Statistica Sinica Preprint No: SS-2025-0478	
Title	Non-parametric Testing for Survival Data with Time-dependent Covariates
Manuscript ID	SS-2025-0478
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0478
Complete List of Authors	Ying Cui, HuiChuan Lai and Limin Peng
Corresponding Authors	Limin Peng
E-mails	lpeng@emory.edu
Notice: Accepted author version.	

Statistica Sinica

Non-parametric Testing for Survival Data With Time-dependent Covariates

Ying Cui[1], HuiChuan J Lai[2] and Limin Peng[1] †

[1] *Emory University and* [2] *University of Wisconsin-Madison*

Abstract: A time-dependent covariate (e.g., time-varying treatment or exposure) is often encountered in survival studies within biomedical research. The varying nature of the time-dependent covariate, evolving alongside the survival outcome, poses extra complications in assessing the covariate-survival association. In this work, we propose a new nonparametric testing framework that is designed to robustly evaluate the effect of a time-dependent covariate on a survival outcome. By adopting the landmark perspective and utilizing a generalized interval quantile correlation index, our testing procedure does not require parametric or semiparametric modeling of the relationship between the time-dependent covariate and the survival outcome, while flexibly accommodating dynamic covariate effects on the survival outcome. We provide theoretical justifications for our proposals. The new method is applied to probe the effect of time-varying breastmilk and infant formula feeding patterns on a key pulmonary outcome of young children with cystic fibrosis in their first 3 years of life.

Key words and phrases: Landmark analysis; Nonparametric hypothesis testing; Survival outcome; Time-dependent covariate.

† Corresponding author: Limin Peng, Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, U.S.A. E-mail: lpeng@emory.edu. ORCID: 0000-0002-8042-3112

1. Introduction

Time-dependent covariates (e.g., time-varying treatment or exposure) are often encountered in biomedical follow-up studies where a survival (or time-to-event) outcome and covariates are both tracked over time. Understanding how a time-dependent covariate is related to the progression to the survival endpoint is often of crucial scientific interests. For example, Feeding Infants Right... from the STart (FIRST) study, initiated in 2012, is a prospective observational study conducted in six cystic fibrosis (CF) centers in the United States (Lai et al., 2025). Under the overarching goal to identify optimal feeding for infants with CF, breastmilk and formula feeding data were prospectively collected since the neonatal period along with many clinical and biological markers. Of particular interest is the long-standing controversy regarding exclusive breastfeeding for infants with CF (Colombo et al., 2021): while breast milk offers superior immunologic benefits against respiratory infections (Meek et al., 2022; Jadin et al., 2011), its caloric and nutrient density may be inadequate to meet the increased requirement resultant from loss via malabsorption in infants with CF (Bronstein et al., 1992; Fustik et al., 2009). It is intriguing to investigate how the complex breastmilk/formula feeding trajectory, modeled as a time-dependent covariate, interplays with one of the most important pulmonary measure in CF, i.e., time to first pseudomonas aeruginosa (PA) infection.

As discussed by Fisher and Lin (1999), studying the relationship between a time-dependent covariate and a time-to-event outcome involves multifold complications. One issue lies in formulating the covariate-outcome relationship, which is not straightforward because the time-

dependent covariate can change in relation to the underlying survival progression and vice versa. Such co-evolution of the time-dependent covariate and the survival outcome also poses challenges for interpretation, for example, in terms of causal or prognosis mechanisms.

Regression modeling is a common strategy to study the relationship between a time-dependent covariate and a survival outcome. A popular approach is to formulate the effect of the time-dependent covariate on the hazard function of the survival outcome, such as the Cox proportional hazards (PH) model with time-dependent covariates (Andersen and Gill, 1982; Therneau and Grambsch, 2000; Cox and Oakes, 2018). A relevant extension is joint modeling of survival and longitudinal data, where the time-dependent covariate itself is separately modeled as a longitudinal outcome, in addition to a hazard-based survival regression model (Wulfsohn and Tsiatis, 1997; Hsieh et al., 2006; Song and Wang, 2008). Some authors have also studied regression models that directly link time-dependent covariates to event time or event time quantiles, for example, the accelerated failure time (AFT) model with time-dependent covariates (Robins and Rotnitzky, 1992; Zeng and Lin, 2007, for example) and the quantile regression model with time-dependent covariates (Gorfine et al., 2017; Cai and Sit, 2020; Chu and Sit, 2023, among others). While accommodating many realistic survival scenarios with time-dependent covariates, these model-based approaches are still prone to issues related to potential model misspecification, which can undermine both the validity and interpretation of the resulting inferences.

The landmark approaches serve as another commonly used approach for handling time-

dependent covariates (Zheng and Heagerty, 2005; Van Houwelingen, 2007; Parast et al., 2012, for example). The basic idea of landmark analysis is to consider a set of landmark time points and treat a time-dependent covariate at each landmark time point as a time-fixed covariate. Compared to the regression modeling approaches mentioned above, landmark approaches are typically more intuitive and easier to interpret, providing a dynamic, local view of the effect of a time-dependent covariate (Putter, 2013; Putter and van Houwelingen, 2017). While conferring a simple and practical perspective for evaluating time-dependent covariate, most of existing landmark analyses involve parametric or semiparametric modeling of the survival time from the landmark onward at multiple landmark time points. This may impose undesriable data restrictions, particularly in the presence of continuous time-dependent covariates, and may compromise the soundness of their effect assessment.

In this paper, we propose a nonparametric hypothesis testing framework to help evaluate the relationship between a time-dependent covariate and a time-to-event outcome. We adopt the landmark perspective and develop a model-free strategy to capture and aggregate the covariate effect at each landmark time through devising generalized interval quantile correlation (GIQC) indices. The new GIQC indices inherit the robustness of the existing interval quantile correlation (IQC) index (Zhu et al., 2018) in capturing quantile-varying association, and are carefully designed to sensibly account for the co-evolution of a time-to-event outcome and a time-dependent covariate (which is allowed to be either continuous or discrete) under the landmark perspective. Constructed based on the GIQC indices, the proposed test

statistics are fully non-parametric and appropriately adjust for censoring to the time-to-event outcome. We establish the limit null distributions of the new test statistics, and prove the testing consistency under a general class of alternative hypotheses. A resampling procedure is developed to approximate the limit null distribution and obtain the p values for the proposed tests.

The details about the proposed testing procedures and theoretical studies are discussed in Section 2. We conduct extensive simulation studies to evaluate the finite-sample performance of our proposals, which are reported in Section 3. In Section 4, we present an application of the proposed method to the FIRST study, which investigates how breastmilk and infant formula feeding history is related to a pulmonary outcome measure of young children with CF in their first 3 years of life. Some remarks are concluded in Section 5.

2. The Proposed Testing Framework

2.1 Formulation of the testing problem

Let T denote the survival time (or time-to-event) and C denote the censoring time. Define $X = \min(T, C)$ and $\delta = I(T \leq C)$. Let $V(t)$ denote a time-dependent covariate (discrete or continuous) or a summary statistic of the time-dependent covariate up to time t. Let $\bar{\boldsymbol{Z}} = \{V(t), 0 \leq t < \infty\}$, which represents the whole time-dependent covariate process.

Let $\mathcal{T} = \{t_k\}_{k=1}^{K}$ denote a set of landmark time points with $K < \infty$. At each landmark time t_k, we define the residual survival time as $T^r(t_k) = T - t_k$, the residual censoring

time as $C^r(t_k) = C - t_k$. The observed residual survival time is then defined as $X^r(t_k) = \min\{T^r(t_k), C^r(t_k)\}$. We assume $\{T^r(t_k), V(t_k)\} \perp C \mid T > t_k$ for any $t_k \in \mathcal{T}$. The data observed at time t_k consist of $\{X_i^r(t_k), \delta_i, V_i(t_k)\}_{\{i;X_i>t_k,i=1\ldots,n\}}$.

Taking the landmark perspective, we assess the relationship between the time-dependent covariate and the survival time by evaluating the influence of the available covariate information on the subsequent progression to the survival endpoint at each landmark time point. Following this rationale, we consider a null hypothesis of the form,

$$H_{0,\mathcal{T}} : Q_{T^r(t_k)|V(t_k),T>t_k}(\tau) = Q_{T^r(t_k)|T>t_k}(\tau) \text{ for } \tau \in \Delta, t_k \in \mathcal{T}, \tag{2.1}$$

where $\Delta \equiv [\tau_L, \tau_U] \subseteq (0, 1)$ is a pre-specified set of quantile levels. The general principle for selecting Δ is to align it with the scientific questions of interest. For example, we may choose $\Delta = [0.4, 0.6]$ to focus on middle-range outcomes, or $\Delta = [0.7, 0.9]$ to target upper-tail outcomes. Here and hereafter, for a random variable Y, we use notation $Q_{Y|W}(\tau)$ to denote the τ-th conditional quantile of Y given W (or the corresponding condition holds). By the formulation in (2.1), $H_{0,\mathcal{T}}$ asserts the interval quantile independence (Zhu et al., 2018) between $T^r(t_k)$ and $V(t_k)$ at each landmark time t_k. A sensible interpretation of $H_{0,\mathcal{T}}$ is that at each landmark time, covariate information in the past does not influence all relevant quantiles of the residual survival time (as captured by Δ).

Of note, $H_{0,\mathcal{T}}$ reduces to the statistical independence between $T^r(t_k)$ and $V(t_k)$ at each landmark time t_k (i.e., $T^r(t_k) \perp V(t_k)$, $t_k \in \mathcal{T}$) when $\Delta = (0, 1)$. In the presence of right censoring to T, $Q_{T^r(t_k)|T>t_k}(\tau)$ may not be identifiable for τ's close to 1. Thus, in

survival settings, it is more tangible to address $H_{0,\mathcal{T}}$ than testing the traditional statistical independence provided that Δ is properly selected to exclude upper tail quantile levels (e.g., $\Delta = [0.2, 0.8]$) . As another practical appeal, targeting $H_{0,\mathcal{T}}$ only requires the information of time-dependent covariates at landmark time points, and thus naturally accommodates practical scenarios with intermittently observed time-dependent covariates. Moreover, $H_{0,\mathcal{T}}$ does not involve any specific modeling of how $V(t_k)$ is associated with the residual survival time $T^r(t_k)$. This model-free feature helps enhance the robustness of the implications derived from testing $H_{0,\mathcal{T}}$.

2.2 Generalized interval quantile correlation indices and the proposed test statistics

To construct test statistics for $H_{0,\mathcal{T}}$, we employ the following key fact:

$$\begin{aligned} & Q_{T^r(t_k)|V(t_k),T>t_k}(\tau_1) = Q_{T^r(t_k)|T>t_k}(\tau_1) \\ \Leftrightarrow & \int_{\Omega_{V(t_k)}} \frac{c^2(\tau_1, v; t_k)}{\tau_1(1-\tau_1)F_2(v;t_k)(1-F_2(v;t_k))} d\mu_2(v) = 0, \end{aligned} \tag{2.2}$$

where $\Omega_{V(t_k)}$ denotes the support of $V(t_k)$, $c(\tau_1, v; t_k) = \text{Cov}\{I(F_1(T^r(t_k); t_k) \le \tau_1), I(V(t_k) \le v) \mid X > t_k\}$, $F_1(s; t_k) = P(T^r(t_k) \le s \mid X > t_k)$, $F_2(v; t_k) = P(V(t_k) \le v \mid X > t_k)$, and μ_2 is a dominating measure for the probability measure induced by $V(t_k)$. The derivation and justification for (2.2) are provided in the Supplementary Material.

The result in (2.2) follows a rationale similar to that of the IQC index, which was proposed by Zhu et al. (2018) to test the interval quantile independence between two continuous random variables. When $T^r(t_k)$ and $V(t_k)$ are both continuous and have non-zero

density functions, an IQC index would be constructed based on the covariance between $I(T^r(t_k) \leq Q_{T^r(t_k)|X>t_k}(\tau_1))$ and $I(V(t_k) \leq Q_{V(t_k)|X>t_k}(\tau_2))$, which are equivalent to $I(F_1(T^r(t_k); t_k) \leq \tau_1)$ and $I(F_2(V(t_k); t_k) \leq \tau_2)$ respectively. In (2.2), we essentially use $I(V(t_k) \leq v)$ in place of $I(V(t_k) \leq Q_{V(t_k)|X>t_k}(\tau_2))$, implicating a main distinction that we examine the association of $V(t_k)$ with $T^r(t_k)$ directly on the original scale of the covariate instead of transforming it to the probability scale. We adopt this modification to facilitate accommodating discrete $V(t_k)$. When $V(t_k)$ is discrete, the quantile function $Q_{V(t_k)|X>t_k}(\tau_2)$ is no longer continuous in τ_2, and it becomes more complicated to assess an integration that involves $Q_{V(t_k)|X>t_k}(\tau_2)$ over the probability scale of τ_2 than directly evaluating its counterpart over the original scale of $V(t_k)$ as in (2.2). As another useful implication of (2.2), assessing whether or not $Q_{T^r(t_k)|V(t_k),T>t_k}(\tau_1) = Q_{T^r(t_k)|T>t_k}(\tau_1)$ can be equivalently translated to assessing an equality that only involves conditional expectations given $X > t_k$, which is an observable subset, in lieu of $T > t_k$, which is not fully known when T is subject to censoring.

Driven by these considerations, we propose to measure the association between $T^r(t_k)$ and $V(t_k)$ by a modified version of IQC index, i.e., $\int_\Delta \int_{\Omega_{V(t_k)}} \frac{c^2(\tau_1,v;t_k)}{\tau_1(1-\tau_1)F_2(v;t_k)(1-F_2(v;t_k))} d\mu_1(\tau_1)d\mu_2(v)$, where μ_1 is a dominating measure for the probability measure induced by $T^r(t_k)$, and then aggregate the modified IQC measure across different landmark time points by taking either maximum or summation. This results in two generalized interval quantile correlation (GIQC)

indices:

$$q_{max}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) = \max_{t_k \in \mathcal{T}} \int_{\Delta} \int_{\Omega_{V(t_k)}} \frac{c^2(\tau_1, v; t_k)}{\tau_1(1-\tau_1)F_2(v; t_k)(1-F_2(v; t_k))} d\mu_2(v) d\mu_1(\tau_1),$$
$$q_{sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) = \sum_{t_k \in \mathcal{T}} \int_{\Delta} \int_{\Omega_{V(t_k)}} \frac{c^2(\tau_1, v; t_k)}{\tau_1(1-\tau_1)F_2(v; t_k)(1-F_2(v; t_k))} d\mu_2(v) d\mu_1(\tau_1).$$

Compared to the existing IQC index, which is confined to addressing a pair of continuous random variables static over time, the new GIQC indices have expanded capacities to accommodate a time-dependent variable, which can be either continuous or discrete. Moreover, it is easy to show that $q_{max}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$ and $q_{sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$ are always nonnegative and are equal to 0 if and only if $H_{0,\mathcal{T}}$ holds. These properties make GIQC indices an ideal foundation for developing test statistics for $H_{0,\mathcal{T}}$.

To construct test statistics based on the GIQC indices, $q_{max}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$ and $q_{sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$, the main task is to estimate $F_1(s; t_k)$, $F_2(v; t_k)$ and $c(\tau_1, v; t_k)$ for any $t_k \in \mathcal{T}$. To this end, we express these quantities as follows:

$$F_1(s; t_k) = \frac{P(T^r(t_k) \le s, X > t_k)}{\pi_{t_k}}, \quad F_2(v; t_k) = \frac{P(V(t_k) \le v, X > t_k)}{\pi_{t_k}},$$
$$c(\tau_1, v; t_k) = \frac{P(F_1(T^r(t_k); t_k) \le \tau_1, V(t_k) \le v, X > t_k)}{\pi_{t_k}} - \frac{P(F_1(T^r(t_k); t_k) \le \tau_1, X > t_k)}{\pi_{t_k}} \cdot F_2(v; t_k)$$

where $\pi_{t_k} = P(X > t_k)$. When there is no censoring, estimators can be readily derived by replacing the unknown probabilities in the above expressions by their empirical counterparts. When there is censoring, we adopt the idea discussed in Lin et al. (1993) to deal with censoring under the assumption that $\{T^r(t_k), V(t_k)\} \perp C \mid T > t_k$ for any $t_k \in \mathcal{T}$. Specifically, define $q_{1,t_k}(\tau_1) = Q_{T^r(t_k)|X>t_k}(\tau_1)$, $S_{12}(s, v; t_k) = P(T^r(t_k) > s, V(t_k) \le v \mid$

$X > t_k$), and $F_{12}(s, v; t_k) = P(T^r(t_k) \leq s, V(t_k) \leq v \mid X > t_k)$. We can show that $P\{T^r(t_k) > x \mid X > t_k\} = P\{X^r(t_k) > x, X > t_k\}/\{\pi_{t_k} G^r_{C,t_k}(x)\}$ and $P\{T^r(t_k) > x, V(t_k) \leq v \mid X > t_k\} = P\{X^r(t_k) > x, V(t_k) \leq v, X > t_k\}/\{\pi_{t_k} G^r_{C,t_k}(x)\}$, where $G^r_{C,t_k}(x) = P\{C^r(t_k) > x \mid X > t_k\}$. These motivate estimating $F_1(s; t_k)$, $F_2(v; t_k)$, $S_{12}(s, v; t_k)$, and $F_{12}(s, v; t_k)$ respectively by

$$F_{n,1}(s; t_k) = 1 - n^{-1} \sum_{i=1} \frac{I(X_i^r(t_k) > s, X_i > t_k)}{\widehat{G}^r_{C,t_k}(s)\widehat{\pi}_{t_k}}, \quad F_{n,2}(v; t_k) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(V_i(t_k) \leq v, X_i > t_k)}{\widehat{\pi}_{t_k}},$$
$$S_{n,12}(s, v; t_k) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(X_i^r(t_k) > s, V_i(t_k) \leq v, X_i > t_k)}{\widehat{G}^r_{C,t_k}(s)\widehat{\pi}_{t_k}},$$
$$F_{n,12}(s, v; t_k) = F_{n,2}(v; t_k) - S_{n,12}(s, v; t_k),$$

where $\widehat{G}^r_{C,t_k}(s)$ is the Kaplan-Meier estimator of the survival function of $C^r(t_k)$ within the subset $X > t_k$, and $\widehat{\pi}_{t_k} = \frac{1}{n}\sum_{i=1}^{n} I(X_i > t_k)$.

Next, estimating $q_{1,t_k}(\tau_1)$ by $\widehat{q}_{1,t_k}(\tau_1) = \inf\{s \geq 0 : \quad F_{n,1}(s; t_k) \geq \tau_1\}$, we derive an estimator of $c(\tau_1, v; t_k)$, given by

$$\widehat{c}(\tau_1, v; t_k) = F_{n,12}(\widehat{q}_{1,t_k}(\tau_1), v; t_k) - F_{n,1}(\widehat{q}_{1,t_k}(\tau_1); t_k) \cdot F_{n,2}(v; t_k).$$

Finally, we propose to estimate the GIQC indices by

$$\widehat{q}_{max}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) = \max_{t_k \in \mathcal{T}} \widehat{q}(T^r(t_k), V(t_k); \Delta),$$
$$\widehat{q}_{sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) = \sum_{t_k \in \mathcal{T}} \widehat{q}(T^r(t_k), V(t_k); \Delta),$$

where $\widehat{q}(T^r(t_k), V(t_k); \Delta) = \int_\Delta \int_{\Omega_{V(t_k)}} \frac{\widehat{c}^2(\tau_1, v; t_k)}{\tau_1(1-\tau_1)F_{n,2}(v; t_k)(1-F_{n,2}(v; t_k))} d\mu_2(v) d\mu_1(\tau_1)$.

In the following, we discuss the calculation of these estimators when setting μ_1 as the Lebesgue measure and setting μ_2 as the Lebesgue measure if $V(t)$ is continuous and as the counting measure if $V(t)$ is discrete. First, consider the case where $V(t)$ is continuous. It is important to note that $F_{n,2}(v;t_k)$ and $c(\tau_1,v;t_k)$ are piecewise constant functions on $\mathcal{A}_2(t_k)$, where $\mathcal{A}_2(t_k)=\{a_1=-\infty,a_2,\ldots,a_{n_2(t_k)}\}$ and $a_2,\ldots,a_{n_2(t_k)}$ correspond to unique values within the set $\{V_i(t_k);X_i>t_k,i=1\ldots,n\}$ in the increasing order for some $n_2(t_k)\leq n$. We may then calculate $\widehat{q}(T^r(t_k),V(t_k);\Delta)$ as

$$\sum_{j_1=1}^{n}\sum_{j_2=1}^{n_2(t_k)-1}\int_{\Delta\cap[\frac{j_1-1}{n},\frac{j_1}{n})}\int_{[a_{j_2},a_{j_2+1})}\frac{\widehat{c}^2(\tau_1,v;t_k)}{\tau_1(1-\tau_1)F_{n,2}(v)(1-F_{n,2}(v))}d\mu_2(v)d\mu_1(\tau_1)$$
$$=\sum_{j_1=1}^{n}\sum_{j_2=1}^{n_2(t_k)-1}\widehat{c}^2(\frac{j_1}{n},a_{j_2};t_k)\int_{\Delta\cap[\frac{j_1-1}{n},\frac{j_1}{n})}\frac{d\mu_1(\tau_1)}{\tau_1(1-\tau_1)}\int_{[a_{j_2},a_{j_2+1})}\frac{d\mu_2(v)}{F_{n,2}(v)(1-F_{n,2}(v))},$$

while utilizing the facts that $\int_{[a,b)}\frac{1}{\tau_1(1-\tau_1)}d\mu_1(\tau_1)=\int_{[a,b)}\frac{1}{\tau_1(1-\tau_1)}d\tau_1=\{\log(b)-\log(1-b)\}-\{\log(a)-\log(1-a)\}$ for any $a<b$, and $\int_{[a_{j_2},a_{j_2+1})}\frac{1}{F_{n,2}(v)(1-F_{n,2}(v))}d\mu_2(v)=\frac{a_{j_2+1}-a_{j_2}}{F_{n,2}(a_{j_2})(1-F_{n,2}(a_{j_2}))}$. When $V(t)$ is discrete and μ_2 is the counting measure, $\widehat{q}(T^r(t_k),V(t_k);\Delta)$ reduces to

$$\sum_{j_1=1}^{n}\sum_{j_2=1}^{n_2(t_k)-1}\left\{\int_{\Delta\cap[\frac{j_1-1}{n},\frac{j_1}{n})}\frac{d\mu_1(\tau_1)}{\tau_1(1-\tau_1)}\right\}\frac{\widehat{c}^2(\frac{j_1}{n},a_{j_2};t_k)}{F_{n,2}(a_{j_2})(1-F_{n,2}(a_{j_2}))}.$$

Therefore, the proposed estimators, $\widehat{q}_{max}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T})$ and $\widehat{q}_{sum}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T})$, can be easily computed with closed-form expressions.

By the nice connection of $q_{max}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T})$ and $q_{sum}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T})$ to $H_{0,\mathcal{T}}$ discussed earlier, we propose to use $\widehat{q}_{max}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T})$ and $\widehat{q}_{sum}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T})$ as the test statistics for $H_{0,\mathcal{T}}$. The rejection region for $H_{0,\mathcal{T}}$ takes the form, $\widehat{q}_{max}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T})>c$ or $\widehat{q}_{sum}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T})>c$,

where the critical value c is determined based on the limit distribution of the adopted test statistic under the null hypothesis $H_{0,\mathcal{T}}$.

2.3 Asymptotic results

We establish the limiting null distribution of the proposed test statistics in Theorem 1. We also investigate the asymptotic behavior of the proposed test statistics under a general class of alternative hypotheses. The findings are stated in Theorem 2.

Define $f_1(s;t_k) = f_{T^r(t_k)|X>t_k}(s)$ as the conditional density function of $T^r(t_k)$ given $X > t_k$. In the Supplementary Material, we provide detailed descriptions of necessary notations as well as the regularity conditions S1 and S2 assumed by Theorem 1 and Theorem 2.

Theorem 1 *Suppose Conditions S1 and S2 in the Supplementary Material hold. Assume that the first derivative of $f_1(q_{1,t_k}(\tau_1);t_k)$ with respect to τ_1 is bounded away from zero and infinity for $\tau_1 \in \Delta$. Under the null hypothesis $H_{0,\mathcal{T}}$, we have*

$$n\widehat{q}_{max}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T}) \to_d \max_{t_k\in\mathcal{T}} \int_\Delta \int_{\Omega_{V(t_k)}} \chi^2(\tau_1,v;t_k)d\mu_1(\tau_1)d\mu_2(v),$$

$$n\widehat{q}_{sum}(T,\bar{\boldsymbol{Z}};\Delta,\mathcal{T}) \to_d \sum_{t_k\in\mathcal{T}} \int_\Delta \int_{\Omega_{V(t_k)}} \chi^2(\tau_1,v;t_k)d\mu_1(\tau_1)d\mu_2(v),$$

where $\chi(\tau_1,v;t_k)$ is a mean-zero Gaussian process with a detailed definition provided in the Supplementary Material.

Theorem 2 *Suppose Conditions S1 and S2 in the Supplementary Material hold. Assume that the first derivative of* $f_1(q_{1,t_k}(\tau_1); t_k)$ *with respect to* τ_1 *is bounded away from zero and infinity for* $\tau_1 \in \Delta$. *We have*

(A) $\widehat{q}_{max}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$ *is a consistent test statistic against the alternative hypothesis*

$$H_{a,max}: \quad q_{max}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) > 0.$$

(B) $\widehat{q}_{sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$ *is a consistent test statistic against the alternative hypothesis*

$$H_{a,sum}: \quad q_{sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) > 0.$$

By Theorem 2, the proposed test statistics may yield power approaching 1 (as n goes to ∞) in some general alternative cases. Given the smoothness of $q_{1,t_k}(\tau_1)$, a general scenario that ensures the consistency of both test statistics can be described as

$\widetilde{H}_a$: There exists an interval $[\tau_1, \tau_2] \subseteq [\tau_L, \tau_U]$ for $v \in \Omega^*_{V(t_k)} \subseteq \Omega_{V(t_k)}$ with $\mu_2(\Omega^*_{V(t_k)}) > 0$ and $t_k \in \mathcal{T}$ such that $|c(\tau, v; t_k)| > 0$ for $\tau \in [\tau_1, \tau_2]$.

This suggests that the proposed tests can yield high power even when the time-dependent covariates influence only a portion of the outcome distribution, rather than the whole, for some v's and t_k's. This is a desirable feature in the presence of varying covariate-outcome associations. The detailed proofs for Theorems 1 and 2 can be found in the Supplementary Material.

2.4 The resampling-based testing procedure

According to Theorem 1, the limit null distributions of the proposed test statistics are quite complex and are not easy to evaluate analytically. Therefore, we propose a resample-based procedure to approximate the limit null distributions and subsequently obtain p values for testing $H_{0,\mathcal{T}}$.

Specifically, under $H_{0,\mathcal{T}}$, as presented in the Proof of Theorem 1, we can show that

$$\frac{\widehat{c}(\tau_1, v; t_k)}{\sqrt{F_{n,2}(v)(1-F_{n,2}(v))}} \frac{F_{n,12}(\widehat{q}_{1,t_k}(\tau_1), v) - F_{n,1}(\widehat{q}_{1,t_k}(\tau_1)) \cdot F_{n,2}(v)}{\sqrt{F_{n,2}(v)(1-F_{n,2}(v))}} \approx -\frac{1}{n}\sum_{i=1}^{n} \xi_i(\tau_1, v; t_k),$$

where $\approx$ means equivalence uniformly in $(\tau_1, v) \in \Delta \times \Omega_{V(t_k)}$, and

$$\xi_i(\tau_1, v; t_k) = \frac{I(X_i > t_k)}{\pi_{t_k}\sqrt{F_2(v)(1-F_2(v))}} \times$$
$$\{I[X_i^r(t_k) > q_{1,t_k}(\tau_1)][I(V_i(t_k) \le v) - F_2(v)]/G_{C,t_k}^r(q_{1,t_k}(\tau_1)) - (1-\tau_1)[I(V_i(t_k) \le v) - F_2(v)]\}.$$

Utilizing this result, we propose the following resampling procedure:

(a) Generate B independent sets of $\{\iota_i^b\}_{i=1}^n$, where $\{\iota_i^b\}_{i=1}^n$ are independent random variables from a standard normal distribution and $b = 1, 2, \ldots, B$.

(b) Compute the estimates for the influence function as

$$\begin{aligned}\widehat{\xi}_i(\tau_1, v; t_k) \;=\; & \frac{I(X_i > t_k)}{\widehat{\pi}_{t_k}\sqrt{F_{n,2}(v)(1-F_{n,2}(v))}} \times \\ & \{[I(X_i^r(t_k) > \widehat{q}_{1,t_k}(\tau_1), V_i(t_k) \le v) - F_{n,2}(v) I(X_i^r(t_k) > \widehat{q}_{1,t_k}(\tau_1))]/ \\ & \widehat{G}_{C,t_k}^r[\widehat{q}_{1,t_k}(\tau_1)] - (1-\tau_1) I[V_i(t_k) \le v] + (1-\tau_1) F_{n,2}(v)\}.\end{aligned}$$

(c) For $b = 1, 2, \ldots, B$, calculate

$$n\widehat{q}_{max,b}(T^r, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) = \max_{t_k \in \mathcal{T}} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n_2(t_k)-1} \{n^{-1/2} \sum_{i=1}^{n} \widehat{\xi}_i(\frac{j_1}{n}, a_{j_2}; t_k)\iota_i^b\}^2 \times \int_{\Delta \cap [\frac{j_1-1}{n}, \frac{j_1}{n})} \frac{1}{\tau_1(1-\tau_1)} d\mu_1(\tau_1),$$

$$n\widehat{q}_{sum,b}(T^r, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) = \sum_{t_k \in \mathcal{T}} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n_2(t_k)-1} \{n^{-1/2} \sum_{i=1}^{n} \widehat{\xi}_i(\frac{j_1}{n}, a_{j_2}; t_k)\iota_i^b\}^2 \times \int_{\Delta \cap [\frac{j_1-1}{n}, \frac{j_1}{n})} \frac{1}{\tau_1(1-\tau_1)} d\mu_1(\tau_1).$$

the empirical distributions of which approximate the corresponding limit null distributions.

(d) The p value is calculated as $\frac{1}{B}\sum_{b=1}^{B} I[n\widehat{q}_{\cdot,b}(T^r, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) > n\widehat{q}_{\cdot}(T^r, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})]$, where "$\cdot$" represents either "max" or "sum".

Similar resampling ideas have been used in other work, for example, Lin et al. (1993), Li and Peng (2014) and Cui and Peng (2022). The rigorous theoretical justification for the presented resampling procedure is provided in the Supplementary Material.

3. Numerical Studies

3.1 Data generation procedure

To simulate data with time-dependent covariates and time-to-event outcome, one option is to adopt the algorithm in Hendry (2014) to generate data from the time-dependent Cox model, for which the key strategy is to perform rejection resampling according to the truncated piece-wise exponential distribution. To further allow for varying covariate effects, we propose a general procedure to generate the survival time sequentially based on the residual quantile

Input: $\bar{\boldsymbol{Z}}_i = \{V_i(t_k)\}_{k=1}^K$, $\eta(\tau)$ and σ

Output: T_i

Initialize with $k \leftarrow 1$, $\tau_{C,i}^{(k)} \leftarrow 0$ and $d \leftarrow 0$;

while $k \leq K$ *and* $d = 0$ **do**

Generate $\tau_{k,i} \sim Uniform(0,1)$;

Define $\xi_{k,i}(\tau; t_k) \leftarrow \eta\left[\tau_{C,i}^{(k)} + (1-\tau_{C,i}^{(k)})\tau\right]$;

Calculate the potential residual survival time at t_k, noted by $T_i^{r*}(t_k)$, with

$T_i^{r*}(t_k) \leftarrow \exp\left\{V_i(t_k)\xi_{k,i}(\tau_{k,i}; t_k) + \sigma\Phi^{-1}(\tau_{k,i})\right\}$;

Obtain $\tau_{C,i}^{(k+1)}$ by solving the equation:

$$t_{k+1} - t_k = \exp\left\{V_i(t_k)\eta\left[\tau_{C,i}^{(k+1)}\right] + \sigma\Phi^{-1}\left(\frac{\tau_{C,i}^{(k+1)} - \tau_{C,i}^{(k)}}{1-\tau_{C,i}^{(k)}}\right)\right\};$$

if $T_i^{r*}(t_k) > t_{k+1} - t_k$ **then**

$k \leftarrow k+1$;

else

$T_i \leftarrow t_k + T_i^{r*}(t_k)$;

$d \leftarrow 1$

Algorithm 1: Simulation Procedure of T_i for any $\bar{\boldsymbol{Z}}_i$.

regression model, which takes the form,

$$Q_{\log\{T^r(t_k)\}|V(t_k)}(\tau) = V(t_k)\xi_k(\tau; t_k) + \sigma\Phi^{-1}(\tau),$$

where $\xi_k(\tau; t_k)$ is the coefficient function of τ given the landmark time t_k, $\Phi^{-1}(\cdot)$ represents the quantile function of standard normal distribution, and σ captures the standard deviation of the error term.

Specifically, for each individual, we first generate $\bar{\boldsymbol{Z}}_i$ on a pre-specified sequence of time points, $\{t_k\}_{k=1}^{K+1}$ with $t_1 = 0$ and $t_{K+1} = \infty$, and then simulate the survival time T_i based on a sequential procedure. In Algorithm 1, we present the proposed sequential procedure to determine T_i for any $\bar{\boldsymbol{Z}}_i$. The basic idea is that we generate the residual survival time sequentially for each t_k. At each t_k, if the residual survival time is longer than $t_{k+1} - t_k$, then we continue to t_{k+1}. Otherwise, we stop and output the simulated event time. We employ an intermediate function $\eta(\tau)$, which is specified as an input, to define the form of $\xi_k(\tau; t_k)$.

After obtaining the simulated T_i, we generate the censoring time C_i independently from a Uniform distribution, and calculate the observed survival time $X_i = \min\{T_i, C_i\}$ and the event status $\delta_i = I(T_i \leq C_i)$.

3.2 Simulation results with continuous time-dependent covariates

Firstly, we considered the setting with continuous time-dependent covariates. The covariate process $\bar{\boldsymbol{Z}}_i = \{V_i(t), 0 \leq t < \infty\}$ was generated to be piecewise constant on a predetermined set of time points, $\{t_1, t_2, \ldots, t_J\} = \{0, 0.5, 1, \ldots, 5\}$, and $\boldsymbol{V}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, where $\boldsymbol{V}_i =$

$\{V_i(t_1), \ldots, V_i(t_J)\}$, $\boldsymbol{\mu}_i = \{\gamma_i \mu(t_1), \ldots, \gamma_i \mu(t_J)\}$ with $\mu(t) = \frac{4t}{5} I(0 \leq t \leq 2.5) + 2I(t > 2.5)$ and $\gamma_i \sim Uniform(-0.5, 0.5)$, and $\boldsymbol{\Sigma} = \{\sigma_{kl}\}_{k,l \in \{1,\ldots,J\}}$ with $\sigma_{kl} = 0.01 I(k = l) + 0.0095 I(k \neq l)$. We considered the following simulation set-ups:

Set-up I: Null case without covariate effects;

Set-up II: A case where the time-dependent Cox model holds;

Set-up III and IV: Cases with dynamic covariate effect.

Specifically, in Set-up II, T_i was generated to follow the standard Cox PH model with the time-dependent covariate $V_i(t)$ based on the procedure described in Hendry (2014). In Set-ups I, III and IV, we adopted the simulation procedures described in Section 3.1. The intermediate function $\eta(\tau)$ was set as $\eta(\tau) \equiv 0$ in Set-up I. In Set-ups III and IV, the $\eta(\tau)$ functions were set to be varying over τ (see Web Figure S1). We set $\sigma = 0.5$. The censoring time C was generated from a Uniform distribution to yield 15% or 30% censoring rate.

With each generated data set, we applied the proposed testing procedure based on the test statistic $\widehat{q}_{max}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$ or $\widehat{q}_{sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$, which is denoted by GIQC(MAX) or GIQC(SUM) respectively. For comparison, we also applied the Wald test based on the Cox PH model with time-dependent covariates, denoted by CPH(Wald). We considered sample sizes, $n = 200$ and 400, three sets of landmark time points, $R_1 = \{0, 2\}$, $R_2 = \{0, 1, 2\}$ and $R_3 = \{0, 0.5, 1, 1.5, 2\}$, and significance levels, $\alpha = 0.05$ and 0.1.

Table 1 summarizes the empirical rejection rates based on 1000 replicates. The results

for Set-up I show that GIQC(MAX), GIQC(SUM) and CPH(Wald) all have empirical sizes quite close to the nominal levels of 0.05 or 0.1, even with the relatively small sample size, 200. When the Cox model holds (i.e., Set-up II), CPH(Wald) exhibits higher empirical power as compared to the proposed methods (e.g., the setting with two landmark time points and $n = 200$). This is not surprising because the proposed tests are nonparametric and do not utilize information tailored to the Cox PH model structure. We note that the advantage of CPH(Wald) in power under Set-up II diminishes as the sample size and the number of landmark time points increase. In Set-ups III and IV, the effect of $\bar{\boldsymbol{Z}}$ is more dynamic and the Cox PH model does not hold. In these cases, CPH(Wald) can have poor power to detect the dynamic effect of $\bar{\boldsymbol{Z}}$ (e.g., 4.5% and 10.7% empirical power in Set-up IV respectively with $\alpha = 0.05$ and 0.1 when $n = 200$ and censoring rate is 15%), while the proposed GIQC(MAX) and GIQC(SUM) can yield much higher power (e.g., >80% power with landmark time set R_3 under the same settings). Moreover, we notice that the power of the proposed tests may increase when more landmark time points are considered. This may be because more information would be utilized to gauge the association between the time-dependent covariate and the survival time. Finally, we note that the two proposed tests, GIQC(MAX) and GIQC(SUM), generally have comparable performance. In the presented simulation settings, GIQC(SUM) seems to be slightly more powerful than GIQC(MAX).

Table 1: Empirical rejection rates with continuous time-dependent covariate.

Set-up	n	α	Method	30% Censoring			15% Censoring		
				R_1	R_2	R_3	R_1	R_2	R_3
I	200	0.05	GIQC(MAX)	0.056	0.053	0.055	0.045	0.044	0.043
			GIQC(SUM)	0.059	0.052	0.054	0.049	0.054	0.048
			CPH(Wald)	0.041	0.041	0.041	0.051	0.051	0.051
		0.1	GIQC(MAX)	0.103	0.098	0.100	0.109	0.107	0.107
			GIQC(SUM)	0.102	0.098	0.102	0.110	0.104	0.091
			CPH(Wald)	0.091	0.091	0.091	0.098	0.098	0.098
	400	0.05	GIQC(MAX)	0.045	0.045	0.042	0.054	0.055	0.053
			GIQC(SUM)	0.044	0.052	0.040	0.052	0.058	0.057
			CPH(Wald)	0.044	0.044	0.044	0.054	0.054	0.054
		0.1	GIQC(MAX)	0.100	0.104	0.100	0.109	0.106	0.106
			GIQC(SUM)	0.100	0.098	0.098	0.100	0.096	0.103
			CPH(Wald)	0.096	0.096	0.096	0.106	0.106	0.106
II	200	0.05	GIQC(MAX)	0.529	0.548	0.543	0.683	0.739	0.734
			GIQC(SUM)	0.718	0.935	0.968	0.864	0.983	0.992
			CPH(Wald)	0.983	0.983	0.983	0.992	0.992	0.992

Continued on next page

Table 1 – continued from previous page

Set-up	n	α	Method	30% Censoring			15% Censoring		
				R_1	R_2	R_3	R_1	R_2	R_3
		0.1	GIQC(MAX)	0.704	0.746	0.738	0.837	0.894	0.890
			GIQC(SUM)	0.868	0.987	0.996	0.949	0.997	0.999
			CPH(Wald)	0.992	0.992	0.992	0.996	0.996	0.996
	400	0.05	GIQC(MAX)	0.923	0.948	0.948	0.979	0.994	0.993
			GIQC(SUM)	0.990	0.999	1.000	0.998	1.000	1.000
			CPH(Wald)	1.000	1.000	1.000	1.000	1.000	1.000
		0.1	GIQC(MAX)	0.967	0.990	0.989	0.997	0.999	0.999
			GIQC(SUM)	0.999	1.000	1.000	1.000	1.000	1.000
			CPH(Wald)	1.000	1.000	1.000	1.000	1.000	1.000
III	200	0.05	GIQC(MAX)	0.290	0.299	0.308	0.452	0.473	0.502
			GIQC(SUM)	0.304	0.423	0.482	0.491	0.639	0.680
			CPH(Wald)	0.106	0.106	0.106	0.328	0.328	0.328
		0.1	GIQC(MAX)	0.439	0.455	0.475	0.625	0.658	0.670
			GIQC(SUM)	0.471	0.608	0.657	0.673	0.787	0.820
			CPH(Wald)	0.195	0.195	0.195	0.463	0.463	0.463
	400	0.05	GIQC(MAX)	0.610	0.632	0.658	0.833	0.867	0.881

Continued on next page

Table 1 – continued from previous page

Set-up	n	α	Method	30% Censoring			15% Censoring		
				R_1	R_2	R_3	R_1	R_2	R_3
			GIQC(SUM)	0.653	0.806	0.839	0.886	0.952	0.970
			CPH(Wald)	0.244	0.244	0.244	0.633	0.633	0.633
		0.1	GIQC(MAX)	0.756	0.787	0.804	0.933	0.953	0.952
			GIQC(SUM)	0.792	0.901	0.923	0.951	0.986	0.986
			CPH(Wald)	0.395	0.395	0.395	0.751	0.751	0.751
IV	200	0.05	GIQC(MAX)	0.563	0.578	0.593	0.803	0.827	0.836
			GIQC(SUM)	0.608	0.744	0.771	0.845	0.929	0.948
			CPH(Wald)	0.051	0.051	0.051	0.045	0.045	0.045
		0.1	GIQC(MAX)	0.705	0.727	0.737	0.887	0.913	0.918
			GIQC(SUM)	0.743	0.838	0.868	0.922	0.973	0.982
			CPH(Wald)	0.097	0.097	0.097	0.107	0.107	0.107
	400	0.05	GIQC(MAX)	0.873	0.900	0.916	0.984	0.994	0.995
			GIQC(SUM)	0.922	0.972	0.987	0.993	1.000	1.000
			CPH(Wald)	0.067	0.067	0.067	0.076	0.076	0.076
		0.1	GIQC(MAX)	0.948	0.962	0.966	0.996	1.000	0.999
			GIQC(SUM)	0.963	0.993	0.997	0.998	1.000	1.000

Continued on next page

Table 1 – continued from previous page

Set-up	n	α	Method	30% Censoring			15% Censoring		
				R_1	R_2	R_3	R_1	R_2	R_3
			CPH(Wald)	0.099	0.099	0.099	0.136	0.136	0.136

3.3 Simulation results with discrete time-dependent covariates

Next, we investigated the case with discrete time-dependent covariates. We generated the covariate process $\bar{\boldsymbol{Z}}_i^* = \{V_i^*(t), 0 \leq t < \infty\}$ in the same way as that for the continuous covariate case under Set-ups I-IV, and then converted $\bar{\boldsymbol{Z}}_i^*$ to $\bar{\boldsymbol{Z}}_i = \{V_i(t), 0 \leq t < \infty\}$ by letting $V_i(t) = (-1) \times I[V_i^*(t) \leq -0.25] + 0 \times I[-0.25 < V_i^*(t) \leq 0] + 1 \times I[0 < V_i^*(t) \leq 0.25] + 2 \times I[V_i^*(t) > 0.25]$. We set $\sigma = 0.5$ and generated the censoring time C from a uniform distribution so that the censoring rate was about 30%. For each generated data, we applied the proposed GIQC(MAX) and GIQC(SUM) and compared the results with those from CPH(Wald). We considered three sets of landmark time points, $R_1 = \{0, 2\}$, $R_2 = \{0, 1, 2\}$ and $R_3 = \{0, 0.5, 1, 1.5, 2\}$, and two different sample sizes, 200 and 400.

Table 2 summarizes the empirical rejection rates based on 1000 replicates. This table shows similar results to those in Table 1, which corresponds to the case with continuous time-dependent covariates. For Set-up I, the empirical sizes are close to the nominal levels. In addition, power increases with larger sample sizes and a greater number of landmark

points. When dynamic covariate effects are present (e.g., Set-up III or IV), the proposed methods demonstrate substantially higher power compared to CPH (Wald).

Table 2: Empirical rejection rates with discrete time-dependent covariate.

Set-up	n	α	Method	30% Censoring		
				R_1	R_2	R_3
I	200	0.05	GIQC(MAX)	0.052	0.051	0.051
			GIQC(SUM)	0.051	0.055	0.054
			CPH(Wald)	0.048	0.048	0.048
		0.1	GIQC(MAX)	0.099	0.100	0.091
			GIQC(SUM)	0.091	0.092	0.087
			CPH(Wald)	0.090	0.090	0.090
	400	0.05	GIQC(MAX)	0.040	0.044	0.040
			GIQC(SUM)	0.037	0.041	0.042
			CPH(Wald)	0.048	0.048	0.048
		0.1	GIQC(MAX)	0.082	0.079	0.085
			GIQC(SUM)	0.077	0.073	0.086
			CPH(Wald)	0.111	0.111	0.111
II	200	0.05	GIQC(MAX)	0.700	0.703	0.719

Continued on next page

Table 2 – continued from previous page

Set-up	n	α	Method	30% Censoring		
				R_1	R_2	R_3
			GIQC(SUM)	0.730	0.735	0.906
			CPH(Wald)	1.000	1.000	1.000
		0.1	GIQC(MAX)	0.805	0.813	0.835
			GIQC(SUM)	0.845	0.848	0.957
			CPH(Wald)	1.000	1.000	1.000
	400	0.05	GIQC(MAX)	0.960	0.964	0.981
			GIQC(SUM)	0.979	0.977	0.999
			CPH(Wald)	1.000	1.000	1.000
		0.1	GIQC(MAX)	0.981	0.980	0.996
			GIQC(SUM)	0.992	0.991	1.000
			CPH(Wald)	1.000	1.000	1.000
III	200	0.05	GIQC(MAX)	0.309	0.309	0.436
			GIQC(SUM)	0.460	0.466	0.608
			CPH(Wald)	0.196	0.196	0.196
		0.1	GIQC(MAX)	0.482	0.484	0.643
			GIQC(SUM)	0.698	0.698	0.842

Continued on next page

Table 2 – continued from previous page

Set-up	n	α	Method	30% Censoring		
				R_1	R_2	R_3
			CPH(Wald)	0.312	0.312	0.312
	400	0.05	GIQC(MAX)	0.597	0.597	0.825
			GIQC(SUM)	0.866	0.880	0.968
			CPH(Wald)	0.344	0.344	0.344
		0.1	GIQC(MAX)	0.818	0.819	0.943
			GIQC(SUM)	0.969	0.964	0.997
			CPH(Wald)	0.461	0.461	0.461
IV	200	0.05	GIQC(MAX)	0.260	0.258	0.306
			GIQC(SUM)	0.438	0.438	0.489
			CPH(Wald)	0.239	0.239	0.239
		0.1	GIQC(MAX)	0.425	0.427	0.526
			GIQC(SUM)	0.676	0.668	0.740
			CPH(Wald)	0.343	0.343	0.343
	400	0.05	GIQC(MAX)	0.587	0.592	0.729
			GIQC(SUM)	0.875	0.869	0.940
			CPH(Wald)	0.484	0.484	0.484

Continued on next page

Table 2 – continued from previous page

Set-up	n	α	Method	30% Censoring		
				R_1	R_2	R_3
		0.1	GIQC(MAX)	0.818	0.811	0.907
			GIQC(SUM)	0.962	0.963	0.988
			CPH(Wald)	0.618	0.618	0.618

In summary, simulation results suggest that the proposed method can robustly and effectively detect association between a time-dependent covariate and a time-to-event outcome subject to censoring.

4. Real Example with FIRST Data

CF is a rare life-shortening genetic disease caused by mutations in the CFTR gene (Bell et al., 2020). It is characterized by gastrointestinal and pulmonary manifestations, and malnutrition and growth faltering are common and occur even in infants (Marcus et al., 1991; Lai et al., 2014). Initiated in 2012, FIRST study is a prospective observational study conducted in 6 collaborating CF centers in US (Lai et al., 2025). Under the overarching goal to identify optimal care for infants with CF, FIRST followed CF infants beginning from the neonatal period to capture data on complete feeding history (initial and progression), along with pulmonary status, nutritional, growth, and biochemical markers. One question of interest is how

the feeding history of a CF infant is related to major pulmonary events, such as the acquisition of PA, which is a common pathogen well documented to accelerate lung function decline (Sanders et al., 2014).

To help address this question, we utilized the data from 121 pancreatic insufficient infants with CF in the FIRST study. The survival outcome T is defined as the time to first PA infection (or age at first positive PA from oropharyngeal cultures). With follow-up up to 36 months, the censoring rate for T is around 60%. The Kaplan-Meier curve for T is shown in Figure 1. Let $Z(t)$ denote the feeding type at time t (i.e., B: breastmilk feeding; P: partial breastmilk partial formula feeding; F: formula feeding) extracted from feeding records. Figure 2 illustrates the time-varying feeding histories of 25 FIRST participants. Since feeding likely influences clinical manifestations of CF infants in a cumulative manner and possibly has a delayed effect, we define the time-varying covariate of interest as the following functional of $Z(\cdot)$ (in lieu of $Z(t)$ itself):

$$V(t;x) = \frac{\int_{\max(1,t-3)}^{t} I[Z(u-1)=x]du}{\min(t-1,3)}, \quad t>1,$$

where x can take the value of "B", "P" or "F". Such defined feeding variable represents the proportion of a given feeding type, delayed by one month, during a period of up to three months prior to time t. The goal of our analysis is to investigate the relationship between the time-varying $V(t;x)$ and T. We note that $V(t;x)$ generally only take several distinct values due to ties and the relatively small sample size. For example, with $x =$"B" and $t = 1.5$ (months), among 120 participants at risk, 28 participants had $V(t;x) = 0$, 85 participants

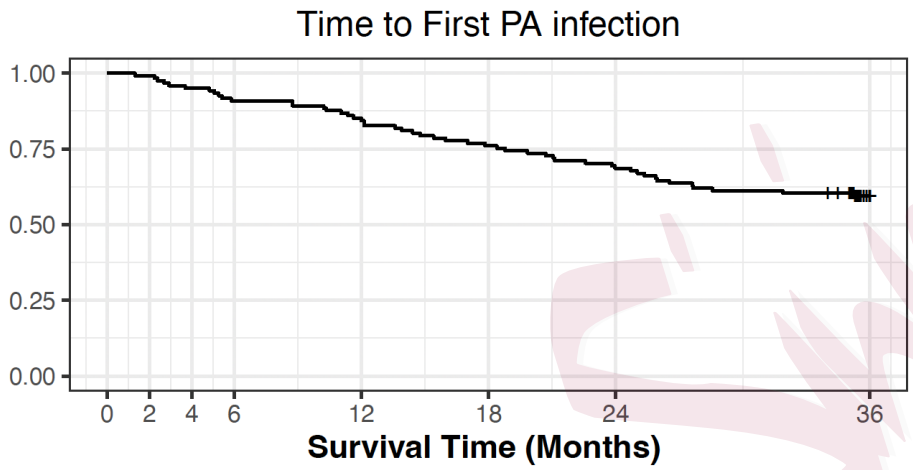

Figure 1: Kaplan-Meier Curve of the Time to first PA infection.

had $V(t;x) = 1$, 4 participants had $V(t;x) = 0.98$, 2 participants had $V(t;x) = .86$, and 1 participant had $V(t;x) = 0.78$. Tailored to this data feature, we set μ_2 as the counting measure when applying the proposed method. We choose landmark times corresponding to the study visit times and set $\mathcal{T} = \{1.5, 2, 3, 4, 5, 6, 7, 9, 11, 13\}$. The size of the risk-set at each landmark time point is reported in Table 3.

Table 3: Number of subjects at risk at each landmark time in the FIRST data example.

Landmark time t_k	1.5	2	3	4	5	6	7	9	11	13
Number at risk	120	120	116	115	114	110	110	108	106	100

We evaluated the relationship between the time-varying feeding variable $V(t;x)$ and the time to first PA infection T using the proposed tests. With $x =$"B", "F", or "P", we denote

4. REAL EXAMPLE WITH FIRST DATA

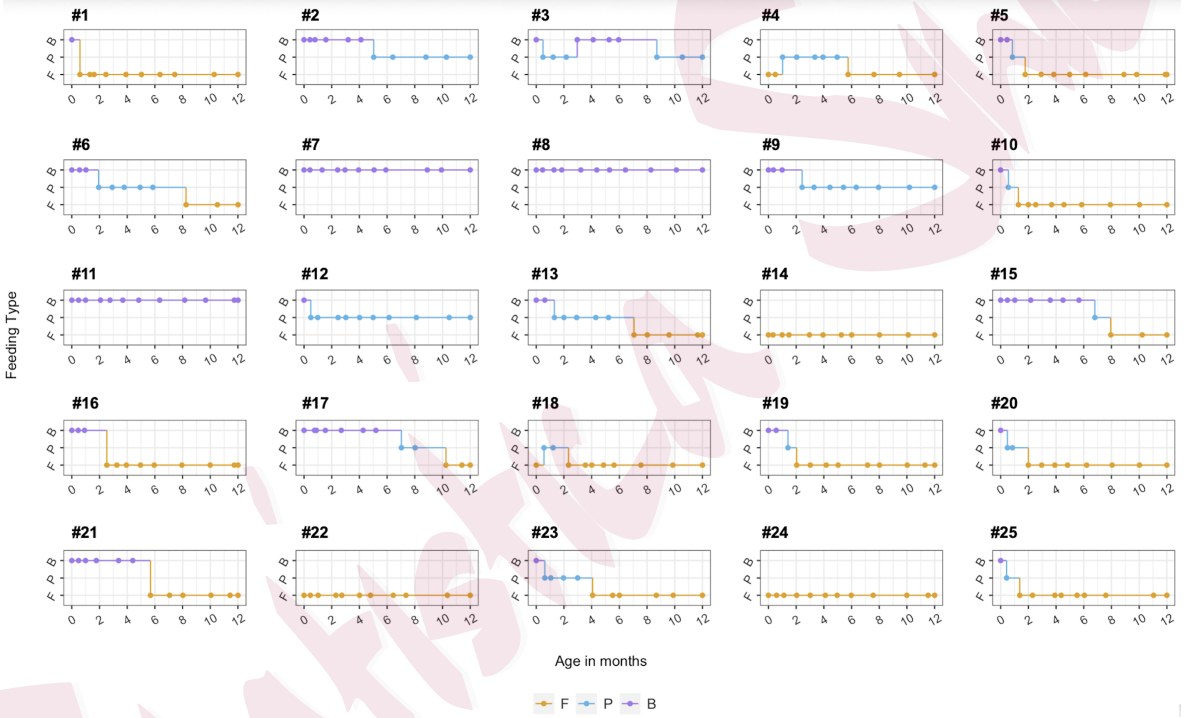

Figure 2: The feeding information (F: formula feeding; P: partial formula feeding; B: breast-milk feeding) over time for the first 25 infants.

the corresponding feeding variable by "P(B)", "P(F)", and "P(P)" respectively. In Table 4, we present the resulting p values with $\Delta = [0.1, 0.4]$, along with the p values obtained by fitting the Cox PH model to T with the time-varying covariate $V(t; x)$. The p values based on GIQC (MAX) suggest that the proportion of breast milk feeding in the past (up to) three months may have some marginal association with the timing of the first PA infection. As indicated by the results of all tests considered, the feeding variable, P(F) or P(P), may have little association with the time to first PA infection.

Note that the landmark time points in $\mathcal{T}$ correspond to scheduled study visit times where the time-dependent feeding covariate $V(t; x)$, representing delayed cumulative feeding history, could also be assessable. We selected $\Delta = [0.1, 0.4]$ primarily based on the observation from Figure 1 that the overall event rate is approximately 40%, suggesting potential identifiability issues for quantiles corresponding to $\tau > 0.4$. With this choice of Δ, we did not encounter the issue related to poorly identified upper-tail residual quantiles. The resulting null hypothesis inquires into the impact of feeding history on early PA acquisitions following the landmark time points and is clinically relevant.

To descriptively examine the direction of the landmark-specific association, we considered the signed summary

$$\rho_x(t_k; \Delta) = \int_{\Delta} \int_{\Omega_{V(t_k;x)}} \frac{c_x(\tau_1, v; t_k)}{\sqrt{\tau_1(1-\tau_1)F_{2,x}(v; t_k)\{1 - F_{2,x}(v; t_k)\}}} \, d\mu_2(v) d\mu_1(\tau_1),$$

where $x \in \{\mathrm{B}, \mathrm{F}, \mathrm{P}\}$, $V(t_k; x)$ is the delayed cumulative feeding-history variable, and $c_x(\tau_1, v; t_k)$ and $F_{2,x}(v; t_k)$ are defined as in Section 2.2 with $V(t_k)$ replaced by $V(t_k; x)$. Unlike the

GIQC index used for testing, which integrates the squared covariance, $\rho_x(t_k;\Delta)$ preserves the sign of the covariance and is used only as a descriptive summary. Figure 3 plots the estimate for $\rho_x(t_k;\Delta)$, denoted by $\widehat{\rho}_x(t_k;\Delta)$, with its pointwise 95% confidence intervals. The figure suggests a quite persistent protecting effect of breastmilk feeding against PA infection as reflected by generally positive values of $\widehat{\rho}_x(t_k;\Delta)$ across different t_k's. Such an effect may reach its maximum at month 4 because the magnitude of $\widehat{\rho}_x(t_k;\Delta)$ is largest when $t_k = 4$ (months) and $x =$"B", and the lower bound of the corresponding 95% confidence interval is above 0. On the other hand, the effects of both formula feeding and partial formula feeding tend to fluctuate around zero, with 0 always included in the 95% confidence intervals. This observation is consistent with the results in Table 4, where suggest neither formula feeding nor partial formula feeding is statistically associated with the time to first PA infection.

To assess the sensitivity of the proposed tests to the choice of landmark time points, we repeated the FIRST data analysis using two alternative specifications of $\mathcal{T}$, while fixing the quantile level interval $\Delta = [0.1, 0.4]$. Specifically, we considered a truncated grid, $\mathcal{T}_1 = \{1.5, 2, 3, 4, 5, 6, 7\}$, which excludes later landmark time points, and a coarser grid, $\mathcal{T}_2 = \{2, 4, 6, 9, 13\}$. The results are reported in Supplementary Table S1. We have similar findings based on these different choices of landmark time points. The delayed cumulative proportion of breast milk feeding, $P(B)$, consistently produced the smallest p values among the three feeding variables. For GIQC(MAX), the p values for $P(B)$ were 0.097, 0.092, and 0.082 under $\mathcal{T}$, $\mathcal{T}_1$, and $\mathcal{T}_2$, respectively. For GIQC(SUM), the corresponding p values were 0.170,

0.158, and 0.180. The p values for $P(F)$ and $P(P)$ remained larger across all specifications. These observations suggest that the testing results based on the proposed method are quite robust to moderate variations in landmark time points.

We also examined the sensitivity to the choice of quantile level interval by repeating the analysis with $\Delta = [0.1, 0.3]$. As shown in Supplementary Table S2, the results follow similar pattern to those obtained with $\Delta = [0.1, 0.4]$. That is, $P(B)$ yielded the smallest p values across $\mathcal{T}$, $\mathcal{T}_1$, and $\mathcal{T}_2$, while the p values for $P(F)$ and $P(P)$ remained larger. Overall, these results demonstrate quite robust performance of the proposed method in the FIRST data example.

The findings from our analyses are consistent with previous studies that explored the link between breastfeeding and pulmonary infections in CF. For example, Colombo et al. (2007) and Jadin et al. (2011) reported fewer PA infections during the first 2-3 years of life in infants who were breastfed for longer periods compared to those who were formula-fed or breastfed for shorter periods. However, the prior analyses typically categorized feeding data without differentiating them before or after a pulmonary infection. In contrast, our analyses properly accommodate the time-varying nature of the feeding data while not relying on a particular modeling of the relationship between feeding and time to first PA infection, and thus are expected to yield results that are more robust and scientifically sound.

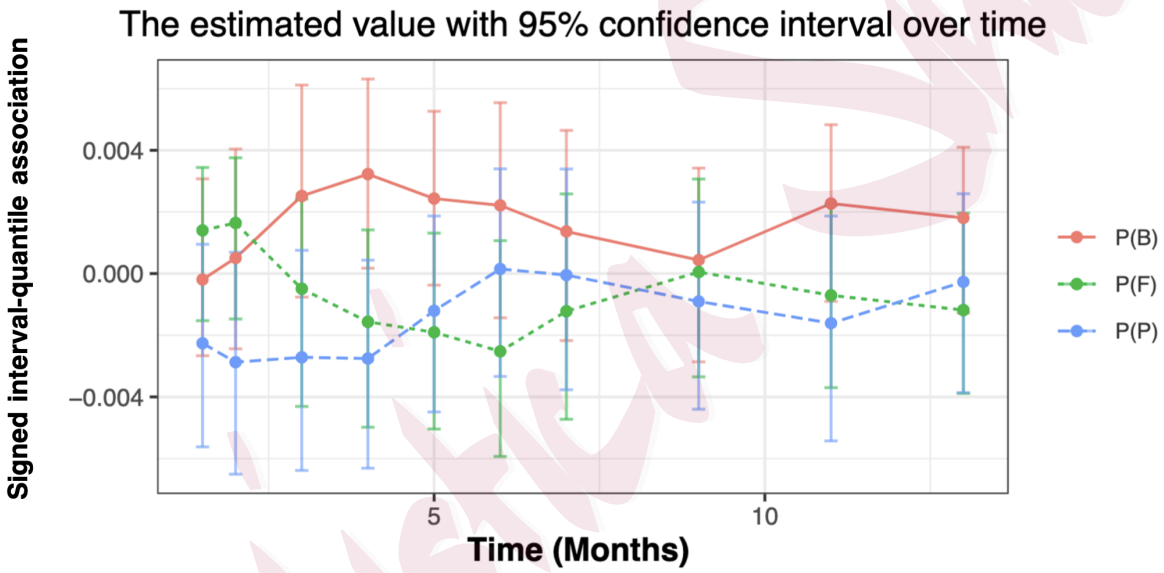

Figure 3: The estimated signed interval-quantile association summary $\widehat{\rho}_x(t_k;\Delta)$ with pointwise 95% confidence intervals at each landmark time point $t_k \in \mathcal{T}$ for B (breast milk feeding), F (formula feeding), and P (partial breast milk/formula feeding).

Table 4: p values from testing the impact of the feeding variable on the time to the first PA infection based on the two proposed methods (GIQC(MAX) and GIQC(SUM)) and the time-dependent Cox model (COX). The covariates include P(B): the delayed cumulative proportion of breast milk feeding; P(P): the delayed cumulative proportion of partial formula feeding; P(F): the delayed cumulative proportion of formula feeding

	GIQC(MAX)	GIQC(SUM)	COX
P(B)	0.097	0.170	0.720
P(F)	0.391	0.479	0.377
P(P)	0.329	0.404	0.107

5. Remarks

In this work, we develop a new nonparametric testing framework that can provide useful insight on the relationship between a time-dependent covariate and a time-to-event outcome. The test statistics have meaningful interpretations pertaining to the new concept of generalized interval quantile independence. The proposed resampling-based testing procedure is easy to implement. Our numerical studies support the empirical advantages of the proposed method over the popular Cox model in assessing the association between a survival outcome and a time-dependent covariate, particularly in data settings with dynamic covariate effects.

Of note, the proposed tests involve the choice of the landmark time set $\mathcal{T}$ and the quantile level interval Δ. In practice, one may choose landmark time points based on scientific considerations. For example, they may correspond to scheduled clinical follow-up visits or key clinical decision times. In addition, we recommend examining the size of the risk set at each selected landmark time point, defined as $n_k = \sum_{i=1}^{n} I(X_i > t_k)$, and checking whether the expected tail counts, $n_k\tau_L$ and $n_k(1-\tau_U)$, exceed a pre-specified threshold (e.g., 5 or 10). Landmark time points that do not meet this criterion may be excluded from the primary analysis, though they can still be considered in sensitivity analyses.

Although the quantile level interval $\Delta = [\tau_L, \tau_U]$ is, in principle, chosen to reflect scientific interest, its selection is subject to intrinsic identifiability constraints induced by censoring. In the current setting, such identifiability constraints apply to multiple landmark time points. In real data applications, we recommend finalizing the selection of Δ in an adaptive manner. For example, one may examine $\widehat{q}_{1,t_k}(\tau_U)$ and $\widehat{G}^r_{C,t_k}\{\widehat{q}_{1,t_k}(\tau_U)\}$ across all landmark time points t_k. When $\widehat{q}_{1,t_k}(\tau_U)$ has unusually large standard error or $\widehat{G}^r_{C,t_k}\{\widehat{q}_{1,t_k}(\tau_U)\}$ is very close to 0, one may reset the value of τ_U to a lower value to avoid these issues while preserving reasonable interpretability.

Note that the proposed GIQC index, $q_{sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$, can be further extended to account for a weighted aggregation of the modified IQC measure across landmark time points rather than simply using an unweighted sum. This more general GIQC index takes the form, $q_{w,sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T}) = \sum_{t_k \in \mathcal{T}} \omega_k q(T^r(t_k), V(t_k); \Delta)$, where the weight $\omega_k > 0$ and can be

selected according to the scientific target of the analysis. For example, one may let ω_k proportional to the landmark-specific risk-set size n_k. The proposed estimation and testing framework based on the unweighted $q_{sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$ can be readily adapted to accommodate the weighted extension using $q_{w,sum}(T, \bar{\boldsymbol{Z}}; \Delta, \mathcal{T})$.

It is also worth noting that, although both GIQC(MAX) and GIQC(SUM) target the same null hypothesis, they adopt different strategies for aggregating the modified IQC measures over time. Specifically, GIQC(MAX) captures the strongest covariate–outcome association, whereas GIQC(SUM) summaries the cumulative association over all the landmark time points. Based on our numerical experiments, GIQC(SUM) may exhibit greater power in finite samples when the covariate–outcome associations display a persistent or relatively homogeneous pattern across landmark time points. In practice, one may apply either or both types of test statistics, and interpret any discrepancies in the results in light of the rationale underlying their constructions.

Supplementary Material

Supplementary Material available online includes technical proofs.

Acknowledgements

We sincerely thank the FIRST cohort of children and their families for participating in our longitudinal study, as well as the funding support from the National Institutes of Health grants

(R01HL113548, R01DK136023, R01DK072126, R01DK109692, and R56DK109692) and the US Cystic Fibrosis Foundation (LAI14A0, LAI15A0 and LAI17A0).

References

Andersen, P. K. and R. D. Gill (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics 10*(4), 1100–1120.

Bell, S. C., M. A. Mall, H. Gutierrez, M. Macek, S. Madge, J. C. Davies, P.-R. Burgel, E. Tullis, C. Castaños, C. Castellani, et al. (2020). The future of cystic fibrosis care: a global perspective. *The Lancet Respiratory Medicine 8*(1), 65–124.

Bronstein, M., R. Sokol, S. Abman, B. Chatfield, K. Hammond, K. Hambidge, C. Stall, and F. Accurso (1992). Pancreatic insufficiency, growth, and nutrition in infants identified by newborn screening as having cystic fibrosis. *The Journal of pediatrics 120*(4), 533–540.

Cai, Z. and T. Sit (2020). Censored quantile regression model with time-varying covariates under length-biased sampling. *Biometrics 76*(4), 1201–1215.

Chu, C. W. and T. Sit (2023). Censored interquantile regression model with time-dependent covariates. *Journal of the American Statistical Association*, 1–12.

Colombo, C., G. Alicandro, V. Dacco, A. Consales, F. Mosca, C. Agostoni, and M. L. Gianni

(2021). Breastfeeding in cystic fibrosis: a systematic review on prevalence and potential benefits. *Nutrients 13*(9), 3263.

Colombo, C., D. Costantini, L. Zazzeron, N. Faelli, M. C. Russo, D. Ghisleni, I. Gatelli, M. Giovannini, E. Riva, R. Zetterström, et al. (2007). Benefits of breastfeeding in cystic fibrosis: a single-centre follow-up survey. *Acta Paediatrica 96*(8), 1228–1232.

Cox, D. R. and D. Oakes (2018). *Analysis of survival data*. Chapman and Hall/CRC.

Cui, Y. and L. Peng (2022). Assessing dynamic covariate effects with survival data. *Lifetime Data Analysis*, 1–25.

Fisher, L. D. and D. Y. Lin (1999). Time-dependent covariates in the cox proportional-hazards regression model. *Annual review of public health 20*(1), 145–157.

Fustik, S., T. Jacovska, L. Spirevska, and S. Koceva (2009). Protein–energy malnutrition as the first manifestation of cystic fibrosis in infancy. *Pediatrics International 51*(5), 678–683.

Gorfine, M., Y. Goldberg, and Y. Ritov (2017). A quantile regression model for failure-time data with time-dependent covariates. *Biostatistics 18*(1), 132–146.

Hendry, D. J. (2014). Data generation for the cox proportional hazards model with time-dependent covariates: a method for medical researchers. *Statistics in medicine 33*(3), 436–454.

Hsieh, F., Y.-K. Tseng, and J.-L. Wang (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics 62*(4), 1037–1043.

Jadin, S. A., G. S. Wu, Z. Zhang, S. M. Shoff, B. M. Tippets, P. M. Farrell, T. Miller, M. J. Rock, H. Levy, and H. J. Lai (2011). Growth and pulmonary outcomes during the first 2 y of life of breastfed and formula-fed infants diagnosed with cystic fibrosis through the wisconsin routine newborn screening program. *The American journal of clinical nutrition 93*(5), 1038–1047.

Lai, H., M. Rock, and P. Farrell (2014). Chapter 48: Nutrition in cystic fibrosis. In R. Kleinman and F. Greer (Eds.), *Pediatric Nutrition*, Chapter 48, pp. 1113–1146. Elk Grove Village, IL: American Academy of Pediatrics.

Lai, H. J., T. R. Bach, T. Miller, C. M. McDonald, K. M. Maguiness, E. E. Seffrood, J. B. Leonard, and P. M. Farrell (2025). Breastfeeding, growth, and lung disease in the first 3 years of life in children with cystic fibrosis. *Journal of Cystic Fibrosis 24*, 507–515.

Li, R. and L. Peng (2014). Varying coefficient subdistribution regression for left-truncated semi-competing risks data. *Journal of multivariate analysis 131*, 65–78.

Lin, D. Y., L.-J. Wei, and Z. Ying (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika 80*(3), 557–572.

Marcus, M., S. Sondel, P. Farrell, A. Laxova, P. Carey, R. Langhough, and E. Mischler

(1991). Nutritional status of infants with cystic fibrosis associated with early diagnosis and intervention. *The American journal of clinical nutrition 54*(3), 578–585.

Meek, J. Y., L. Noble, and S. on Breastfeeding (2022). Policy statement: breastfeeding and the use of human milk. *Pediatrics 150*(1), e2022057988.

Parast, L., S.-C. Cheng, and T. Cai (2012). Landmark prediction of long-term survival incorporating short-term event time information. *Journal of the American Statistical Association 107*(500), 1492–1501.

Putter, H. (2013). Landmarking. In *Handbook of Survival Analysis*, Chapter 21, pp. 441–456. Boca Raton: Chapman & Hall/CRC.

Putter, H. and H. C. van Houwelingen (2017). Understanding landmarking and its relation with time-dependent cox regression. *Statistics in biosciences 9*, 489–503.

Robins, J. M. and A. Rotnitzky (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS epidemiology*, pp. 297–331. Springer.

Sanders, D. B., Z. Li, A. Laxova, M. J. Rock, H. Levy, J. Collins, C. Ferec, and P. M. Farrell (2014). Risk factors for the progression of cystic fibrosis lung disease throughout childhood. *Annals of the American Thoracic Society 11*(1), 63–72.

Song, X. and C. Wang (2008). Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics 64*(2), 557–566.

REFERENCES

Therneau, T. M. and P. M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Springer.

Van Houwelingen, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics 34*(1), 70–85.

Wulfsohn, M. S. and A. A. Tsiatis (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 330–339.

Zeng, D. and D. Lin (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association 102*(480), 1387–1396.

Zheng, Y. and P. J. Heagerty (2005). Partly conditional survival models for longitudinal data. *Biometrics 61*(2), 379–391.

Zhu, L., Y. Zhang, K. Xu, et al. (2018). Measuring and testing for interval quantile dependence. *The Annals of Statistics 46*(6A), 2683–2710.

Emory University

E-mail: cuiyingbeicheng@outlook.com

University of Wisconsin-Madison,

E-mail: hlai@wisc.edu

Emory University

E-mail: lpeng@emory.edu