American Journal of Applied Mathematics
2022; 10(2): 59-85
http://www.sciencepublishinggroup.com/j/ajam
doi: 10.11648/j.ajam.20221002.15
ISSN: 2330-0043 (Print); ISSN: 2330-006X (Online)

Methodology Article

Interval Estimation of the Absolute Risk of an Event with Competing Risks Using Proportional Regression of Cause-Specific Hazards

Michael Richard Crager*, Jerome Victor Braun

Department of Biostatistics, Exact Sciences Corporation, Redwood City, USA

Email address:

mcrager@exactsciences.com (M. R. Crager), jbraun@exactsciences.com (J. V. Braun)

*Corresponding author

To cite this article:

Michael Richard Crager, Jerome Victor Braun. Interval Estimation of the Absolute Risk of an Event with Competing Risks Using Proportional Regression of Cause-Specific Hazards. *American Journal of Applied Mathematics*. Vol. 10, No. 2, 2022, pp. 59-85.
doi: 10.11648/j.ajam.20221002.15

Received: April 4, 2022; **Accepted**: April 19, 2022; **Published**: April 28, 2022

Abstract: It is often important to account for the effects of a competing risk when estimating the risk of a particular event of interest by estimating its absolute risk. Available methodology for interval estimation of the absolute risk using the proportional regression of cause-specific hazards (CSH) has been limited to situations with time-invariant covariates and a single random censoring mechanism, without accommodation of cohort sampling study designs. Here we derive asymptotic pointwise confidence intervals in closed form for the absolute risk of an event at a specified time (the value of the cumulative incidence function) in the presence of competing risks using proportional CSH regression, accommodating external time-dependent covariates, cohort sampling study designs and multiple censoring mechanisms. Different covariates may be used for the event of interest and the various competing risks. Consistent with the definition of absolute risk, the CSH method produces absolute risk estimates that are less than or approximately equal to corresponding "conditional" risk estimates that do not account for competing risks. An example shows that this property is not necessarily shared by methods based on subdistribution hazard regression. Simulation studies indicate that the CSH method confidence intervals computed on the log cumulative hazard or the risk scale have coverage probabilities that approximate the nominal level for small and moderate samples, provided that the number of events per covariate is at least 10 and, when using cohort sampling, the ratio of patients without events to patients with events is at least 2:1.

Keywords: Absolute Risk, Cause-Specific Hazards, Cohort Sampling, Competing Risks, Cumulative Incidence Function, Interval Estimation, Time-Dependent Covariates

1. Introduction

When individual study participants or members of a cohort are followed for an event of interest, they may experience other events that prevent the occurrence of the event of interest or its observation. These other events may be regarded as competing risks. For example, in a clinical study of recurrence of cancer after surgical removal of the primary tumor, death without recurrence may be considered a competing risk. It is often useful to estimate the absolute (unconditional) risk of the event of interest, accounting for the effect of the competing risk. Analyses adjusting for competing risks can also be used to estimate the risk of the event of interest as a first event among several others. This is useful for composite endpoints such as disease-free survival: a competing risks analysis could be used to describe the treatment effect on each of the components of the composite.

The absolute risk of the event of interest as a function of time is known as the cumulative incidence function (CIF). If M is the event type of interest and $D_1, D_2, \ldots, D_K$ are

competing risk event types, then the CIF at time t_0 is $CIF(t_0) = \int_0^{t_0} S(t)\,\mathrm{d}\,\Lambda_M(t)$ where $S(t) = \exp\left\{-\Lambda_M(t) - \sum_{k=1}^{K} \Lambda_{D_k}(t)\right\}$ is the probability of remaining free of events of any type through time *t* and where $\Lambda_M(t)$ and the $\Lambda_{D_k}(t)$ are the event-type-specific (or, thinking of the various event types as distinct causes of failure, "cause-specific") cumulative hazards at time *t* without prior occurrence of any event type [1, 2]. If we treat observations with event type D_k as censored at the time of occurrence of D_k, and if the time to occurrence of D_k can reasonably be assumed to be independent of time to occurrence of M, our estimate of the risk of event *M* at time t_0 is effectively conditional on no event of type D_k occurring before time t_0. Since events of type D_k prevent the occurrence of events of type *M*, the absolute risk of *M* is lower than the risk of *M* conditional on the absence of events of type $D_k, k = 1, 2, \ldots, K$. The difference between the conditional and absolute risks may be substantial if D_k events are frequent [3, 4].

In this work, we discuss estimation of the absolute risk using Cox proportional hazards regression applied to the cause-specific hazards. We consider "external" time-dependent covariates, which arise when the effect of a covariate measured just once is not constant over time. We allow that the various event types may have distinct censoring mechanisms. We also consider stratified cohort sampling studies, in which a random sample is drawn from the study cohort with sampling fractions that vary across strata [5]. Typically, all study subjects with an event are included in the sample, but only a fraction of subjects without an event. Such designs are particularly useful when events are rare and assessment of necessary covariates is expensive or logistically difficult.

Point estimation of the absolute risk using proportional CSH regression is described by Prentice *et al.* [1] and by Kalbfleisch and Prentice [2], but neither variance estimation nor confidence intervals for the CIF are mentioned. Andersen *et al.* [6] note that the CSH approach to competing risks analysis is a special case of counting process methods for multi-state models (with each of the event types being an absorbing state), but do not discuss methods for weighted analyses of stratified cohort sampling designs. The SAS macros CumInc and CumIncV [7], which estimate cumulative incidence functions based on these methods, do not accommodate stratified cohort sampling weighting or time-dependent covariates. Putter and colleagues [8] give a very helpful review of methods for estimating the absolute risk, but do not discuss variance estimation. Benichou and Gail [9] describe an approach to estimating the variance of an estimator of a generalization of the CIF when proportional hazards regression of the cause-specific hazard is used, but do not provide a closed form solution. Large sample pointwise confidence intervals (and simultaneous confidence bands) for the CIF are provided in closed form by Cheng, Fine and Wei [10] using proportional regression of CSH with time-invariant covariates in studies not using cohort sampling, but they provide no simulations to evaluate the small- and moderate-sample coverage probabilities of the intervals.

Here we use the delta method to derive asymptotically valid confidence intervals in closed form for the absolute risk of an event at a specified time (value of the CIF at that time) using proportional regression of CSH, accommodating external time-dependent covariates, stratified cohort sampling designs and censoring mechanisms that vary across event types. The methods also accommodate stratified analyses in which the cumulative hazard function may vary freely between strata. Using an example data set, we compare point and interval estimates from the CSH approach to the point and interval estimates derived from a regression analysis for competing risks based on modeling the subdistribution hazard [11]. We study the small and moderate sample coverage probabilities of the CSH-method confidence intervals using simulations.

2. Interval Estimation of the Absolute Risk with a Single Competing Risk

Suppose we have possibly right-censored data on the time to an event of interest *M* and a single competing risk event *D*, with corresponding vectors of covariates z_M and z_D. These may be "external" time-dependent covariates $\mathrm{z}_M = \mathrm{z}_M(t)$ and $\mathrm{z}_D = \mathrm{z}_D(t)$ for $t \geq 0$, as defined by Kalbfleisch and Prentice [2]. Such covariates arise when the effect of a covariate assessed at baseline is time-dependent, as the model $\lambda(t) = \exp\left\{\beta^{\mathrm{T}}(t)\mathrm{z}\right\}\lambda_0(t)$ is equivalent to a model $\lambda(t) = \exp\left\{\beta^{\mathrm{T}}\mathrm{z}(t)\right\}\lambda_0(t)$ with fixed regression coefficients and time-dependent covariates with $\mathrm{z}(t)$ being a fixed vector function of time. For example, if the elements of $\mathrm{z}(t)$ are indicator functions of time, a piecewise constant hazard ratio model results. The elements of $\mathrm{z}(t)$ might also be continuous functions of time, such as the basis functions of a spline. Current patient age, defined as age at study entry plus elapsed time on study, is an external time-dependent covariate. Other examples are environmental factors not associated with individual study subjects. While model-fitting is straightforward with "internal" time-dependent covariates resulting from repeated measures over time of a covariate for each subject, the estimation of the CIF with such covariates is complex [2, 6, 12]. We will not consider internal time-dependent covariates here. In the following, we will use z_M and z_D to denote the vector functions of time; if only time-invariant covariates are used, the functions are constants. Using proportional hazards regression and the cause-specific hazards approach to competing risks analysis [1], we model the hazard function

for event type M as $\lambda_M(t;\mathrm{z}_M)=\lambda_{0M}(t)\exp\left(\beta_M^{\mathrm{T}}\mathrm{z}_M(t)\right)$, where $\beta_M=\left(\beta_{M1},\beta_{M2},\ldots,\beta_{Mp_M}\right)^{\mathrm{T}}$ is a vector of regression parameters associated with the covariate vector function z_M. Similarly, we model the hazard function for event type D as $\lambda_D(t;\mathrm{z}_D)=\lambda_{0D}(t)\exp\left(\beta_D^{\mathrm{T}}\mathrm{z}_D(t)\right)$, with $\beta_D=\left(\beta_{D1},\beta_{D2},\ldots,\beta_{Dp_D}\right)^{\mathrm{T}}$. The covariate vector functions z_M and z_D for event types M and D may be identical, have elements in common, or be completely distinct. For notational convenience, define $\mathrm{z}(t)=\left(\mathrm{z}_M^{\mathrm{T}}(t),\mathrm{z}_D^{\mathrm{T}}(t)\right)^{\mathrm{T}}$. No assumption on the relationship of the M-type and D-type events is required; in particular, the times to the two event types need not be independent [1]. If the study design uses stratified cohort sampling, let w_i be the sampling weight for subject $i=1,2,\ldots,n,$ which is the inverse of the sampling fraction in the subject's cohort sampling stratum. If the study design does not use stratified cohort sampling, set $w_i\equiv 1$. Without cohort sampling, the likelihood covering both event types M and D factors into a product of separate likelihoods for event types M and D, which implies that we can obtain maximum partial likelihood estimators $\hat{\beta}_M$ and $\hat{\beta}_D$ for β_M and β_D by fitting separate proportional hazards model for each event type, censoring on occurrence of the other event type [1]. The same result holds for the pseudolikelihood that is maximized when analyzing data from a stratified cohort sampling design [5],

$$\begin{aligned}L&=\prod_{i=1}^{n}\left(\frac{\exp\left\{\hat{\beta}_{\varepsilon_i}^{\mathrm{T}}\mathrm{z}_{\varepsilon_i i}(t_i)\right\}}{\sum_{j=1}^{n}w_jY_{\varepsilon_i j}(t_i)\exp\left\{\hat{\beta}_{\varepsilon_i}^{\mathrm{T}}\mathrm{z}_{\varepsilon_i j}(t_i)\right\}}\right)^{w_i\delta_{\varepsilon_i}}\\&=\left[\prod_{i=1}^{n}\left(\frac{\exp\left\{\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_{Mi}(t_i)\right\}}{\sum_{j=1}^{n}w_jY_{\varepsilon_i j}(t_i)\exp\left\{\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_{Mj}(t_i)\right\}}\right)^{w_i\delta_{Mi}}\right]\\&\quad\times\left[\prod_{i=1}^{n}\left(\frac{\exp\left\{\hat{\beta}_D^{\mathrm{T}}\mathrm{z}_{Di}(t_i)\right\}}{\sum_{j=1}^{n}w_jY_{\varepsilon_i j}(t_i)\exp\left\{\hat{\beta}_D^{\mathrm{T}}\mathrm{z}_{Dj}(t_i)\right\}}\right)^{w_i\delta_{Di}}\right],\end{aligned}\tag{1}$$

where for $i=1,2,\ldots,n,$ $\varepsilon_i\in\{M,D\}$ is the event type observed at time t_i, δ_ε equals 1 if an event of type ε occurred at this time and 0 if not, $\delta_{\varepsilon i}$ is an indicator for whether patient i had an event of type ε, $\mathrm{z}_{\varepsilon i}(t)$ is the covariate vector functions of time for event type ε for subject i, and $Y_{\varepsilon i}(t)$ is an indicator function for whether subject i is in the risk set for event type ε at time t.

When we fit the cause-specific models, we are modeling the occurrence of event type M without prior occurrence of event type D, and *vice versa*. This requires concurrent follow-up for both event types. The right-censoring times might differ between event types. In a clinical trial, for example, if cancer recurrence is the event of interest M and death without cancer recurrence is the competing risk D, time to event M or censoring may be determined exclusively by clinical trial visits whereas time to event D or censoring may also include post-study follow-up using registries. Subjects who die may have the censoring date for M recorded as the last clinical evaluation date before death. In this situation, it is appropriate for the analysis to include deaths occurring within a reasonably short window of time after the last follow-up for M, perhaps guided by the planned follow-up time interval for event M. For competing risk follow-up outside this window, the M-status is unknown, so we cannot determine whether D has occurred without prior M. In these cases, the time to D without prior M should be censored at the M censoring time plus the window. When neither event is observed, a subject's censoring time for both the M and D analyses should be the minimum censoring time for the two event types in that subject; when using a time window for event type M, we can consider that the subject is still being simultaneously followed for M and D until censoring for D if the censoring occurs within the time window, or until the end of the window if it does not. These considerations determine the risk set indicators $Y_{\varepsilon i}(t)$.

If stratified cohort sampling is used, the inverted Fisher information matrices (based on the pseudolikelihood) are not valid estimators of the covariance matrices of the regression parameter estimators $\hat{\beta}_M$ and $\hat{\beta}_D$; appropriate covariance matrix estimators $\hat{\mathrm{V}}_M$ and $\hat{\mathrm{V}}_D$ are given by the "robust" covariance sandwich method of Lin and Wei [13]. The Lin-Wei estimator approximates the jackknife estimator of variance [14, Section 7.2]. If stratified cohort sampling is not used, either the covariance sandwich estimators or the inverted Fisher information matrices are appropriate for $\hat{\mathrm{V}}_M$ and $\hat{\mathrm{V}}_D$. Let $N_{Mi}(t)$ and $N_{Di}(t)$ be the event-counting processes and let t_{Mi} and t_{Di} be the times to event or censoring for events M and D for subject i. Let $\mathrm{z}=\left(\mathrm{z}_M^{\mathrm{T}},\mathrm{z}_D^{\mathrm{T}}\right)^{\mathrm{T}}$ be a covariate function of time for which we wish to estimate the absolute risk of event M occurring by time t_0 (the cumulative incidence function value at time t_0 for event type M). Following Cheng *et al.* [10], a consistent estimator is given by

$$\hat{CIF}_M(t_0;\mathrm{z})=\int_0^{t_0}\hat{S}(t;\mathrm{z})\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)\mathrm{d}\hat{\Lambda}_{0M}(t),\tag{2}$$

where $\hat{S}(t;\mathrm{z})=\hat{S}_M(t;\mathrm{z}_M)\hat{S}_D(t;\mathrm{z}_D)$, and, for $\varepsilon\in\{M,D\}$, $\hat{S}_\varepsilon(t;\mathrm{z}_\varepsilon)=\exp\left\{-\int_0^t\exp\left(\hat{\beta}_\varepsilon^{\mathrm{T}}\mathrm{z}_\varepsilon(s)\right)\mathrm{d}\hat{\Lambda}_{0\varepsilon}(s)\right\}$, with

$$\hat{\Lambda}_{0\varepsilon}(t)=\int_0^t\frac{\sum_{i=1}^{n}w_i\,\mathrm{d}N_{\varepsilon i}(s)}{\sum_{j=1}^{n}w_jY_{\varepsilon j}(s)\exp\left(\hat{\beta}_\varepsilon^{\mathrm{T}}\mathrm{z}_{\varepsilon j}(s)\right)},\tag{3}$$

the last quantity being the estimated baseline cause-specific cumulative hazard function.

For large samples, the asymptotic variance of the estimator of the risk at a specific time can be obtained using the delta method. The *M*- and *D*-type event-counting processes jump at the same time with probability 0, so they are orthogonal [15]. The regression parameter estimators $\hat{\beta}_M$ and $\hat{\beta}_D$ are asymptotically independent of the jumps $\mathrm{d}N_{Mi}(t_{Mi})$ and $\mathrm{d}N_{Di}(t_{Di})$ in the *M* and *D* event counting processes [15, 16]. Hence, the variance of $C\hat{I}F_M(t_0;\mathrm{z})$ is consistently estimated by

$$\begin{aligned}\hat{\mathrm{Var}}\left\{C\hat{I}F_M(t_0;\mathrm{z})\right\} = &\left\{\nabla_{\hat{\beta}_M} C\hat{I}F_M(t_0;\mathrm{z})\right\}^{\mathrm{T}} \hat{\mathrm{V}}_M \left\{\nabla_{\hat{\beta}_M} C\hat{I}F_M(t_0;\mathrm{z})\right\} \\ &+\left\{\nabla_{\hat{\beta}_D} C\hat{I}F_M(t_0;\mathrm{z})\right\}^{\mathrm{T}} \hat{\mathrm{V}}_D \left\{\nabla_{\hat{\beta}_D} C\hat{I}F_M(t_0;\mathrm{z})\right\} \\ &+\sum_{i=1}^{n}\left\{\frac{\partial C\hat{I}F_M(t_0;\mathrm{z})}{\partial \mathrm{d}N_{Mi}(t_{Mi})}\right\}^2 \mathrm{d}N_{Mi}(t_{Mi}) \\ &+\sum_{i=1}^{n}\left\{\frac{\partial C\hat{I}F_M(t_0;\mathrm{z})}{\partial \mathrm{d}N_{Di}(t_{Di})}\right\}^2 \mathrm{d}N_{Di}(t_{Di}),\end{aligned} \tag{4}$$

where

$$\nabla_{\hat{\beta}_M} = \left(\partial/\partial\hat{\beta}_{M1}, \partial/\partial\hat{\beta}_{M2}, \ldots, \partial/\partial\hat{\beta}_{Mp_M}\right)^{\mathrm{T}}$$

and

$$\nabla_{\hat{\beta}_D} = \left(\partial/\partial\hat{\beta}_{D1}, \partial/\partial\hat{\beta}_{D2}, \ldots, \partial/\partial\hat{\beta}_{Dp_D}\right)^{\mathrm{T}}$$

are gradient operators, and, as shown in Appendix A1,

$$\begin{aligned}\nabla_{\hat{\beta}_M} C\hat{I}F_M(t_0;\mathrm{z}) = &\int_0^{t_0} \hat{S}(t;\mathrm{z})\left\{\mathrm{z}_M(t) - \xi_M(t) - \psi_M(t;\mathrm{z}_M)\right\} \\ &\times \mathrm{d}\hat{\Lambda}_M(t;\mathrm{z}_M),\end{aligned} \tag{5}$$

$$\nabla_{\hat{\beta}_D} C\hat{I}F_M(t_0;\mathrm{z}) = -\int_0^{t_0} \hat{S}(t;\mathrm{z})\,\psi_D(t;\mathrm{z}_D)\,\mathrm{d}\hat{\Lambda}_M(t;\mathrm{z}_M), \tag{6}$$

$$\begin{aligned}\frac{\partial C\hat{I}F_M(t_0;\mathrm{z})}{\partial \mathrm{d}N_{Mi}(t_{Mi})} = &\left[\hat{S}(t_{Mi};\mathrm{z}) - \left\{C\hat{I}F_M(t_0;\mathrm{z}) - C\hat{I}F_M(t_{Mi}-;\mathrm{z})\right\}\right] \\ &\times \mathrm{d}\hat{\Lambda}_M(t_{Mi};\mathrm{z}_M),\end{aligned} \tag{7}$$

and

$$\begin{aligned}\frac{\partial C\hat{I}F_M(t_0;\mathrm{z})}{\partial \mathrm{d}N_{Di}(t_{Di})} = &-\mathrm{d}\hat{\Lambda}_D(t_{Di};\mathrm{z}_D) \\ &\times\left\{C\hat{I}F_M(t_0;\mathrm{z}) - C\hat{I}F_M(t_{Di}-;\mathrm{z})\right\},\end{aligned} \tag{8}$$

where

$$\hat{S}(t;\mathrm{z}) = \exp\left\{-\int_0^t \mathrm{d}\hat{\Lambda}_M(s;\mathrm{z}_M) - \int_0^t \mathrm{d}\hat{\Lambda}_D(s;\mathrm{z}_D)\right\}, \tag{9}$$

and, for $\varepsilon \in \{M, D\}$,

$$\psi_\varepsilon(t;\mathrm{z}_\varepsilon) = \int_0^t \left\{\mathrm{z}_\varepsilon(s) - \xi_\varepsilon(s)\right\}\mathrm{d}\hat{\Lambda}_\varepsilon(s;\mathrm{z}_\varepsilon), \tag{10}$$

$$\xi_\varepsilon(t) = \frac{\sum_{i=1}^n Y_{\varepsilon i}(t) w_i \exp\left(\hat{\beta}_\varepsilon^{\mathrm{T}} \mathrm{z}_{\varepsilon i}(t)\right)\mathrm{z}_{\varepsilon i}(t)}{\sum_{i=1}^n Y_{\varepsilon i}(t) w_i \exp\left(\hat{\beta}_\varepsilon^{\mathrm{T}} \mathrm{z}_{\varepsilon i}(t)\right)}, \tag{11}$$

$$\mathrm{d}\hat{\Lambda}_\varepsilon(t;\mathrm{z}_\varepsilon(\bullet)) = \frac{\exp\left(\hat{\beta}_\varepsilon^{\mathrm{T}} \mathrm{z}_\varepsilon(t)\right)\sum_{i=1}^n w_i\,\mathrm{d}N_{\varepsilon i}(t)}{\sum_{i=1}^n Y_{\varepsilon i}(t) w_i \exp\left(\hat{\beta}_\varepsilon^{\mathrm{T}} \mathrm{z}_{\varepsilon i}(t)\right)}. \tag{12}$$

The asymptotic normality of the estimator $C\hat{I}F_M(t_0;\mathrm{z})$ follows from a first order Taylor series expansion based on $\nabla_{\hat{\beta}_M} C\hat{I}F_M(t_0;\mathrm{z})$, $\nabla_{\hat{\beta}_D} C\hat{I}F_M(t_0;\mathrm{z})$, $\partial C\hat{I}F_M(t_0;\mathrm{z})/\partial\,\mathrm{d}N_{Mi}(t_{Mi})$ and $\partial C\hat{I}F_M(t_0;\mathrm{z})/\partial\,\mathrm{d}N_{Di}(t_{Di})$, the orthogonality of the *M* and *D* counting process, the asymptotic multivariate normality of $\hat{\beta}_M$ and $\hat{\beta}_D$ [6, 16], the mutual independence among the random variables $\mathrm{d}N_{Mi}(t_{Mi}), i = 1, 2, \ldots n$ and the mutual independence among the random variables $\mathrm{d}N_{Di}(t_{Di}), i = 1, 2, \ldots n$.

To form a confidence interval for the absolute risk that is restricted to the unit interval (0,1), we can transform to the log cumulative hazard scale, defining $\hat{\rho}_{LL}(t_0;\mathrm{z}) = \ln\left[-\ln\left\{1 - C\hat{I}F_M(t_0;\mathrm{z})\right\}\right]$. Using the delta method, the standard deviation of $\hat{\rho}(T;\mathrm{z})$ is consistently estimated by $\hat{\mathrm{STD}}\left\{\hat{\rho}_{LL}(t_0;\mathrm{z})\right\} = \sqrt{\hat{\mathrm{Var}}\left\{C\hat{I}F_M(t_0;\mathrm{z})\right\}} \times\left|\left\{1 - C\hat{I}F_M(t_0;\mathrm{z})\right\}\ln\left\{1 - C\hat{I}F_M(t_0;\mathrm{z})\right\}\right|^{-1}$. Back-transforming to the risk scale gives an asymptotic level α confidence interval for $CIF_M(t_0;\mathrm{z})$ with endpoints $1 - \exp\left(-\exp\left[\hat{\rho}_{LL}(t_0;\mathrm{z}) \pm \Phi^{-1}(1-\alpha/2)\hat{\mathrm{STD}}\left\{\rho_{LL}(t_0;\mathrm{z})\right\}\right]\right)$, where Φ^{-1} is the inverse cumulative distribution function (CDF) of the standard normal distribution. We will refer to this as the "log-log" confidence interval. We can similarly compute the confidence intervals transforming to the cumulative hazard scale, giving an interval with endpoints $1 - \exp\left[-\hat{\rho}_L(t_0;\mathrm{z}) \pm \Phi^{-1}(1-\alpha/2)\hat{\mathrm{STD}}\left\{\hat{\rho}_L(t_0;\mathrm{z})\right\}\right]$, where $\hat{\rho}_L(t_0;\mathrm{z}) = -\ln\left\{1 - C\hat{I}F_M(t_0;\mathrm{z})\right\}$. and $\hat{\mathrm{STD}}\left\{\hat{\rho}_L(t_0;\mathrm{z})\right\} = \sqrt{\hat{\mathrm{Var}}\left\{C\hat{I}F_M(t_0;\mathrm{z})\right\}}\left\{1 - C\hat{I}F_M(t_0;\mathrm{z})\right\}^{-1}$. We

will refer to this as the "log" confidence interval. A confidence interval can also be computed on the risk scale with endpoints $C\hat{I}F_M(t_0;\mathrm{z}) \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\mathrm{Var}}\left\{C\hat{I}F_M(t_0;\mathrm{z})\right\}}$, but this interval may contain values less than 0 or greater than 1. We will refer to this as the "linear" interval.

The derivations of the "log-log" and "log" method intervals assume that $0 < C\hat{I}F_M(t_0;\mathrm{z}) < 1$. By its definition, the estimated absolute risk is never exactly equal to 1. If we estimate the CIF at a time t_0 that precedes the first observed *M*-type event, the estimated CIF will be 0, in which case the "log-log" and "log" methods do not give a confidence interval.

It is possible to estimate the absolute risk for event type *M* using proportional CSH regression on a set of covariates z_M but without covariates for the competing risk *D*. In this case, the term in the variance formula (4) associated with $\nabla_{\hat{\beta}_D} C\hat{I}F_M(t_0;\mathrm{z})$ vanishes and the hazard function estimate for event type *D* in (12) becomes $\mathrm{d}\hat{\Lambda}_D(t) = \sum_{i=1}^{n} w_i \,\mathrm{d}N_{Di}(t) \Big/ \sum_{j=1}^{n} w_j Y_{Dj}(t)$.

Ties in the observed times to event of the same type can be handled using the calculation that Efron uses to adjust the partial likelihood for ties [14, 17 Section 10.2.2.]. If *k* subjects (for notational convenience, denote them as subjects $i = 1,2,\ldots,k$) have events of type ε simultaneously at time *t*, we replace each of the terms $w_i \exp\left(\hat{\beta}_\varepsilon^{\mathrm{T}} \mathrm{z}_{\varepsilon i}(t)\right)$, $i = 1,2,\ldots,k$, in the numerator and denominator of (11) and the denominator of (12) by the average value $\sum_{j=1}^{k} w_j \exp\left(\hat{\beta}_\varepsilon^{\mathrm{T}} \mathrm{z}_{\varepsilon j}(t)\right) \Big/ k$, and we replace each of the *k* weights w_i in the numerator of (12) by $\sum_{i=1}^{k} w_i \Big/ k$.

3. Estimation of the Absolute Risk with Multiple Competing Risks

If we have an event of interest *M* and several competing risk event types $D_1, D_2, \ldots, D_K$, one option is to combine the competing risk events into a single event *D* with time to event $T_D = \min_{k \in \{1,2,\ldots,K\}} \left\{T_{D_k}\right\}$. This requires the covariate vector functions to be identical for all the competing risk event types $D_1, D_2, \ldots, D_K$, which may not be desirable. For example, if the event of interest was distant cancer recurrence in breast cancer patients and there were two competing risks, local/regional breast cancer recurrence and death without cancer recurrence, we would likely want to use different covariates for the two competing risks.

Fortunately, there is no need to combine competing risk event types, as the methods described above generalize easily to the case of multiple competing risks. Denote by z_{D_k} the covariate vector for event type D_k. The absolute risk at time t_0 is estimated consistently by $C\hat{I}F_M(t_0;\mathrm{z}) = \int_0^{t_0} \hat{S}(t;\mathrm{z}) \exp\left(\hat{\beta}_M^{\mathrm{T}} \mathrm{z}_M(t)\right) \mathrm{d}\hat{\Lambda}_{0M}(t)$, where we now have $\hat{S}(t;\mathrm{z}) = \hat{S}_M(t;\mathrm{z}_M) \prod_{k=1}^{K} \hat{S}_{D_k}(t;\mathrm{z}_{D_k})$. The arguments used previously show that the variance of $C\hat{I}F_M(t_0;\mathrm{z})$ is estimated consistently by

$$\begin{aligned}
\hat{\mathrm{Var}}\left\{C\hat{I}F_M(t_0;\mathrm{z})\right\} &= \left\{\nabla_{\hat{\beta}_M} C\hat{I}F_M(t_0;\mathrm{z})\right\}^{\mathrm{T}} \hat{\mathrm{V}}_M \left\{\nabla_{\hat{\beta}_M} C\hat{I}F_M(t_0;\mathrm{z})\right\} \\
&+ \sum_{k=1}^{K} \left\{\nabla_{\hat{\beta}_{D_k}} C\hat{I}F_M(t_0;\mathrm{z})\right\}^{\mathrm{T}} \hat{\mathrm{V}}_{D_k} \left\{\nabla_{\hat{\beta}_{D_k}} C\hat{I}F_M(t_0;\mathrm{z})\right\} \\
&+ \sum_{i=1}^{n} \left\{\frac{\partial C\hat{I}F_M(t_0;\mathrm{z})}{\partial \mathrm{d}N_{Mi}(t_{Mi})}\right\}^2 \mathrm{d}N_{Mi}(t_{Mi}) \\
&+ \sum_{k=1}^{K} \sum_{i=1}^{n} \left\{\frac{\partial C\hat{I}F_M(t_0;\mathrm{z})}{\partial \mathrm{d}N_{D_k i}(t_{D_k i})}\right\}^2 \mathrm{d}N_{D_k i}(t_{D_k i}),
\end{aligned} \tag{13}$$

with, for all $k = 1,2,\ldots,K$,

$$\begin{aligned}
&\nabla_{\hat{\beta}_M} C\hat{I}F_M(t_0;\mathrm{z}) = \\
&\int_0^{t_0} \hat{S}(t;\mathrm{z}) \left\{\mathrm{z}_M(t) - E_M(t) - \Psi_M(t;\mathrm{z}_M)\right\} \\
&\times \mathrm{d}\hat{\Lambda}_M(t;\mathrm{z}_M),
\end{aligned} \tag{14}$$

$$\begin{aligned}
&\nabla_{\hat{\beta}_{D_k}} C\hat{I}F_M(t_0;\mathrm{z}) = \\
&-\int_0^{t_0} \hat{S}(t;\mathrm{z}) \Psi_{D_k}(t;\mathrm{z}_{D_k}) \mathrm{d}\hat{\Lambda}_M(t;\mathrm{z}_M),
\end{aligned} \tag{15}$$

$$\begin{aligned}
\frac{\partial C\hat{I}F_M(t_0;\mathrm{z})}{\partial \mathrm{d}N_{Mi}(t_{Mi})} &= \left[\hat{S}(t_{Mi};\mathrm{z}) - \left\{C\hat{I}F_M(t_0;\mathrm{z}) - C\hat{I}F_M(t_{Mi};\mathrm{z})\right\}\right] \\
&\times \mathrm{d}\hat{\Lambda}_M(t_{Mi};\mathrm{z}_M),
\end{aligned} \tag{16}$$

and

$$\begin{aligned}
&\frac{\partial C\hat{I}F_M(t_0;\mathrm{z})}{\partial \mathrm{d}N_{D_k i}(t_{D_k i})} = \\
&-\mathrm{d}\hat{\Lambda}(t_{D_k i};\mathrm{z}_D) \left\{C\hat{I}F_M(t_0;\mathrm{z}) - C\hat{I}F_M(t_{D_k i} -;\mathrm{z})\right\},
\end{aligned} \tag{17}$$

where

$$\hat{S}(t;\mathrm{z}) = \exp\left\{-\int_0^t \mathrm{d}\hat{\Lambda}_M(s;\mathrm{z}_M) - \sum_{k=1}^{K} \int_0^t \mathrm{d}\hat{\Lambda}_{D_k}(s;\mathrm{z}_{D_k})\right\} \tag{18}$$

and, for $\varepsilon = M, D_1, D_2, \ldots, D_K$,

$$\psi_\varepsilon(t;\mathrm{z}_\varepsilon) = \int_0^t \left\{\mathrm{z}_\varepsilon(s) - \xi_\varepsilon(s)\right\} \mathrm{d}\hat{\Lambda}_\varepsilon(s;\mathrm{z}_\varepsilon), \tag{19}$$

$$\xi_{\varepsilon}(t)=\frac{\sum_{i=1}^{n} w_i Y_{\varepsilon i}(t)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon i}(t)\right)\mathrm{z}_{\varepsilon i}(t)}{\sum_{i=1}^{n} w_i Y_{\varepsilon i}(t)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon i}(t)\right)}, \tag{20}$$

with

$$\mathrm{d}\hat{\Lambda}_{\varepsilon}\left(t;\mathrm{z}_{\varepsilon}\right)=\frac{\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon}(t)\right)\sum_{i=1}^{n} w_i\,\mathrm{d}N_{\varepsilon i}(t)}{\sum_{i=1}^{n} w_i Y_{\varepsilon i}(t)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon i}(t)\right)}. \tag{21}$$

This formulation permits the use of different covariates for different competing risk event types. The practitioner can thus avoid using covariates that theoretically or empirically are not related to a particular event type.

With multiple competing risks, we should restrict the times included in the analysis to when there is concurrent follow-up for all event types. If we apply a window to the main event type, we count the first event among the competing risks that occurs within that window, if any; otherwise, all competing risk events are considered censored at the end of the window. The recoding of censoring times in an analysis of multiple competing risks is diagrammed schematically in Figure 1.

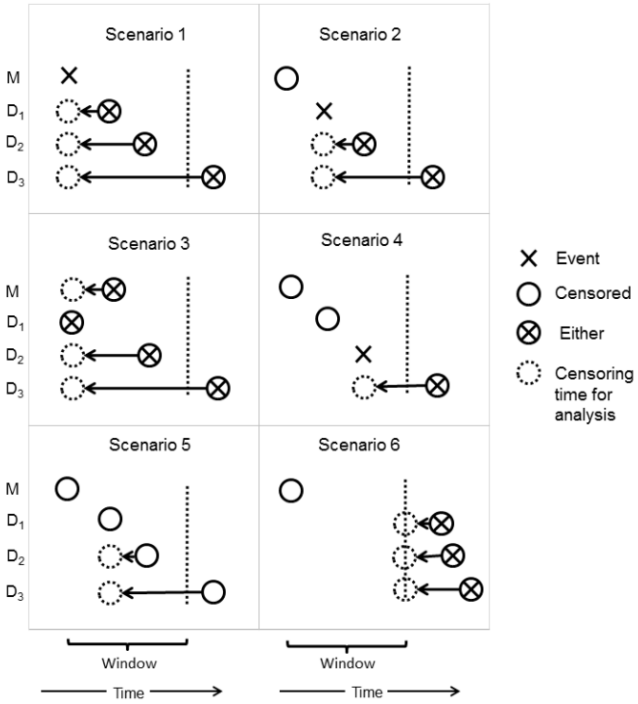

Figure 1. *Schematic showing recoding of censoring status and times for analysis in six different scenarios. In each scenario, the first line represents the event of interest (M). The next lines represent competing risks* (D_1, D_2, D_3). *The dotted vertical line shows the edge of the window within which occurrence of a competing risk is assumed to be the reason for censoring of the event of interest.*

4. Stratified Analyses

Sometimes it is appropriate to stratify the study population and allow different cause-specific baseline hazard functions across strata. (Note that this concept is distinct from stratified cohort sampling, which we discussed earlier. Stratified analyses can be done without stratified cohort sampling, and the reverse is also true.) For example, in a study of cancer patients, we might stratify on the stage of cancer. Stratified proportional cause-specific hazard regression models are easily accommodated using the methods described above. All the derivations hold except that the baseline hazard estimator and the subject-specific hazard estimators (12) and (21) are based solely on observations within each stratum. Letting $S_1, S_2, \ldots, S_L$ be the exhaustive, mutually exclusive subsets of the sample population $\{1,2,\ldots,n\}$ created by the joint distribution of the stratification variables in the study sample, and letting $I_{\{i\in S_l\}}$ be the indicator function for membership in subset S_l, the baseline cumulative hazard estimator for stratum S_l is

$$\hat{\Lambda}_{0\varepsilon}^{(l)}\left(t_0\right)=\int_0^{t_0}\frac{\sum_{i=1}^{n} I_{\{i\in S_l\}} w_i\,\mathrm{d}N_{\varepsilon i}(s)}{\sum_{i=1}^{n} I_{\{i\in S_l\}} Y_{\varepsilon i}(s) w_i \exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon i}(s)\right)},$$

and the patient-specific hazard estimator for stratum S_l is

$$\mathrm{d}\hat{\Lambda}_{\varepsilon}^{(l)}\left(t;\mathrm{z}_{\varepsilon}\right)=\frac{\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon}(t)\right)\sum_{i=1}^{n} I_{\{i\in S_l\}} w_i\,\mathrm{d}N_{\varepsilon i}(t)}{\sum_{i=1}^{n} I_{\{i\in S_l\}} Y_{\varepsilon i}(t) w_i \exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon i}(t)\right)}.$$

5. Relationship of the Absolute and Conditional Risk Estimates

Recall that if the competing risk events times and the censoring time are assumed to be independent of the time to the event of interest, the risk estimate obtained by censoring on the competing risks has the interpretation of the probability of the event of interest in the scenario that none of the other event types occurs. We call such risk estimates "conditional". As is often pointed out (see, for example, [8]), the independence assumption is generally not testable from the data. However, the assumption might be reasonable in some circumstances, for example if disease recurrence was the event of interest M and the competing risk D was death from unrelated causes.

The true absolute risk is always less than or equal to the true conditional risk. Consistent with this inequality, within limits of approximations due to the discreteness of the estimated cumulative hazard function, the proportional CSH regression estimate of the absolute risk at a given time is less than or equal to the corresponding conditional risk estimate at the same time, since

$$\begin{aligned}
&\hat{CIF}_M\left(t_0;\mathrm{z}\right)= \\
&\int_0^{t_0}\hat{S}_M\left(t;\mathrm{z}_M\right)\left\{\prod_{k=1}^{K}\hat{S}_{D_k}\left(t;\mathrm{z}_{D_k}\right)\right\} \\
&\times\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)\mathrm{d}\hat{\Lambda}_{0M}\left(t\right) \\
&\le\int_0^{t_0}\hat{S}_M\left(t;\mathrm{z}_M\right)\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)\mathrm{d}\hat{\Lambda}_{0M}\left(t\right) \\
&=\int_0^{t_0}\exp\left\{-\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)\right\}\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right) \\
&\approx 1-\exp\left\{-\hat{\Lambda}_M\left(T;\mathrm{z}_M\right)\right\} \\
&=1-\hat{S}_M\left(t_0;\mathrm{z}_M\right)=\hat{r}_M\left(t_0;\mathrm{z}_M\right),
\end{aligned}$$

where $\hat{S}_M\left(t;\mathrm{z}_M\right)$ is the Breslow-method estimate of the survival function for event M, censoring on occurrence of the competing risks $D_1, D_2, \ldots, D_K$, so $\hat{r}_M\left(t_0;\mathrm{z}_M\right)$ is the conditional risk estimate for event M at time t_0 under the scenario that no competing risk event has yet occurred. The approximation

$$\int_0^{t_0}\exp\left\{-\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)\right\}\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)\approx 1-\exp\left\{-\hat{\Lambda}_M\left(t_0;\mathrm{z}_M\right)\right\}$$

is not exact because the empirical cumulative hazard function estimate is not continuous. The near approximation allows the estimated CIF to be very slightly higher than the conditional risk estimate when the competing risks are highly unlikely.

Absolute risk estimates produced by methods other than CSH may be noticeably higher than the corresponding conditional risk estimates, as illustrated by the example calculation in the next section.

6. Example Calculation with Comparison to Methods Based on the Subdistribution Hazard

For an example calculation, we use the data from the National Cancer Institute for 506 Stage 3 and 4 prostate cancer patients described by Byar and Green [18]. Andrews and Herzberg published the data [19]. The patients were randomly assigned to four treatments: placebo, 0.2 mg diethylstilbestrol (DES), 1 mg DES, or 5 mg DES. We followed the recommendation of Byar and Green [18] to combine the placebo and 0.2 mg groups as "low dose DES" and the 1 and 5 mg groups as "high dose DES". Patients were followed for survival, with cause of death categorized as prostate cancer, heart or vascular disease, cerebrovascular accident, pulmonary embolus, other cancer, respiratory disease, or other. There are several covariates in the data set, including patient age at study entry, size of primary tumor in cm^2 as estimated from a digital rectal exam, and history of cardiovascular disease. Our analysis excluded observations with missing covariate values. We considered prostate cancer death (PCD) as the event of interest M (129 events), CV death (heart or vascular disease or cerebrovascular accident) as competing risk D_1 (126 events), and death from any other cause as competing risk D_2 (96 events). The proportional regression of cause-specific hazards used the following factors:

1) For PCD: tumor size as a linear covariate with separate effects for each stage, high or low dose DES, and an external time-dependent factor for stage with the time-dependence described by a 2-degree-of-freedom natural cubic spline applied to follow-up time with knots at 0, 2, and 7 years. Time-dependent modeling for stage was chosen because the diagnostic method of Lin and colleagues [22] suggested a departure from constant proportional hazards for this factor.
2) For the competing risk of CV death: a time-invariant factor for prior history of cardiovascular disease.
3) For the competing risk of death due to any other cause: an external time-dependent effect for current patient age, defined as age at entry plus the current follow-up time.

Figure 2 shows estimates and pointwise 95% confidence intervals for the one-year risk of PCD from the CSH-method competing risks model and corresponding conditional risk estimates for PCD censoring on death due to causes other than prostate cancer. Patients with a history of cardiovascular disease are at higher risk for cardiovascular death, so the absolute risk of prostate cancer death is lower for those patients relative to patients without prior cardiovascular disease. Similarly, as patient age at study entry increases, the absolute risk of prostate cancer death decreases due to increasing mortality from other causes. The case of a 30-year old stage 4 patient provides an example in which the absolute risk estimate is slightly higher than the conditional risk estimate due to the discreteness of the empirical cumulative hazard function estimate.

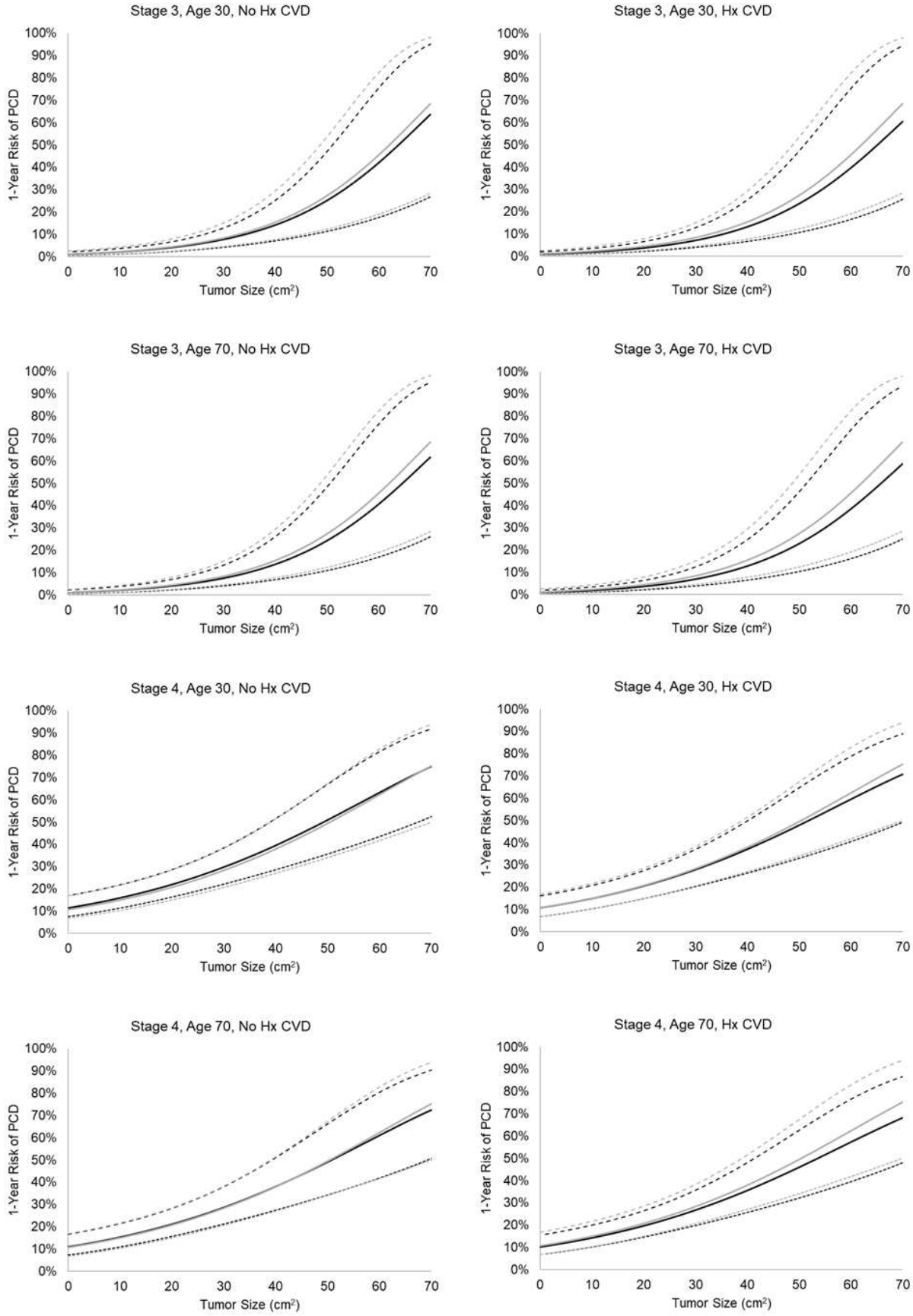

Figure 2. *Estimates and 95% confidence intervals for the one-year risk of death due to prostate cancer using CSH analysis of competing risks (black) and the conditional estimate censoring on other causes of death (gray), for prostate cancer patients taking low-dose DES.*

Among the patients with stage 3 disease, 35 experienced PCD and 257 did not. For stage 4 disease, 95 experienced PCD and 119 did not. To illustrate the analysis of a stratified cohort sampling study, we randomly sampled 100 of the 257 stage 3 patients without PCD and included them in the analysis with a weight of $257/100 = 2.57$. All other patients were assigned a weight of 1. The results are shown in Figure 3. The risk estimates are similar to the estimates using the full cohort, and the confidence intervals are slightly wider. With this relatively small data and the resulting wide confidence intervals, repeated sampling of the non-event patients produces estimates with substantial variability about the full-sample estimates.

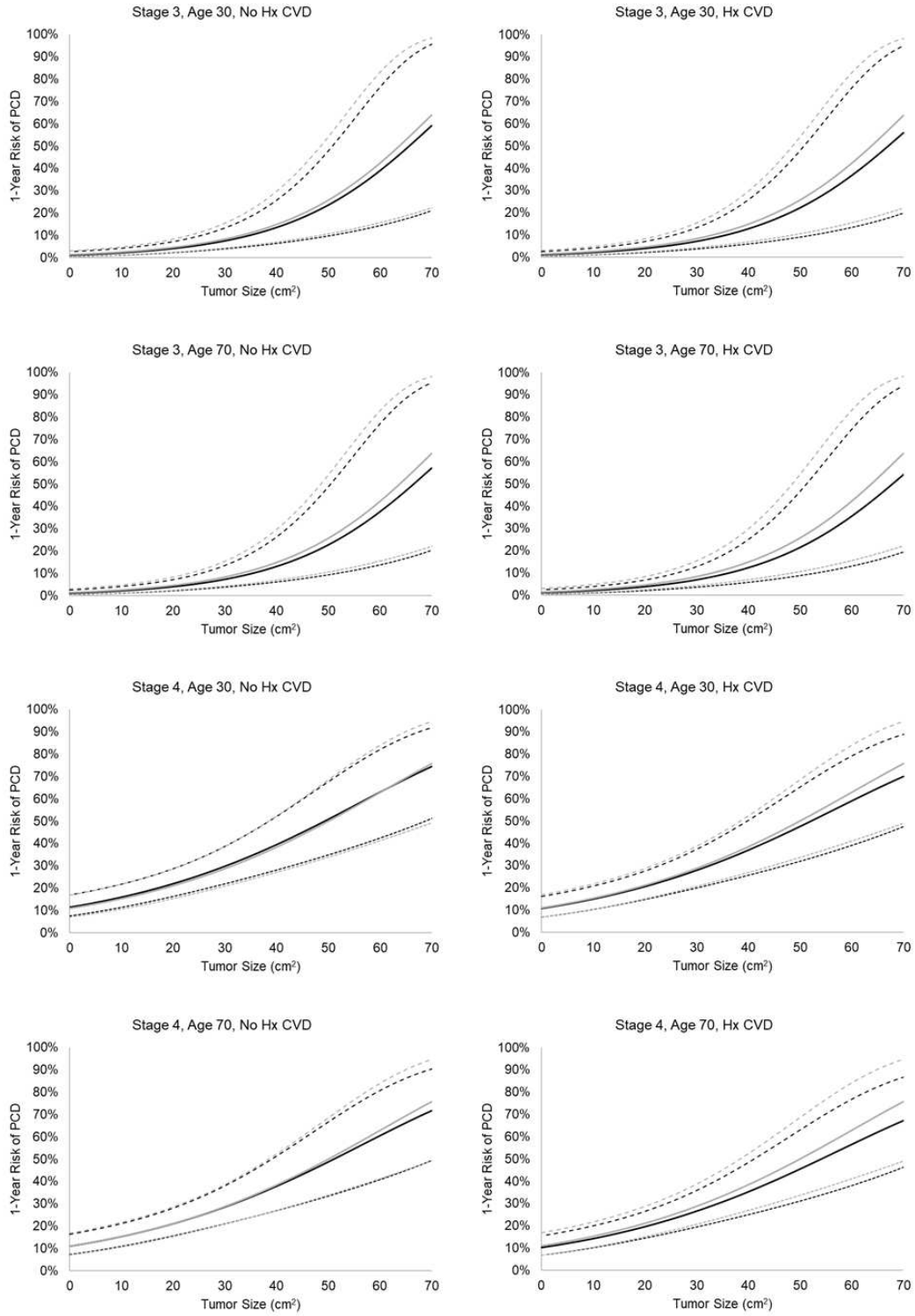

Figure 3. *Estimates and 95% confidence intervals for the one-year risk of death due to prostate cancer using CSH analysis of competing risks (black) and the conditional estimate censoring on other causes of death (gray), for prostate cancer patients taking low-dose DES. Stratified cohort sampling of the 100 / 257 stage 3 patients without prostate cancer death.*

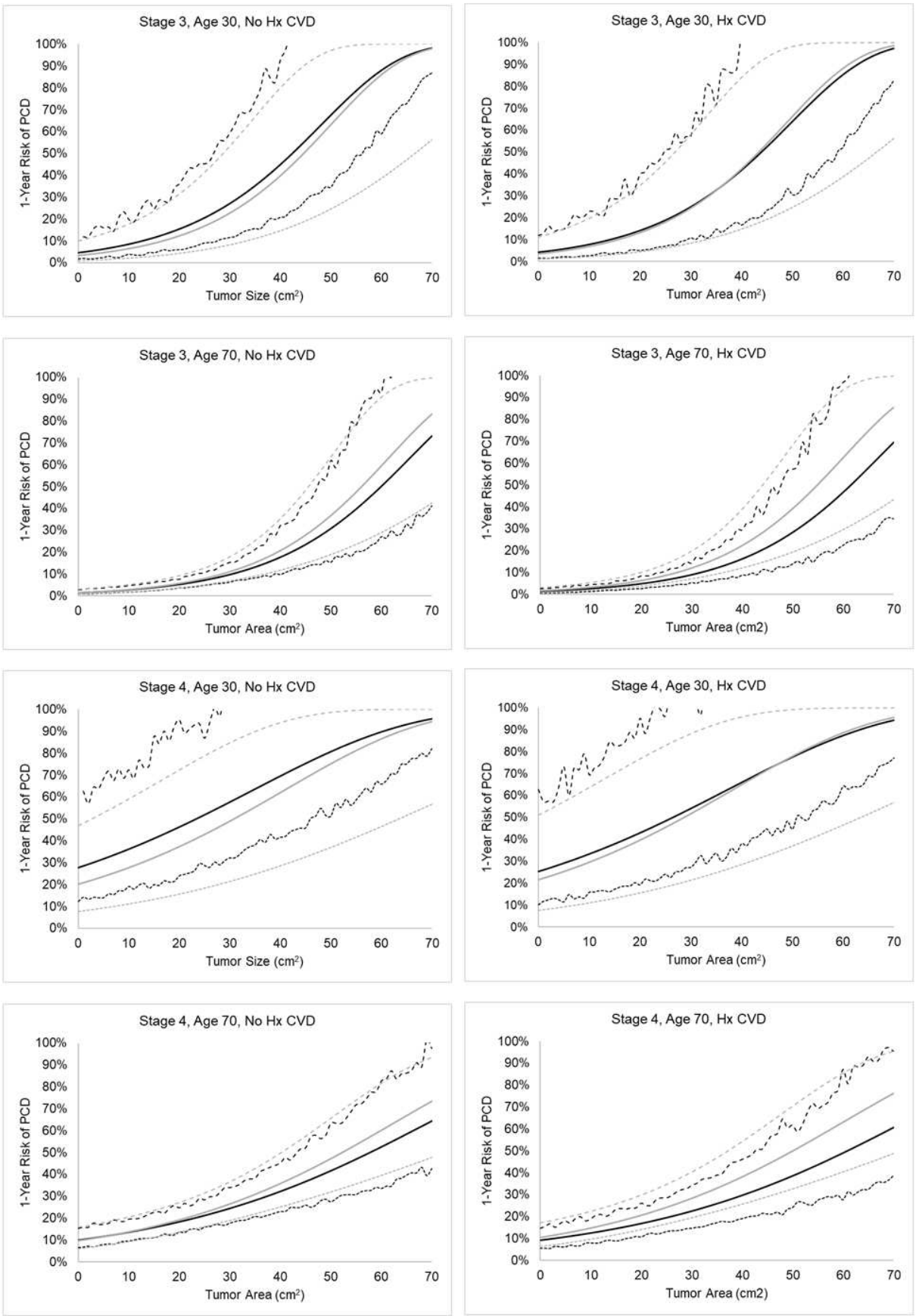

Figure 4. *Estimates and 95% confidence intervals for the one-year risk of death due to prostate cancer using the Fine-Gray method with time-invariant effect for stage for the analysis of competing risks (black), and the conditional risk estimate with the same covariates censoring on other causes of death (gray), for stage 3 patients taking low-dose DES.*

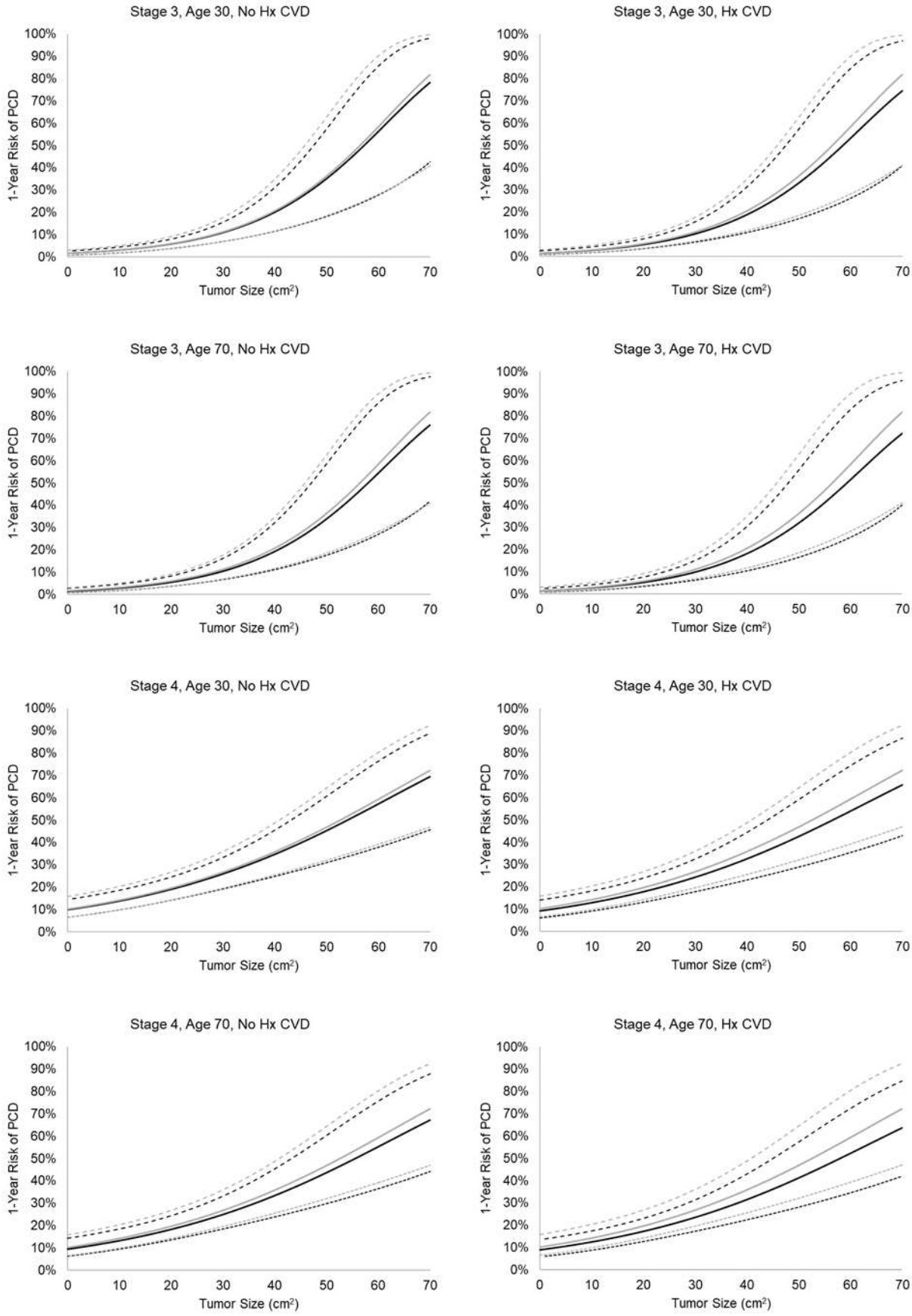

Figure 5. *Estimates and 95% confidence intervals for the one-year risk of death due to prostate cancer using the CSH method with time-invariant effect for stage for the analysis of competing risks (black), and the conditional risk estimate with the same covariates censoring on other causes of death (gray), for prostate cancer patients taking low-dose DES.*

Fine and Gray (1999) developed an alternative to the cause-specific hazard approach to proportional hazards regression analysis in the presence of competing risks. They defined the "subdistribution hazard" for event type M as $\lambda_M(t;\mathrm{z}) = CIF'_M(t;\mathrm{z})/\{1-CIF_M(t;\mathrm{z})\}$, where $CIF'_M(t;\mathrm{z})$ is the derivative of the CIF with respect to time, and applied the proportional regression model $\lambda_M(t;\mathrm{z}) = \lambda_{0M}(t;\mathrm{z})\exp\left(\beta^{\mathrm{T}}\mathrm{z}\right)$. Geskus developed a method of fitting this model using ordinary Cox proportional hazards regression with time-dependent weights [21]. To fit the Fine-Gray model, the usual proportional hazards regression partial likelihood is modified by keeping subjects who experience the competing risk event in the risk set until the future time at which they would have been censored, if this is known (for example, if only administrative censoring is present), or using time-dependent weights in the presence of random censoring.

The proportional subdistribution hazard assumption of the Fine-Gray model imposes a different constraint on the relation between the covariates and the risk of the event of interest than does the model of proportional cause specific hazards. This can lead to differences in the absolute risk point and interval estimates produced by the two methods.

Figure 4 shows one-year PCD risk estimates from the Fine-Gray model applied to the example data set with time-invariant covariates for high dose DES, stage, tumor size (separate effect for each stage), history of CVD, and patient age at study entry. Figure 5 shows estimates from the CSH method using the same covariates as for the analysis that generated Figure 2, except that a time-invariant effect was used for stage. The risk estimates with time-invariant effect for stage 3 tend to be higher than with time-dependent stage effect, suggesting a potential bias due to non-proportional hazards. Some of the Fine-Gray model risk estimates for the absolute risk are substantially higher than the corresponding conditional risk estimates.

7. Simulation Studies

The confidence intervals described here use asymptotic approximations assuming large samples. To examine their small- and moderate-sample coverage probabilities, we performed a series of simulation studies. Details of the simulation methods and results are given in the Supplemental Material. Covariate vectors z_{Mi} and $\mathrm{z}_{Di}, i = 1,2,\ldots,n$ for M and D events were generated from multivariate normal distributions. True log hazard ratio vectors β_M and β_D were set. For each simulated observation i, times to M and D events were generated from the exponential distribution with intensity parameters $\lambda_{M0}\exp\left(\beta_M^{\mathrm{T}}\mathrm{z}_{Mi}\right)$ and $\lambda_{D0}\exp\left(\beta_D^{\mathrm{T}}\mathrm{z}_{Di}\right)$, with a random censoring time from an independent exponential distribution. The sample size n was varied to give a wide range of expected events counts. Individual simulated data sets were included in the summaries only if they had at least 2 first events of type M and at least 2 first events of type D. For covariate values of z_M and z_D, the true CIF at a fixed time t_0 was calculated as $CIF\left(t_0;\mathrm{z}\right) = \int_0^{t_0}\exp\left\{-\lambda_{M0}\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M\right)t - \lambda_{D0}\exp\left(\hat{\beta}_D^{\mathrm{T}}\mathrm{z}_D\right)t\right\}\lambda_{M0}\exp\left(\beta_M^{\mathrm{T}}\mathrm{z}_M\right)\mathrm{d}t$. We generated 1000 replications of each scenario, giving a standard error of 0.7% for coverage probabilities close to the nominal level.

Scenarios examined by simulation had (I) 3 covariates for M, 1 covariate for D, (II), 6 covariates for M, 2 covariates for D, and (III) 1 covariate for M and no covariate for D. The covariate covariance matrices, true log hazard ratio vectors, and covariate values (designated as Case 1, 2, and 3) at which to estimate the risk of M are specified in the Supplemental Material. Simulations using 3 covariates for M-type events and 1 covariate for D-type events (Scenario I) were conducted for situations in which the D-type and M-type events occurred with similar frequency (Appendix A3 Figure A1) and when the D-type event was substantially less frequent than the M-type event (Appendix A3 Figure A2).

The model-based and Lin-Wei robust estimators of the cause-specific log hazard ratio covariance matrices gave similar coverage probabilities. The simulation study results indicate that the "log-log" and "linear" method confidence intervals have actual coverage close to the nominal level even for small sample sizes. In contrast, the "log" method confidence intervals have lower-than-nominal coverage probability even for moderate sample sizes, tending to "miss low", with the upper limit below the true value.

Simulations using 6 covariates for M-type events and 2 covariates for D-type events (Scenario II) showed 95% confidence interval coverage probabilities close to the nominal level using the "log-log" and "linear" methods, provided the mean number of events was at least 10 times the number of covariates (Appendix A3 Figure A3). This is a reminder that the ratio of events to covariates in proportional hazards regression models should be at least 10:1 [22]. The "log-log" and "linear" confidence interval methods again gave better coverage than the "log" method. Scenario III simulation results using a single covariate for M events and none for D events were similar (Appendix A3 Figure A4).

Simulations using stratified cohort sampling were conducted using the covariate distributions of Scenario I (3 covariates for M-type events, 1 covariate for D-type events). Within each replication of the simulation, a random sample was drawn from the simulated observations without either an M-type or a D-type event. The ratio of the number of non-event observations sampled to the number of observations with events varied from 1:1 to 4:1. The results are shown in Appendix A3 Figures A5 and A6. The true coverage probabilities are close to the nominal level for ratios of 3:1 and 4:1. Coverage probabilities for some of the covariate cases were slightly below the nominal level for the ratio 2:1 and substantially below nominal level for ratio 1:1. This appears to be due primarily to minor biases in the regression parameter estimators, which decrease as the sample size increases. At the 2:1 ratio, the relative bias (the bias

divided by the true value) in one of the three covariates was 8.4% for the smallest samples and 2.4% for the largest samples in the simulation. For the 3:1 ratio, the bias was reduced to 1.1% for the largest samples and for 4:1 it was less than 1%. To understand how this bias might arise, recall that the regression parameter estimators are obtained by maximizing the pseudolikelihood (1), which is equivalent to finding the regression parameter vector values at which the associated score function is **0**. The weighting gives a consistent estimator of the true regression parameter vector β_ε for each event type ε because the score function,

$$U_\varepsilon^{(w)}(\beta_\varepsilon)=\sum_{i\in\tilde{C}} w_i\delta_{\varepsilon i}\left(z_{\varepsilon i}(t_i)-\frac{\sum_{j\in\tilde{C}} z_{\varepsilon j}(t_i)w_j Y_{\varepsilon j}(t_i)\exp\left\{\beta_\varepsilon^{\mathrm{T}} z_{\varepsilon j}(t_i)\right\}}{\sum_{j\in\tilde{C}} w_j Y_{\varepsilon j}(t_i)\exp\left\{\beta_\varepsilon^{\mathrm{T}} z_{\varepsilon j}(t_i)\right\}}\right) \quad (22)$$

where $\tilde{C}\subset\{1,2,\ldots,n\}$ denotes the stratified random sample of the cohort, is an unbiased estimator of the score function using the full cohort without sampling,

$$U_\varepsilon(\beta_\varepsilon)=\sum_{i=1}^{n}\delta_{\varepsilon i}\left(z_{\varepsilon i}(t_i)-\frac{\sum_{j=1}^{n} z_{\varepsilon j}(t_i)Y_{\varepsilon j}(t_i)\exp\left\{\beta_\varepsilon^{\mathrm{T}} z_{\varepsilon j}(t_i)\right\}}{\sum_{j=1}^{n} Y_{\varepsilon j}(t_i)\exp\left\{\beta_\varepsilon^{\mathrm{T}} z_{\varepsilon j}(t_i)\right\}}\right).$$

Observations without an event contribute heavily to the last term of (22), which is a weighted average of the covariate vector values among all observations in the risk set at time t_i. Therefore, particularly for smaller sample sizes, the variance of $U_\varepsilon^{(w)}(\beta_\varepsilon)-U_\varepsilon(\beta_\varepsilon)$ increases as the number of non-event observations in the stratified cohort sample $\tilde{C}$ decreases due to lowering the sampling fraction. The score function $U_\varepsilon^{(w)}(\beta_\varepsilon)$ is a non-linear function of the regression parameter vector β_ε and in general, for $d\neq 0$, $U_\varepsilon^{(w)}(\beta_\varepsilon+d)-U_\varepsilon^{(w)}(\beta_\varepsilon) \neq U_\varepsilon^{(w)}(\beta_\varepsilon)-U_\varepsilon^{(w)}(\beta_\varepsilon-d)$. Therefore, although $U_\varepsilon^{(w)}(\beta_\varepsilon)$ is unbiased for $U_\varepsilon(\beta_\varepsilon)$, the increased variability in $U_\varepsilon^{(w)}(\beta_\varepsilon)$ due to a lower sampling fraction might well induce a non-negligible bias in the maximum partial pseudolikelihood estimator $\hat{\beta}_\varepsilon$ obtained by solving the score equation $U_\varepsilon^{(w)}(\beta_\varepsilon)=0$. However, the results of this simulation study indicate that if the number of sampled subjects without events is at least twice the number of subjects with events, the bias is small enough that the coverage probability of the CSH-method confidence intervals for the absolute risk is close to the nominal level.

Information on the bias of the CSH point estimator of absolute risk and its estimated standard error was also captured from the simulation studies. The relative bias was defined as the mean difference between the estimated and true absolute risks divided by the true absolute risk. The performance of variance estimators was summarized as the ratio of the mean estimated standard error of the absolute risk to the standard deviation of the absolute risk estimates in the population of 1000 simulations.

The relative bias of the CSH point estimator is described in Appendix A3 Figure A7 for studies without cohort sampling. Relative biases were generally in the range of 1-2% except for small sample sizes, where some relative biases approached 10%. Relative biases of the CSH point estimator in cohort sampling studies (Appendix A3 Figure A8) were similar except that more notable relative biases occurred for non-event: event ratios of 2:1 and 1:1.

The standard error estimates for non-cohort sampling studies were close to the true standard deviation, with a ratio generally remaining within the range 0.95-1.05 when there were at least 10 events per covariate (Appendix A3 Figure A9). Similar results were obtained for cohort sampling studies with non-event:event ratios of 3:1 or higher. More substantial differences between the estimated and true standard errors were noted for lower non-event:event ratios, particularly for the 1:1 ratio, in which the standard errors for small samples in some cases were more than 10% below the true value.

8. Discussion

Confidence intervals computed on the cumulative hazard scale ("log" method intervals) consistently performed worse than the confidence intervals computed on the log cumulative hazard scale ("log-log" method intervals) or the risk scale ("linear" method intervals) in our simulation studies. We therefore recommend using either the "log-log" or the "linear" intervals. The "log-log" intervals may be preferable since the confidence interval limits always lie within the unit interval.

For time-invariant covariates in studies without cohort sampling, expression (4) reduces to the calculation given by Cheng and colleagues [10]. In the absence of a competing risk *D* and without stratified cohort sampling or time-dependent covariates, expression (4) reduces to the expressions given by Tsiatis [16] for the variance of a risk estimator from Cox regression. The confidence interval widths for the conditional and absolute risks in our example calculations are similar. This is because the *D*-type events influence the CIF only through the function $\hat{S}_D(t;z_D)$, which is relatively stable if there are few *D*-type events.

An important feature of the CSH model is that different covariates may be used for the event of interest and the various competing risks. For example, covariates related to cancer are not generally useful for predicting the risk of death to other causes and need not be included in the model for the competing risk. In general, it is better to avoid using non-informative covariates, as they add noise to the estimation.

In contrast to the proportional CSH regression model, the Fine-Gray proportional subdistribution hazard model provides a single measure, the subdistribution hazard ratio, of the

strength of association of the covariate with the event of interest, accounting for the competing risks. The value of the subdistribution hazard ratio corresponds to the estimates of risk of the event of interest as a function of the covariate in a familiar way [8, 20]. If a proportional CSH regression model is fit using the same covariate for both the event of interest and the competing risk and the covariate has a strong association with both event types, the estimated risk of the event of interest as a function of the covariate value may not correspond to the hazard ratio for the event of interest viewed in isolation. The hazard ratios for both the event of interest and the competing risk are required to understand absolute risk estimates derived from proportional CSH regression. It is possible to construct a standardized subdistribution hazard ratio estimate using the CSH method (see Appendix A2). However, if estimation of the subdistribution hazard ratio is a primary goal of the analysis, it is probably best to obtain the estimate using the Fine-Gray method.

In a clinical setting, the conditional risk of an event may better reflect a patient's understanding than absolute risk. For example, a patient's risk of cancer recurrence within 10 years assuming the patient's survival for that period would seem more intuitive to the patient than the absolute risk, which is discounted for the probability of non-cancer death. From an epidemiologic perspective or in a health economic analysis, however, the absolute risk may be more relevant, as it provides an estimate of the population rate of cancer recurrence.

If both the conditional and absolute risk estimates are computed, the CSH method absolute risk estimate will be less than or equal to the conditional risk estimate using the same covariates for the event of interest (or, at most, very slightly higher due to approximation of the continuous baseline cumulative hazard with the discrete empirical estimate). As shown in Figure 2, the Fine-Gray method based on proportional regression of the subdistribution hazard may occasionally produce absolute risk estimates that are substantially higher than the corresponding conditional risk estimates. When this occurs, the most likely explanation is misspecification of either the proportional hazards model used to estimate the conditional risk or the proportional subdistribution hazards model used to estimate the absolute risk (or both). Diagnostic tests for non-proportional hazards might indicate which model is mis-specified. Although mis-specified models will not cause inconsistency between the conditional and absolute risk estimates using proportional CSH regression, the model assumptions should still be checked. Martingale residual methods (for example, [20]) can be used for the event of interest and each of the competing risks, censoring each on the occurrence of events of other types.

9. Software

A SAS macro that calculates the proportional CSH regression interval estimate of absolute risk using these methods is available in the repository https://mcrager.github.io/SAS-macros/. The calculations are also available in the `R` package cshCIF. Example calculations with the Fine-Gray model used base SAS Version 9.4 with SAS/STAT version 14.1 [25].

10. Conclusions

We have provided closed form, pointwise confidence intervals for the absolute risk (cumulative incidence function at a fixed time point) using proportional cause-specific hazard regression, accommodating stratified cohort sampling designs, external time-dependent covariates and multiple censoring mechanisms. The CSH model allows different covariates for the event of interest and the various competing risks. In contrast to methods using subdistribution hazard regression, the CSH regression model always gives absolute risk estimates that are less than or approximately equal to the corresponding risk estimates conditional on non-occurrence of competing risks obtained by censoring the event of interest on occurrence of a competing risk. Simulation studies indicate that the proportional CSH regression confidence intervals for the CIF computed on the log cumulative hazard scale ("log-log" intervals) or the risk scale ("linear" intervals) have coverage probabilities that approximate nominal levels for small and moderate samples, provided there are at least 10 events per covariate and, with stratified cohort sampling, the ratio of non-events to events is at least 2:1. Software to perform the calculations is freely available.

Conflict of Interest

The authors declare that they have no conflicting interest.

Data Availability

The data used in the example analyses are published in Andrews and Herzberg (1985).

Acknowledgements

The authors wish to thank Mark Segal, Carl Yoshizawa and Steven Shak for helpful comments and suggestions.

Appendix

Appendix A1. Asymptotic Approximation for the Variance of the Absolute Risk Estimator

Define for $\varepsilon = M, D$ the estimated cause-specific cumulative hazard function estimators

$$\hat{\Lambda}_{\varepsilon}\left(t;\mathrm{z}_{\varepsilon}\right)=\int_{0}^{t}\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon}(s)\right)\mathrm{d}\,\hat{\Lambda}_{0\varepsilon}\left(s\right)=\int_{0}^{t}\frac{\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon}(s)\right)\sum_{i=1}^{n}w_{i}\,\mathrm{d}\,N_{\varepsilon i}(s)}{\sum_{j=1}^{n}w_{j}Y_{\varepsilon j}(s)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon j}(s)\right)},$$

and the hazard function estimators

$$\mathrm{d}\,\hat{\Lambda}_{\varepsilon}\left(t;\mathrm{z}_{\varepsilon}\right)=\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon}(t)\right)\mathrm{d}\,\hat{\Lambda}_{0\varepsilon}\left(t;\mathrm{z}_{\varepsilon}\right)=\frac{\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon}(t)\right)\sum_{i=1}^{n}w_{i}\,\mathrm{d}\,N_{\varepsilon i}(t)}{\sum_{j=1}^{n}w_{j}Y_{\varepsilon j}(t)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon j}(t)\right)}.$$

Define also

$$\xi_{\varepsilon}(t)=\frac{\sum_{i=1}^{n}w_{i}Y_{\varepsilon i}(t)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon i}(t)\right)\mathrm{z}_{\varepsilon i}(t)}{\sum_{i=1}^{n}w_{i}Y_{\varepsilon i}(t)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon i}(t)\right)}.$$

This is a consistent estimator of the expected value of the covariate vector given the occurrence of an event [23].

The gradient of the baseline cumulative hazard estimator with respect to the regression parameter estimate vector is

$$\nabla_{\hat{\beta}_{\varepsilon}}\,\mathrm{d}\,\hat{\Lambda}_{0\varepsilon}\left(s\right)=\nabla_{\hat{\beta}_{\varepsilon}}\,\frac{\sum_{i=1}^{n}w_{i}\,\mathrm{d}\,N_{\varepsilon i}(t)}{\sum_{j=1}^{n}w_{j}Y_{\varepsilon j}(t)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon j}(t)\right)}=-\frac{\left\{\sum_{j=1}^{n}w_{j}Y_{\varepsilon j}(t)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon j}(t)\right)\mathrm{z}_{\varepsilon j}(t)\right\}\sum_{i=1}^{n}w_{i}\,\mathrm{d}\,N_{\varepsilon i}(t)}{\left\{\sum_{j=1}^{n}w_{j}Y_{\varepsilon j}(t)\exp\left(\hat{\beta}_{\varepsilon}^{\mathrm{T}}\mathrm{z}_{\varepsilon j}(t)\right)\right\}^{2}}$$

$$=-\xi_{\varepsilon}(t)\,\mathrm{d}\,\hat{\Lambda}_{0\varepsilon}\left(s\right).$$

We first account for the variability in the regression parameter estimates. The gradient of $\hat{S}_{M}\left(t;\mathrm{z}_{M}\right)$ with respect to the parameter estimate $\hat{\beta}_{M}$ is

$$\nabla_{\hat{\beta}_{M}}\hat{S}_{M}\left(t;\mathrm{z}_{M}\right)=\nabla_{\hat{\beta}_{M}}\,\exp\left\{-\int_{0}^{t}\exp\left(\hat{\beta}_{M}^{\mathrm{T}}\mathrm{z}_{M}(s)\right)\mathrm{d}\,\hat{\Lambda}_{0M}\left(s\right)\right\}$$

$$=-\left[\exp\left\{-\int_{0}^{t}\exp\left(\hat{\beta}_{M}^{\mathrm{T}}\mathrm{z}_{M}(s)\right)\mathrm{d}\,\hat{\Lambda}_{0M}\left(s\right)\right\}\right]\left\{\int_{0}^{t}\exp\left(\hat{\beta}_{M}^{\mathrm{T}}\mathrm{z}_{M}(s)\right)\mathrm{z}_{M}(s)\,\mathrm{d}\,\hat{\Lambda}_{0M}\left(s\right)+\int_{0}^{t}\exp\left(\hat{\beta}_{M}^{\mathrm{T}}\mathrm{z}_{M}(s)\right)\nabla_{\hat{\beta}_{M}}\,\mathrm{d}\,\hat{\Lambda}_{0M}\left(s\right)\right\}$$

$$=-\hat{S}_{M}\left(t;\mathrm{z}_{M}\right)\left\{\int_{0}^{t}\exp\left(\hat{\beta}_{M}^{\mathrm{T}}\mathrm{z}_{M}(s)\right)\mathrm{z}_{M}(s)\,\mathrm{d}\,\hat{\Lambda}_{0M}\left(s\right)\;-\int_{0}^{t}\exp\left(\hat{\beta}_{M}^{\mathrm{T}}\mathrm{z}_{M}(s)\right)\xi_{M}(s)\,\mathrm{d}\,\hat{\Lambda}_{0M}\left(s\right)\right\}$$

$$=-\hat{S}_{M}\left(t;\mathrm{z}_{M}\right)\psi_{M}\left(t;\mathrm{z}_{M}\right),$$

where

$$\psi_{M}\left(t;\mathrm{z}_{M}\right)=\int_{0}^{t}\exp\left(\hat{\beta}_{M}^{\mathrm{T}}\mathrm{z}_{M}(s)\right)\left\{\mathrm{z}_{M}(s)-\xi_{M}(s)\right\}\mathrm{d}\,\hat{\Lambda}_{0M}\left(s\right)\;=\int_{0}^{t}\left\{\mathrm{z}_{M}(s)-\xi_{M}(s)\right\}\mathrm{d}\,\hat{\Lambda}_{M}\left(s;\mathrm{z}_{M}\right).$$

We can pass the gradient under the integral sign because the integrand is continuously differentiable (or alternately, because the Stieltjes integrals actually represent finite sums).

Using the same methods for the *D* event type, we have

$$\nabla_{\hat{\beta}_{D}}\hat{S}_{D}\left(t;\mathrm{z}_{D}\right)=-\hat{S}_{D}\left(t;\mathrm{z}_{D}\right)\psi_{D}\left(t;\mathrm{z}_{D}\right),$$

where

$$\psi_D(t;\mathrm{z}_D)=\int_0^t\{\mathrm{z}_D(s)-\xi_D(s)\}\,\mathrm{d}\hat{\Lambda}_D(s;\mathrm{z}_D).$$

The gradient of $\hat{CIF}_M(t_0;\mathrm{z})$, the absolute risk estimator at time t_0 with covariate vector function z, with respect to $\hat{\beta}_M$ is thus

$$\nabla_{\hat{\beta}_M}\hat{CIF}_M(t_0;\mathrm{z})=\int_0^{t_0}\hat{S}_D(t;\mathrm{z})\Big[\Big\{\nabla_{\hat{\beta}_M}\hat{S}_M(t;\mathrm{z}_M)\Big\}\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)$$
$$+\hat{S}_M(t;\mathrm{z}_M)\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)\mathrm{z}_M(t)\Big]\mathrm{d}\hat{\Lambda}_{0M}(t)+\int_0^{t_0}\hat{S}(t;\mathrm{z})\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)\nabla_{\hat{\beta}_M}\,\mathrm{d}\hat{\Lambda}_{0M}(t)$$

$$=\int_0^{t_0}\hat{S}_D(t;\mathrm{z}_D)\Big[\Big\{-\hat{S}_M(t;\mathrm{z}_M)\psi_M(t;\mathrm{z})\Big\}\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)+\hat{S}_M(t;\mathrm{z}_M)\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)\mathrm{z}_M(t)\Big]\mathrm{d}\hat{\Lambda}_{0M}(t)$$
$$-\int_0^{t_0}\hat{S}(t;\mathrm{z})\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)\xi_M(t)\,\mathrm{d}\hat{\Lambda}_{0M}(t)$$

$$=\int_0^{t_0}\hat{S}(t;\mathrm{z})\big\{\mathrm{z}_M(t)-\xi_M(t)-\psi_M(t;\mathrm{z}_M)\big\}\,\mathrm{d}\hat{\Lambda}_M(t;\mathrm{z}_M).$$

The gradient of $\hat{CIF}_M(T;\mathrm{z})$ with respect to $\hat{\beta}_D$ is

$$\nabla_{\hat{\beta}_D}\hat{CIF}_M(T;\mathrm{z})=\int_0^{t_0}\hat{S}_M(t;\mathrm{z}_M)\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)\Big\{\nabla_{\hat{\beta}_D}\hat{S}_D(t;\mathrm{z}_D)\Big\}\,\mathrm{d}\hat{\Lambda}_{0M}(t)$$

$$=\int_0^{t_0}\hat{S}_M(t;\mathrm{z}_M)\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t)\right)\Big\{-\hat{S}_D(t;\mathrm{z}_D)\psi_D(t;\mathrm{z}_D)\Big\}\,\mathrm{d}\hat{\Lambda}_{0M}(t)\;=-\int_0^{t_0}\hat{S}(t;\mathrm{z})\psi_D(t;\mathrm{z}_D)\,\mathrm{d}\hat{\Lambda}_M(t;\mathrm{z}_M).$$

These gradients will be combined with the estimated covariance matrices of the regression parameter estimates to capture the contribution of the variances of the regression parameter vector estimators to the variance of the cumulative incidence function.

Next, we account for the variability in the number of jumps $\mathrm{d}N_{Mi}(t_{Mi})$ in the cumulative hazard function estimate for event *M*. Since these jumps are based on a random sample of subjects, the $\mathrm{d}N_{Mi}(t_{Mi})$ are independent random variables. We can locally approximate the counting process $\bar{N}_M(t)=\int_0^t\sum_{i=1}^n\mathrm{d}N_{Mi}(s)$ by a Poisson process, for which the mean and variance are equal, so the method of moments estimate of both the mean and variance is $\mathrm{d}N_{Mi}(t_{Mi})$ [14]. Hence, a consistent estimate for this contribution to the variance of the CIF is

$$\sum_{i=1}^n\left\{\frac{\partial\hat{CIF}_M(t_0;\mathrm{z})}{\partial\,\mathrm{d}N_{Mi}(t_{Mi})}\right\}^2\hat{\mathrm{Var}}\big\{\mathrm{d}N_{Mi}(t_{Mi})\big\}=\sum_{i=1}^n\left\{\frac{\partial\hat{CIF}_M(t_0;\mathrm{z})}{\partial\,\mathrm{d}N_{Mi}(t_{Mi})}\right\}^2\mathrm{d}N_{Mi}(t_{Mi}).$$

Now for all i for which there is a *M* event, that is, for which $\mathrm{d}N_{Mi}(t_{Mi})=1$,

$$\frac{\partial\hat{\Lambda}_M(t;\mathrm{z}_M)}{\partial\,\mathrm{d}N_{Mi}(t_{Mi})}=I_{\{t_{Mi}\le t\}}\frac{w_i\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_M(t_{Mi})\right)}{\sum_{j=1}^n w_jY_{Mj}(t_{Mi})\exp\left(\hat{\beta}_M^{\mathrm{T}}\mathrm{z}_{Mj}(t_{Mi})\right)}=I_{\{t_{Mi}\le t\}}\,\mathrm{d}\hat{\Lambda}_M(t_{Mi};\mathrm{z}_M),$$

where I_S denotes the indicator function for the set *S*, and similarly,

$$\frac{\partial\,\mathrm{d}\hat{\Lambda}_M(t;\mathrm{z}_M)}{\partial\,\mathrm{d}N_{Mi}(t_{Mi})}=I_{\{t_{Mi}=t\}}\,\mathrm{d}\hat{\Lambda}_M(t_{Mi};\mathrm{z}_M)=I_{\{t_{Mi}=t\}}\,\mathrm{d}\hat{\Lambda}_M(t;\mathrm{z}_M).$$

Hence

$$\frac{\partial\hat{CIF}_M(t_0;\mathrm{z})}{\partial\,\mathrm{d}N_{Mi}(t_{Mi})}=\frac{\partial}{\partial\,\mathrm{d}N_{Mi}(t_{Mi})}\int_0^{t_0}\hat{S}_D(t;\mathrm{z}_D)\hat{S}_M(t;\mathrm{z}_M)\,\mathrm{d}\hat{\Lambda}_M(t;\mathrm{z}_M)$$

$$= \int_0^{t_0} \hat{S}_D\left(t;\mathrm{z}_D\right)\left[(-1)\exp\left\{-\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)\right\}\frac{\partial\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)}{\partial \mathrm{d}N_{Mi}\left(t_{Mi}\right)}\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)+\exp\left\{-\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)\right\}\frac{\partial \mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)}{\partial \mathrm{d}N_{Mi}\left(t_{Mi}\right)}\right]$$

$$= \int_0^{t_0} \hat{S}_D\left(t;\mathrm{z}_D\right)\left[(-1)\exp\left\{-\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)\right\}I_{\{t_{Mi}\le t\}}\,\mathrm{d}\hat{\Lambda}_M\left(t_{Mi};\mathrm{z}_M\right)\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)+\exp\left\{-\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)\right\}I_{\{t_{Mi}=t\}}\,\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)\right]$$

$$= \int_{t_{Mi}}^{t_0} \hat{S}_D\left(t;\mathrm{z}_D\right)(-1)\hat{S}_M\left(t;\mathrm{z}_M\right)\mathrm{d}\hat{\Lambda}_M\left(t_{Mi};\mathrm{z}_M\right)\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)+\hat{S}\left(t_{Mi};\mathrm{z}\right)\mathrm{d}\hat{\Lambda}_M\left(t_{Mi};\mathrm{z}_M\right)$$

$$= \hat{S}\left(t_{Mi};\mathrm{z}\right)\mathrm{d}\hat{\Lambda}_M\left(t_{Mi};\mathrm{z}_M\right)-\mathrm{d}\hat{\Lambda}_M\left(t_{Mi};\mathrm{z}_M\right)\int_{t_{Mi}}^{t_0}\hat{S}\left(t;\mathrm{z}\right)\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)$$

$$= \left[\hat{S}\left(t_{Mi};\mathrm{z}\right)-\left\{C\hat{I}F_M\left(T;\mathrm{z}\right)-C\hat{I}F_M\left(t_{Mi}-;\mathrm{z}\right)\right\}\right]\mathrm{d}\hat{\Lambda}_M\left(t_{Mi};\mathrm{z}_M\right).$$

Accounting for the variance of the jumps $\mathrm{d}N_{Di}(t_{Di})$ in the estimated cumulative hazard function for the competing risk event D, for all i for which there is a D event, that is, for which $\mathrm{d}N_{Di}\left(t_{Di}\right)=1$,

$$\frac{\partial\hat{\Lambda}_D\left(t;\mathrm{z}_D\right)}{\partial \mathrm{d}N_{Di}\left(t_{Di}\right)} = I_{\{t_{Di}\le t\}}\frac{w_i\exp\left(\hat{\beta}_D^{\mathrm{T}}\mathrm{z}_D(t_{Di})\right)}{\sum_{j=1}^{n} w_j Y_j(t_{Di})\exp\left(\hat{\beta}_D^{\mathrm{T}}\mathrm{z}_{Dj}(t_{Di})\right)} = I_{\{t_{Di}\le t\}}\,\mathrm{d}\hat{\Lambda}_D\left(t_{Di};\mathrm{z}_D\right),$$

so that

$$\frac{\partial C\hat{I}F_M\left(t_0;\mathrm{z}\right)}{\partial \mathrm{d}N_{Di}\left(t_{Di}\right)} = \frac{\partial}{\partial \mathrm{d}N_{Di}\left(t_{Di}\right)}\int_0^{t_0}\exp\left\{-\hat{\Lambda}_D\left(t;\mathrm{z}_D\right)\right\}\hat{S}_M\left(t;\mathrm{z}_M\right)\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)$$

$$= -\int_0^{t_0}\hat{S}_D\left(t;\mathrm{z}\right)\frac{\partial\hat{\Lambda}_D\left(t;\mathrm{z}_D\right)}{\partial \mathrm{d}N_{Di}\left(t_{Di}\right)}\hat{S}_M\left(t;\mathrm{z}_M\right)\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)$$

$$= -\int_{t_{Di}}^{t_0}\hat{S}\left(t;\mathrm{z}\right)\mathrm{d}\hat{\Lambda}_D\left(t_{Di};\mathrm{z}_D\right)\mathrm{d}\hat{\Lambda}_M\left(t;\mathrm{z}_M\right)$$

$$= -\mathrm{d}\hat{\Lambda}_D\left(t_{Di};\mathrm{z}_D\right)\left\{C\hat{I}F_M\left(t_0;\mathrm{z}\right)-C\hat{I}F_M\left(t_{Di}-;\mathrm{z}\right)\right\}.$$

The M and D counting processes jump at the same time with probability 0, so they are orthogonal [15]. The parameter estimates $\hat{\beta}_M$ and $\hat{\beta}_D$ are asymptotically independent of the jumps $\mathrm{d}N_{Mi}(t_{Mi})$ and $\mathrm{d}N_{Di}(t_{Di})$ [6, 16]. Hence the variance of $C\hat{I}F_M\left(t_0;\mathrm{z}\right)$ is consistently estimated by

$$\hat{\mathrm{Var}}\left\{C\hat{I}F_M\left(t_0;\mathrm{z}\right)\right\} = \left\{\nabla_{\hat{\beta}_M}C\hat{I}F_M\left(t_0;\mathrm{z}_M\right)\right\}^{\mathrm{T}}\hat{\mathrm{V}}_M\left\{\nabla_{\hat{\beta}_M}C\hat{I}F_M\left(t_0;\mathrm{z}_M\right)\right\}+\left\{\nabla_{\hat{\beta}_D}C\hat{I}F_M\left(t_0;\mathrm{z}_D\right)\right\}^{\mathrm{T}}\hat{\mathrm{V}}_D\left\{\nabla_{\hat{\beta}_D}C\hat{I}F_M\left(t_0;\mathrm{z}_D\right)\right\}$$
$$+\sum_{i=1}^{n}\left\{\frac{\partial C\hat{I}F_M\left(t_0;\mathrm{z}\right)}{\partial \mathrm{d}N_{Mi}(t_{Mi})}\right\}^2\mathrm{d}N_{Mi}\left(t_{Mi}\right)+\sum_{i=1}^{n}\left\{\frac{\partial C\hat{I}F_M\left(t_0;\mathrm{z}\right)}{\partial \mathrm{d}N_{Di}(t_{Di})}\right\}^2\mathrm{d}N_{Di}\left(t_{Di}\right),$$

which is equation (4) in the main text. Equations (5)–(12) are derived above.

Appendix A2. Estimating the Time-Dependent Standardized Subdistribution Hazard Ratio from the CSH Model

It is possible to use the CSH model to construct a measure similar to a standardized Fine-Gray subdistribution hazard ratio. In contrast to the Fine-Gray subdistribution hazard ratio, the CSH measure would be time-dependent. In a conventional Cox proportional hazards regression, the absolute standardized log hazard ratio is the standard deviation of the risk score $\hat{\beta}^{\mathrm{T}}\mathrm{z}$ in the study population, where $\hat{\beta}$ is the regression parameter estimate vector, and $\mathbf{z}$ is the covariate vector [24]. Transforming the cumulative incidence function to the log cumulative hazard scale, define the CSH time-averaged absolute standardized subdistribution log

hazard ratio at time t_0 as

$$\bar{B}_{M\sigma}(t_0) = \mathrm{SD}\left(\ln\left[-\ln\left\{1-CIF_M(t_0;z)\right\}\right]\right),$$

where SD denotes the standard deviation. Averaging over time

$$\hat{\bar{B}}_{M\sigma}(t_0) = \sqrt{\frac{\sum_{i=1}^{n} w_i\left(\ln\left[-\ln\left\{1-\hat{CIF}_M(t_0;z_i)\right\}\right]-\hat{\mu}_M(t_0)\right)^2}{\sum_{i=1}^{n} w_i - 1}},$$

where

$$\hat{\mu}_M(t_0) = \frac{\sum_{i=1}^{n} w_i \ln\left[-\ln\left\{1-\hat{CIF}_M(t_0;z_i)\right\}\right]}{\sum_{i=1}^{n} w_i}.$$

If the covariates modulate the hazard for the event of interest and the competing risks similarly, resulting in very little differentiation of risk in the population, then $\hat{\bar{B}}_{M\sigma}(t_0)$ will tend to be close to 0. The estimate of the time-dependent average absolute standardized subdistribution hazard ratio is $\exp\left\{\hat{\bar{B}}_{M\sigma}(t_0)\right\}$.

occurs here even if all the covariates are time-invariant, since the CIF is influenced by the survival curves for the M and D_k event types. A consistent estimator of $\bar{B}_{M\sigma}(t_0)$ is

Appendix A3. Details of Simulation Studies

Simulation scenarios used the following true log hazard ratio vectors and covariate covariance matrices with absolute risk estimated at specified covariate values designated as Case 1, 2, and 3:

I. $\beta_M = \begin{pmatrix} 0.2 \\ -0.3 \\ 0.4 \end{pmatrix}$, $\Sigma_{z_M} = \begin{pmatrix} 1 & -0.5 & 0 \\ -0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\beta_D = (0.2)$, $\Sigma_{z_D} = (1)$

Case 1: $z_M = (1,-0.5,1)^{\mathrm{T}}$, $z_D = (0)$

Case 2: $z_M = (2,1,-1)^{\mathrm{T}}$, $z_D = (1)$

Case 3: $z_M = (0,0,0)^{\mathrm{T}}$, $z_D = (-1)$

II. $\beta_M = \begin{pmatrix} 0.3 \\ 0.3 \\ 0.4 \\ -0.4 \\ 0.5 \\ 0.5 \end{pmatrix}$, $\Sigma_{z_M} = \begin{pmatrix} 1 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -0.6 & 0 & 0 \\ 0 & 0 & -0.6 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0.7 \\ 0 & 0 & 0 & 0 & 0.7 & 1 \end{pmatrix}$, $\beta_D = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}$, $\Sigma_{z_D} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Case 1: $z_M = (1,1.2,0.7,-0.8,0.1,0.2)^{\mathrm{T}}$, $z_D = (1.5,-0.4)^{\mathrm{T}}$

Case 2: $z_M = (1.4,-1.3,0.2,0.4,-0.3,0.2)^{\mathrm{T}}$, $z_D = (0.7,-0.6)^{\mathrm{T}}$

Case 3: $z_M = (0,0,0,0,0,0)^{\mathrm{T}}$, $z_D = (0,0)^{\mathrm{T}}$

III. $\beta_M = (0.4)$, $\Sigma_{z_M} = (1)$, no covariate for D

Case 1: $z_M = -1$

Case 2: $z_M = 0$

Case 3: $z_M = 2$

The results of the simulation studies are given in Figures A1 through A6. Actual coverage probabilities of the 95% confidence intervals are plotted throughout as a function of the total sample size. The left column in each figure shows results using the model-based (Fisher information) variance estimate; the right column shows results using the Lin-Wei (1989) robust variance estimate. The plot at the bottom center of each page shows the mean number of M and D events as a function of the total sample size.

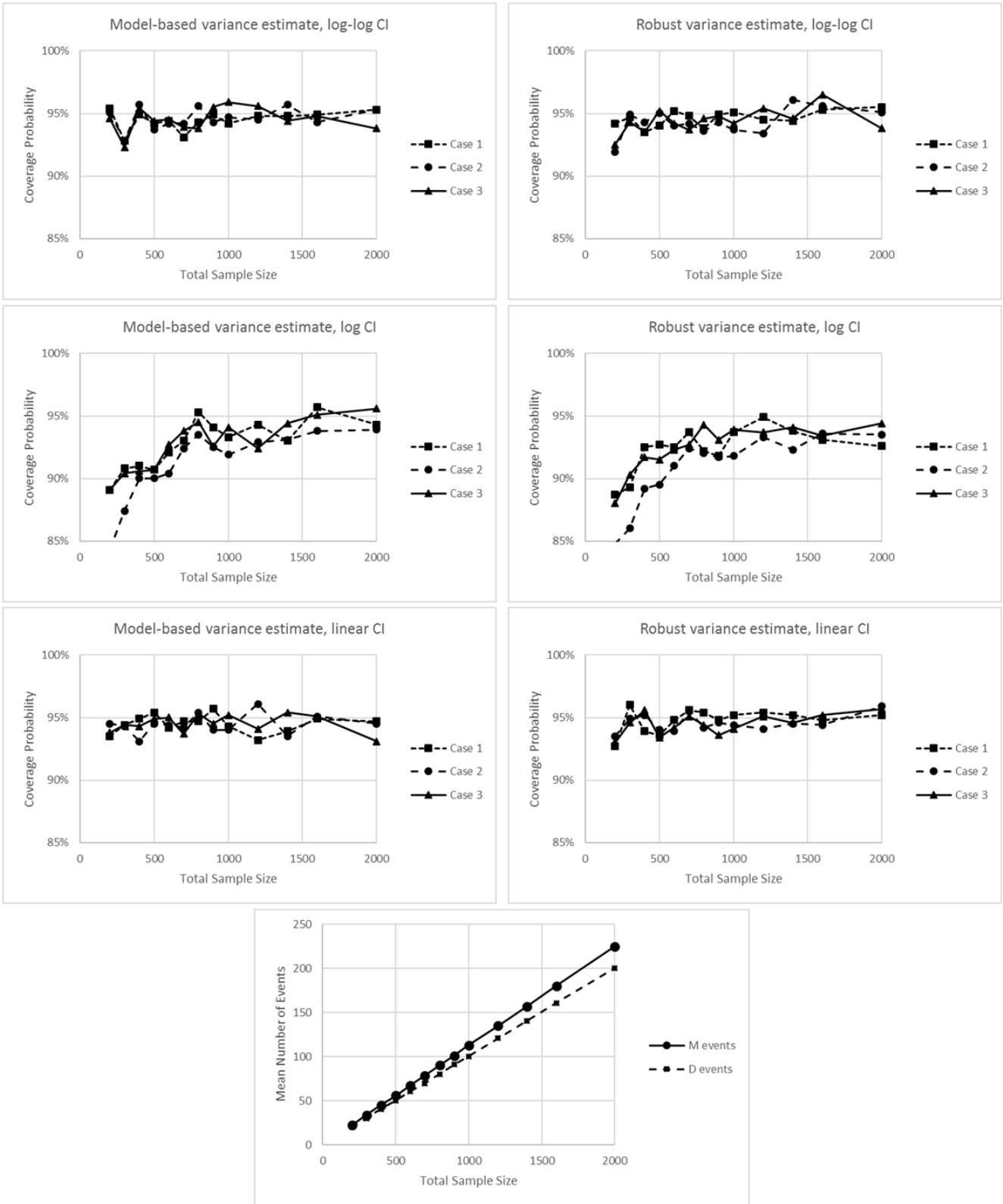

Figure A1. *Coverage probabilities of 95% confidence intervals in simulation studies. Similar expected event counts for M and D. Three (3) covariates for M events, one (1) covariate for D events.*

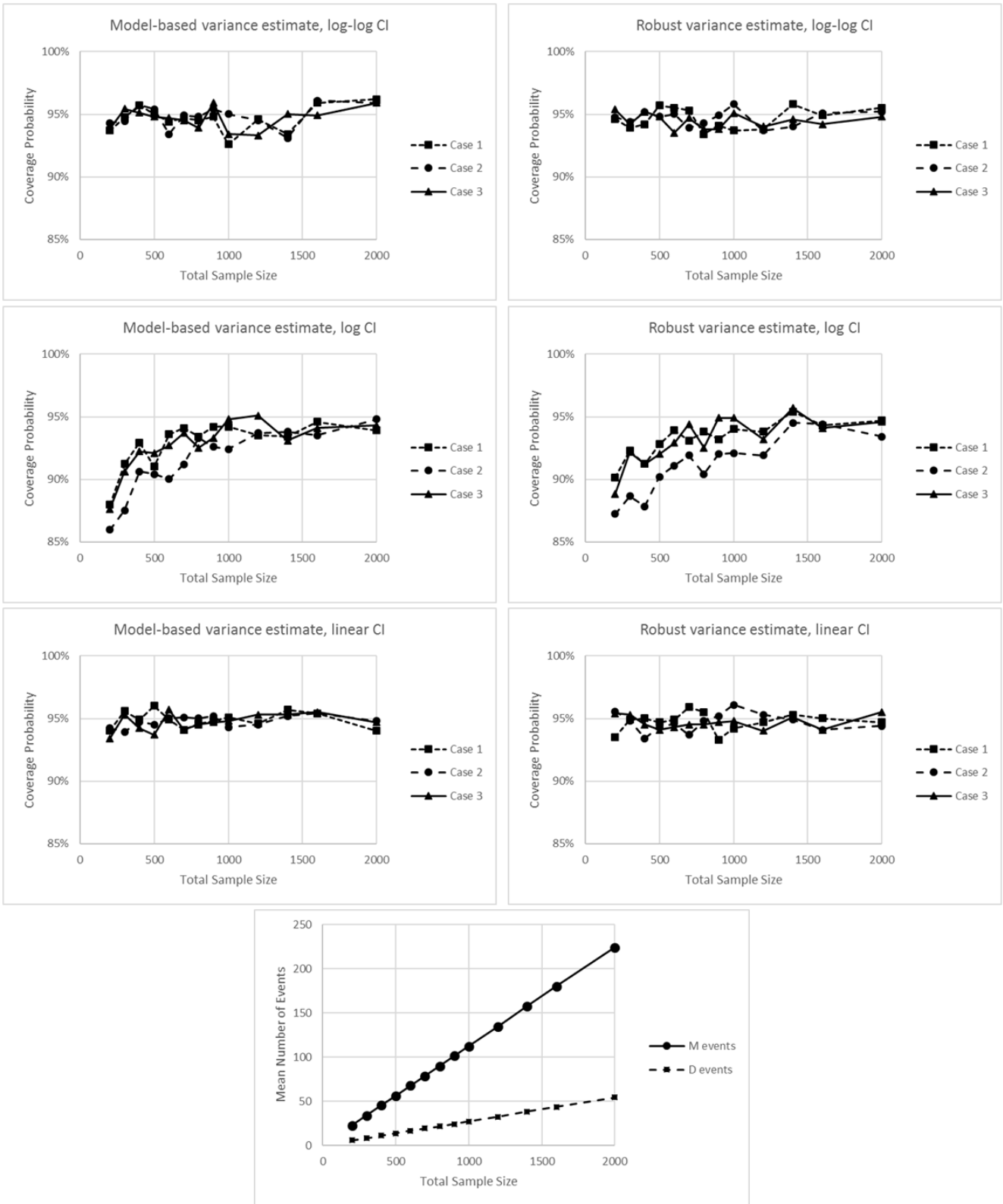

Figure A2. *Coverage probabilities of 95% confidence intervals in simulation studies. D events expected to be less frequent than M events. Three (3) covariates for M events, one (1) covariate for D events.*

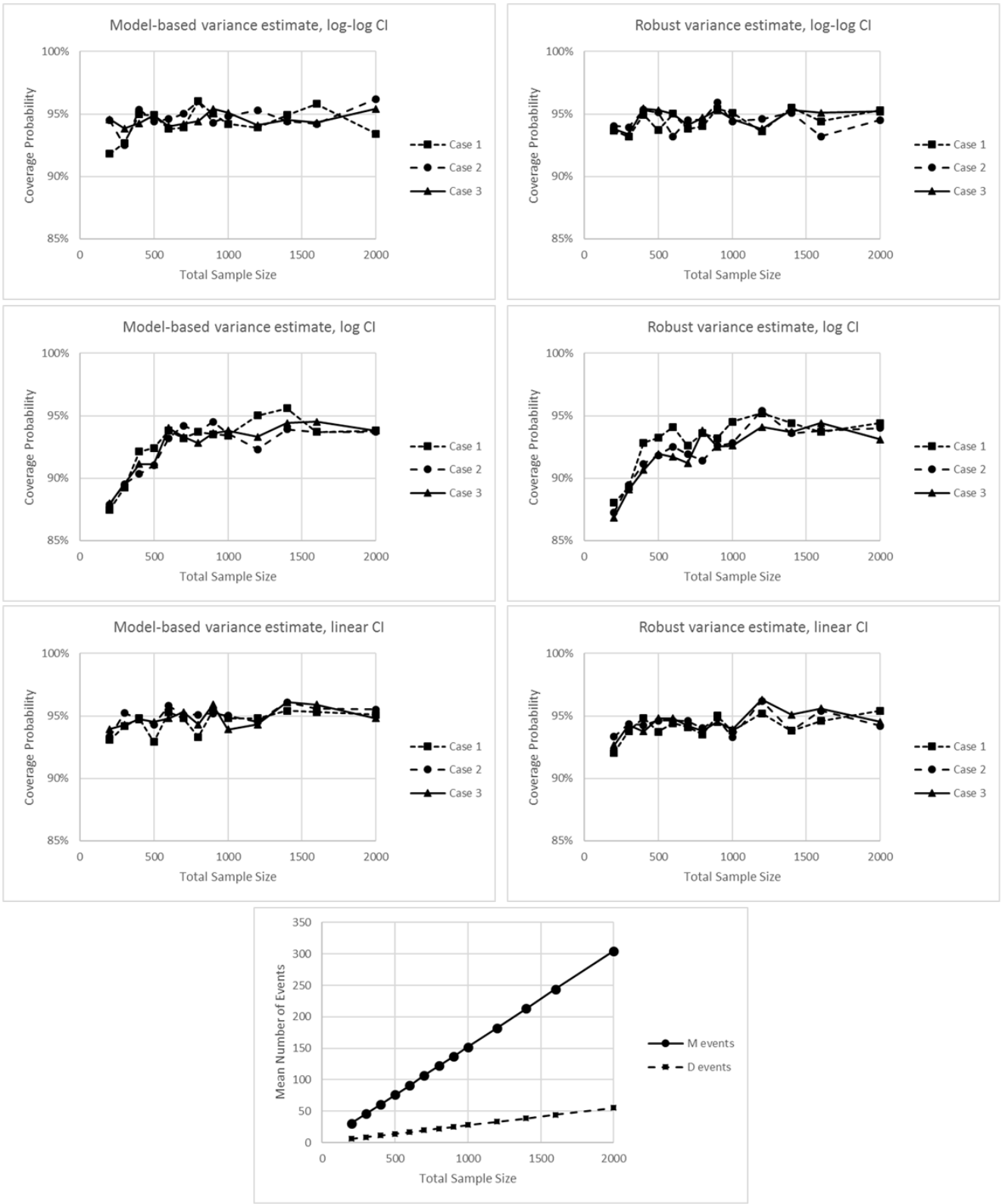

Figure A3. *Coverage probabilities of 95% confidence intervals in simulation studies. D events expected to be less frequent than M events. Six (6) covariates for M events, 2 covariates for D events.*

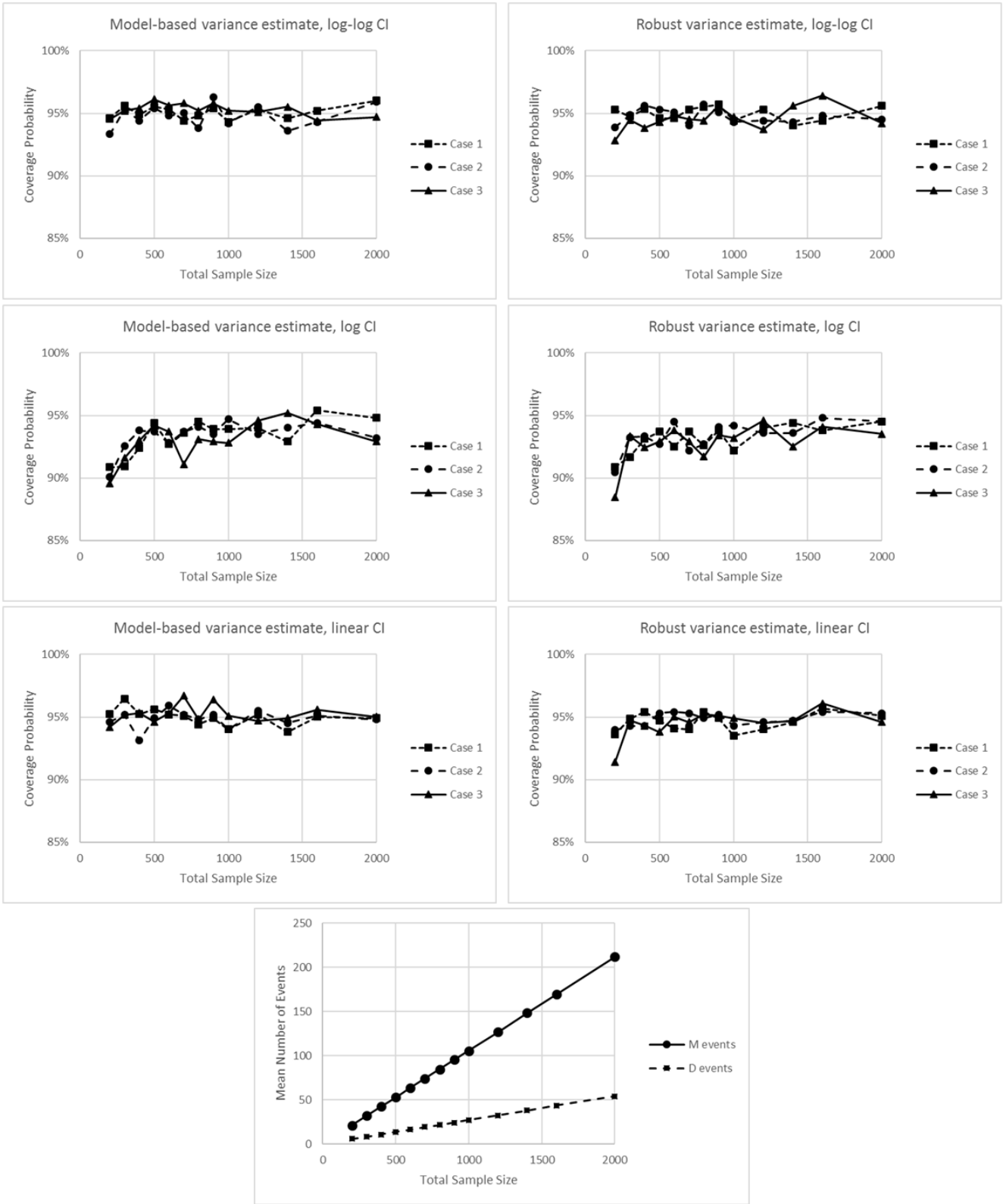

Figure A4. *Coverage probabilities of 95% confidence intervals in simulation studies. D events expected to be less frequent than M events. One (1) covariate for M events, none for D events.*

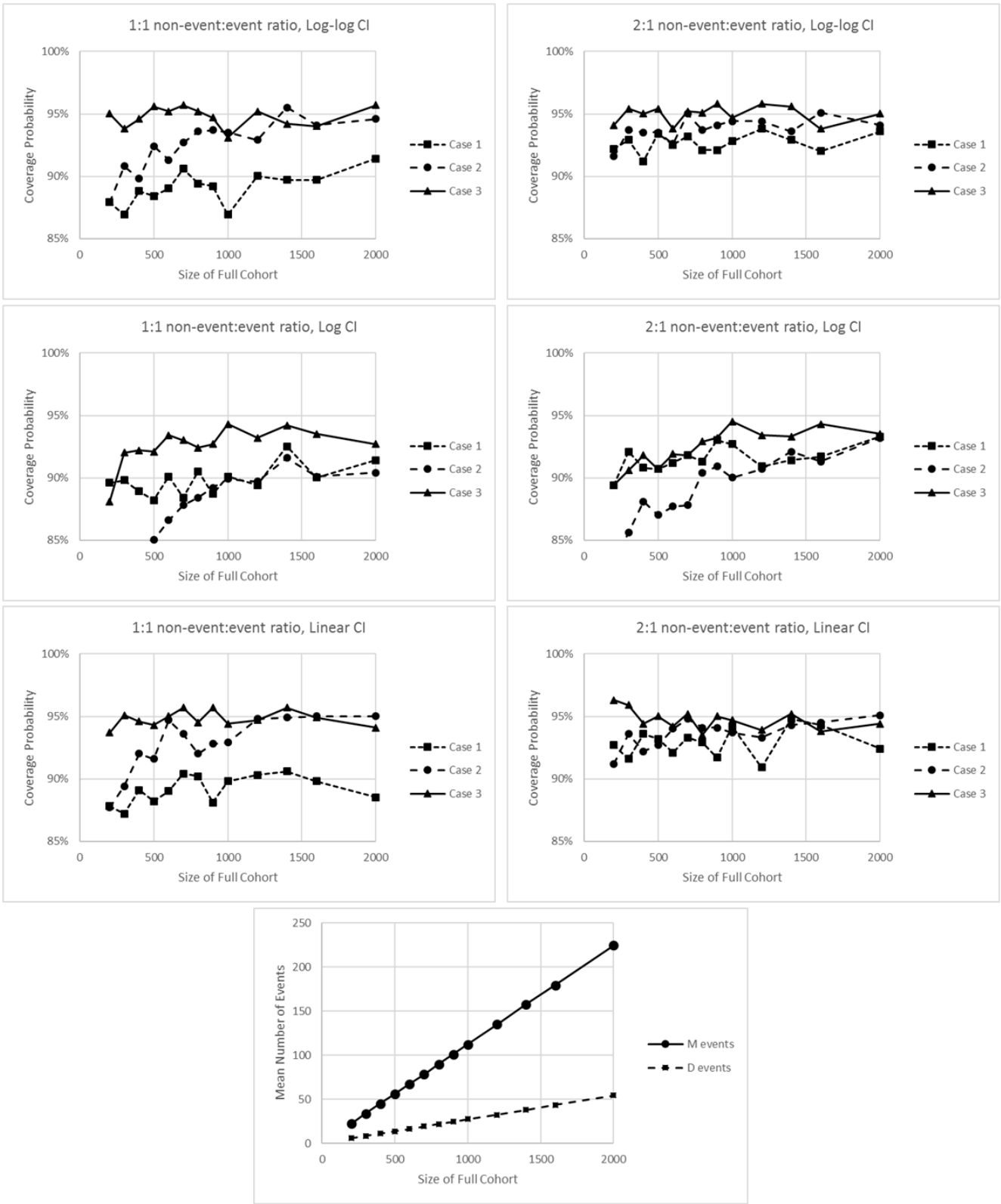

Figure A5. *Coverage probabilities of 95% confidence intervals in simulation studies. D events expected to be less frequent than M events. Three (3) covariates for M events, 1 covariate for D events. Cohort sampling with ratio of non-events to events 1:1 (left column) and 2:1 (right column). The horizontal axis shows the size of the full cohort before sampling.*

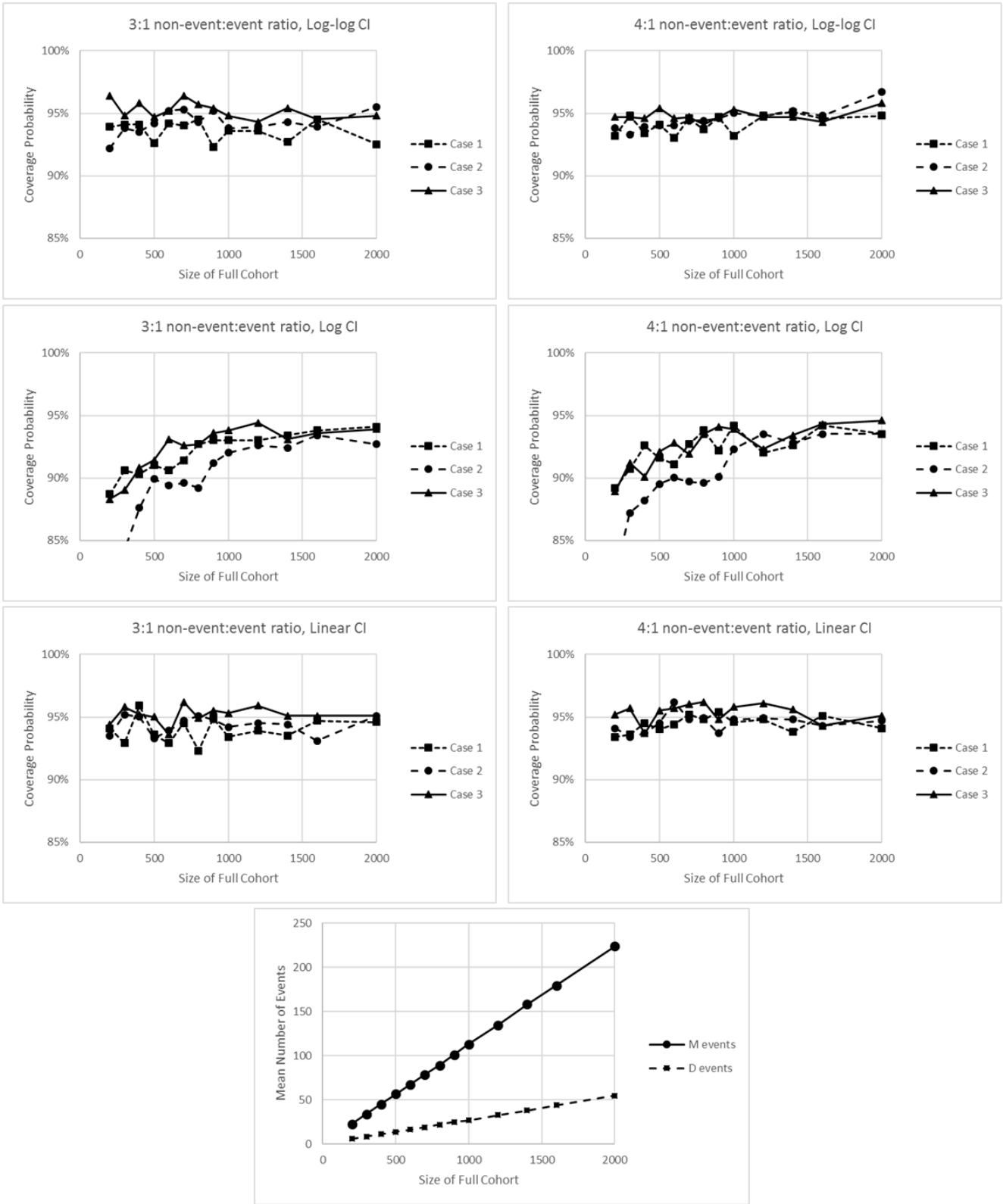

Figure A6. *Coverage probabilities of 95% confidence intervals in simulation studies. D events expected to be less frequent than M events. Three (3) covariates for M events, 1 covariate for D events. Cohort sampling with ratio of non-events to events 3:1 (left column) and 4:1 (right column). The horizontal axis shows the size of the full cohort before sampling.*

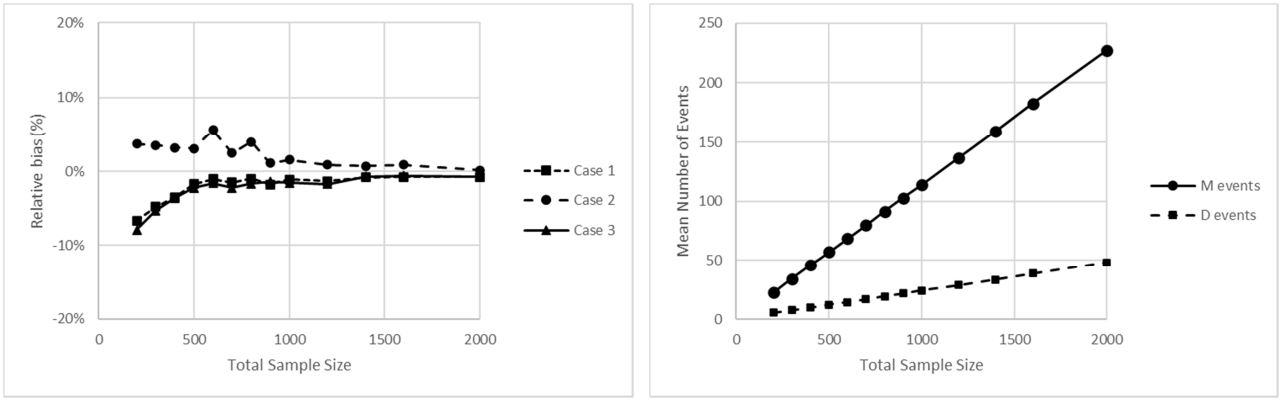

Figure A7. *Relative bias of the point estimate for risk as a percentage of the true value in simulations. D events expected to be less frequent than M events. Three (3) covariates for M events, 1 covariate for D events. The horizontal axis shows the size of the full cohort before sampling.*

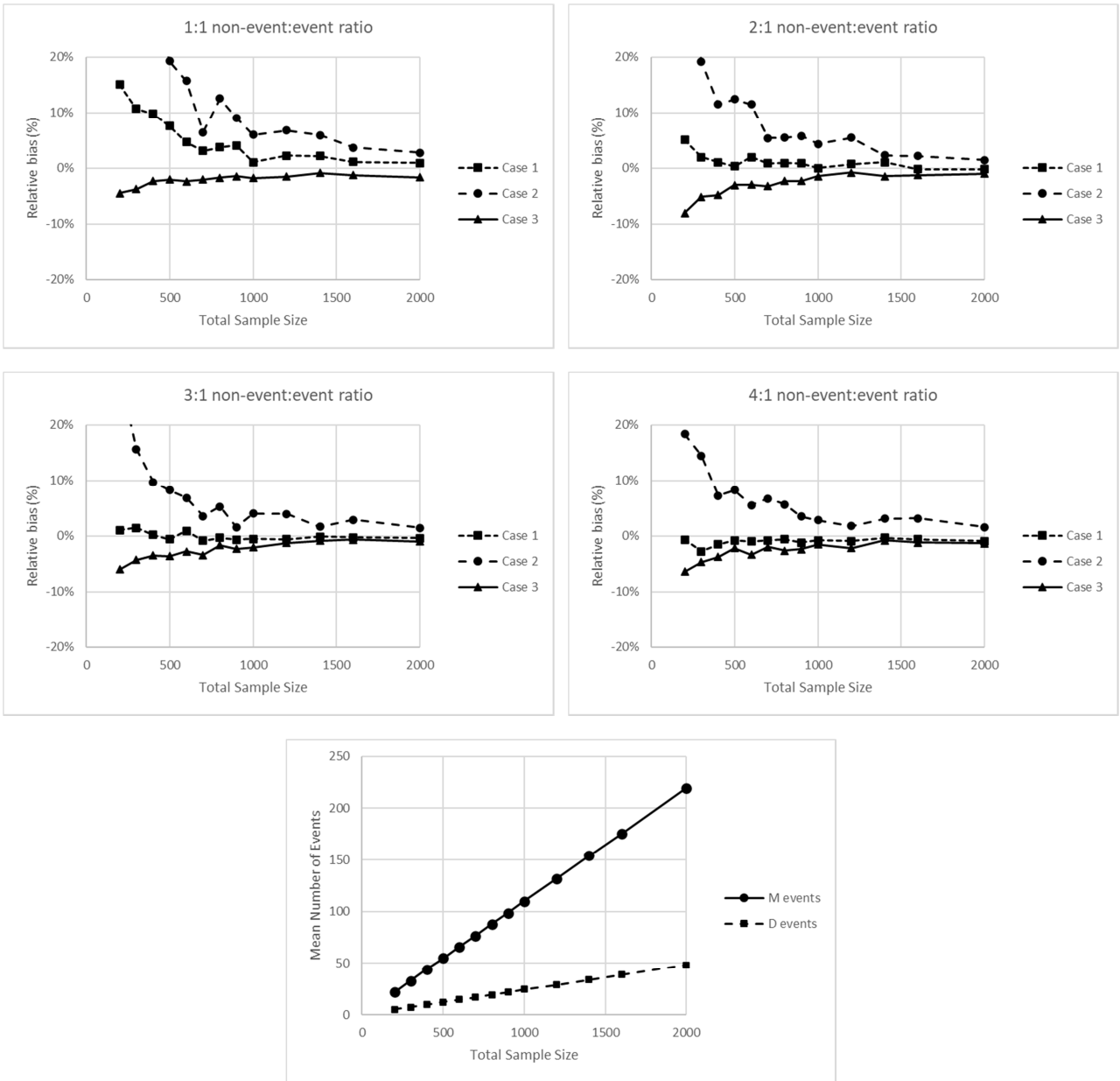

Figure A8. *Relative bias of the point estimate for risk as a percentage of the true value in simulations. D events expected to be less frequent than M events. Three (3) covariates for M events, 1 covariate for D events. Cohort sampling with ratio of non-events to events specified. The horizontal axis shows the size of the full cohort before sampling.*

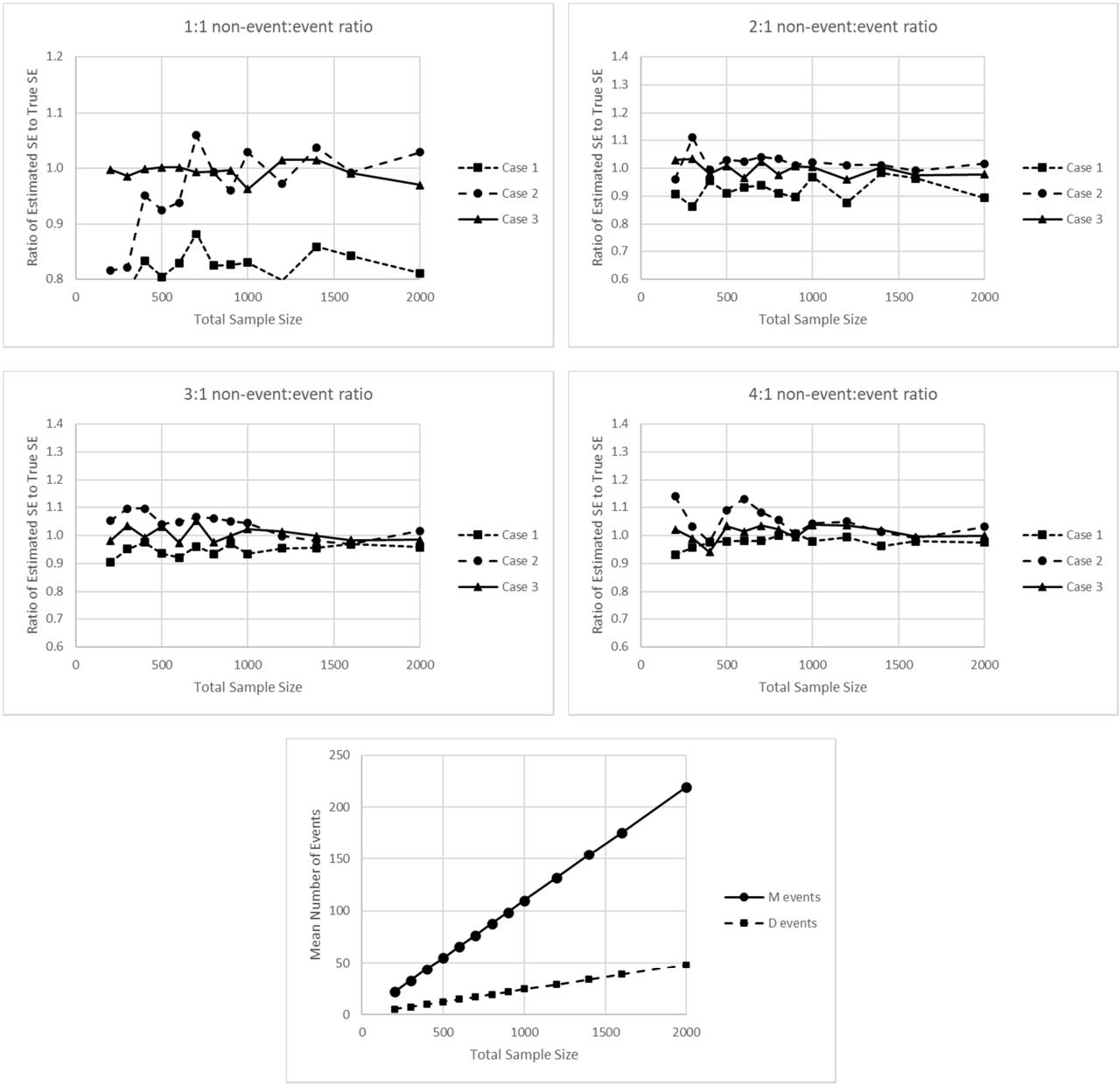

Figure A9. *Ratios of estimated standard errors of the estimate to true standard errors in simulations. D events expected to be less frequent than M events. Three (3) covariates for M events, 1 covariate for D events. Cohort sampling with ratio of non-events to events specified. The horizontal axis shows the size of the full cohort before sampling.*

References

[1] Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. Biometrics, 34, 541–554.

[2] Kalbfleisch, J. D. & Prentice, R. L. (2002). The Statistical Analysis of Failure Time Data, 2nd ed. Wiley, New York. Chapter 8, Competing Risks and Multistate Models.

[3] Gooley, T. A., Leisenring, W., Crowley, J., & Storer, B. E. (1999). Estimation of Failure Probabilities in the Presence of Competing Risks: New Representations of Old Estimators. Statistics in Medicine, 18, 695–706.

[4] Satagopan, J. M., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D., & Auerbach, A. D. (2004). A Note on Competing Risks in Survival Data Analysis. British Journal of Cancer, 91, 1229–1235.

[5] Gray, R. (2009). Weighted Analysis for Cohort Sampling Designs. Lifetime Data Analysis, 15, 24–40.

[6] Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (1993). Statistical Models Based on Counting Processes. Spring-Verlag, New York. Chapter VII: Regression Models.

[7] Rosthøj, S., Andersen, P. K., & Abildstrom, S. Z. (2004). SAS Macros for Estimation of the Cumulative Incidence Functions Based on a Cox Regression. Computer Methods and Programs in Biomedicine, 74, 69–75.

[8] Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in Biostatistics: Competing Risks and Multi-State Models. Statistics in Medicine, 26, 2389–2430.

[9] Benichou, J., & Gail, M. (1990). Estimates of Absolute Cause-Specific Risk in Cohort Studies. Biometrics, 46, 813–826.

[10] Cheng, S. C., Fine, J. P., & Wei, L. J. (1998). Prediction of Cumulative Incidence Function Under the Proportional Hazards Model. Biometrics, 54, 219–228.

[11] Fine, J. P., & Gray, R. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association, 94, 496–509.

[12] Cortese, G., & Andersen, P. K. (2009). Competing Risks and Time-Dependent Covariates. Biometrical Journal, 51: 138–158.

[13] Lin, D. Y., & Wei, L. J. (1989). The Robust Inference for the Cox Proportional Hazards Model. Journal of the American Statistical Association, 84, 1074–1078.

[14] Therneau, T. M., Grambsch and P. M. (2000). Modeling Survival Data: Extending the Cox Model. Wiley, New York.

[15] Bryant, J., & Dignam, J. (2004). Semiparametric Models for Cumulative Incidence Functions. Biometrics, 6, 182–190.

[16] Tsiatis, A. (1981) A Large Sample Study of Cox's Regression Model. The Annals of Statistics 9, 93–108.

[17] Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. Journal of the American Statistical Association, 72, 557–565.

[18] Byar, D. P., & Green, S. B. (1980). The Choice of Treatment for Cancer Patients Based on Covariate Information: Application to Prostate Cancer. Bulletin Cancer, Paris, 67, 477–488.

[19] Andrews, D. F., & Herzberg, A. M. (1985). Data: A Collection of Problems from Many Fields for the Student and Research Worker. Springer-Verlag, New York.

[20] Lin, D. Y., Wei, L. J., & Ying, Z. (1993). Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. Biometrika 80, 557–572.

[21] Geskus, R. B. (2011) Cause-Specific Cumulative Incidence Estimation and the Fine and Gray Model Under Both Left Truncation and Right Censoring. Biometrics, 67, 39–49.

[22] Peduzzi, P., Concato, J., Feinstein, A. R., & Holford R. T. (1995). Importance of Events per Independent Variable in Proportional Hazards Regression Analysis II. Accuracy and precision of regression estimates. Journal of Clinical Epidemiology, 48, 1503–1510.

[23] Xu, R., & O'Quigley, J. (2000). Proportional Hazards Estimate of the Conditional Survival Function. Journal of the Royal Statistical Society, Series B, Methodological, 62, 667–680.

[24] Crager, M. R. (2012). Generalizing The Standardized Hazard Ratio to Multivariate Proportional Hazards Regression, with an Application to Clinical Genomic Studies. Journal of Applied Statistics, 39, 399–417.

[25] SAS Institute Inc. (2015). SAS/STAT® 14.1 User's Guide. SAS Institute, Inc., Cary, North Carolina.