PSEUDO-VALUE APPROACH FOR CONDITIONAL QUANTILE RESIDUAL LIFETIME ANALYSIS FOR CLUSTERED SURVIVAL AND COMPETING RISKS DATA WITH APPLICATIONS TO BONE MARROW TRANSPLANT DATA[1]

BY KWANG WOO AHN AND BRENT R. LOGAN

Medical College of Wisconsin

Quantile residual lifetime analysis is conducted to compare remaining lifetimes among groups for survival data. Evaluating residual lifetimes among groups after adjustment for covariates is often of interest. The current literature is limited to comparing two groups for independent data. We propose a pseudo-value approach to compare quantile residual lifetimes given covariates between multiple groups for independent and clustered survival data. The proposed method considers clustered event times and clustered censoring times in addition to independent event times and censoring times. We show that the method can also be used to compare multiple groups on the cause-specific residual life distribution in the competing risk setting, for which there are no current methods which account for clustering. The empirical Type I errors and statistical power of the proposed study are examined in a simulation study, which shows that the proposed method controls Type I errors very well and has higher power than an existing method. The proposed method is illustrated by a bone marrow transplant data set.

1. Introduction. A blood and marrow transplant is one of the most widely used procedures to treat cancers, including leukemia, lymphoma and multiple myeloma. As patients survive longer, pre- and post-transplant exposures have the potential to compromise life expectancy and can contribute to development of late complications [Majhail and Rizzo (2013)]. Thus, clinicians in bone marrow transplantation are often interested in studying survival outcomes for patients who survived for at least some specific period after transplant. For example, Martin et al. (2010) studied residual life expectancy in patients surviving more than 5 years after allogeneic or autologous hematopoietic cell transplantation from 1970 through 2002. In practice, survival outcomes and censoring times of bone marrow transplant data are often clustered due to a transplant center effect. Thus, developing statistical models for clustered data is essential in studying residual life: remaining lifetime of a patient given that the patient survived at least to time t.

Quantile residual lifetime analysis is often preferred when the distribution of the residual lifetime is skewed [Ma and Wei (2012)]. Statistical methods for residual

lifetime analysis with survival data have been recently developed as in Kim, Zhou and Jeong (2012) and Lin, Zhang and Zhou (2015). Although they incorporate covariates into the models, they are restricted to independent survival data.

To illustrate some existing methods, consider independent survival data with sample size n. Let T_i, C_i and $\mathbf{Z}_i$ be the event time, censoring time and covariate vector of individual i, respectively, for $i = 1, \ldots, n$. For simplicity, we assume that $\mathbf{Z}_i$'s are fixed over time. The observed time of individual i is defined by $X_i = \min(T_i, C_i)$. Let $\Delta_i = I(T_i \leq C_i)$. Define q_τ as the τth conditional quantile residual lifetime given survival to time t_0 for covariate $\mathbf{Z} = \mathbf{z}_0$. Then q_τ satisfies

$$(1.1) \qquad S(t_0 + q_\tau | \mathbf{z}_0) = (1 - \tau) S(t_0 | \mathbf{z}_0),$$

where $S(t|\mathbf{Z})$ is a survival probability at time t given the covariate $\mathbf{Z} = \mathbf{z}_0$. The Cox proportional hazards model [Cox (1972)] and Breslow estimator [Breslow (1972)] can be used to consistently estimate $S(t|\mathbf{Z})$ as in Lin (2007) and Zhao et al. (2015). Lin, Zhang and Zhou (2015) estimated a solution to (1.1) by solving

$$(1.2) \qquad \hat{U}(q_\tau) = \hat{S}(t_0 + q_\tau | \mathbf{z}_0) - (1 - \tau) \hat{S}(t_0 | \mathbf{z}_0) = 0.$$

Let $\hat{q}_\tau$ be a solution to (1.2). Lin, Zhang and Zhou (2015) studied the asymptotics of $\hat{q}_\tau$ and proposed one-sample and two-sample tests based on $\hat{q}_\tau$ for comparing conditional residual lifetimes given $\mathbf{Z} = \mathbf{z}_0$. However, they did not discuss methods for comparing more than two groups, and their method is only applicable for independent survival data. In addition, estimating the variance of $\hat{q}_\tau$ requires simulation-based methods such as resampling methods and bootstrap in practice [Zeng and Lin (2008)]. Thus, it is desirable to develop a nonsimulation-based method for multiple group comparisons with independent and clustered data.

Jeong and Fine (2009) proposed analyzing the cause-specific residual life distribution for the competing risks setting. For failure time T and cause of failure ε, let the cumulative incidence function for cause k be $F_k(t) = P(T \leq t, \varepsilon = k)$. They defined the residual cumulative incidence function given survival to t_0 for cause k as follows:

$$\frac{F_k(t + t_0) - F_k(t_0)}{S(t_0)}.$$

Let Q_τ be the τth quantile of the cause k residual cumulative incidence function given survival to t_0. Then Q_τ is defined as a solution to

$$\frac{F_k(t_0 + Q_\tau) - F_k(t_0)}{S(t_0)} = \tau.$$

Note that if there is at least an event for each cause, then $F_k(\infty) < 1$ for all k. Although the above equation allows that $F_k(\infty) < 1$, Q_τ may not exist for some extreme τ. In this paper, we only consider a case that Q_τ exists. A nonparametric test was developed for testing one sample and two samples by Jeong and Fine

(2013). Like Lin, Zhang and Zhou (2015), they did not discuss methods for comparing more than two groups, and their method is only applicable to independent data. Moreover, their proposed method does not take covariates into account.

We propose a pseudo-value-based method to compare quantile residual lifetimes or residual cumulative incidence given covariates between multiple groups for clustered survival and competing risks data. The pseudo-values can be used as a response variable under the generalized estimating equations (GEE) setting. Thus, statistical inference on clustered data can be readily handled via the GEE. Thanks to this technical convenience, the GEE using pseudo-values has been widely used to make inference on survival or competing risks data for independent and clustered data [Andersen, Klein and Rosthøj (2003); Logan, Zhang and Klein (2011)]. Our proposed method using pseudo-values is flexible and the strategy can be applied to independent or clustered data, as well as to residual lifetimes or cumulative incidence with minimal modifications. Standard errors and inference are also straightforward, without requiring simulation-based techniques. Importantly, it allows for covariate specific inference for comparing residual cumulative incidence functions, which has not previously been accomplished. We describe our motivating data in Section 2. Then, in Section 3 we extend the pseudo-value technique, which previously was restricted to clustered event times and independent censoring times, to also handle clustered censoring times. In Section 4, we propose a test statistic based on pseudo-values and study its asymptotic distribution. A simulation study is conducted in Section 5. A bone marrow transplant example is illustrated in Section 6. We give a brief conclusion in Section 7.

2. Data. The data for this application were collected by the Center for International Blood and Marrow Transplant Research [Shaw et al. (2010)]. It consists of pediatric patients (<18 years) undergoing allogeneic T-replete, myeloablative bone marrow transplantation between 1993 and 2006. We study relapse and disease-free survival (DFS) of patients with severe disease conditions in this paper: intermediate or advanced disease status. The data contains 847 patients from 99 transplant centers. A significant center effect exists in DFS rates (p-value = 0.038) and relapse (p-value = 0.042) at a significance level of 0.05 using the random effect score test of Commenges and Andersen (1995). The censoring times of both endpoints are also correlated because their p-values from the score test are 0.050 and 0.008 for DFS and relapse, respectively.

We consider four variables: disease status, donor type, disease type and recipient age at transplant. They were all identified as significant factors for DFS in the marginal Cox proportional hazards model [Lee, Wei and Amato (1992)]. There are 619 and 228 patients with intermediate disease status and advanced disease status, respectively. Donor type has three groups: 80 patients with one-antigen mismatched related donors or phenotypically matched nonsibling related donors, 583 patients with human leukocyte antigen (HLA) identical sibling donors, and 184 patients with 8/8 allele-matched unrelated donors. Three disease types are
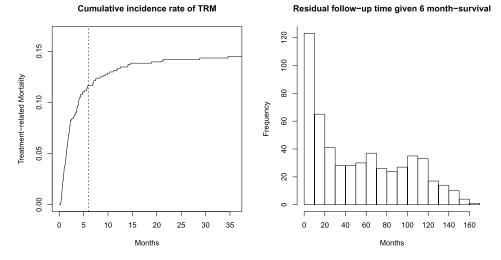
FIG. 1. *Cumulative incidence rate of treatment-related mortality (left plot) and histogram of residual follow-up time (right plot). The dotted vertical line of the left plot indicates 6 months after transplant.*

considered: 170 patients with acute myeloid leukemia (AML), 547 patients with acute lymphoblastic leukemia (ALL), and 130 patients with myelodysplastic syndrome (MDS). Two groups are studied for recipient age as in Shaw et al. (2010): 554 patients who are younger than or equal to 10 years old and 293 who are older than 10 years old.

Many of the deadly treatment-related complications occur within 3 months after bone marrow transplant. Early complications include acute graft-versus-host disease, engraftment failure and various early infections. The left plot of Figure 1 shows the cumulative incidence rate of treatment-related mortality (TRM). The dotted vertical line indicates 6 months after transplant. The majority of the treatment-related deaths occurred within 6 months post transplant due to early complications. Thus, it is of interest to study the residual lifetime of patients who survived disease-free to at least 6 months after transplant. There were 547 patients from 86 centers who survived disease-free to at least 6 months. The right plot of Figure 1 shows the histogram of the last follow-up times of patients who survived disease-free to at least 6 months. The median last follow-up time and the longest last follow-up time are 15 months and 171 months, respectively. For patients who survived disease-free to at least 6 months, 48% of relapse and 47% of events in DFS occurred between 6 months and 1 year after transplant. An additional 21% of relapse and 20% of events in DFS occurred between 12 months and 18 months post transplant. These suggest that the distributions of event times for patients who survived disease-free to at least 6 months after transplant are skewed in relapse and DFS. Therefore, we will study conditional quantile residual lifetime of pa-

tients who survived disease-free to at least 6 months in relapse and DFS based on the four variables that we mentioned in the previous paragraph.

3. Pseudo-value approach. In this section, we extend the pseudo-value approach to clustered events and clustered censoring times for competing risks and survival settings. We first consider the competing risks setting with two causes of failure $\varepsilon \in \{1, 2\}$. We assume that there are m clusters and each cluster has ℓ individuals, where $n = m \times \ell$ is the total sample size. Note that the clusters may have different sizes by defining censoring times as zero when observed times are missing [Spiekerman and Lin (1998)]. Let T_{ij}, C_{ij}, ε_{ij} and $\mathbf{Z}_{ij}$ be the event time, censoring time, cause of failure and covariate vector of individual j in cluster i, respectively, for $i = 1, \ldots, m$ and $j = 1, \ldots, \ell$. Let $\mathbf{T}_i = \{T_{ij}, j = 1, \ldots, \ell\}$, $\mathbf{C}_i = \{C_{ij}, j = 1, \ldots, \ell\}$, $\boldsymbol{\varepsilon}_i = \{\varepsilon_{ij}, j = 1, \ldots, \ell\}$, and $\mathbf{Z}_i = \{\mathbf{Z}_{ij}, j = 1, \ldots, \ell\}$. Suppose that $(\mathbf{T}_i, \boldsymbol{\varepsilon}_i, \mathbf{C}_i, \mathbf{Z}_i)$ are independent and identically distributed (i.i.d.). We assume that the C_{ij}'s do not depend on the $\mathbf{Z}_{ij}$'s and the $(T_{ij}, \varepsilon_{ij})$'s are independent of the C_{ij}'s given $\mathbf{Z}_{ij}$'s for $i = 1, \ldots, m$ and $j = 1, \ldots, \ell$. This setting allows that the event times may be correlated within the same cluster. Similarly, the censoring times may be correlated within the same cluster. The C_{ij}'s are assumed to have a common distribution $G(t) = P(C \geq t)$, where C is a censoring time. Let $X_{ij} = \min(T_{ij}, C_{ij})$ be the observed time and $\Delta_{ij} = I(T_{ij} \leq C_{ij})$.

To define a pseudo-value, we first consider the unadjusted marginal cumulative incidence function for cause 1 without loss of generality. We adjust for covariates under the GEE framework after obtaining the pseudo-values. Let $F_1(t) = P(T \leq t, \varepsilon = 1)$ and $N_{kij}(t) = I(T_{ij} \leq t)I(\varepsilon_{ij} = k)\Delta_{ij}$ for $k = 1, 2$. Define $N_{ij}(t) = N_{1ij}(t) + N_{2ij}(t)$. Let $Y_{ij}(t) = I\{t \leq X_{ij}\}$ and $Y(t) = \sum_{i=1}^{m}\sum_{j=1}^{\ell} Y_{ij}$. The cumulative incidence estimate can be estimated by $\hat{F}_1(t) = \int_0^t \hat{S}(u-)\, d\hat{H}_1(u)$, where $\hat{S}(u-)$ is the Kaplan–Meier estimate of event-free survival and $\hat{H}_1(t)$ is the estimated cause-specific cumulative hazard function given by

$$d\hat{H}_1(t) = \sum_{i=1}^{m}\sum_{j=1}^{\ell} \frac{dN_{1ij}(t)}{Y(t)}.$$

Zhou et al. (2012) showed that $\hat{F}_1(t)$ is a consistent estimator of $F_1(t)$ for clustered competing risks event times and clustered censoring times.

A pseudo-value at time t of the jth individual in the ith cluster for $F_1(t)$ is defined by

$$P_{ij}^f(t) = n\hat{F}_1(t) - (n-1)\hat{F}_1^{-ij}(t), \qquad i = 1, \ldots, m, j = 1, \ldots, \ell,$$

where $\hat{F}_1^{-ij}(t)$ is the cumulative incidence estimate obtained by omitting the jth individual in the ith cluster. Let $N_{ij}^c(t) = I(C_{ij} \leq t)$ and $H^c(t)$ be the cumulative hazard function by treating censored observations as events. Let $M_{ij}^c(t) =$

$N_{ij}^c(t) - \int_0^t I(X_{ij} \geq u) \, dH^c(u)$ and T is an event time of any cause. Then, as in the Supplementary Materials [see Ahn and Logan (2016)], we can show

$$(3.1) \quad P_{ij}^f(t) = \frac{N_{1ij}(t)}{G(X_{ij})} + \int_0^{X_{ij}} \frac{P(T \leq t, \varepsilon = 1 | T \geq u)}{G(u)} \, dM_{ij}^c(u) + O_p(m^{-1/2}),$$

which implies that $\mathbf{P}_i^f(t) = \{P_{i1}^f(t), \ldots, P_{i\ell}^f(t)\}^T$'s asymptotically depend on i. Thus, $\mathbf{P}_i^f(t)$'s are asymptotically independent with $\lim_{m \to \infty} E(P_{ij}^f(t)|\mathbf{Z}_{ij}) = F_1(t|\mathbf{Z}_{ij})$. A generalized estimating equation (GEE) setting can be used by treating pseudo-values as a response variable [Andersen, Klein and Rosthøj (2003); Logan, Zhang and Klein (2011); Klein and Andersen (2005)]. We illustrate the use of the GEE at a fixed time point t. To model the marginal cumulative incidence function at time t, we assume the cumulative incidence is related to covariates through a link function $h(\cdot)$ so that $h\{F_1(t|\mathbf{Z}_{ij})\} = \boldsymbol{\beta}^T \mathbf{Z}_{ij}$. Let $\nu_{ij}(t) = F_1(t|\mathbf{Z}_{ij}) = h^{-1}(\boldsymbol{\beta}^T \mathbf{Z}_{ij})$ and $\boldsymbol{\nu}_i = \{\nu_{i1}(t), \ldots, \nu_{i\ell}(t)\}^T$ for $i = 1, \ldots, m$ and $j = 1, \ldots, \ell$. Then the GEE is defined as follows:

$$\sum_i \left(\frac{\partial \boldsymbol{\nu}_i}{\partial \boldsymbol{\beta}}\right)^T \mathbf{V}_i^{-1}(\mathbf{P}_i^f - \boldsymbol{\nu}_i) = \mathbf{0},$$

where $\mathbf{V}_i$ is a working covariance matrix for $\mathbf{P}_i^f$. Statistical inference on $\boldsymbol{\beta}$ can be readily handled by using the sandwich estimator in GEE [Liang and Zeger (1986)], which is consistent even if the working correlation matrix is misspecified [Zeger, Liang and Albert (1988)]. This sandwich estimator may slightly overestimate the variance of $\hat{\boldsymbol{\beta}}$ as Jacobsen and Martinussen (2014) pointed out. However, they argued that this overestimation should be minor in many applications because their simulation study showed that the sandwich estimator estimated the variance of $\hat{\boldsymbol{\beta}}$ very well unless $\boldsymbol{\beta}$ was too large. In addition, Logan, Zhang and Klein (2011), Ahn and Mendolia (2014), and Klein and Andersen (2005) showed that the sandwich estimator worked well for the pseudo-value approach under their various simulation settings.

For the survival setting, the pseudo-value for survival at time t is defined as $P_{ij}^s(t) = n\hat{S}(t) - (n-1)\hat{S}^{-ij}(t)$ for $i = 1, \ldots, m$ and $j = 1, \ldots, \ell$, where $\hat{S}^{-ij}(t)$ is the Kaplan–Meier estimate obtained by omitting the jth individual in the ith cluster. As in the competing risks setting, we can show

$$(3.2) \qquad P_{ij}^s(t) = \frac{N_{ij}(t)}{G(X_{ij})} + \int_0^{X_{ij}} \frac{P(T \leq t | T \geq u)}{G(u)} \, dM_{ij}^c(u) + O_p(m^{-1/2}),$$

where T is event time of interest. This shows that $\lim_{m \to \infty} E\{P_{ij}^s(t)|\mathbf{Z}_{ij}\} = S(t|\mathbf{Z}_{ij})$ and $\mathbf{P}_i^s(t) = (P_{i1}^s(t), \ldots, P_{i\ell}^s(t))^T$'s are asymptotically independent. The GEE setting can be similarly used as in the competing risks setting to directly model the marginal survival function at time t in the presence of clustered survival and censoring data.

4. Method. In this section, we propose a test to compare the conditional residual quantiles between multiple groups for independent and clustered data given a fixed covariate level $\mathbf{Z} = \mathbf{z}_0$. We first propose a test for the competing risks setting where cause 1 is of interest.

We denote $Q_{g\tau}$, $g = 1, \ldots, \zeta$ as the τth conditional quantiles of the cause 1 residual cumulative incidence given survival to time t_0 for group g when the covariate level is $\mathbf{Z} = \mathbf{z}_0$. Then $Q_{g\tau}$ is defined as a solution to

$$\mathcal{U}_g(Q_{g\tau}) = F_{1g}(t_0 + Q_{g\tau}|\mathbf{z}_0) - F_{1g}(t_0|\mathbf{z}_0) - \tau S_g(t_0|\mathbf{z}_0) = 0,$$

where $F_{1g}(t|\mathbf{z}_0)$ and $S_g(t|\mathbf{z}_0)$ are the cumulative incidence and survival probability at time t given $\mathbf{Z} = \mathbf{z}_0$ for group g. Assuming that $F_{1g}(t|\mathbf{z}_0)$ is absolutely continuous and $f_{1g}(t|\mathbf{z}_0) = dF_{1g}(t|\mathbf{z}_0)/dt$ is positive on some neighborhood of $t_0 + Q_{g\tau}$, $\mathcal{U}_g(Q_{g\tau})$ has a unique solution. We are interested in testing $H_0 : Q_{1\tau} = \cdots = Q_{\zeta\tau} \equiv Q_\tau$. Testing H_0 is equivalent to testing $H_0' : \mathcal{U}_1(Q_\tau) = \cdots = \mathcal{U}_\zeta(Q_\tau) = 0$ due to the unique solution of $\mathcal{U}_g(Q_\tau)$ for $g = 1, \ldots, \zeta$. To test H_0', we follow the steps below:

1. Obtain a consistent estimator $\hat{Q}_\tau$ of Q_τ based on the pooled data;
2. Compute pseudo-values $P_{ij}^f(t_0 + \hat{Q}_\tau)$, $P_{ij}^f(t_0)$ and $P_{ij}^s(t_0)$ for $i = 1, \ldots, m$ and $j = 1, \ldots, \ell$;
3. Fit GEE using the pseudo-values to estimate $F_{1g}(t_0 + \hat{Q}_\tau|\mathbf{z}_0)$, $F_{1g}(t_0|\mathbf{z}_0)$ and $S_g(t_0|\mathbf{z}_0)$ for $g = 1, \ldots, \zeta$;
4. Estimate $\mathcal{U}_g(Q_\tau)$ for $g = 1, \ldots, \zeta$ and their covariance matrix using the estimates of the GEE;
5. Obtain a quadratic form test statistic to test H_0'.

For step 1, we first study how to obtain consistent estimates of $S(t|\mathbf{Z})$ and $F_1(t|\mathbf{Z})$. A consistent estimate of $S(t|\mathbf{Z})$ even for clustered data is obtained by the marginal Cox model and Breslow estimator. See the Supplementary Materials for details of the marginal Cox model and Breslow estimator for clustered data, and the asymptotic distribution of $S(t|\mathbf{Z})$. To obtain a consistent estimate of $F_1(t|\mathbf{Z})$, we use a marginal proportional subdistribution hazard model proposed by Zhou et al. (2012). Here the subdistribution hazard $\lambda_1(t|\mathbf{Z}) = dF_1(t|\mathbf{Z})/\{1 - F_1(t|\mathbf{Z})\}$ follows a proportional hazards representation as

$$\lambda_1(t|\mathbf{Z}) = \lambda_{10}(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}).$$

Under this model, the cumulative subdistribution hazard for cause 1 can be estimated by

$$\hat{\Lambda}_1(t|\mathbf{Z}) = \int_0^t \exp(\hat{\boldsymbol{\beta}}^T \mathbf{Z}) \, d\hat{\Lambda}_{10}(u)$$

$$= \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^\ell \int_0^t \frac{\exp(\hat{\boldsymbol{\beta}}^T \mathbf{Z})}{\frac{1}{n} \sum_{a=1}^m \sum_{b=1}^\ell \hat{w}_{ab}(u) I(X_{ab} \geq u) \exp\{\hat{\boldsymbol{\beta}}^T \mathbf{Z}\}}$$

$$\times \hat{w}_{ij}(u) \, dN_{1ij}(u),$$

where $\hat{w}_{ij}(t) = I\{C_{ij} \geq \min(X_{ij}, t)\}\hat{G}(t)/\hat{G}\{\min(X_{ij}, t)\}$ and $\hat{G}(t)$ is the Kaplan–Meier estimate at time t for the censoring distribution. Zhou et al. (2012) showed the consistency of $\hat{F}_1(t|\mathbf{Z}) = 1 - \exp\{-\hat{\Lambda}_1(t|\mathbf{Z})\}$ for clustered event times and clustered censoring times.

Let $\hat{F}_1(t|\mathbf{z}_0)$ and $\hat{S}(t|\mathbf{z}_0)$ be consistent estimators of $F_1(t|\mathbf{z}_0)$ and $S(t|\mathbf{z}_0)$ based on the pooled data. Then, under the null hypothesis, Q_τ can be consistently estimated by solving $\hat{F}_1(t_0 + Q_\tau|\mathbf{z}_0) - \hat{F}_1(t_0|\mathbf{z}_0) - \tau\hat{S}(t_0|\mathbf{z}_0) = 0$. We estimate $\hat{Q}_\tau$ as the smallest Q_τ at which $\hat{F}_1(t_0 + Q_\tau|\mathbf{z}_0) - \hat{F}_1(t_0|\mathbf{z}_0) - \tau\hat{S}(t_0|\mathbf{z}_0)$ crosses 0. Due to the consistency of $\hat{F}_1(t|\mathbf{z}_0)$ and $\hat{\tilde{S}}(t|\mathbf{z}_0)$, $\hat{Q}_\tau$ is also consistent.

For step 2, we calculate the pseudo-values $P_{ij}^f(t_0 + \hat{Q}_\tau)$, $P_{ij}^f(t_0)$ and $P_{ij}^s(t_0)$ for individual j of cluster i as described in Section 3.

Because each $\mathcal{U}_g(\hat{Q}_\tau)$ depends on the components $F_{1g}(t_0 + \hat{Q}_\tau|\mathbf{z}_0)$, $F_{1g}(t_0|\mathbf{z}_0)$ and $S_g(t_0|\mathbf{z}_0)$, we use a single pseudo-value regression model to obtain estimates of these components for each group in step 3. One main advantage of using a single pseudo-value regression model is that it provides direct estimates of the covariance matrix across these components in the complex clustered data setting using the sandwich variance estimate, leading to straightforward estimation of the variance of the final test statistic. We parameterize a joint model using a link function $h(\cdot)$ so that

$$\mu_{ij}^f(t_0) = F_{1g}(t_0|\mathbf{Z}_{ij}) = h^{-1}(\phi + \phi_1 + \boldsymbol{\gamma}_1^T\mathbf{Z}_{ij} + \theta_{1g}),$$

$$\mu_{ij}^f(t_0 + \hat{Q}_\tau) = F_{1g}(t_0 + \hat{Q}_\tau|\mathbf{Z}_{ij}) = h^{-1}(\phi + \phi_2 + \boldsymbol{\gamma}_2^T\mathbf{Z}_{ij} + \theta_{2g}),$$

$$\mu_{ij}^s(t_0) = S_g(t_0|\mathbf{Z}_{ij}) = h^{-1}(\phi + \boldsymbol{\gamma}_3^T\mathbf{Z}_{ij} + \theta_{3g}),$$

where $\theta_{1\zeta} = \theta_{2\zeta} = \theta_{3\zeta} = 0$. Here ϕ, $\phi + \phi_1$ and $\phi + \phi_2$ are intercept terms for $h\{\mu_{ij}^s(t_0)\}, h\{\mu_{ij}^f(t_0)\}$ and $h\{\mu_{ij}^f(t_0 + \hat{Q}_\tau)\}$, respectively. And θ_{1g}, θ_{2g} and θ_{3g} are parameters for an indicator function of group $g = 1, \ldots, \zeta - 1$ for $h\{\mu_{ij}^f(t_0)\}, h\{\mu_{ij}^f(t_0 + \hat{Q}_\tau)\}$ and $h\{\mu_{ij}^s(t_0)\}$, respectively. In matrix notation, consider the response variable vector $\mathbf{P}_i = \{\mathbf{P}_i^f(t_0), \mathbf{P}_i^f(t_0 + \hat{Q}_\tau), \mathbf{P}_i^s(t_0)\}^T$, where $\mathbf{P}_i^f(t) = \{P_{i1}^f(t), \ldots, P_{i\ell}^f(t)\}^T$ and $\mathbf{P}_i^s(t) = \{P_{i1}^s(t), \ldots, P_{i\ell}^s(t)\}^T$ for $i = 1, \ldots, m$. Assuming $N_{1ij}(x)$ is continuous at $x = Q_\tau + t_0$ with probability one, we can show that $\mathbf{P}_i^f(t_0 + \hat{Q}_\tau)$ converges in probability to $\mathbf{P}_i^f(t_0 + Q_\tau)$ as in the Supplementary Materials. Define the parameter vector $\boldsymbol{\alpha} = (\phi, \phi_1, \phi_2, \boldsymbol{\gamma}_1^T, \boldsymbol{\theta}_1^T, \boldsymbol{\gamma}_2^T, \boldsymbol{\theta}_2^T, \boldsymbol{\gamma}_3^T, \boldsymbol{\theta}_3^T)^T$, where $\boldsymbol{\theta}_k = (\theta_{k1}, \ldots, \theta_{k\zeta-1})^T$ for $k = 1, 2, 3$. Define the mean vector as $\boldsymbol{\mu}_i = \{\boldsymbol{\mu}_i^f(t_0), \boldsymbol{\mu}_i^f(t_0 + \hat{Q}_\tau), \boldsymbol{\mu}_i^s(t_0)\}^T$ for $i = 1, \ldots, m$, where $\boldsymbol{\mu}_i^f(t) = (\mu_{i1}^f(t), \ldots, \mu_{i\ell}^f(t))^T$ and $\boldsymbol{\mu}_i^s(t) = (\mu_{i1}^s(t), \ldots, \mu_{i\ell}^s(t))^T$ for $i = 1, \ldots, m$. This joint model can be fitted using the GEE

$$\sum_i \left(\frac{\partial\boldsymbol{\mu}_i}{\partial\boldsymbol{\alpha}}\right)^T \mathbf{V}_i^{-1}(\mathbf{P}_i - \boldsymbol{\mu}_i) \equiv \sum_i \mathbf{A}_i(\boldsymbol{\alpha}) = \mathbf{0},$$

where $\mathbf{V}_i$ models the covariance matrix of $\mathbf{P}_i$. Then $\sqrt{m}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ converges in distribution to $N(\mathbf{0}, \Sigma_f)$ given $\hat{Q}_\tau$. The sandwich estimator is used to estimate Σ_f as follows:

$$\hat{\Sigma}_f = \mathcal{I}(\hat{\boldsymbol{\alpha}})^{-1} \widehat{\mathrm{Var}}\{\mathbf{A}(\hat{\boldsymbol{\alpha}})\} \mathcal{I}(\hat{\boldsymbol{\alpha}})^{-1},$$

where

$$\mathcal{I}(\boldsymbol{\alpha}) = \sum_i \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\alpha}} \right)^T \mathbf{V}_i^{-1} \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\alpha}} \right), \qquad \widehat{\mathrm{Var}}\{\mathbf{A}(\hat{\boldsymbol{\alpha}})\} = \sum_i \mathbf{A}_i(\hat{\boldsymbol{\alpha}}) \mathbf{A}_i(\hat{\boldsymbol{\alpha}})^T.$$

This GEE is used to estimate the parameter vector $\boldsymbol{\alpha}$ and obtain the covariance matrix of the estimate; however, what we are ultimately interested in is testing H_0' through $\mathcal{U}$, which is specific to a particular covariate value $\mathbf{z_0}$.

Thus, in step 4, we estimate $\mathcal{U}_g(\hat{Q}_\tau)$ under the null hypothesis by

$$\hat{\mathcal{U}}_g(\hat{Q}_\tau) = h^{-1}(\hat{\phi} + \hat{\phi}_2 + \hat{\boldsymbol{\gamma}}_2^T \mathbf{z}_0 + \hat{\theta}_{2g}) - h^{-1}(\hat{\phi} + \hat{\phi}_1 + \hat{\boldsymbol{\gamma}}_1^T \mathbf{z}_0 + \hat{\theta}_{1g})$$

$$- \tau h^{-1}(\hat{\phi} + \hat{\boldsymbol{\gamma}}_3^T \mathbf{z}_0 + \hat{\theta}_{3g}), \qquad \text{if } g = 1, \dots, \zeta - 1;$$

$$\hat{\mathcal{U}}_\zeta(\hat{Q}_\tau) = h^{-1}(\hat{\phi} + \hat{\phi}_2 + \hat{\boldsymbol{\gamma}}_2^T \mathbf{z}_0) - h^{-1}(\hat{\phi} + \hat{\phi}_1 + \hat{\boldsymbol{\gamma}}_1^T \mathbf{z}_0) - \tau h^{-1}(\hat{\phi} + \hat{\boldsymbol{\gamma}}_3^T \mathbf{z}_0),$$

$$\text{if } g = \zeta.$$

Using the delta method, we can show that under the null hypothesis $\{\hat{\mathcal{U}}_1(\hat{Q}_\tau), \dots, \hat{\mathcal{U}}_\zeta(\hat{Q}_\tau)\}^T$ converges in distribution to $N(\mathbf{0}, \Omega_f \Sigma_f \Omega_f^T)$, where

$$\Omega_f = \begin{pmatrix} \dfrac{\partial \mathcal{U}_1(\hat{Q}_\tau)}{\partial \phi} & \dfrac{\partial \mathcal{U}_1(\hat{Q}_\tau)}{\partial \phi_1} & \cdots & \dfrac{\partial \mathcal{U}_1(\hat{Q}_\tau)}{\partial \theta_{3\zeta - 1}} \\[2mm] \dfrac{\partial \mathcal{U}_2(\hat{Q}_\tau)}{\partial \phi} & \dfrac{\partial \mathcal{U}_2(\hat{Q}_\tau)}{\partial \phi_1} & \cdots & \dfrac{\partial \mathcal{U}_2(\hat{Q}_\tau)}{\partial \theta_{3\zeta - 1}} \\[2mm] \vdots & \vdots & \ddots & \vdots \\[2mm] \dfrac{\partial \mathcal{U}_\zeta(\hat{Q}_\tau)}{\partial \phi} & \dfrac{\partial \mathcal{U}_\zeta(\hat{Q}_\tau)}{\partial \phi_1} & \cdots & \dfrac{\partial \mathcal{U}_\zeta(\hat{Q}_\tau)}{\partial \theta_{3\zeta - 1}} \end{pmatrix}.$$

We can obtain $\hat{\Omega}_f$ by plugging $\hat{\boldsymbol{\alpha}}$ into Ω_f.

Finally, in step 5, we construct a quadratic form test based on an estimate of the vector

$$\mathcal{U} = \{\mathcal{U}_1(\hat{Q}_\tau), \dots, \mathcal{U}_\zeta(\hat{Q}_\tau)\}^T,$$

which should have a mean vector of 0 under the null hypothesis. Let n_g be the sample size of group g and $n = \sum_{g=1}^\zeta n_g$. Define the weighted mean of $\hat{\mathcal{U}}_g(\hat{Q}_\tau)$'s as

$$\overline{\mathcal{U}}(\hat{Q}_\tau) = \frac{1}{n} \sum_{g=1}^\zeta n_g \hat{\mathcal{U}}_g(\hat{Q}_\tau),$$

which converges in probability to zero under the null hypothesis. Then, under the null hypothesis, $\{\hat{\mathcal{U}}_1(\hat{Q}_\tau) - \overline{\mathcal{U}}(\hat{Q}_\tau), \ldots, \hat{\mathcal{U}}_\zeta(\hat{Q}_\tau) - \overline{\mathcal{U}}(\hat{Q}_\tau)\}^T$ converges in distribution to $N(\mathbf{0}, \mathbf{B}\Omega_f \Sigma_f \Omega_f^T \mathbf{B}^T)$, where

$$\mathbf{B} = \mathbf{I}_\zeta - \frac{1}{n}\mathcal{M},$$

$\mathbf{I}_\zeta$ is a $\zeta \times \zeta$ identity matrix and

$$\mathcal{M} = \begin{pmatrix} n_1 & n_2 & \cdots & n_\zeta \\ n_1 & n_2 & \cdots & n_\zeta \\ \vdots & \vdots & \ddots & \vdots \\ n_1 & n_2 & \cdots & n_\zeta \end{pmatrix}.$$

Note that $\mathbf{B}\Omega_f \Sigma_f \Omega_f^T \mathbf{B}^T$ is not invertible. To test H_0', the final test statistic is

$$X^2 = m\{\hat{\mathcal{U}}_1(\hat{Q}_\tau) - \overline{\mathcal{U}}(\hat{Q}_\tau), \ldots, \hat{\mathcal{U}}_\zeta(\hat{Q}_\tau) - \overline{\mathcal{U}}(\hat{Q}_\tau)\}^T (\mathbf{B}\hat{\Omega}_f \hat{\Sigma}_f \hat{\Omega}_f^T \mathbf{B}^T)^-$$
$$\times \{\hat{\mathcal{U}}_1(\hat{Q}_\tau) - \overline{\mathcal{U}}(\hat{Q}_\tau), \ldots, \hat{\mathcal{U}}_\zeta(\hat{Q}_\tau) - \overline{\mathcal{U}}(\hat{Q}_\tau)\},$$

where $(\mathbf{B}\hat{\Omega}\hat{\Sigma}\hat{\Omega}^T \mathbf{B}^T)^-$ is a generalized inverse of $\mathbf{B}\hat{\Omega}\hat{\Sigma}\hat{\Omega}^T \mathbf{B}^T$. Under the null hypothesis, X^2 converges in distribution to a chi-squared distribution with degrees of freedom $\zeta - 1$.

The proposed method can be easily adapted to the one-sample setting where we are interested in testing if $Q_\tau = Q_{0\tau}$ for a prespecified $Q_{0\tau}$. Let

$$\omega_f = \left\{ \frac{\partial \mathcal{U}_1(Q_{0\tau})}{\partial \phi}, \frac{\partial \mathcal{U}_1(Q_{0\tau})}{\partial \phi_1}, \ldots, \frac{\partial \mathcal{U}_1(Q_{0\tau})}{\partial \theta_{3\zeta - 1}} \right\}^T.$$

Let $\hat{\omega}_f$ be an estimator of ω_f by plugging $\hat{\boldsymbol{\alpha}}$ into ω_f. We propose a test statistic $m\hat{\mathcal{U}}_1(Q_{0\tau})^2/\hat{\mathcal{D}}$, where $\hat{\mathcal{D}} = \hat{\omega}_f^T \hat{\Sigma}_f \hat{\omega}_f$. Under the null hypothesis, $m\hat{\mathcal{U}}_1(Q_{0\tau})^2/\hat{\mathcal{D}}$ follows a chi-squared distribution with degree of freedom 1.

The proposed method for the survival setting can be similarly constructed; see the Supplementary Materials for details.

5. Simulation.

5.1. *Survival setting.* In this section, we perform a simulation study for the survival setting. We consider a one-sample test first. A positive stable frailty is used to generate correlated event times as in Logan, Zhang and Klein (2011). Independent of event times, correlated censoring times are generated. We consider $m = 100, 200$ and 400 with cluster size $\ell = 4$. For each cluster, two independent random effects w and w_c are generated from a positive stable frailty distribution with parameter ψ, where the Laplace transformation of the standard positive stable distribution is $L(s) = \exp(-s^\psi)$. Three ψ values are used: 0.25 and 0.5 for

clustered data and 1 for independent data. These values of ψ would correspond to values of Kendall's τ equal to 0.75, 0.5 and 0, respectively [Logan, Zhang and Klein (2011)]. For each cluster, three binary covariates are considered: (i) Z_1 is a cluster level covariate. For each cluster, a random binary number is generated with probability 0.5 and it is assigned to all individuals within a cluster; (ii) Z_2 is independently generated with probability 0.5 for each individual within a cluster; and (iii) Z_3 represents a matched pair design. Thus, two individuals have 0's and the other two have 1's. Following Logan, Zhang and Klein (2011), we generate event and censoring times for each cluster from

$$(5.1) \qquad S(t|w) = \exp\{-w\exp(\boldsymbol{\beta}^T \mathbf{Z})t\}, \qquad G(t|w) = \exp(-w_c\lambda_c t),$$

where $\mathbf{Z} = (Z_1, Z_2, Z_3)^T$ and $\psi\boldsymbol{\beta} = (1, -0.5, 0.5)^T$. We select λ_c to generate a 50% censoring rate. We consider the median residual lifetime given $\mathbf{z}_0 = (1, 1, 1)^T$ conditional on survival to t_0, where $S(t_0|\mathbf{z}_0) = 0.8$, which yields $t_0 = -\{\log(0.8)\}^{1/\psi}/\exp(1/\psi)$. Let $q_{0\tau}$ be the true conditional median lifetime under the null hypothesis where $\tau = 0.5$. Then $q_{0\tau} = 0.013, 0.107$ and 0.255 for $\psi = 0.25, 0.5$ and 1, respectively.

To test $q_\tau = q_{0\tau}$, we test if $S(t_0 + q_{0\tau}|\mathbf{z}_0) - 0.5S(t_0|\mathbf{z}_0) = 0$ using GEE. The identity link and logit link functions with the independence working correlation matrix are considered. Table 1 shows the empirical Type I error probabilities with 5000 iterations at a significance level of 0.05. For each link function, two GEE models are fitted: (i) GEE ignoring within-cluster correlation, that is, assuming independent data; and (ii) GEE accounting for within-cluster correlation. "PVI–Ind," "PVI–Cluster," "PVL–Ind" and "PVL–Cluster" indicate the pseudo-value approach with identity link function assuming independent data, identity link function accounting for within-cluster correlation, logit link function assuming independent data and logit link function accounting for within-cluster correlation, respectively. As we can see in Table 1, the pseudo-value approach ignoring within-cluster correlation performs well for independent data. However, it becomes more liberal as the within-cluster correlation increases. Thus, the presence of a cluster effect leads to underestimation of the variance when the cluster effect is ignored under our simulation setting. On the other hand, the pseudo-value approach accounting for within-cluster correlation controls Type I error very well for clustered data. As m increases, its empirical Type I errors become closer to 0.05 in general. It appears that the pseudo-value approach with the identity link function controls Type I errors slightly better than that with the logit link function.

Next, we consider the three group comparison. A cluster size $\ell = 6$ is considered with two binary covariates: (i) Z_1 is a cluster level covariate so that it is generated with probability 0.5; and (ii) Z_2 is independently generated with probability 0.5 for each individual within a cluster. Each group has two individuals from each cluster. We generate event and censoring times from (5.1) with $\mathbf{Z} = (Z_1, Z_2)^T$ and $\psi\boldsymbol{\beta} = (1, 0.5)^T$. We choose λ_c to generate a 50% censoring rate. We consider the

TABLE 1
Empirical Type I error probabilities for the survival setting: "PVI–Ind," "PVI–Cluster," "PVL–Ind" and "PVL–Cluster" indicate the pseudo-value approach with identity link function assuming independent data, identity link function accounting for within-cluster correlation, logit link function assuming independent data and logit link function accounting for within-cluster correlation, respectively

ψ	m	PVI–Ind	PVI–Cluster	PVL–Ind	PVL–Cluster
			One-sample test		
1	50	0.0600	0.0646	0.0658	0.0714
	100	0.0506	0.0516	0.0530	0.0578
	200	0.0506	0.0534	0.0508	0.0520
	400	0.0502	0.0514	0.0504	0.0516
0.5	50	0.0918	0.0656	0.0990	0.0790
	100	0.0872	0.0600	0.0902	0.0640
	200	0.0868	0.0566	0.0832	0.0592
	400	0.0862	0.0516	0.0848	0.0558
0.25	50	0.1172	0.0734	0.1286	0.0850
	100	0.1148	0.0600	0.1100	0.0622
	200	0.1096	0.0514	0.1046	0.0532
	400	0.1122	0.0500	0.1024	0.0512
			Three-sample test		
1	50	0.0604	0.0732	0.0650	0.0812
	100	0.0556	0.0584	0.0586	0.0626
	200	0.0444	0.0504	0.0512	0.0524
	400	0.0484	0.0514	0.0494	0.0520
0.5	50	0.0166	0.0714	0.0250	0.081
	100	0.0146	0.0594	0.0204	0.0614
	200	0.0148	0.0536	0.0178	0.0514
	400	0.0124	0.0514	0.0168	0.0506
0.25	50	0.0032	0.0592	0.0126	0.0770
	100	0.0018	0.0586	0.0038	0.0610
	200	0.0016	0.0524	0.0034	0.0546
	400	0.0010	0.0468	0.0024	0.0514

median residual lifetime given $\mathbf{z}_0 = (1, 1)^T$ conditional on survival to t_0, where $S(t_0|\mathbf{z}_0) = 0.8$, which leads t_0 to be $-\{\log(0.8)\}^{1/\psi} / \exp(1.5/\psi)$. Table 1 shows the empirical Type I errors from the three-sample test with 5000 iterations. As in the one-sample test setting, the pseudo-value approach accounting for within-cluster correlation controls Type I errors very well for clustered data. As m increases, the empirical Type I errors become closer to 0.05. The pseudo-value approach ignoring within-cluster correlation works properly only for independent data as expected. It is somewhat conservative for clustered data. This is because our three-group-comparison setting is more like a matched pairs setting, where

ignoring the cluster effect amounts to ignoring the reduction in variance due to the within-cluster group comparison. Therefore, the variance is overestimated and the type I error is conservative. A similar phenomenon was seen in Logan, Zhang and Klein (2011). We also examined various $\tau = 0.1, 0.25, 0.75$ and 0.9. In addition, we performed a simulation assuming Z_2 follows $N(0, 1)$ to study the median residual lifetime given $\mathbf{z}_0 = (0.5, 1)^T$ conditional on survival to t_0. In all scenarios, the proposed method performed as well as shown in Table 1. Thus, the results were omitted.

We also compare the proposed method to Lin, Zhang and Zhou (2015). Because the method of Lin, Zhang and Zhou (2015) is restricted to the two-sample test, the two-group comparison is considered. Let $q_{g\tau}$ be the conditional median residual lifetime of group g for $g = 1$ and 2. To test $q_{1\tau} = q_{2\tau}$, Lin, Zhang and Zhou (2015) proposed a test statistic

$$L = \frac{\hat{q}_{1\tau} - \hat{q}_{2\tau}}{\hat{\sigma}_1^2/n_1 + \hat{\sigma}_2^2/n_2},$$

where $\hat{q}_{i\tau}$ is obtained by solving (1.2) for group i and $\hat{\sigma}_i^2/n_i$ is the estimated variance of $\hat{q}_{i\tau}$ for $i = 1, 2$. Under the null hypothesis, L follows a chi-squared distribution with degree of freedom 1. The bootstrap with 500 iterations is used to estimate σ_i^2/n_i for $i = 1, 2$. We also conducted the bootstrap with 1000 iterations for some of the simulation scenarios and obtained very similar results to those with 500 iterations. The setting for examining empirical Type I errors is the same as the three-group comparison's except $\ell = 4$. Thus, each cluster consists of two groups instead of three groups. We compare two median residual lifetimes given $\mathbf{z}_0 = (1, 1)^T$ conditional on survival to t_0, where $S(t_0|\mathbf{z}_0) = 0.8$. We examine $m = 100$ and $m = 200$. Table 2 shows the summary of 5000 iterations. For independent data, all methods control Type I errors well, although the pseudo-value approaches ignoring within-cluster correlation seem to work slightly better than the others. However, Lin, Zhang and Zhou (2015) and the pseudo-value approaches ignoring within-cluster correlation are conservative for clustered data. To compare power, survival and censoring times are generated from

$$S_1(t|w) = \exp\left\{-w \sum_{i=1}^{2} \beta_i Z_i t\right\}, \qquad S_2(t|w) = \exp\left\{-w \sum_{i=1}^{3} \beta_i Z_i t\right\},$$

$$G(t|w) = \exp(-w_c \lambda_c t),$$

where $S_i(t|w)$ is the survival probability for group i, $\psi\boldsymbol{\beta} = (1, 0.5, 0.5)^T$ and Z_i's have the same definitions as in the three-group comparison setting. As can be seen from Table 2, the pseudo-value approach has higher statistical power than Lin, Zhang and Zhou (2015). It appears that the pseudo-value approach with the identity link has the highest power among the five methods.

TABLE 2

Two-sample setting with survival data and comparison to Lin, Zhang and Zhou (2015): "PVI–Ind," "PVI–Cluster," "PVL–Ind" and "PVL–Cluster" indicate the pseudo-value approach with identity link function assuming independent data, identity link function accounting for within-cluster correlation, logit link function assuming independent data and logit link function accounting for within-cluster correlation, respectively

ψ	m	PVI–Ind	PVI–Cluster	PVL–Ind	PVL–Cluster	Lin et al.
			Empirical Type I errors			
1	100	0.0552	0.0598	0.0540	0.0552	0.0594
	200	0.0492	0.0486	0.0486	0.0506	0.0612
0.5	100	0.0204	0.0588	0.0238	0.0526	0.0382
	200	0.0214	0.0558	0.0278	0.0558	0.0398
0.25	100	0.0046	0.0574	0.0074	0.0590	0.0304
	200	0.0030	0.0560	0.0046	0.0540	0.0270
			Empirical statistical power			
1	100	0.6714	0.6742	0.5614	0.5712	0.2002
	200	0.9304	0.9298	0.8554	0.8572	0.5256
0.5	100	0.6960	0.8216	0.5788	0.6584	0.0536
	200	0.9572	0.9840	0.8832	0.9230	0.2616
0.25	100	0.7412	0.9054	0.6154	0.7400	0.0022
	200	0.9732	0.9986	0.8986	0.9522	0.0190

5.2. *Competing risks setting.* In this section, we conduct a simulation study for the competing risks setting. Consider the one-sample test first. As in Section 5.1, a positive stable frailty is used to generate correlated event times and censoring times. We consider $m = 100, 200$ and 400 with cluster size $\ell = 4$. For each cluster, two independent random effects w and w_c are generated from a positive stable frailty distribution with parameter $\psi = 0.25, 0.5$ and 1. Three binary covariates are considered where the definitions of Z_1, Z_2 and Z_3 are the same as in Section 5.1. Similarly to Logan, Zhang and Klein (2011), we generate competing risks event times and censoring times for each cluster from

$$F_1(t|w) = 1 - \{1 - p(1 - e^{-t})\}^{w \exp(\boldsymbol{\beta}^T \mathbf{Z})},$$

(5.2)
$$F_2(t|w) = (1 - p)^{w \exp(\boldsymbol{\beta}^T \mathbf{Z})}(1 - e^{-t \exp(\boldsymbol{\beta}^T \mathbf{Z})}),$$

$$G(t|w_c) = \exp(-w_c \lambda_c t),$$

where $\mathbf{Z} = (Z_1, Z_2, Z_3)^T$ and $\psi \boldsymbol{\beta} = (1, -0.5, 0.5)^T$. We select p and λ_c to generate 30% cause 1 events, 30% cause 2 events and 40% censoring. We consider the 0.25th quantile of the cause 1 residual cumulative incidence for $\mathbf{z}_0 = (1, 1, 1)^T$ conditional on survival to t_0, where $S(t_0|\mathbf{z}_0) = 0.8$. Let $Q_{0\tau}$ be the true conditional quantile under the null hypothesis, where $\tau = 0.25$. Then, $Q_{0\tau} = 0.068, 0.194$

TABLE 3
Empirical Type I error probabilities for the competing risks setting: "PVI–Ind," "PVI–Cluster," "PVL–Ind" and "PVL–Cluster" indicate the pseudo-value approach with identity link function assuming independent data, identity link function accounting for within-cluster correlation, logit link function assuming independent data and logit link function accounting for within-cluster correlation, respectively

ψ	m	PVI–Ind	PVI–Cluster	PVL–Ind	PVL–Cluster
			One-sample test		
1	50	0.0560	0.0650	0.0914	0.0970
	100	0.0544	0.0580	0.0658	0.0680
	200	0.0522	0.0588	0.0586	0.0590
	400	0.0460	0.0476	0.0462	0.0488
0.5	50	0.0936	0.0724	0.1374	0.1286
	100	0.0898	0.0640	0.0930	0.0772
	200	0.0856	0.0568	0.0816	0.0612
	400	0.0872	0.0522	0.0782	0.0514
0.25	50	0.1246	0.0836	0.1638	0.1406
	100	0.1092	0.0622	0.1068	0.0750
	200	0.1042	0.0556	0.0958	0.0598
	400	0.1082	0.0540	0.0952	0.0534
			Three-sample test		
1	50	0.0652	0.0816	0.0684	0.092
	100	0.0566	0.0620	0.0582	0.0644
	200	0.0532	0.0550	0.0520	0.0572
	400	0.0514	0.0536	0.0496	0.0506
0.5	50	0.0226	0.0664	0.0584	0.0978
	100	0.0222	0.0626	0.0314	0.0678
	200	0.0220	0.0572	0.0288	0.0578
	400	0.0190	0.0542	0.0256	0.0522
0.25	50	0.0007	0.0553	0.0555	0.0975
	100	0.0064	0.0544	0.0142	0.0538
	200	0.0046	0.0498	0.0086	0.0486
	400	0.0060	0.0516	0.0080	0.0462

and 0.341 for $\psi = 0.25, 0.5$ and 1, respectively. To test if $Q_\tau = Q_{0\tau}$, we test if $\mathcal{U}_1(Q_{0\tau}) = F_1(Q_{0\tau} + t_0|\mathbf{z}_0) - F_1(t_0|\mathbf{z}_0) - \tau S(t_0|\mathbf{z}_0) = 0$ using GEE. The identity link and logit link functions with the independence working correlation matrix are considered. Table 3 shows the empirical Type I error probabilities with 5000 iterations at a significance level of 0.05. The pseudo-value approaches ignoring within-cluster correlation become more liberal as the within-cluster correlation increases, although they work well for independent data. On the other hand, the pseudo-value approach accounting for within-cluster correlation controls Type I error very well for clustered data. As m increases, its empirical Type I errors become closer to

0.05 in general. It appears that the pseudo-value approach with the identity link function controls Type I errors slightly better than that with the logit link function as in the survival setting.

Next, we consider the three-group comparison. A cluster size of $\ell = 6$ is considered with two binary covariates Z_1 and Z_2. Each group has two individuals from each cluster. We generate event and censoring times from (5.2) with $\mathbf{Z} = (Z_1, Z_2)^T$ and $\psi \boldsymbol{\beta} = (1, 0.5)^T$. We choose p and λ_c to generate 30% of cause 1, 30% of cause 2 and 40% of censoring rate. We compare the 0.25th conditional quantile of the cause 1 residual lifetimes given $\mathbf{z}_0 = (1, 1)^T$ survival to t_0, where $S(t_0|\mathbf{z}_0) = 0.8$. Table 3 shows the empirical Type I errors from the three-sample test with 5000 iterations. As in the one-sample test setting, the pseudo-value approach accounting for within-cluster correlation controls Type I errors very well for clustered data. As m increases, the empirical Type I errors become closer to 0.05 in general. The pseudo-value approach ignoring within-cluster correlation works satisfactorily only for independent data. It is conservative for clustered data. As in the survival setting, we also examined various $\tau = 0.1, 0.25, 0.75$ and 0.9. In addition, we performed a simulation assuming Z_2 follows $N(0, 1)$. In all scenarios, the proposed method performed as well as shown in Table 3. Thus, the results were omitted.

6. Example.   We revisit our motivating data [Shaw et al. (2010)] of Section 2. As we discussed in Section 2, event times and censoring times in relapse and DFS are clustered. The independence working correlation matrix was used for GEE. We consider four variables: disease status, donor type, disease type and recipient age at transplant. The censoring distribution of relapse and DFS did not depend on any of the four variables at the significance level 0.05.

First of all, we compare the conditional median residual disease-free survival lifetimes given disease-free survival to 6 months between the two recipient age groups for patients having intermediate disease status, HLA identical sibling donors and AML. The left plot of Figure 2 shows two conditional residual disease-free survival distributions given survival to 6 months for the patients of interest, that is,

$$\frac{\hat{S}(t + 6|\text{intermediate disease status, HLA identical sibling donors and AML})}{\hat{S}(6|\text{intermediate disease status, HLA identical sibling donors and AML})}.$$

The dotted horizontal line indicates 50% of the conditional residual disease-free survival probability (CRDFS). The estimated conditional median residual lifetimes were 73 months and 23 months for the recipient age ≤ 10 and > 10, respectively. It appears that the two median residual lifetimes were different. The proposed method with the identity link function and the independence working correlation matrix found a statistically significant difference with a p-value of 0.004. The proposed method with the logit link function also found a statistical significance with
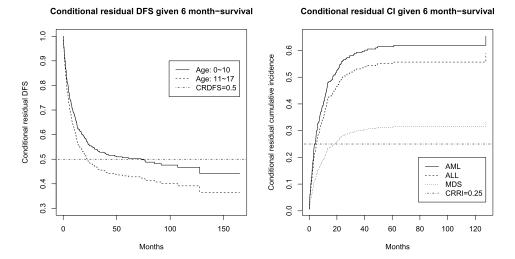
Conditional residual DFS given 6 month−survival   **Conditional residual CI given 6 month−survival**

FIG. 2. *Estimated conditional residual DFS and cumulative incidence. "CI" of the right plot stands for cumulative incidence.*

a p-value 0.007. However, Lin, Zhang and Zhou (2015) did not find a statistically significant difference at a significance level of 0.05 with a p-value of 0.257. In addition, when adjustment for center effects was ignored, the proposed method with the identity link function and the logit link function did not find statistical significance with p-values of 0.081 and 0.138, respectively. As Jacobsohn (2015) recently pointed out, there are only a few pediatric late effects studies. Nonetheless, our finding is interesting because Ferry et al. (2007) did not find a significant age effect on long-term survival for pediatric patients with hematological malignancies. Our result suggests that there might be another group of pediatric patients that have a recipient age effect on long-term survival.

Next, we consider disease groups and disease status for the conditional quantile of the residual cumulative incidence of relapse. Treatment-related mortality is the competing risk for relapse. We compare the conditional 0.25th residual cumulative incidence of relapse given disease-free survival to 6 months between the three disease groups of AML, ALL and MDS. The right plot of Figure 2 shows the conditional residual cumulative incidence given disease-free survival to 6 months of the three disease groups with intermediate disease status, that is,

$$\frac{\{\hat{F}(t+6|\text{intermediate disease status}) - \hat{F}(6|\text{intermediate disease status})\}}{\hat{S}(6|\text{intermediate disease status})}.$$

The dotted horizontal line indicates 25% of the conditional residual relapse incidence (CRRI). The estimated conditional residual relapse incidence quantiles were 4 months, 5 months and 17 months for AML, ALL and MDS, respectively. It appears that the conditional 0.25th residual cumulative incidence of relapse for

MDS is different from those for AML and ALL. We found a statistically significant result with a p-value of 0.037 using the proposed method. When we ignored adjustment for center effects, we did not find statistical significance with a p-value of 0.482. This result is consistent to the finding of a recent adult study: Eapen et al. (2015) studied long-term outcomes including survival, relapse and TRM for adult patients with hematological malignancies that had a bone marrow transplant. They also found a significant disease effect on relapse.

7. Conclusion. A pseudo-value approach has been proposed to test conditional quantile residual lifetimes for survival and competing risks data. Using Zhou et al. (2012), we extended the method of Logan, Zhang and Klein (2011) to clustered event times and clustered censoring times. The proposed method uses GEE with a sandwich variance estimator to account for correlation within a cluster. The simulation studies show that the proposed method controls Type I errors well and has higher power than an existing method. A bone marrow transplant data set was provided as an example. Although we used the Cox proportional hazards model and the marginal Fine–Gray model to estimate $S(t|\mathbf{Z})$ and $F_1(t|\mathbf{Z})$, respectively, other methods such as the additive hazards models [Yin and Cai (2004)] and the cause-specific hazards models [Prentice et al. (1978)] can also be used to estimate them. It needs a further investigation via extensive simulation studies to examine the performance of the proposed method with various techniques to estimate $S(t|\mathbf{Z})$ and $F_1(t|\mathbf{Z})$. Nonetheless, using those models to estimate $S(t|\mathbf{Z})$ and $F_1(t|\mathbf{Z})$ adds additional assumptions to be satisfied to use the proposed method.

There are several areas for future research. We accounted for covariates by comparing covariate-specific residual quantiles; a method which aggregates this inference across multiple levels of covariates would also be useful. The methods focus on testing, but estimation of confidence intervals for the residual quantiles in the clustered data setting are also important for further investigation. Selecting the most appropriate link function for the pseudo-value approach is a hard problem in practice. Extensive simulation study may be needed to compare the performance of link functions in the future. The proposed method is a marginal model. Developing a pseudo-value approach dealing with random effects is another interesting problem. Finally, the proposed pseudo-value method assumes that censoring is independent of covariates. He et al. (2016) and Binder, Gerds and Andersen (2014) have recently studied the subdistribution hazards model and pseudo-values, respectively, allowing covariate-dependent censoring. Using these methods, the proposed method may be extended for covariate-dependent censoring data.

SUPPLEMENTARY MATERIAL

Supplemental materials (DOI: 10.1214/16-AOAS927SUPP; .pdf). The online Supplementary Materials are available with this paper at the *Annals of Applied Statistics* website.

REFERENCES

AHN, K. W. and LOGAN, B. R. (2016). Supplement to "Pseudo-value approach for conditional quantile residual lifetime analysis for clustered survival and competing risks data with applications to bone marrow transplant data." DOI:10.1214/16-AOAS927SUPP.

AHN, K. W. and MENDOLIA, F. (2014). Pseudo-value approach for comparing survival medians for dependent data. *Stat. Med.* **33** 1531–1538. MR3240767

ANDERSEN, P. K., KLEIN, J. P. and ROSTHØJ, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* **90** 15–27. MR1966547

BINDER, N., GERDS, T. A. and ANDERSEN, P. K. (2014). Pseudo-observations for competing risks with covariate dependent censoring. *Lifetime Data Anal.* **20** 303–315. MR3181016

BRESLOW, N. E. (1972). Discussion of the paper by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34** 216–217.

COMMENGES, D. and ANDERSEN, P. K. (1995). Score test of homogeneity for survival data. *Lifetime Data Anal.* **1** 145–159. MR1353846

COX, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220. MR0341758

EAPEN, M., LOGAN, B. R., APPELBAUM, F. R., ANTIN, J. H., ANASETTI, C., COURIEL, D. R., CHEN, J., MAZIARZ, R. T., McCARTHY, P. L., NAKAMURA, R., RATANATHARATHORN, V., VIJ, R. and CHAMPLIN, R. E. (2015). Long-term survival after transplantation of unrelated donor peripheral blood or bone marrow hematopoietic cells for hematologic malignancy. *Biol. Blood Marrow Transplant.* **21** 55–59.

FERRY, C., GEMAYEL, G., ROCHA, V., LABOPIN, M., ESPEROU, H., ROBIN, M., DE LATOUR, R. P., RIBAUD, P., DEVERGIE, A., LEBLANC, T., BARUCHEL, E. G. A. and SOCIE, G. (2007). Long-term outcomes after allogeneic stem cell transplantation for children with hematological malignancies. *Bone Marrow Transplant.* **40** 219–224.

HE, P., ERIKSSON, F., SCHEIKE, T. H. and ZHANG, M. J. (2016). A proportional hazards regression model for the subdistribution with covariates–adjusted censoring weight for competing risks data. *Scand. J. Stat.* **43** 103–122.

JACOBSEN, M. and MARTINUSSEN, T. (2014). A note on the large sample properties of estimators based on generalized linear models for correlated pseudo-observations. Research Report 14/4. Dept. Biostatistics, Univ. Copenhagen.

JACOBSOHN, D. A. (2015). Outcomes of pediatric bone marrow transplantation for leukemia and myelodysplasia using matched sibling, mismatched related, or matched unrelated donor. *Bone Marrow Transplant.* **50** 749–750.

JEONG, J.-H. and FINE, J. P. (2009). A note on cause-specific residual life. *Biometrika* **96** 237–242. MR2482149

JEONG, J.-H. and FINE, J. P. (2013). Nonparametric inference on cause-specific quantile residual life. *Biom. J.* **55** 68–81. MR3042385

KIM, M.-O., ZHOU, M. and JEONG, J.-H. (2012). Censored quantile regression for residual lifetimes. *Lifetime Data Anal.* **18** 177–194. MR2903719

KLEIN, J. P. and ANDERSEN, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* **61** 223–229. MR2135864

LEE, E. W., WEI, L. J. and AMATO, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis*: *State of the Art* (*Columbus*, *OH*, 1991). *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* **211** 237–247. Kluwer Academic, Dordrecht. MR1175646

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

LIN, D. Y. (2007). On the Breslow estimator. *Lifetime Data Anal.* **13** 471–480. MR2416534

LIN, C., ZHANG, L. and ZHOU, Y. (2015). Conditional quantile residual lifetime models for right censored data. *Lifetime Data Anal.* **21** 75–96.

LOGAN, B. R., ZHANG, M.-J. and KLEIN, J. P. (2011). Marginal models for clustered time-to-event data with competing risks using pseudovalues. *Biometrics* **67** 1–7. MR2898811

MA, Y. and WEI, Y. (2012). Analysis on censored quantile residual life model via spline smoothing. *Statist. Sinica* **22** 47–68. MR2933167

MAJHAIL, N. S. and RIZZO, J. D. (2013). Surviving the cure: Long term followup of hematopoietic cell transplant recipients. *Bone Marrow Transplant.* **48** 1145–1151.

MARTIN, P. J., COUNTS, G. W., APPELBAUM, F. R., LEE, S. J., SANDERS, J. E., DEEG, H. J., FLOWERS, M. E. D., SYRJALA, K. L., HANSEN, J. A., STORB, R. F. and STORER, B. E. (2010). Life expectancy in patients surviving more than 5 years after hematopoietic cell transplantation. *J. Clin. Oncol.* **28** 1011–1016.

PRENTICE, R. L., KALBFLEISCH, J. D., PETERSON, A. V. JR., FLOURNOY, N., FAREWELL, V. T. and BRESLOW, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34** 541–554.

SHAW, P. J., KAN, F., AHN, K. W., SPELLMAN, S. R., ALJURF, M., AYAS, M., BURKE, M., CAIRO, M. S., CHEN, A. R., DAVIES, S. M., FRANGOUL, H., GAJEWSKI, J., GALE, R. P., GODDER, K., HALE, G. A., HEEMSKERK, M. B. A., HORAN, J., KAMANI, N., KASOW, K. A., CHAN, K. W., LEE, S. J., LEUNG, W. H., LEWIS, V. A., MIKLOS, D., OUDSHOORN, M., PETERSDORF, E. W., RINGDÉN, O., SANDERS, J., SCHULTZ, K. R., SEBER, A., SETTERHOLM, M., WALL, D. A., YU, L. and PULSIPHER, M. A. (2010). Outcomes of pediatric bone marrow transplantation for leukemia and myelodysplasia using matched sibling, mismatched related, or matched unrelated donors. *Blood* **116** 4007–4015.

SPIEKERMAN, C. F. and LIN, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.* **93** 1164–1175. MR1649210

YIN, G. and CAI, J. (2004). Additive hazards model with multivariate failure time data. *Biometrika* **91** 801–818. MR2126034

ZEGER, S. L., LIANG, K.-Y. and ALBERT, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44** 1049–1060. MR0980999

ZENG, D. and LIN, D. Y. (2008). Efficient resampling methods for nonsmooth estimating functions. *Biostat.* **9** 355–363.

ZHAO, Y. Q., ZENG, D., LABER, E. B., SONG, R., YUAN, M. and KOSOROK, M. R. (2015). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika* **102** 151–168. MR3335102

ZHOU, B., FINE, J., LATOUCHE, A. and LABOPIN, M. (2012). Competing risks regression for clustered data. *Biostat.* **13** 371–383.

DIVISION OF BIOSTATISTICS
MEDICAL COLLEGE OF WISCONSIN
8701 WATERTOWN PLANK ROAD
MILWAUKEE, WISCONSIN 53226
USA
E-MAIL: kwooahn@mcw.edu
          blogan@mcw.edu