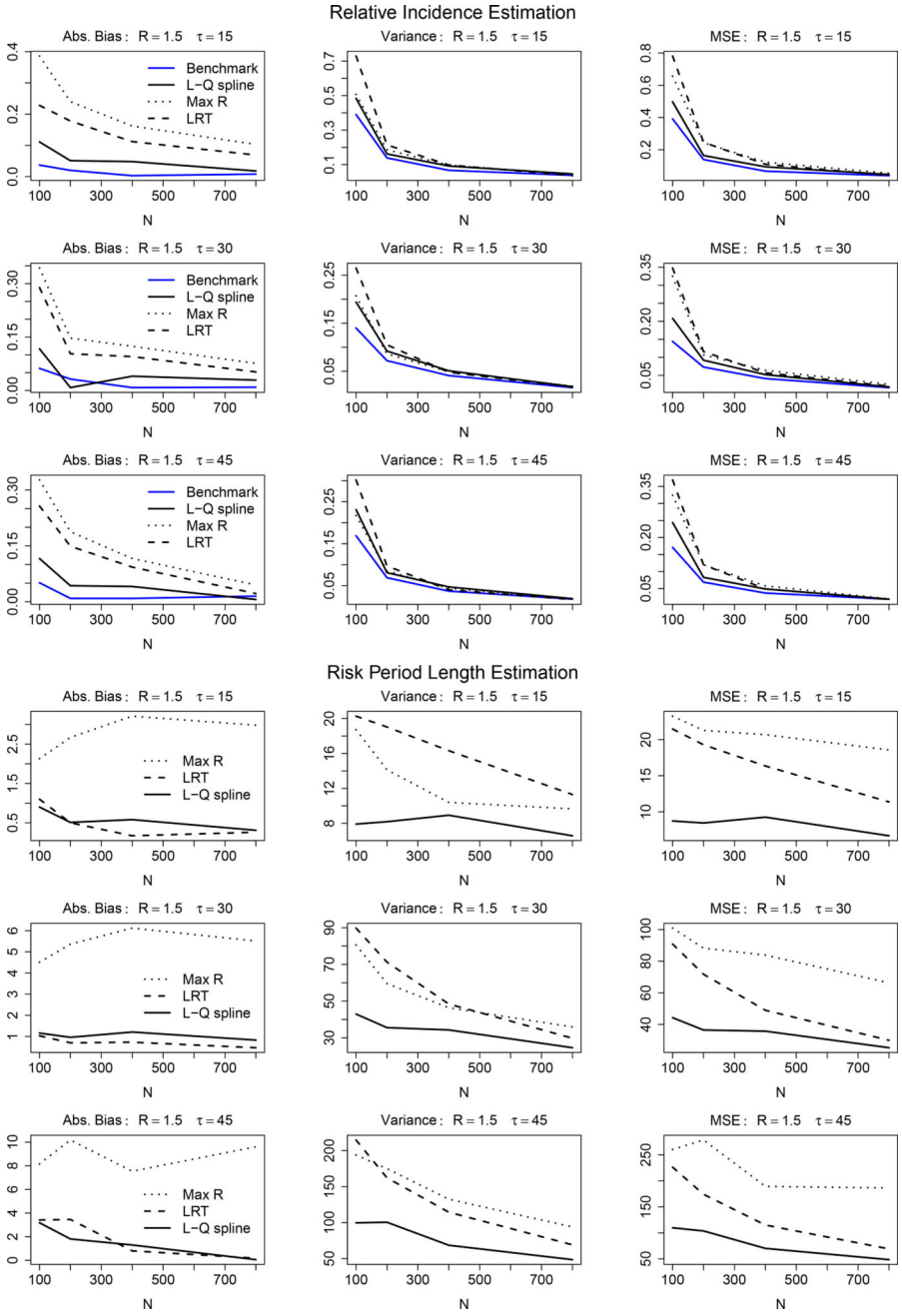

Figure 4.

Performance: Bias (absolute), variance and mean square error for estimation of the true relative incidence R (R = 1.5, rows 1–3) and risk period length τ (τ = 15, 30, 45, rows 4–6) corresponding to Tables 1 – 2.

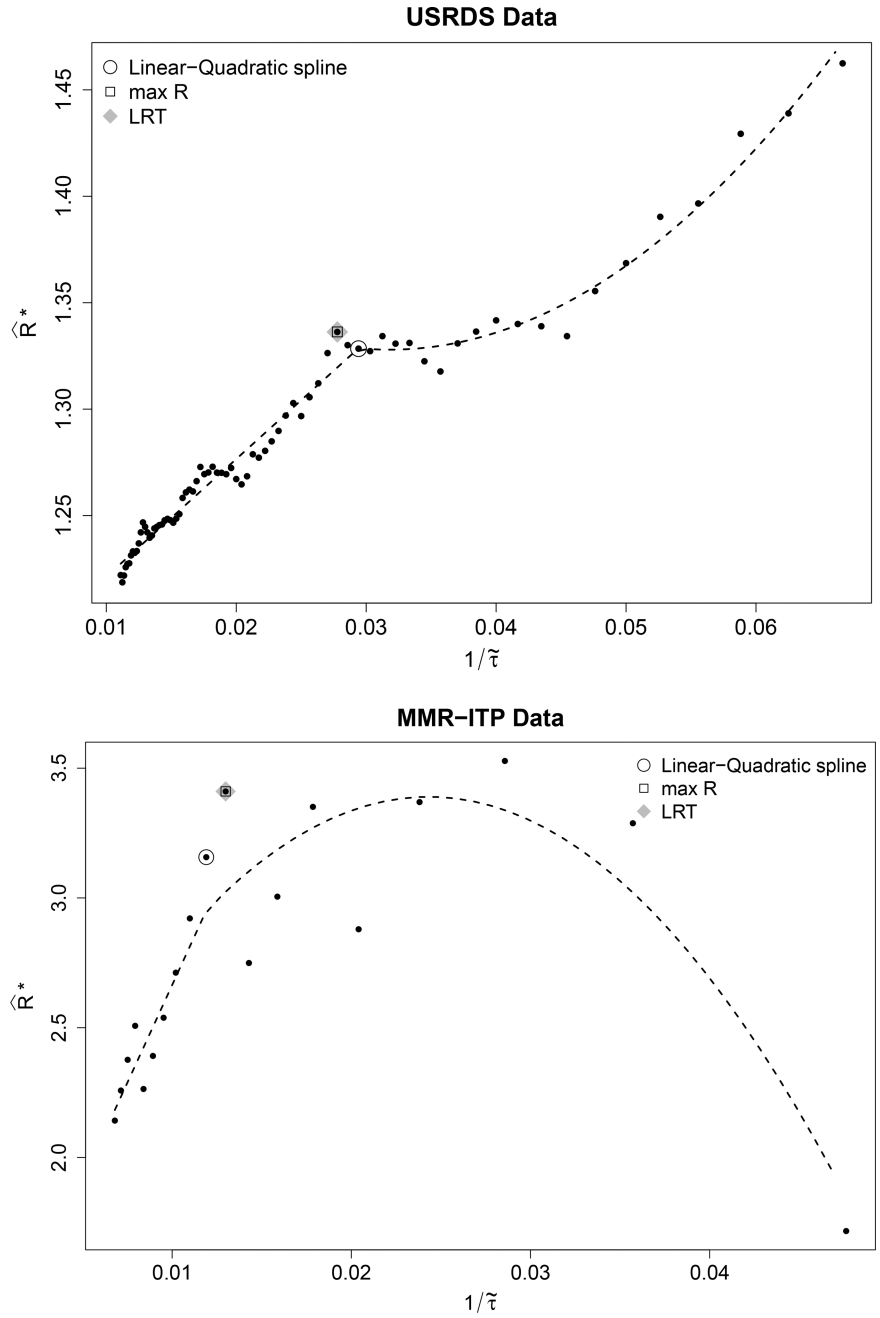

Figure 5.

Estimation of the relative incidence of (a: top) cardiovascular events following infection-related hospitalizations in patients on dialysis from USRDS data and (b: bottom) idiopathic thrombocytopenic purpura (ITP) after measles-mumps-rubella (MMR) vaccination using the proposed linear-quadratic spline method and the currently available methods ($\hat{R}_{\max}$ approach and scan likelihood ratio test (LRT) statistic approach). Given is the scatterplot (solid circles) of $\hat{R}^*$ versus $1/\tilde{\tau}$. The $\hat{R}_{\max}$ approach (square) and the LRT method (gray diamond) resulted in the same estimates.

Table 1

Relative incidence estimation of $R = \exp(\beta) = 1.5$ based on (a) SCCS benchmark where the true risk period τ is used; (b) proposed linear-quadratic spline method; (c) $\hat{R}_{\max}$ approach; and (d) scan likelihood ratio test (LRT) statistic approach. Given are mean estimate (Est.), absolute bias (Bias), variance (Var.), and mean square error (MSE) over 200 simulated datasets.

		(a) Benchmark				(b) Linear-quadratic spline				(c) $\hat{R}_{\max}$ approach				(d) Scan LRT			
τ	N	Est.	Bias	Var.	MSE	Est.	Bias	Var.	MSE	Est.	Bias	Var.	MSE	Est.	Bias	Var.	MSE
15	100	1.537	0.037	0.390	0.391	1.611	0.111	0.484	0.497	1.886	0.386	0.507	0.656	1.728	0.228	0.729	0.781
15	200	1.520	0.020	0.139	0.139	1.551	0.051	0.162	0.164	1.740	0.240	0.186	0.243	1.678	0.178	0.214	0.246
15	400	1.503	0.003	0.067	0.067	1.548	0.048	0.092	0.094	1.662	0.162	0.097	0.123	1.612	0.112	0.098	0.110
15	800	1.508	0.008	0.038	0.039	1.518	0.018	0.046	0.046	1.603	0.103	0.043	0.054	1.569	0.069	0.036	0.041
30	100	1.562	0.062	0.140	0.144	1.617	0.117	0.194	0.208	1.844	0.344	0.207	0.325	1.789	0.289	0.265	0.348
30	200	1.468	0.032	0.072	0.073	1.508	0.008	0.092	0.092	1.647	0.147	0.086	0.107	1.603	0.103	0.105	0.116
30	400	1.508	0.008	0.041	0.041	1.540	0.040	0.051	0.052	1.624	0.124	0.049	0.064	1.595	0.095	0.048	0.057
30	800	1.509	0.009	0.016	0.016	1.529	0.029	0.018	0.018	1.576	0.076	0.019	0.025	1.552	0.052	0.017	0.020
45	100	1.551	0.051	0.169	0.171	1.616	0.116	0.231	0.244	1.827	0.327	0.217	0.324	1.757	0.257	0.302	0.369
45	200	1.509	0.009	0.069	0.069	1.543	0.043	0.081	0.083	1.688	0.188	0.085	0.120	1.649	0.149	0.097	0.120
45	400	1.509	0.009	0.037	0.037	1.541	0.041	0.047	0.049	1.616	0.116	0.044	0.058	1.593	0.093	0.040	0.049
45	800	1.485	0.015	0.018	0.019	1.494	0.006	0.019	0.019	1.545	0.045	0.018	0.020	1.522	0.022	0.017	0.018

Table 2

Estimation of the true period risk length, is τ = 15, 30 and 45, based on (b) proposed linear-quadratic spline fit; (c) $\hat{R}_{\max}$ approach; and (d) scan likelihood ratio test (LRT) statistic approach. Given are mean estimate (Est.), absolute bias (Bias), variance (Var.), and mean square error (MSE) over 200 simulated datasets.

		(b) Linear-quadratic spline				(c) $\hat{R}_{\max}$ approach				(d) Scan LRT			
τ	***N***	**Est.**	**Bias**	**Var.**	**MSE**	**Est.**	**Bias**	**Var.**	**MSE**	**Est.**	**Bias**	**Var.**	**MSE**
15	100	14.10	0.90	7.92	8.73	12.87	2.13	18.71	23.25	13.90	1.10	20.27	21.47
15	200	14.49	0.51	8.19	8.45	12.33	2.67	14.13	21.28	14.50	0.50	19.03	19.29
15	400	14.42	0.58	8.93	9.26	11.79	3.21	10.40	20.68	14.83	0.17	16.32	16.35
15	800	14.69	0.31	6.58	6.68	12.02	2.98	9.66	18.57	14.74	0.26	11.30	11.37
30	100	28.84	1.16	42.94	44.27	25.49	4.51	80.56	100.92	28.97	1.03	89.77	90.83
30	200	29.04	0.96	35.54	36.47	24.64	5.36	59.57	88.26	29.30	0.70	71.23	71.73
30	400	28.79	1.21	34.29	35.75	23.88	6.12	46.37	83.83	29.27	0.73	48.43	48.96
30	800	29.17	0.83	24.61	25.30	24.49	5.51	35.92	66.31	29.53	0.47	29.77	29.98
45	100	41.82	3.18	99.71	109.80	36.84	8.16	193.79	260.39	41.59	3.41	214.61	226.22
45	200	43.19	1.81	100.46	103.75	34.82	10.18	174.91	278.46	41.54	3.46	162.15	174.14
45	400	43.70	1.30	68.55	70.23	37.46	7.54	132.50	189.39	44.20	0.80	114.45	115.10
45	800	45.06	0.06	48.58	48.59	35.39	9.61	93.95	186.26	45.18	0.18	69.18	69.22