

A Review of the Definition of Time-Dependent Covariate in Survival Analysis

Changyong Feng^{1,2}, Hongyue Wang¹

¹Department of Biostatistics and Computational Biology

²Department of Anesthesiology and Perioperative Medicine

University of Rochester, Rochester, NY 14642, USA

e-mail: changyong_feng@urmc.rochester.edu

SUMMARY

Time-dependent covariates frequently arise in biomedical and epidemiologic research when subject characteristics or exposures evolve over time. Appropriate handling of such covariates is essential for valid inference in survival analysis. This paper reviews the theoretical foundations of time-dependent covariates, emphasizing the critical distinction between external and internal processes. We define conditional cumulative hazard and survival functions, establish their relationships, and discuss key probabilistic properties. When covariates are external, the conditional cumulative hazard at time t depends only on the covariate history up to t , and the classical product-integral representation between the conditional cumulative hazard and survival functions remains valid. The framework presented provides conceptual clarity and a foundation for time-varying Cox regression, landmark analysis, and modern joint modeling techniques.

Key words: Conditional cumulative hazard function; conditional survival function; product integral; external and internal covariates; joint modeling.

1. Introduction

Time-to-event data are ubiquitous in clinical, epidemiological, and reliability studies, where the primary objective is to evaluate the duration until an event occurs and identify predictors that influence survival (Aalen et al., 2008; Andersen et al., 1993; Cox, 1972; Fleming and Harrington, 1991; Martinussen and Scheike, 2006).

In many applications, predictors evolve over the course of follow-up. Such variables are referred to as *time-dependent* or *time-varying covariates* (Kalbfleisch and Prentice, 1980, 2002). Classic examples include fluctuating biomarker measurements (e.g., blood pressure, CD4 cell counts), as well

as time-varying treatment regimes or exposure levels. Properly accounting for the temporal evolution of these covariates is critical; treating them as baseline values or ignoring their time-varying nature can lead to substantial bias and invalid statistical inferences (Austin et al., 2020; Fisher and Lin, 1991; Rizopoulos, 2012; Therneau and Grambsch, 2000; Tsiatis and Davidian, 2004; van Houwelingen and Putter, 2011; Zabor and Assel, 2023).

Let T denote a non-negative survival time, and define the upper bound of the support as

$$\tau = \inf\{t : P(T \leq t) = 1\}.$$

The marginal survival function of T is $S(t) = P(T > t)$ for $0 \leq t \leq \tau$, and its cumulative hazard function is defined as

$$\Lambda(t) = \int_{[0,t]} P\{u \leq T < u + du \mid T \geq u\}, \quad 0 \leq t \leq \tau.$$

A fundamental result in survival analysis is the one-to-one correspondence between the cumulative hazard function and the survival function via the product-integral representation:

$$\Lambda(t) = - \int_{[0,t]} \frac{dS(u)}{S(u-)}, \quad (1)$$

$$S(t) = \prod_{[0,t]} (1 - d\Lambda(u)), \quad (2)$$

for $0 \leq t \leq \tau$ (see, e.g., Fleming and Harrington, 1991; Andersen et al., 1993).

A natural and crucial question arises when covariates are time-dependent: does this fundamental one-to-one correspondence between the cumulative hazard and survival functions—expressed in equations (1) and (2)—extend naturally to their conditional counterparts given the covariate history?

To address this question, Kalbfleisch and Prentice (1980, 2002) introduced the critical distinction between *external* (or exogenous) and *internal* (or endogenous) time-dependent covariates. A modern and comprehensive account of this classification from a counting-process perspective is given by Aalen et al. (2008).

In this paper, we systematically review these definitions, investigate their underlying probabilistic properties, and formalize the exact conditions

under which conditional hazard representations remain valid in the presence of time-varying covariates.

2. Definition of external covariates

For each $u \geq 0$, let $x(u)$ be a q -dimensional random covariate vector. The covariate process is defined as $X = \{x(u) : 0 \leq u < \infty\}$. For $0 \leq t < s$, let $X(t, s] = \{x(u) : t < u \leq s\}$, representing the segment of the covariate path between times t and s . We denote the complete covariate path as $X = X(0, \infty)$, and the covariate history up to and including time t as $X(0, t]$. For brevity, we denote $X(0, t]$ simply as $X(t)$.

Definition 1. Given the complete covariate process X , the conditional cumulative hazard function of T given X , denoted by $\Lambda(\cdot | X)$, is

$$\Lambda(t | X) = \int_{[0, t]} P\{u \leq T < u + du | T \geq u, X\}, \quad 0 \leq t \leq \tau.$$

Definition 2. For a fixed $s \geq 0$, the conditional cumulative hazard function of T given the covariate history up to time s , denoted by $\Lambda[\cdot | X(s)]$, is defined as

$$\Lambda[t | X(s)] = \int_{[0, t]} P\{u \leq T < u + du | T \geq u, X(s)\}, \quad 0 \leq t \leq \tau.$$

In particular, when $s = t$,

$$\Lambda[t | X(t)] = \int_{[0, t]} P\{u \leq T < u + du | T \geq u, X(t)\}.$$

Remark 1. It is important to note that, in general,

$$\Lambda[t | X(t)] \neq \int_{[0, t]} P\{u \leq T < u + du | T \geq u, X(u)\}.$$

The integrand on the left conditions on the full history $X(t)$ up to time t for all integration steps, whereas the integrand on the right updates the conditioning set to $X(u)$ dynamically at each time $u \leq t$.

With these definitions established, we formally define external covariates.

Definition 3. The covariate process X is said to be *external* (or *exogenous*) to the survival time T if, for all $t \in [0, \tau]$,

$$\Lambda(t \mid X) = \Lambda[t \mid X(t)].$$

Otherwise, X is said to be *internal* (or *endogenous*) to T .

This definition formalizes the notion that for an external covariate, the conditional hazard at any time t depends only on the covariate history up to time t . Consequently, future values of X beyond time t provide no additional information regarding the conditional probability of the event occurring at or before time t .

3. Properties

Theorem 1. *If the covariate process X is external to the survival time T , then for all $s \geq t$,*

$$\Lambda[t \mid X(s)] = \Lambda[t \mid X(t)].$$

Proof. From the definition of externality, for $0 \leq u \leq t \leq s$,

$$P\{u \leq T < u + du \mid T \geq u, X\} = P\{u \leq T < u + du \mid T \geq u, X(u)\}.$$

This equality implies

$$\begin{aligned} & P\{u \leq T < u + du \mid X\} P\{T \geq u \mid X(u)\} \\ &= P\{u \leq T < u + du \mid X(u)\} P\{T \geq u \mid X\}. \end{aligned}$$

Taking conditional expectations on $X(s)$ on both sides yields

$$\begin{aligned} & P\{u \leq T < u + du \mid X(s)\} P\{T \geq u \mid X(u)\} \\ &= P\{u \leq T < u + du \mid X(u)\} P\{T \geq u \mid X(s)\}, \end{aligned}$$

which simplifies to

$$\begin{aligned} & P\{u \leq T < u + du \mid T \geq u, X(s)\} \\ &= P\{u \leq T < u + du \mid T \geq u, X(u)\}. \end{aligned}$$

Similarly, we have

$$\begin{aligned} & P\{u \leq T < u + du \mid T \geq u, X(t)\} \\ &= P\{u \leq T < u + du \mid T \geq u, X(u)\}. \end{aligned} \quad (3)$$

Thus,

$$\begin{aligned} & P\{u \leq T < u + du \mid T \geq u, X(s)\} \\ &= P\{u \leq T < u + du \mid T \geq u, X(t)\}. \end{aligned}$$

Integrating over $[0, t]$ gives the desired result. $\square$

Remark 2. From equation (3) we can see that if X is external, then, for $0 \leq u \leq t$,

$$\begin{aligned} & P\{u \leq T < u + du \mid T \geq u, X(t)\} \\ &= P\{u \leq T < u + du \mid T \geq u, X(u)\}. \end{aligned}$$

This matches the definition of externality provided in Kalbfleisch and Prentice (2002, p. 196, Eq. 6.8). Furthermore, it follows that

$$\begin{aligned} \Lambda[t \mid X(t)] &= \int_0^t P\{u \leq T < u + du \mid T \geq u, X(t)\} \\ &= \int_0^t P\{u \leq T < u + du \mid T \geq u, X(u)\}. \end{aligned}$$

Theorem 2. *If the covariate process X is external to T , then for any $s > t$ and any event $A \in \sigma(X(u) : t < u \leq s)$, we have*

$$P\{A, T = t \mid T \geq t, X(t)\} = P\{A \mid T \geq t, X(t)\} P(T = t \mid T \geq t, X(t)).$$

Proof. For $s > t$, if $A \in \sigma(X(t, s])$, then

$$\begin{aligned} P\{A, t \leq T < t + dt \mid X(t)\} &= \mathbb{E}[P(A, t \leq T < t + dt \mid X(s)) \mid X(t)] \\ &= \mathbb{E}[\mathbf{1}_A P(t \leq T < t + dt \mid X(s)) \mid X(t)]. \end{aligned}$$

By the definition of an external covariate process,

$$\begin{aligned} & P\{t \leq T < t + dt \mid T \geq t, X(s)\} \\ &= P\{t \leq T < t + dt \mid T \geq t, X(t)\}. \end{aligned}$$

From the above we have

$$\begin{aligned} & P\{t \leq T < t + dt \mid X(s)\} \\ &= P\{t \leq T < t + dt \mid T \geq t, X(t)\} P\{T \geq t \mid X(s)\}. \end{aligned}$$

Substituting this back into the expectation gives:

$$\begin{aligned} & P\{A, t \leq T < t + dt \mid X(t)\} \\ &= \mathbb{E}[\mathbf{1}_A P\{t \leq T < t + dt \mid T \geq t, X(t)\} P\{T \geq t \mid X(s)\} \mid X(t)] \\ &= P\{t \leq T < t + dt \mid T \geq t, X(t)\} \mathbb{E}[\mathbf{1}_A P\{T \geq t \mid X(s)\} \mid X(t)]. \end{aligned}$$

Since $A \in \sigma(X(t, s])$, we have $\mathbf{1}_A P\{T \geq t \mid X(s)\} = P\{A, T \geq t \mid X(s)\}$. Taking the conditional expectation with respect to $X(t)$ yields:

$$\begin{aligned} & P\{A, t \leq T < t + dt \mid X(t)\} \\ &= P\{t \leq T < t + dt \mid T \geq t, X(t)\} P\{A, T \geq t \mid X(t)\}. \end{aligned}$$

Dividing both sides by $P(T \geq t \mid X(t))$ completes the proof:

$$\begin{aligned} & P\{A, t \leq T < t + dt \mid T \geq t, X(t)\} \\ &= P\{t \leq T < t + dt \mid T \geq t, X(t)\} P\{A \mid T \geq t, X(t)\}. \quad \square \end{aligned}$$

Remark 3. Theorem 2 formalizes the condition that the future path of an external covariate is conditionally independent of event occurrence at time t , given the historical process $X(t)$ and survival up to time t (Kalbfleisch and Prentice, 2002, Eq. 6.9).

4. Conditional survival functions

Equations (1) and (2) establish a one-to-one relationship between the marginal cumulative hazard and survival functions. We now examine the analogous relationship between their conditional counterparts.

Definition 4. The conditional survival function of T given the complete covariate process X , denoted by $S(\cdot \mid X)$, is defined as

$$S(t \mid X) = P\{T > t \mid X\}, \quad 0 \leq t \leq \tau.$$

Definition 5. For each fixed $s \geq 0$, the conditional survival function of T

given the covariate history up to time s , denoted by $S[\cdot | X(s)]$, is defined as

$$S[t | X(s)] = P\{T > t | X(s)\}, \quad 0 \leq t \leq \tau.$$

The challenge of internal covariates

If X is an internal covariate, then for any $t \in [0, \tau]$, the existence of $X(t)$ is often contingent upon the subject being alive at time t (i.e., $T \geq t$). This implies that $T \geq t$ is effectively contained within the information set $\sigma(X(t))$. Thus,

$$P\{T \geq t | X(t)\} = 1.$$

In such cases, for internal covariates, $S[t | X(t)]$ ceases to be a meaningful measure of survival probability. For internal covariates, failure rates are instead defined through intensity-based approaches or joint models (Tsiatis and Davidian, 2004; Rizopoulos, 2012).

Theorem 3. *If the covariate process X is external to T , then for all $t \in [0, \tau]$, the conditional survival function of T given X satisfies*

$$S(t | X) = \prod_{[0, t]} (1 - d\Lambda(u | X)) = \prod_{[0, t]} (1 - d\Lambda[u | X(u)]).$$

Proof. For any $t \in [0, \tau]$, we have $S(t | X) = P(T > t | X)$. For any $u \in [0, t]$,

$$-dS(u | X) = P\{u \leq T < u + du | X\}.$$

It follows that

$$\begin{aligned} \frac{-dS(u | X)}{S(u- | X)} &= \frac{P\{u \leq T < u + du | X\}}{P\{T \geq u | X\}} \\ &= P\{u \leq T < u + du | T \geq u, X\} \\ &= d\Lambda(u | X). \end{aligned}$$

By the definition of an external covariate, $\Lambda(u | X) = \Lambda[u | X(u)]$, which implies $d\Lambda(u | X) = d\Lambda[u | X(u)]$. Consequently,

$$dS(u | X) = -S(u- | X)d\Lambda(u | X).$$

Applying the product-integral representation theorem (Andersen et al., 1993,

Theorem II.6.1) yields

$$S(t | X) = \prod_{[0,t]} (1 - d\Lambda(u | X)) = \prod_{[0,t]} (1 - d\Lambda[u | X(u)]). \quad \square$$

Remark 4. Theorem 3 confirms that under the condition of externality, the conditional survival and hazard functions maintain the standard product-integral relationship. This preserves the interpretability of the survival function as:

$$S[t | X(t)] = \prod_{[0,t]} (1 - d\Lambda[u | X(u)]).$$

Corollary 4. *The covariate process X is external to T if and only if $S[t | X(t)] = S(t | X)$ for all $t \in [0, \tau]$.*

5. Numerical example

In this section, we analyze the classic Stanford Heart Transplant dataset (Crowley and Hu, 1977) to illustrate the practical implications of time-dependent covariates and the severe biases that arise when their dynamic nature is ignored. The dataset contains records for 103 patients accepted into the heart transplant program after physicians determined that they were suitable candidates. As donor organs become available over time, suitable candidates receive a transplant, causing a patient's treatment status to change dynamically during follow-up from waiting status to recipient status.

Transplant status at time t is naturally modeled as a time-dependent binary indicator $z(t)$:

$$z(t) = \begin{cases} 0 & \text{if the patient has not received a transplant by time } t, \\ 1 & \text{if the patient received a transplant at or prior to time } t. \end{cases}$$

Because receiving a transplant requires the patient to survive until a suitable donor heart is identified, $z(t)$ is intrinsically an *internal* (or endogenous) covariate process: the event path $z(t)$ is generated by the patient's ongoing survival process up to time t .

Of the 103 patients in the study, 69 eventually received a transplant. For each subject, the primary date variables are:

- Bir_Date: Date of birth;

- Acc_Date: Date of acceptance into the transplant program;
- Ter_Date: Terminal date (date of death or study closure).

Total survival time in days is defined as $\text{Surv_Time} = \text{Ter_Date} - \text{Acc_Date}$. Survival status is classified as uncensored ($\text{Status} = 1$) or censored ($\text{Status} = 0$), depending on whether Ter_Date represents death or administrative censoring. Patient age at acceptance in years is given by $\text{Acc_Age} = (\text{Acc_Date} - \text{Bir_Date})/365.25$. For transplant recipients, Xpl_Date records the date of transplantation, and the waiting time in days is calculated as $\text{Wait_Time} = \text{Xpl_Date} - \text{Acc_Date}$.

To demonstrate the empirical impact of mischaracterizing time-dependent covariates, we compare two estimation strategies for fitting a Cox proportional hazards model with transplant status and Acc_Age as predictors:

- **Method 1: Correct specification (time-dependent model):** Treatment status is modeled dynamically as a time-dependent covariate $z(t)$ using counting process formulation ($t_{\text{start}}, t_{\text{stop}}$ intervals). The estimation results are presented in Table 1.

Table 1. Cox proportional hazards model with correctly specified time-dependent transplant status

Variable	Estimate	SE	Hazard ratio	<i>p</i> -value
Transplant ($z(t)$)	−0.068	0.306	0.934	0.825
Acc_Age	0.031	0.014	1.032	0.030

When correctly specified, transplantation is associated with a modest reduction in mortality risk (hazard ratio = 0.934), though this effect is not statistically significant ($p = 0.825$).

- **Method 2: Incorrect specification (time-fixed model):** Treatment status is erroneously treated as a static baseline variable, classifying any subject who ever received a transplant during follow-up as $z = 1$ starting at baseline ($t = 0$). The estimation results are presented in Table 2.

Failing to account for the time-dependent nature of transplantation introduces severe *immortal time bias*: subjects in the transplant group are artificially credited with surviving the period between acceptance

Table 2. Cox proportional hazards model incorrectly treating transplant status as time-fixed

Variable	Estimate	SE	Hazard ratio	p -value
Transplant (z)	-1.791	0.272	0.167	< 0.001
Acc_Age	0.060	0.015	1.062	< 0.001

into the program and actual transplantation. This flawed specification leads to the spurious conclusion that heart transplantation dramatically reduces mortality risk by over 83% (hazard ratio = 0.167, $p < 0.001$).

6. Conclusions

Time-dependent covariates are ubiquitous in survival analysis, but their mathematical treatment requires careful handling depending on whether they evolve exogenously or endogenously. In this review, we have systematically formalized the probabilistic foundations of time-varying covariates, focusing on the essential boundary between external and internal processes.

The theoretical results synthesized in this paper provide clear guidance regarding the validity of conditional survival representations:

- **Theorem 1** verifies that for external covariates, conditioning on future history $X(s)$ for $s \geq t$ adds no additional information beyond current history $X(t)$, formalizing the classical foundational intuition laid out by Kalbfleisch and Prentice (1980, 2002).
- **Theorem 2** explicitly establishes the conditional independence between future covariate trajectories $X(t, s]$ and event occurrence at time t , given observed history $X(t)$ and survival up to t .
- **Theorem 3 and Corollary 4** prove that the classical product-integral relationship $S(t | X) = \prod_{[0, t]} (1 - d\Lambda[u | X(u)])$ holds if and only if the covariate process is external. This provides the exact theoretical threshold at which standard conditional survival calculations remain mathematically sound.

Beyond pure theory, these results carry critical practical implications for empirical survival analysis. As illustrated by our analysis of the Stanford Heart Transplant dataset in section 5, treating internal covariates as

time-fixed baseline variables severely violates the externality conditions of Theorem 3, resulting in catastrophic immortal time bias (reducing mortality hazard ratios spuriously from 0.934 to 0.167). While external covariates can be directly modeled via time-varying Cox proportional hazards models, internal covariates require specialized methods—such as joint longitudinal-survival models or landmark analysis—to properly handle the feedback mechanism between the covariate and survival processes (Martinussen and Scheike, 2006; Rizopoulos, 2012).

Looking forward, as biomedical studies increasingly generate high-dimensional, continuous-time streams from wearable sensors and electronic health records (van Houwelingen and Putter, 2011; Zabor and Assel, 2023), distinguishing between external environmental factors and internal physiological markers will become even more pivotal. The probabilistic framework reinforced in this manuscript provides the necessary theoretical footing for sound model specification and inference in modern time-to-event research.

REFERENCES

- Aalen O., Borgan Ø., Gjessing H.K. (2008): *Survival and Event History Analysis: A Process Point of View*. New York: Springer.
- Andersen P.K., Borgan Ø., Gill R.D., Keiding N. (1993): *Statistical Models Based on Counting Processes*. New York: Springer.
- Austin P.C., Latouche A., Fine J.P. (2020): A review of the use of time-varying covariates in the Fine-Gray subdistribution hazard competing risk regression model. *Stat Med.* 39(2): 103–113.
- Cox D.R. (1972): Regression models and life-tables (with discussion). *J R Stat Soc Ser B* 34: 187–220.
- Crowley J., Hu M. (1977): Covariance Analysis of Heart Transplant Survival Data. *Journal of the American Statistical Association* 72: 27–36.
- Fisher L.D., Lin D.Y. (1999): Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health* 20: 145–157.
- Fleming T.R., Harrington D.P. (1991): *Counting Processes and Survival Analysis*. New York: Wiley
- Kalbfleisch J.D., Prentice R.L. (1980): *The Statistical Analysis of Failure Time Data* (1st ed.) New York: Wiley.
- Kalbfleisch J.D., Prentice R.L. (2002): *The Statistical Analysis of Failure Time Data* (2nd ed.) New York: Wiley.
- Martinussen T., Scheike T.H. (2006): *Dynamic Regression Models for Survival Data*. New York: Springer.
- Rizopoulos D. (2012): *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Boca Raton, FL: Chapman & Hall/CRC.

- Therneau T.M., Grambsch P.M. (2000): Modeling Survival Data: Extending the Cox Model. New York: Springer.
- Tsiatis A.A., Davidian M. (2004): Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* 14: 809–834.
- Van Houwelingen H.C., Putter H. (2011): Dynamic Prediction in Clinical Survival Analysis. Boca Raton, FL: CRC Press
- Zabor E.C., Assel M. (2023): On the Need for Landmark Analysis or Time-dependent Covariates. *J Urol.* 209(6): 1060–1062.