AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the accepted manuscript version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Overgaard, M, Parner, ET, Pedersen, J. Estimating the variance in a pseudo-observation scheme with competing risks. *Scand J Statist.* 2018; 45: 923– 940. <https://doi.org/10.1111/sjos.12328>

Publication metadata

Title:	Estimating the variance in a pseudo-observation scheme with competing risks
Author(s):	M. Overgaard, E.T. Parner, J. Pedersen
Journal:	Scandinavian Journal of Statistics
DOI/Link:	https://doi.org/10.1111/sjos.12328
Document version:	Accepted manuscript (post-print)

This is the peer reviewed version of the following article Overgaard, M, Parner, ET, Pedersen, J. Estimating the variance in a pseudo-observation scheme with competing risks. Scand J Statist. 2018; 45: 923– 940. <https://doi.org/10.1111/sjos.12328> . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

Estimating the variance in a pseudo-observation scheme with competing risks

MORTEN OVERGAARD¹, ERIK THORLUND PARNER¹,
and JAN PEDERSEN²

¹*Department of Public Health, Aarhus University*

²*Department of Mathematics, Aarhus University*

January 23, 2018

Abstract

The Huber–White sandwich variance estimator of the variance of estimates from a regression using jack-knife pseudo-observations from the Aalen–Johansen estimator in a competing risks situation is known to be biased. In this paper, an expression of the asymptotic bias of the Huber–White variance estimator is found, and the Huber–White variance estimator is seen to be biased upwards. An alternative variance estimator, obtained by plugging in empirical values in the true variance expression, is studied by simulation and its performance is

compared to the performance of the biased Huber–White variance estimator. Based on the simulation study, recommendations of its use are given. The alternative variance estimator works well in large samples, but, since it estimates the asymptotic variance, may be less useful on small samples.

Keywords: Aalen–Johansen functional, functional differentiability, generalized estimating equation, jack-knife pseudo-values, time-to-event analysis.

1 Introduction

Regression analyses using jack-knife pseudo-observations from some suitable estimator instead of potentially unobserved outcomes were proposed in Andersen *et al.* (2003). In the competing risks setting, an outcome of interest could be an indication that the person under observation has had a certain event before a certain time, the mean of which would be interpreted as a risk or expected cumulative incidence proportion of that event. This outcome may well be left unobserved by censoring before the time point of interest and before the event of interest. In symbols, we let T denote the time of event and Δ denote the type of event. If t is the time point of interest and event 1 is the event of interest, $1(T \leq t, \Delta = 1)$ is the mentioned outcome. Say, given covariates Z , we model $F_{Z,1}(t) := P(T \leq t, \Delta = 1 \mid Z)$ by $\mu(\beta_0; Z)$ for some suitable function μ and some unknown parameter of interest β_0 . This could be a linear model $\mu(\beta_0; Z) = \beta_0^\top Z$ with β_0 corresponding to the difference in the risk of event 1 per difference in Z . Based on a sample of

size n of independent observations in this setting, the pseudo-observation method, as proposed by Andersen *et al.* (2003), would suggest calculating jack-knife pseudo-observations $\hat{\theta}_{n,i} = n\hat{\theta}_n - (n-1)\hat{\theta}_n^{(i)}$, $i = 1, \dots, n$ based on the Aalen–Johansen estimate of $\theta := F_1(t) := P(T \leq t, \Delta = 1)$ on the overall sample, $\hat{\theta}_n$, and on leave-one-out samples, the $\hat{\theta}_n^{(i)}$ s, and solving an estimating equation like

$$\sum_{i=1}^n A(\beta; Z_i)(\hat{\theta}_{n,i} - \mu(\beta; Z_i)) = 0 \quad (1)$$

to obtain an estimate of β_0 . Here, A is a column vector function of the same dimension as β_0 , for instance $A(\beta; Z) = \frac{\partial}{\partial \beta} \mu(\beta; Z)^\top$, that is, $A(\beta; Z) = Z$ if $\mu(\beta; Z) = \beta^\top Z$. Under certain conditions on the censoring mechanism, this will produce consistent estimates, $\hat{\beta}_n$, if covariates Z and functions μ and A are sufficiently regular as was proven in Overgaard *et al.* (2017), where further details can be found.

The original suggestion by Andersen *et al.* (2003), seemingly backed by the results of Graw *et al.* (2009), was to estimate the variance of the resulting estimates, $\hat{\beta}_n$, by a Huber–White sandwich estimate, here corresponding to $\frac{1}{n} \hat{M}^{-1} \hat{\Sigma} (\hat{M}^{-1})^\top$ with

$$\hat{M} = \frac{1}{n} \sum_{i=1}^n \left(A(\hat{\beta}_n; Z_i) \frac{\partial}{\partial \beta} \mu(\beta; Z_i) \Big|_{\beta=\hat{\beta}_n} \right) \quad (2)$$

and

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n A(\hat{\beta}_n; Z_i) A(\hat{\beta}_n; Z_i)^\top (\hat{\theta}_{n,i} - \mu(\hat{\beta}_n; Z_i))^2. \quad (3)$$

Considering pseudo-observation schemes based on the Kaplan–Meier esti-

mator, Jacobsen & Martinussen (2016) proved that this type of sandwich estimator should be expected to be biased upwards from the asymptotic variance of estimates $\hat{\beta}_n$. In Overgaard *et al.* (2017) it was seen that this type of sandwich estimator should be expected to be biased in more general schemes as well, including schemes based on the Aalen–Johansen estimator for competing risks such as considered here, and a plug-in variance estimator was suggested instead.

Another suggestion by Andersen *et al.* (2003), as an alternative to the Huber–White sandwich variance estimator, was to estimate the variance by the bootstrap method on the given set of pseudo-observations.

In this paper, we apply the results of Overgaard *et al.* (2017) to the competing risks setting by considering the Aalen–Johansen estimator as an inverse probability of censoring weighted estimator and functional, as was done in Overgaard *et al.* (2017) following the idea of Graw *et al.* (2009). This allows us to give an explicit expression of the asymptotic bias of the Huber–White sandwich variance estimator in terms of event time and censoring time distributions and we see that the variance estimate is asymptotically biased upwards. This result generalizes the result of Jacobsen & Martinussen (2016) from the one-cause to the competing risks setting, but, unlike the approach of Jacobsen & Martinussen (2016), builds on the inverse probability weighted form of the estimator. The expression of the asymptotic bias reveals how the bias disappears when there is no censoring or no effect of the covariates before the time point of interest, and suggests that the bias will be small when the amount of censoring and the effect of covariates are low. We present a simulation study with modest sample sizes, but a large amount of

replications, which examines the performance of the variance estimator suggested by Overgaard *et al.* (2017) compared to the Huber–White sandwich variance estimator. The simulation study is in agreement with the conclusion based on the expression of the asymptotic bias by showing that the Huber–White sandwich variance estimator has an inflated coverage probability primarily when the amount of censoring and the effect of covariates are large. The simulation study shows that the variance estimator proposed by Overgaard *et al.* (2017) has an excellent performance in larger samples, but a poor performance in small samples with a large amount of censoring, where estimating the asymptotic variance may be inappropriate. The Huber–White variance estimator does well when the amount of censoring is low or the effect of covariates is low. We also present a smaller simulation study looking into the behavior of the bootstrap variance estimate, which suggests that the bootstrap variance behaves similarly to the Huber–White variance estimate.

In Section 2 we introduce the setting and state the main results, in Section 3 the main results are proven, in Section 4 the plug-in variance estimator is considered, and in Section 5 the results of our simulation study are presented, and we provide a rough guideline on which variance estimate to use. We have some concluding remarks in Section 6.

2 Setting and main results

We have already introduced the event time, $T \in [0, \infty)$, and the event type indicator, $\Delta \in \{1, \dots, d\}$. The observation of (T, Δ) may be prevented by censoring. We assume the censoring mechanism is given by a censoring time

C , such that what we are able to observe of (T, Δ) is $\tilde{T} = T \wedge C \in [0, \infty)$ and $\tilde{\Delta} = \Delta 1(T \leq C) \in \{0, 1, \dots, d\}$. We will denote the pair by $X = (\tilde{T}, \tilde{\Delta})$. The information we have to work with is then observations of (X, Z) with Z being the set of covariates. We let $H_j(s) := P(\tilde{T} \leq s, \tilde{\Delta} = j)$, $j = 0, 1, \dots, d$ be the probability of observing the event of type j , $j = 1, \dots, d$ or censoring, $j = 0$, before time s and let $H(s) := P(\tilde{T} \geq s)$ be the probability of being at risk at time s . These functions relate to the observable $(\tilde{T}, \tilde{\Delta})$ and can be estimated by their empirical versions $\hat{H}_n(s) := \frac{1}{n} \sum_{i=1}^n 1(\tilde{T}_i > s)$ and, for $j = 0, 1, \dots, d$, $\hat{H}_{n,j}(s) := \frac{1}{n} \sum_{i=1}^n 1(\tilde{T} \leq s, \tilde{\Delta} = j)$. If the censoring, C , is assumed to be independent of (T, Δ) , the observable cumulative censoring hazard, given by $\Lambda_0(s) := \int_0^s \frac{1}{H(u)} dH_0(u)$, equals the actual cumulative censoring hazard and the equation $G(s) = \prod_0^s (1 - d\Lambda_0(u))$ holds, where $G(s) = P(C > s)$ and $\prod$ is the product integral, see the appendix. Natural estimators are therefore the Nelson-Aalen type $\hat{\Lambda}_{n,0}(s) := \int_0^s \frac{1}{\hat{H}_n(u)} d\hat{H}_{n,0}(u)$ and the Kaplan-Meier type $\hat{G}_n(s) := \prod_0^s (1 - d\hat{\Lambda}_{n,0}(u))$. Under the independent censorings assumption, we also have $H_1(s) = \int_0^s G(u-) dF_1(u)$, where $F_1(s) = P(T \leq s, \Delta = 1)$ is the probability of having event 1 before time s , and thus $F_1(s) = \int_0^s \frac{1}{G(u-)} dH_1(u)$ when $G(s-) > 0$. This suggests the estimator $\hat{F}_{n,1}(s) := \int_0^s \frac{1}{\hat{G}_n(u-)} d\hat{H}_{n,1}(u)$, which is in fact the Aalen-Johansen estimator of $F_1(s)$ in its inverse probability of censoring weighting form, which was used in the study of pseudo-observations by Overgaard *et al.* (2017) based on the idea by Graw *et al.* (2009), who referenced Satten & Datta (2001) and Jewell *et al.* (2007) for the expression.

So, we have seen how F_1 can be obtained using H , H_0 and H_1 and that the Aalen-Johansen estimator is obtained by substituting these functions

for empirical versions $\hat{H}_n$, $\hat{H}_{n,0}$ and $\hat{H}_{n,1}$. In Overgaard *et al.* (2017), this substitution was formalized and an Aalen–Johansen functional was introduced to study the asymptotics of the pseudo-observation method. For a triplet $f = (f_*, f_0, f_1)$ of suitable functions, let $\psi(s; f) = \int_0^s \frac{1}{f_*(u)} df_0(u)$ and $\chi(s; f) = \prod_{[0,s)} (1 - \psi(du; f))$. Then an Aalen–Johansen functional is given by

$$\phi(s; f) = \int_0^s \frac{1}{\chi(u; f)} df_1(u). \quad (4)$$

With $F = (H, H_0, H_1)$ and $F_n = (\hat{H}_n, \hat{H}_{n,0}, \hat{H}_{n,1})$ we see that $\phi(s; F) = F_1(s)$ and $\phi(s; F_n) = \hat{F}_{n,1}(s)$. A first order derivative at f in direction g is defined by $\phi'_f(s; g) := \frac{d}{du} \phi(s; f + ug) \Big|_{u=0}$ and a second order derivative at f in directions g and h is similarly defined by $\phi''_f(s; g, h) := \frac{d^2}{du dv} \phi(s; f + ug + vh) \Big|_{u=0, v=0}$.

The first order influence function is defined by $\dot{\phi}(s; x) := \phi'_F(s; \delta_x - F)$ and similarly the second order influence function is defined by $\ddot{\phi}(s; x_1, x_2) := \phi''_F(s; \delta_{x_1} - F, \delta_{x_2} - F)$, where δ_x is the Dirac- x measure version of F for an observation $x = (\tilde{t}, \tilde{\delta})$, i.e. $\delta_x = (Y_x, N_{x,0}, N_{x,1})$, where $Y_x(s) = 1(\tilde{t} \geq s)$ and $N_{x,j}(s) = 1(\tilde{t} \leq s, \tilde{\delta} = j)$ for $j = 0, 1$.

In addition to certain regularity conditions, Overgaard *et al.* (2017) assumed completely independent censorings, that is $C \perp\!\!\!\perp (T, \Delta, Z)$, continuity of H , H_0 and H_1 , and positivity $H(t) > 0$, such that also $H(s) > 0$ up to the time point of interest, t . The result of Overgaard *et al.* (2017) was that when the estimating equation in (1) is considered with pseudo-observations $\hat{\theta}_{n,i}$ based on the Aalen–Johansen estimate of $F_1(t)$ for a time point of interest, t , the resulting estimate $\hat{\beta}_n$ will be reasonable in the sense that $\sqrt{n}(\hat{\beta}_n - \beta_0)$ will asymptotically follow a normal distribution with mean zero and a variance

of $M^{-1}\Sigma(M^{-1})^\top$, where

$$M = \mathbb{E} \left(A(\beta_0; Z) \frac{\partial}{\partial \beta} \mu(\beta; Z) \Big|_{\beta=\beta_0} \right), \quad (5)$$

$$\Sigma = \text{Var} (h_0(X, Z) + h_1(X)), \quad (6)$$

with

$$h_0(x, z) = A(\beta_0; z)(F_1(t) + \dot{\phi}(t; x) - \mu(\beta_0; z)), \quad (7)$$

$$h_1(x) = \mathbb{E} \left(A(\beta_0; Z) \ddot{\phi}(t; X, x) \right). \quad (8)$$

The estimator of M in (2) is consistent, but the Huber–White estimator of Σ in (3) estimates the variance of $h_0(X, Z)$ consistently and thus cannot in general be expected to estimate Σ consistently since $h_1(X)$ is not taken into account. This is the case since $r_{n,i} := \hat{\theta}_{n,i} - (F_1(t) + \dot{\phi}(t; X_i))$ converges to 0 in probability uniformly in i for $n \rightarrow \infty$, which can be established by the results of Overgaard *et al.* (2017).

If we define a vector function, W , by

$$W(s) = \mathbb{E} \left(A(\beta_0; Z) \left((F_{Z,1}(t) - F_{Z,1}(s)) - \frac{H_Z(s)}{H(s)} (F_1(t) - F_1(s)) \right) \right), \quad (9)$$

where $H_Z(s) := \mathbb{P}(\tilde{T} > s \mid Z)$, we have the following results under the assumptions mentioned.

Theorem 1. *The variance of $h_1(X)$ is*

$$\text{Var}(h_1(X)) = \int_0^t W(s)W(s)^\top \frac{1}{H(s)} d\Lambda_0(s) \quad (10)$$

and is thus 0 if and only if $W(s) = 0$ for Λ_0 -almost all $s \in [0, t]$.

Theorem 2. *We have that*

$$\text{Cov}(h_0(X, Z), h_1(X)) = -\text{Var}(h_1(X)) \quad (11)$$

and consequently

$$\Sigma = \text{Var}(h_0(X, Z)) - \text{Var}(h_1(X)). \quad (12)$$

The main implication of Theorem 1 and Theorem 2 is that, in matrix terms, $\Sigma \leq \text{Var}(h_0(X, Z))$ with equality if and only if $W(s) = 0$ for Λ_0 -almost all $s \in [0, t]$. This happens for instance when $F_{Z,1} = F_1$ and $H_Z = H$, i.e. when Z has no effect of interest, or if there is no risk of censoring. Specifically, the same inequality holds for each of the pairs of variances on the diagonal of the matrices, i.e. for entry (k, k) we have $\Sigma_{kk} \leq \text{Var}(h_0(X, Z))_{kk}$ with equality if and only if $W(s)_k = 0$ for Λ_0 -almost all $s \in [0, t]$. Furthermore, (10) gives us an expression of the difference between $\text{Var}(h_0(X, Z))$ and Σ , which is the expected bias of the Huber–White estimator of Σ . In the one-cause, Kaplan–Meier case, the last statement in our Theorem 1 reduces to the statement in Theorem 1 of Jacobsen & Martinussen (2016).

As a referent to the expression of the bias in (10), we have also obtained an expression of $\text{Var}(h_0(X, Z))$.

Theorem 3. *The variance of $h_0(X, Z)$ is*

$$\begin{aligned} \mathbb{E} \Big(A(\beta_0; Z) A(\beta_0; Z)^\top & \Big(\int_0^t \frac{1}{G(s)} dF_{Z,1}(s) - F_{Z,1}(t)^2 \\ & + \int_0^t (F_1(t) - F_1(s))^2 \frac{H_Z(s)}{H(s)} \frac{1}{H(s)} d\Lambda_0(s) \\ & - 2 \int_0^t (F_1(t) - F_1(s))(F_{Z,1}(t) - F_{Z,1}(s)) \frac{1}{H(s)} d\Lambda_0(s) \Big) \Big). \end{aligned} \quad (13)$$

The proofs of these theorems follow in the next section.

3 Proofs

As a first step, let us express what the first and second order derivative of the Aalen–Johansen functional look like. The first order derivative can be expressed by

$$\begin{aligned} \phi'_f(s; g) &= \int_0^s \frac{1}{\chi(u; f)} dg_1(u) \\ &+ \int_0^s \frac{1}{\chi(u; f)} \int_{[0, u]} \frac{1}{1 - \Delta\psi(v; f)} \psi'_f(dv; g) df_1(u), \end{aligned} \quad (14)$$

where $\psi'_f(v; g)$ is similarly the derivative of $\psi(v; f)$ in direction g and is given by

$$\psi'_f(s; g) = \int_0^s \frac{1}{f_*(u)} dg_0(u) - \int_0^s \frac{g_*(u)}{f_*(u)^2} df_0(u). \quad (15)$$

Similar expressions can be found in Overgaard *et al.* (2017) and we will not go into details on derivation. With the results on the product integral from the appendix, the expressions can be obtained by the chain rule. The second

order derivative is given by

$$\begin{aligned}
\phi_f''(s; g, h) = & \int_0^s \frac{1}{\chi(u; f)} \int_{[0, u)} \frac{1}{1 - \Delta\psi(v; f)} \psi_f'(\mathrm{d}v; h) \mathrm{d}g_1(u) \\
& + \int_0^s \frac{1}{\chi(u; f)} \int_{[0, u)} \frac{1}{1 - \Delta\psi(v; f)} \psi_f'(\mathrm{d}v; g) \mathrm{d}h_1(u) \\
& + \int_0^s \frac{1}{\chi(u; f)} \int_{[0, u)} \frac{1}{1 - \Delta\psi(v; f)} \psi_f'(\mathrm{d}v; g) \\
& \quad \cdot \int_{[0, u)} \frac{1}{1 - \Delta\psi(w; f)} \psi_f'(\mathrm{d}w; h) \mathrm{d}f_1(u) \\
& - \int_0^s \frac{1}{\chi(u; f)} \int_{[0, u)} \frac{1}{1 - \Delta\psi(v; f)} \frac{h_*(v)}{f_*(v)} \psi_f'(\mathrm{d}v; g) \mathrm{d}f_1(u) \\
& - \int_0^s \frac{1}{\chi(u; f)} \int_{[0, u)} \frac{1}{1 - \Delta\psi(v; f)} \frac{g_*(v)}{f_*(v)} \psi_f'(\mathrm{d}v; h) \mathrm{d}f_1(u) \\
& + \int_0^s \frac{1}{\chi(u; f)} \sum_{v \leq u} \frac{1}{(1 - \Delta\psi(v; f))^2} \Delta\psi_f'(v; g) \\
& \quad \cdot \Delta\psi_f'(v; h) \mathrm{d}f_1(u).
\end{aligned} \tag{16}$$

With the derivatives in place, we are ready for the proofs of Theorem 1 and Theorem 2. The proofs can be considered adaptations of the proofs of the similar statements in the one-cause scenario given in Jacobsen & Martinussen (2016).

Proof of Theorem 1. Since $h_1(x) = \mathbb{E}(A(\beta_0; Z) \mathbb{E}(\ddot{\phi}(t; x, X) | Z))$, we will first determine $\mathbb{E}(\ddot{\phi}(t; x, X) | Z)$. To do this we consider $\dot{\psi}(s; X)$. By inserting $f = F$ and $g = \delta_X - F$, the expression in (15) leads to

$$\dot{\psi}(s; X) = \int_0^s \frac{1}{H(u)} \mathrm{d}M_{X,0}(u), \tag{17}$$

where $M_{X,0} = N_{X,0} - \int Y_X \mathrm{d}\Lambda_0$. Due to the completely independent censorings assumption, $M_{X,0}$ and thereby $s \mapsto \dot{\psi}(s; X)$ are martingales with respect to

the filter given by $\mathcal{F}_s^Z = \sigma(Z, N_{X,0}(u), \dots, N_{X,d}(u) \mid u \leq s)$. Since it is also 0 in 0 this implies that $E(\dot{\psi}(s; X) \mid Z) = 0$ for any relevant s . Because of the linear structure of the terms in (16), this means that terms including $\psi'_f(h)$ are eliminated when plugging in $f = F, h = \delta_X - F$ and taking the conditional mean. This leaves us with terms two and four, that due to continuity of $u \mapsto \psi(u; F)$ lead to

$$\begin{aligned}
E(\ddot{\phi}(t; x, X) \mid Z) &= \int_0^t \frac{1}{G(s)} \dot{\psi}(s; x) d(H_{Z,1} - H_1)(s) \\
&\quad - \int_0^t \frac{1}{G(s)} \int_0^s \frac{H_Z(u) - H(u)}{H(u)} \dot{\psi}(du; x) dH_1(s) \\
&= \int_0^t \frac{1}{G(s)} \dot{\psi}(s; x) dH_{Z,1}(s) \\
&\quad - \int_0^t \frac{1}{G(s)} \int_0^s \frac{H_Z(u)}{H(u)} \dot{\psi}(du; x) dH_1(s) \\
&= \int_0^t \dot{\psi}(s; x) dF_{Z,1}(s) - \int_0^t \int_0^s \frac{H_Z(u)}{H(u)} \dot{\psi}(du; x) dF_1(s).
\end{aligned} \tag{18}$$

A change in the order of integration means this is equal to

$$\int_0^t \left((F_{Z,1}(t) - F_{Z,1}(s)) - \frac{H_Z(s)}{H(s)} (F_1(t) - F_1(s)) \right) \dot{\psi}(ds; x). \tag{19}$$

By the definition of W in (9), we see that

$$h_1(x) = \int_0^t W(s) \dot{\psi}(ds; x). \tag{20}$$

The compensator of $s \mapsto \dot{\psi}(s; X)^2$ is $s \mapsto \int_0^s \frac{Y_x(u)}{H(u)^2} d\Lambda_0(u)$ according to (17),

and this leads to the conclusion

$$\begin{aligned}\text{Var}(h_1(X)) &= \text{E}(h_1(X)h_1(X)^\top) = \text{E}\left(\int_0^t W(s)W(s)^\top \frac{Y_X(s)}{H(s)^2} d\Lambda_0(s)\right) \\ &= \int_0^t W(s)W(s)^\top \frac{1}{H(s)} d\Lambda_0(s).\end{aligned}\tag{21}$$

Since $H(s) > 0$ for all $s \leq t$, this is equal to 0 if and only if $W(s) = 0$ for Λ_0 -almost all $s \leq t$. $\square$

For the proof of Theorem 2, note that by using (14) the influence function of the Aalen–Johansen estimator can be expressed by

$$\begin{aligned}\dot{\phi}(t; x) &= \int_0^t \frac{1}{G(s)} dN_{x,1}(s) - F_1(t) + \int \frac{1}{G(s)} \dot{\psi}(s; x) dH_1(s) \\ &= \int_0^t \frac{1}{G(s)} dN_{x,1}(s) - F_1(t) + \int_0^t (F_1(t) - F_1(s)) \dot{\psi}(ds; x).\end{aligned}\tag{22}$$

Proof of Theorem 2. Recall that $h_0(x, z) = A(\beta_0; z)(F_1(t) + \dot{\phi}(t; x) - \mu(\beta_0; z))$.

We want to show

$$\text{E}(h_0(X, Z)h_1(X)^\top) = -\text{E}(h_1(X)h_1(X)^\top).\tag{23}$$

As $h_1(X)$ is the value at time t of an $(\mathcal{F}_s^Z)$ -martingale by (20) and since $A(\beta_0; Z)\mu(\beta_0; Z)$ is $\sigma(Z)$ -measurable, the $\text{E}(A(\beta_0; Z)\mu(\beta_0; Z)h_1(X))$ part of $\text{E}(h_0(X, Z)h_1(X)^\top)$ is 0. The parts of $h_0(X, Z)$ that we need to consider are

$$A(\beta_0; Z) \int_0^t \frac{1}{G(s)} dN_{X,1}(s),\tag{24}$$

and

$$A(\beta_0; Z) \int_0^t (F_1(t) - F_1(s)) \dot{\psi}(ds; X).\tag{25}$$

Concerning (25), we see that

$$\begin{aligned}
& \mathbb{E} \left(A(\beta_0; Z) \int_0^t (F_1(t) - F_1(s)) \dot{\psi}(\mathrm{d}s; X) \int_0^t W(s)^\top \dot{\psi}(\mathrm{d}s; X) \right) \\
&= \mathbb{E} \left(A(\beta_0; Z) \int_0^t (F_1(t) - F_1(s)) W(s)^\top \frac{Y_X(s)}{H(s)^2} \mathrm{d}\Lambda_0(s) \right) \\
&= \mathbb{E} \left(A(\beta_0; Z) \int_0^t \frac{H_Z(s)}{H(s)} (F_1(t) - F_1(s)) W(s)^\top \frac{1}{H(s)} \mathrm{d}\Lambda_0(s) \right).
\end{aligned} \tag{26}$$

As for (24), we use that

$$\begin{aligned}
& \int_0^t \frac{1}{G(s)} \mathrm{d}N_{X,1}(s) \int_0^t W(s)^\top \dot{\psi}(\mathrm{d}s; X) \\
&= \int_0^t \int_0^s \frac{1}{G(u)} \mathrm{d}N_{X,1}(u) W(s)^\top \dot{\psi}(\mathrm{d}s; X) \\
&\quad + \int_0^t \int_0^s W(u)^\top \dot{\psi}(\mathrm{d}u; X) \frac{1}{G(s)} \mathrm{d}N_{X,1}(s),
\end{aligned} \tag{27}$$

which applies since $N_{X,1}$ and $\dot{\psi}(X)$ have no jumps in common and both are 0 in 0. The first term is an $(\mathcal{F}_s^Z)$ -martingale and has conditional mean 0 given Z and as such does not contribute to $\mathbb{E}(h_0(X, Z)h_1(X)^\top)$. The last term can be rewritten as

$$- \int_0^t \int_0^s W(u)^\top \frac{1}{H(u)} \mathrm{d}\Lambda_0(u) \frac{1}{G(s)} \mathrm{d}N_{X,1}(s), \tag{28}$$

since a jump in $N_{X,1}$ at a time point s implies $Y_X(u) = 1$ and $N_{X,0}(u) = 0$

for $u \leq s$. Now,

$$\begin{aligned}
& \mathbb{E} \left(-A(\beta_0; Z) \int_0^t \int_0^s W(u)^\top \frac{1}{H(u)} d\Lambda_0(u) \frac{1}{G(s)} dN_{X,1}(s) \right) \\
&= \mathbb{E} \left(-A(\beta_0; Z) \int_0^t \int_0^s W(u)^\top \frac{1}{H(u)} d\Lambda_0(u) dF_{Z,1}(s) \right) \\
&= \mathbb{E} \left(-A(\beta_0; Z) \int_0^t (F_{Z,1}(t) - F_{Z,1}(s)) W(s)^\top \frac{1}{H(s)} d\Lambda_0(s) \right).
\end{aligned} \tag{29}$$

Combining (26) and (29), we see that

$$\begin{aligned}
& \mathbb{E}(h_0(X, Z)h_1(X)^\top) \\
&= \mathbb{E} \left(A(\beta_0; Z) \int_0^t \frac{H_Z(s)}{H(s)} (F_1(t) - F_1(s)) W(s)^\top \frac{1}{H(s)} d\Lambda_0(s) \right) \\
&\quad + \mathbb{E} \left(-A(\beta_0; Z) \int_0^t (F_{Z,1}(t) - F_{Z,1}(s)) W(s)^\top \frac{1}{H(s)} d\Lambda_0(s) \right) \\
&= -\mathbb{E} \left(\int_0^t W(s) W(s)^\top \frac{1}{H(s)} d\Lambda_0(s) \right) \\
&= -\mathbb{E}(h_1(X)h_1(X)^\top),
\end{aligned} \tag{30}$$

which was to be proven. □

Proof of Theorem 3. Recall again that

$$h_0(x, z) = A(\beta_0; z)(F_1(t) + \dot{\phi}(t; x) - \mu(\beta_0; z)) \tag{31}$$

and we want to find an expression for the variance of $h_0(X, Z)$. As the expectation $\mathbb{E}(h_0(X, Z))$ is 0, we find

$$\text{Var}(h_0(X, Z)) = \mathbb{E} \left(A(\beta_0; Z) A(\beta_0; Z)^\top (F_1(t) + \dot{\phi}(t; X) - \mu(\beta_0; Z))^2 \right). \tag{32}$$

So, to prove the theorem, we will be concerned with $(F_1(t) + \dot{\phi}(t; X) -$

$\mu(\beta_0; Z))^2$ and its conditional expectation given Z . Note that, under the model assumption $\mu(\beta_0; Z) = F_{Z,1}(t)$ and using (22), we have

$$\begin{aligned} F_1(t) + \dot{\phi}(t; X) - \mu(\beta_0; Z) &= \int_0^t \frac{1}{G(s)} dN_{X,1}(s) - F_{Z,1}(t) \\ &+ \int_0^t (F_1(t) - F_1(s)) \dot{\psi}(ds; X). \end{aligned} \quad (33)$$

Squaring the expression and considering the different parts, we find that $(\int_0^t \frac{1}{G(s)} dN_{X,1}(s) - F_{Z,1}(t))^2 = \int_0^t \frac{1}{G(s)^2} dN_{X,1}(s) - 2 \int_0^t \frac{1}{G(s)} dN_{X,1}(s) F_{Z,1}(t) + F_{Z,1}(t)^2$ has conditional expectation equal to $\int_0^t \frac{1}{G(s)^2} dH_{Z,1}(s) - F_{Z,1}(t)^2 = \int_0^t \frac{1}{G(s)} dF_{Z,1}(s) - F_{Z,1}(t)^2$ and that, due to the martingale property of $s \mapsto \psi(s; X)$, $(\int_0^t (F_1(t) - F_1(s)) \dot{\psi}(ds; X))^2$ has conditional expectation $\int_0^t (F_1(t) - F_1(s))^2 \frac{H_Z(s)}{H(s)} \frac{1}{H(s)} d\Lambda_0(s)$. Also due to the martingale property, the double product term $2(\int_0^t \frac{1}{G(s)} dN_{X,1}(s) - F_{Z,1}(t))(\int_0^t (F_1(t) - F_1(s)) \dot{\psi}(ds; X))$ has conditional expectation $-2 \int_0^t \frac{1}{G(u)} \int_0^u (F_1(t) - F_1(s)) \frac{1}{H(s)} d\Lambda_0(s) dH_{Z,1}(u) = -2 \int_0^t (F_1(t) - F_1(s))(F_{Z,1}(t) - F_{Z,1}(s)) \frac{1}{H(s)} d\Lambda_0(s)$. Putting it all together establishes the theorem. □

4 Consistently estimating the variance

Recall that the asymptotic variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is $M^{-1}\Sigma(M^{-1})^\top$ with M and Σ from (5) and (6), respectively. As noted in Overgaard *et al.* (2017), these expressions suggest estimating the variance by $\hat{M}_n^{-1}\hat{\Sigma}_n(\hat{M}_n^{-1})^\top$, where

$$\hat{M}_n = \frac{1}{n} \sum_{i=1}^n A(\hat{\beta}_n; Z_i) \frac{\partial}{\partial \beta} \mu(\beta; Z_i) \Big|_{\beta=\hat{\beta}_n}, \quad (34)$$

as in (2), and

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n (\hat{h}_{n,0}(X_i, Z_i) + \hat{h}_{n,1}(X_i)) (\hat{h}_{n,0}(X_i, Z_i) + \hat{h}_{n,1}(X_i))^\top, \quad (35)$$

with

$$\hat{h}_{n,0}(x, z) = A(\hat{\beta}_n; z) (\hat{F}_{n,1}(t) + \phi'_{F_n}(t; \delta_x - F_n) - \mu(\hat{\beta}_n; z)), \quad (36)$$

$$\hat{h}_{n,1}(x) = \frac{1}{n} \sum_{i=1}^n (A(\hat{\beta}_n; Z_i) \phi''_{F_n}(t; \delta_{X_i} - F_n, \delta_x - F_n)). \quad (37)$$

Using (14) and making a change in the order of integration, $\hat{h}_{n,0}(x, z)$ can be expressed and calculated by

$$\begin{aligned} \hat{h}_{n,0}(x, z) = A(\hat{\beta}_n; z) & \left(\int_0^t \frac{1}{\hat{G}_n(s-)} dN_{x,1}(s) - \mu(\hat{\beta}_n; z) \right. \\ & \left. + \int_0^t \frac{\hat{F}_{n,1}(t) - \hat{F}_{n,1}(s)}{(1 - \Delta \hat{\Lambda}_{n,0}(s)) \hat{H}_n(s)} d\hat{M}_{n,x,0}(s) \right), \end{aligned} \quad (38)$$

where $\hat{G}_n(s-) = \chi(s; F_n)$, $\hat{\Lambda}_{n,0}(s) = \psi(s; F_n)$, and $\hat{M}_{n,x,0}(s) = N_{x,0}(s) - \int_0^s Y_x(u) d\hat{\Lambda}_{n,0}(u)$.

Introducing

$$A\hat{M}_{n,0}(s) = \frac{1}{n} \sum_{i=1}^n A(\hat{\beta}_n; Z_i) \hat{M}_{n,X_i,0}(s), \quad (39)$$

$$A\hat{F}_{n,1}(s) = \frac{1}{n} \sum_{i=1}^n A(\hat{\beta}_n; Z_i) \int_0^s \frac{1}{\hat{G}_n(u-)} dN_{X_i,1}(u), \quad (40)$$

$$A\hat{H}_n(s) = \frac{1}{n} \sum_{i=1}^n A(\hat{\beta}_n; Z_i) Y_{X_i}(s), \quad (41)$$

we can express and calculate $\hat{h}_{n,1}(x)$ by

$$\begin{aligned}
\hat{h}_{n,1}(x) = & \int_0^t \frac{1}{\hat{G}_n(s-)} \int_{[0,s)} \frac{1}{(1 - \Delta \hat{\Lambda}_{n,0}(u)) \hat{H}_n(u)} dA \hat{M}_{n,0}(u) dN_{x,1}(s) \\
& + \int_0^t \frac{A \hat{F}_{n,1}(t) - A \hat{F}_{n,1}(s)}{(1 - \Delta \hat{\Lambda}_{n,0}(s)) \hat{H}_n(s)} d\hat{M}_{n,x,0}(s) \\
& + \int_0^t \int_s^t \int_0^u \frac{1}{(1 - \Delta \hat{\Lambda}_{n,0}(v)) \hat{H}_n(v)} dA \hat{M}_{n,0}(v) d\hat{F}_{n,1}(u) \\
& \quad \cdot \frac{1}{(1 - \Delta \hat{\Lambda}_{n,0}(s)) \hat{H}_n(s)} d\hat{M}_{n,x,0}(s) \quad (42) \\
& - \int_0^t \frac{\hat{F}_{n,1}(t) - \hat{F}_{n,1}(s)}{(1 - \Delta \hat{\Lambda}_{n,0}(s)) \hat{H}_n(s)} \frac{A \hat{H}_n(s)}{\hat{H}_n(s)} d\hat{M}_{n,x,0}(s) \\
& - \int_0^t \frac{\hat{F}_{n,1}(t) - \hat{F}_{n,1}(u)}{(1 - \Delta \hat{\Lambda}_{n,0}(u))} \frac{Y_x(u)}{\hat{H}_n(u)^2} dA \hat{M}_{n,0}(u) \\
& + \sum_{s \in [0,t)} \frac{\hat{F}_{n,1}(t) - \hat{F}_{n,1}(s)}{(1 - \Delta \hat{\Lambda}_{n,0}(s))^2 \hat{H}_n(s)} \Delta A \hat{M}_{n,0}(s) \Delta \hat{M}_{n,x,0}(s).
\end{aligned}$$

Since $A \hat{M}_{n,0}$ is expected to be close to 0, confer the proof of Theorem 1, the important terms are number two and four. Here integration of or with respect to a vector is to be understood as the corresponding vector of integrals.

Using these expressions in (35) yields the estimator. A similar, but not identical, estimator of Σ in the one-cause, Kaplan–Meier case was suggested by Jacobsen & Martinussen (2016) and studied in a simulation study with good results in terms of coverage. In the next section we will look into the performance of the variance estimator considered in the present section in a simulation study.

5 Simulation study

Examining the performance of the proposed and Huber–White variance estimators, we are primarily interested in the ability of corresponding confidence intervals to cover the true parameter value.

We consider a dichotomous covariate, $Z \in \{0, 1\}$, with $P(Z = 1) = 0.5$ in a scenario where $t = 1$ is of interest and

$$F_{Z,1}(s) = (\alpha_0 + \alpha_1 Z)s, \quad s \in [0, 1] \quad (43)$$

for the event of interest, and $F_{Z,2}(s) = \eta s$ for $s \in [0, 1]$ for the competing events. In words, η is the risk of having event 2 before time 1 in either Z -group, α_0 is the risk of having event 1 before time 1 in the $Z = 0$ group, and $\alpha_0 + \alpha_1$ is the risk of having event 1 before time 1 in the $Z = 1$ group such that α_1 is the risk difference between the groups at the time point of interest, time 1. Completely independent censorings, C , are introduced, and they will follow an exponential distribution, $C \sim e(\lambda_C)$. The intensity λ_C is controlled by $\gamma = P(C < T \wedge 1)$, the expected proportion of censored information at time 1.

In our simulations, we have fixed $\alpha_0 = \eta = 0.2$ and consider the choices $\alpha_1 = 0.05, 0.30, 0.55$, corresponding to risk differences $RD = 0.05, 0.30, 0.55$, risk ratios $RR = 1.25, 2.50, 3.75$ and odds ratios $OR = 1.33, 4.00, 12.00$ at time 1.

We let $(1, Z)^\top$ play the role of the Z in the previous sections. We consider the link functions $g(x) = x, g(x) = \log(x), g(x) = \text{logit}(x)$, and use $\mu(\beta; Z) = g^{-1}(\beta^\top(1, Z)^\top)$, and $A(\beta; Z) = \frac{\partial}{\partial \beta} \mu(\beta; Z)$ in (1). When solving

the estimating equation, the parameter β_1 in $\beta = (\beta_0, \beta_1)^\top$ has the interpretation of the risk difference, the log risk ratio, and the log odds ratio with the three choices of link function.

For a given link function, an experiment involves simulating from the distributions described above, calculating pseudo-observations based on the Aalen–Johansen estimator, estimating β_1 by solving (1), and estimating the variance of $\hat{\beta}_1$ by each of the two variance estimators: the Huber–White and the proposed variance estimator. In addition to considering different choices of link functions and different effects of the covariate, α_1 , we also consider different choices of the amount of censoring, γ , and the number of observations, n .

To examine the coverage properties of the variance estimate, we have taken the following approach. For each choice of link function (identity, log, logit), $\alpha_1 = 0.05, 0.30, 0.55$, $n = 200, 600, 1000, 1600, 2400, 3200, 4000$, and censoring parameter $\gamma = 0.25, 0.50, 0.75$, run 100 000 replications of the experiment and record the estimate $\hat{\beta}_1$, the number still at risk at time 1 in each group, the number of events in each group, the Huber–White variance estimate, the proposed variance estimate, the coverage of a 95 % confidence interval based on the Huber–White variance estimate, and the coverage of a 95 % confidence interval based on the proposed variance estimate. Most of the results of this approach are reported in the supplement. Since all three choices of link function paint the same picture, we present here only an excerpt in the case with logit link, which can be found in Table 1.

[Table 1 around here (see page 30).]

To further illustrate the convergence of the variance estimates over n and the variability of both variance estimates, we have also done the following simulations. For each configuration of link function (identity, log, logit), $\alpha_1 = 0.05, 0.55$ and $\gamma = 0.25, 0.75$ run 50 000 replications for each step of 100 between $n = 100$ and $n = 1000$. Record estimate $\hat{\beta}_1$, Huber–White variance estimate, and proposed variance estimate. In Figure 1, 2, and 3, a scaled median squared deviation of the obtained $\hat{\beta}_1$ estimates, the median and the 10th and 90th percentiles of the obtained Huber–White variance estimates and the median and the 10th and 90th percentiles of the obtained proposed variance estimates have been plotted against n for each of the three link functions considered.

Confidence intervals are in all cases based on the property of asymptotic normality. We report median instead of mean and a scaled median squared deviation, scaled to match the variance under normality, instead of the actual variance due to their robustness to outliers.

[Figure 1 around here (see page 31).]

[Figure 2 around here (see page 32).]

[Figure 3 around here (see page 33).]

Looking at Table 1 and the tables of the supplement, we see that the Huber–White variance estimate indeed overestimates the variance resulting in too large coverage, but primarily in cases with large β_1 and γ . This observation matches the results of Theorem 1 and Theorem 2. We also see that the proposed variance estimator can perform poorly with limited

amounts of data, especially when most of that data is censored. The coverage produced by the proposed variance estimator is however mostly between 94 % and 96 % and only falls below that range in the more extreme cases with large amounts of censoring and low numbers of observations.

Figures 1, 2, and 3 reveal small differences overall in the situation with small effect of Z and small amount of censoring. When the effect of Z is small and the amount of censoring is large, the Huber–White variance estimator does best at resembling the observed variance with the sample sizes used here. When the effect of Z is large, the bias of the Huber–White variance estimator is apparent and the proposed variance estimator does best. In our setting, and when we have a large effect of Z and a large amount of censoring, the actual variance seemingly converges to its asymptotic value from above, whereas the proposed variance estimate converges to the same value from below. This helps explain why the coverage proportions seen in the tables are sometimes below 95 %.

In scenarios that are believed to be similar to the setting in our simulation study, the following recommendations seem to be in order.

- (i) For large sample sizes, the proposed variance estimator reaches the right level and will be advantageous to the potentially biased Huber–White estimator. In this case, the proposed variance estimator should be used.
- (ii) For small sample sizes and when the effect of covariates is believed to be small and the amount of censoring is limited, the Huber–White estimator better avoids the underestimation seen for the proposed variance

estimator and the theoretical overestimation of the Huber–White estimator is expected to be small. In this case the Huber–White estimator should be used.

- (iii) When covariates are believed to have a strong effect on the cumulative incidence and a considerable amount of censoring is observed, the bias of the Huber–White estimator is expected to be large. In cases with a decent amount of data, the proposed variance estimator should be used.

How well these recommendations carry over to other scenarios has not been studied here, but the recommendations can be considered a rough guideline.

A suggestion by Andersen *et al.* (2003) was to use the bootstrap method on the given set of pseudo-observations as an alternative way to obtain variance estimates. Due to the computational effort needed for this method, we have conducted a simulation study that is smaller in scale to take a look at the bootstrap variance estimate. The results are summarized by Figure 4 in line with Figure 1.

[Figure 4 around here (see page 34).]

The results suggest the bootstrap variance will be similar to the Huber–White variance estimate with an added variability in the variance estimates, which is likely influenced by the number of bootstrap replications made. This would mean that the bootstrap variance estimate and the Huber–White variance estimate are equally biased.

6 Concluding remarks

We have argued that the Huber–White variance estimator will be upwards biased from the asymptotic variance, except in special cases with no effect of the covariates or no censoring as examples. Our simulation study was consistent with this finding showing a bias only when the effect of Z was large and the amount of censoring was large. The proposed, and consistent, variance estimator apparently can struggle to produce the 95 % coverage for extreme cases with small sample sizes, but works well for larger sample sizes.

Our result that the Huber–White variance estimator is biased upwards generalizes the identical result for the Kaplan–Meier-based scheme by Jacobsen & Martinussen (2016). It seems likely that the result will hold for other pseudo-observation schemes as well.

The result hinges on the assumption of completely independent censorings. In particular, the censoring time is assumed independent of covariates. This assumption is a requirement for unbiased estimation of the regression parameters when using the pseudo-observation method with pseudo-observations from the Aalen–Johansen estimator, as found by Graw *et al.* (2009) and Overgaard *et al.* (2017), but the assumption may well be inappropriate in applications. To overcome the problem, Binder *et al.* (2014) has suggested the use of pseudo-observations based on inverse probability weighted estimators that differ from the Aalen–Johansen estimator by taking the effect of covariates on censoring into account. The approach seemingly corrects the problem. The theoretical justification of the approach is a topic of ongoing research. The results of Overgaard *et al.* (2017) again suggests

bias of the Huber–White sandwich variance estimator in this setting.

Although we have described a setting with one time point of interest, the pseudo-observation method is designed, as it was presented in Andersen *et al.* (2003), to handle several time points of interest. The method has been considered in a multiple time point setting with competing risks in i.a. Klein & Andersen (2005) and Andersen & Klein (2007). Overgaard *et al.* (2017) considered only a single time point of interest, but suggests the results carry over easily to the multiple time points setting.

Other variance estimates may be preferable to the one suggested here. A weighted average of the Huber–White and proposed variance estimators with the weight of the Huber–White variance estimator vanishing as $n \rightarrow \infty$ could be useful to avoid small sample problems of the proposed variance estimate. In Jacobsen & Martinussen (2016) it is suggested to instead estimate the bias of the Huber–White variance estimator and subtract it from the Huber–White variance estimate. Even though the bootstrap variance estimate based on the given set of pseudo-observations does not seem to give correct coverage, it would likely be appropriate, though computationally demanding, to bootstrap the entire procedure, including calculation of the pseudo-observations. None of these approaches have been explored here. We have instead focused on making many replications of few, simple scenarios to get a detailed picture of these particular examples.

Acknowledgements

The research of Jan Pedersen is supported by the Danish Council for Independent Research (Grant DFF - 4002-00003).

Supporting information

Additional information for this article is available online. This includes tables with results from the simulation study.

References

- Andersen, P. K. & Klein, J. P. (2007). Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scand. J. Stat.* **34**, 3–16.
- Andersen, P. K., Klein, J. P. & Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* **90**, 15–27.
- Binder, N., Gerds, T. A. & Andersen, P. K. (2014). Pseudo-observations for competing risks with covariate dependent censoring. *Lifetime Data Anal.* **20**, 303–315.
- Dudley, R. M. & Norvaiša, R. (1999). *Differentiability of six operators on nonsmooth functions and p-variation*, Springer-Verlag, Berlin.

- Dudley, R. M. & Norvaiša, R. (2011). *Concrete functional calculus*, Springer, New York.
- Gill, R. D. & Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Annals of Statistics* **18**, 1501–1555.
- Graw, F., Gerds, T. A. & Schumacher, M. (2009). On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal.* **15**, 241–255, 2009.
- Jacobsen, M. & Martinussen, T. (2016). A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations. *Scand. J. Stat.* **43**, 845–862.
- Jewell, N. P., Lei, X., Ghani, A. C., Donnelly, C. A., Leung, G. M., Ho, L. & Cowling, B. J. (2007). Non-parametric estimation of the case fatality ratio with competing risks data: An application to severe acute respiratory syndrome (sars). *Stat. Med.* **26**, 1982–1998.
- Klein, J. P. & Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* **61**, 223–229.
- Overgaard, M., Parner, E. T. & Pedersen, J. (2017). Asymptotic theory of generalized estimating equations based on jack-knife pseudo-observations. *Annals of Statistics* **45**, 1988–2015.

Satten, G. A. & Datta, S. (2001). The Kaplan—Meier estimator as an inverse-probability-of-censoring weighted average. *Amer. Statist.* **55**, 207–210.

Address

Morten Overgaard

Department of Public Health, Aarhus University

Bartholins Allé 2 - Building 1260, DK-8000 Aarhus C

E-mail: *moov@ph.au.dk*

Appendix: The product integral

For a càdlàg function $\alpha: [a, b] \rightarrow \mathbb{R}$, the product integral of α can be defined as

$$\prod_a^s (1 + d\alpha(u)) = \lim_{\max |u_i - u_{i-1}| \rightarrow 0} \prod_i (1 + \alpha(u_i) - \alpha(u_{i-1})) \quad (44)$$

where $a = u_0 < u_1 < \dots < u_m = s$ is a partition of $[a, s]$ for an $s \in (a, b]$. This definition is given and many properties of the product integral are proven in Gill & Johansen (1990).

In Dudley & Norvaiša (2011) and Dudley & Norvaiša (1999), the operator $\alpha \mapsto \prod_a^{(\cdot)} (1 + d\alpha)$ is defined and studied as a functional between spaces of right-continuous functions of bounded p -variation, and smoothness of the operator was established. Let $\mu(J) = \prod_J (1 + d\alpha)$ for any interval $J \subseteq [a, b]$. We need not concern ourselves with the technicalities of the p -variation setting, but the following expressions are of importance to us.

- i.* The first order derivative of the product integral operator at α in direction β can be written as

$$\int_a^{(\cdot)} \mu([a, s)) \mu((s, (\cdot)]) d\beta(s), \quad (45)$$

and can also be expressed by

$$\mu([a, (\cdot)]) \int_a^{(\cdot)} \frac{1}{1 + \Delta\alpha(s)} d\beta(s) \quad (46)$$

when $\Delta\alpha(s) \neq -1$ for the s in the integration.

- ii.* The second order derivative at α in directions β and γ is given by

$$\begin{aligned} & \int_a^{(\cdot)} \int_{[a, u)} \mu([a, s)) \mu((s, u)) \mu((u, (\cdot)]) d\beta(s) d\gamma(u) \\ & + \int_a^{(\cdot)} \int_{[a, s)} \mu([a, u)) \mu((u, s)) \mu((s, (\cdot)]) d\gamma(u) d\beta(s), \end{aligned} \quad (47)$$

or alternatively

$$\begin{aligned} & \mu([a, (\cdot)]) \int_a^{(\cdot)} \frac{1}{1 + \Delta\alpha(s)} d\beta(s) \int_a^{(\cdot)} \frac{1}{1 + \Delta\alpha(u)} d\gamma(u) \\ & - \mu([a, (\cdot)]) \sum_{s \in [a, (\cdot)]} \frac{1}{(1 + \Delta\alpha(s))^2} \Delta\beta(s) \Delta\gamma(s) \end{aligned} \quad (48)$$

when $\Delta\alpha(s) \neq -1$ for the s in the integrations or sum.

Table 1: Simulation results of 100 000 replications of each configuration of (n, β_1, γ) with logit link. Across replications, med. $\hat{\beta}_1$ is the median of the β_1 estimates, Var_{obs} is a scaled median squared deviation estimating the variance of $\sqrt{n}\hat{\beta}_1$, $\hat{\text{Var}}_{\text{new}}$ is the median of the proposed variance estimator, $\hat{\text{Var}}_{\text{HW}}$ is the median of the Huber–White variance estimator, cov_{new} is the coverage proportion of 95 % confidence intervals based on the proposed variance estimator, and cov_{HW} is the coverage proportion of 95 % confidence intervals based on the Huber–White variance estimator.

(n, β_1, γ)	med. $\hat{\beta}_1$	Var_{obs}	$\hat{\text{Var}}_{\text{new}}$	$\hat{\text{Var}}_{\text{HW}}$	cov_{new}	cov_{HW}
(200, 0.29, 0.25)	0.2915	29.381	29.056	29.389	0.953	0.954
(1600, 0.29, 0.25)	0.2886	28.457	28.552	28.605	0.951	0.951
(4000, 0.29, 0.25)	0.2880	28.166	28.503	28.534	0.949	0.949
(200, 2.48, 0.25)	2.5148	31.521	30.991	34.904	0.953	0.967
(1600, 2.48, 0.25)	2.4901	30.377	30.176	33.576	0.951	0.963
(4000, 2.48, 0.25)	2.4863	30.077	30.085	33.443	0.949	0.961
(200, 0.29, 0.75)	0.2996	95.257	79.483	89.018	0.930	0.956
(1600, 0.29, 0.75)	0.2884	84.480	83.592	84.834	0.948	0.950
(4000, 0.29, 0.75)	0.2895	84.154	83.847	84.447	0.950	0.951
(200, 2.48, 0.75)	2.6776	222.091	146.029	264.757	0.945	0.948
(1600, 2.48, 0.75)	2.5077	155.270	146.606	212.422	0.944	0.982
(4000, 2.48, 0.75)	2.4947	148.926	146.467	207.373	0.948	0.980

Figure 1: Simulation results using identity link. Median (solid line), 10th percentile and 90th percentile (short-dashed lines) of the Huber–White variance estimate of $\sqrt{n}\hat{\beta}_1$ in light gray, median (solid line), 10th percentile and 90th percentile (short-dashed lines) of the proposed variance estimate of $\sqrt{n}\hat{\beta}_1$ in dark gray, and a scaled median squared deviation estimating the variance of the $\sqrt{n}\hat{\beta}_1$ s in black (long-dashed line) in four different scenarios plotted against the number of observations. Based on 50 000 replications at each point $n = 100, 200, \dots, 1000$.

Figure 2: Simulation results using log link. Median (solid line), 10th percentile and 90th percentile (short-dashed lines) of the Huber–White variance estimate of $\sqrt{n}\hat{\beta}_1$ in light gray, median (solid line), 10th percentile and 90th percentile (short-dashed lines) of the proposed variance estimate of $\sqrt{n}\hat{\beta}_1$ in dark gray, and a scaled median squared deviation estimating the variance of the $\sqrt{n}\hat{\beta}_1$ s in black (long-dashed line) in four different scenarios plotted against the number of observations. Based on 50 000 replications at each point $n = 100, 200, \dots, 1000$.

Figure 3: Simulation results using logit link. Median (solid line), 10th percentile and 90th percentile (short-dashed lines) of the Huber–White variance estimate of $\sqrt{n}\hat{\beta}_1$ in light gray, median (solid line), 10th percentile and 90th percentile (short-dashed lines) of the proposed variance estimate of $\sqrt{n}\hat{\beta}_1$ in dark gray, and a scaled median squared deviation estimating the variance of the $\sqrt{n}\hat{\beta}_1$ s in black (long-dashed line) in four different scenarios plotted against the number of observations. Based on 50 000 replications at each point $n = 100, 200, \dots, 1000$.

Figure 4: Simulation results using identity link. Median (solid line), 10th percentile and 90th percentile (short-dashed lines) of the Huber–White variance estimate of $\sqrt{n}\hat{\beta}_1$ in light gray, median (solid line), 10th percentile and 90th percentile (short-dashed lines) of the proposed variance estimate of $\sqrt{n}\hat{\beta}_1$ in dark gray, a scaled median squared deviation estimating the variance of the $\sqrt{n}\hat{\beta}_1$ s in black (long-dashed line), and median (solid line), 10th percentile and 90th percentile (short-dashed lines) of the bootstrap variance estimate of $\sqrt{n}\hat{\beta}_1$ based on 50 bootstrap replications in gray in four different scenarios plotted against the number of observations. Based on 10 000 replications at each point $n = 100, 200, \dots, 1000$.