Received: 20 April 2020 | Revised: 5 January 2021 | Accepted: 9 January 2021

DOI: 10.1002/sim.8888

RESEARCH ARTICLE

Statistics in Medicine WILEY

Displaying survival of patient groups defined by covariate paths: Extensions of the Kaplan-Meier estimator

Melissa Jay[1] | **Rebecca A. Betensky**[2]

[1]Department of Biostatistics, The University of Iowa, Iowa City, Iowa

[2]Department of Biostatistics, NYU School of Global Public Health, New York City, New York

Correspondence
Melissa Jay, Department of Biostatistics, The University of Iowa, Iowa City, IA, USA.
Email: melissa-jay@uiowa.edu

Funding information
National Institutes of Health, Grant/Award Numbers: R01NS094610, T32CA009337; National Science Foundation, Grant/Award Number: NSF GRFP Grant No. 000390183

Extensions of the Kaplan-Meier estimator have been developed to illustrate the relationship between a time-varying covariate of interest and survival. In particular, Snapinn et al and Xu et al developed estimators to display survival for patients who always have a certain value of a time-varying covariate. These estimators properly handle time-varying covariates, but their clinical interpretation is limited. It is of greater clinical interest to display survival for patients whose covariates lie along certain defined paths. In this article, we propose extensions of Snapinn et al and Xu et al's estimators, providing crude and covariate-adjusted estimates of the survival function for patients defined by covariate paths. We also derive analytical variance estimators. We demonstrate the utility of these estimators with medical examples and a simulation study.

KEYWORDS
survival distribution, time-dependent covariates, time-varying covariates

1 | INTRODUCTION

The Kaplan-Meier estimator is widely used in the medical literature to display survival distributions and to compare these between two or more patient groups of interest.[1] It often accompanies a Cox proportional hazards model, or another time-to-event regression model, to provide visual evidence of survival function differences. The Cox proportional hazards model can appropriately incorporate time-varying covariate values in hazard estimation while the standard Kaplan-Meier estimator cannot. To this end, extensions of the Kaplan-Meier estimator have been developed to properly present the relationship between a time-varying covariate value of interest and survival.[2-10]

Common incorrect accommodations of time-varying covariates in Kaplan-Meier estimates include the last covariate value approach and the adjusted time-zero approach.[4,5,9] These approaches are illustrated simply with an example drawn from oncology.[4,9] Suppose researchers conduct a study to understand if a new cancer therapy prolongs survival. While the primary endpoint is survival, data are additionally collected on whether or not the patients achieve tumor response, and if so, when the response occurs, while being treated with the new therapy. In a secondary analysis, the investigators aim to understand whether tumor response is associated with an improvement in survival. Tumor response is a time-varying covariate that can change for each patient throughout the study follow-up. Patients may transition from being a "tumor non-responder" to a "tumor responder" if their tumors shrink during the study.

The last covariate value approach defines patients as either responders or non-responders on the basis of the last measured value of the time-varying tumor response covariate. Those who respond at any time during follow-up are included in the responder Kaplan-Meier curve, and those who never respond are included in the non-responder curve. This approach is problematic because it introduces *guarantee-time bias*.[4,11] Individuals who are included in the group of

tumor responders have necessarily survived long enough to experience tumor response. Furthermore, those who respond quickly after the onset of treatment are not differentiated from those who have a long delay until response. Ignoring these features of the ever-responders results in an uninterpretable and misleading estimate, and in conclusions that may differ from those from a Cox model, which accounts for time of response in the estimation. Alternatively, the adjusted time-zero method takes the time origin to be the time of response for responders and compares the curve based on the original time origin for those who never responded to that based on the updated time origin for those who responded. This approach also disregards the time of response, that is, the time responders spent as non-responders, and is subject to guarantee-time bias, as well. For an illustration of the biases associated with these incorrect approaches, see figure 2 in Simon and Makuch.[4]

Several methods have been developed to display survival while appropriately incorporating time-varying covariate information. Dubin et al[7] created a graphical device for displaying the relationship between a time-varying covariate and patient survival. They proposed an event history graph, which incorporates individual-level information into the traditional Kaplan-Meier plot. The Kaplan-Meier estimator is plotted, and color-coded horizontal bars are included under the Kaplan-Meier curve, representing each patient in the study. The length of each bar depicts each patient's follow-up time, while the colors of the bar display the time-varying covariate values corresponding to each patient at each time point. This allows readers to visually examine which colors are associated with longer survival times and which colors may be related to poorer survival.

Simon and Makuch[4] and Feuer et al[8] developed methods for depicting the impact of a binary, non-reversible covariate, such as response, on survival. Tumor response is an example of a non-reversible binary covariate: if a patient has a tumor response, they cannot return to the tumor non-response state during the study. Simon and Makuch's[4] method displays survival of two groups (those who advance to the second live state and those who do not) after conditioning on a specific time at which the majority of patients have moved to the second live state. Their method correctly incorporates each patient's "at-risk" person-time by updating patient groups at each event time, but it does not extend to situations in which the time-varying covariate is reversible. Feuer et al[8] proposed two methods: one in which survival estimates are based on prospectively conditioning on a sequence of state changes and one in which the overall survival estimator of all patients is compared to the survival estimator that would occur with the removal of the second live state. The first of these methods fails to remove guarantee-time bias, and the second method is best used to evaluate the effect of an overall program (ie, having a heart transplantation program vs. not having one) rather than evaluating individual effects on survival (ie, the effect of a heart transplant on an individual patient's survival). Li et al[10] proposed a method to compare survival for two levels of a non-reversible time-varying treatment covariate. They applied matching methods, along with inverse probability of censoring weighting, to obtain the estimated survival probabilities for treated patients, and the counterfactual survival probabilities that would have occurred if those patients were untreated. They used this method to obtain an estimate of the average effect of treatment on the treated.

Survival estimators that appropriately accommodate reversible time-varying covariates were developed by Snapinn et al[5] and Xu et al.[6] One example of a possibly reversible time-varying covariate is treatment status. At any point during follow-up of a study, a patient may be compliant or non-compliant with treatment, or the patient may temporarily discontinue treatment due to adverse events. The estimators developed by Snapinn et al[5] and Xu et al[6] are unweighted and weighted extensions of the Kaplan-Meier estimator that illustrate the impact of a time-varying covariate on survival. These extensions require the interpretation that each survival curve corresponds to a group of patients who always have a constant value of the time-varying covariate throughout their follow-up.

Smith et al[12] created a method to display survival probabilities in the presence of a clinically relevant intervening event. The goal of their work was to create a visual representation of the effect of a clinically relevant state change (such as receiving a liver transplant) on survival at a pre-specified time point. To accomplish this, they used fitted survival probabilities from a Cox model with time-dependent indicators. Each patient's time-varying covariate values were coded as zeros until they reached the intervening event, at which time they were converted to ones. They also proposed a Cox model with an added time-dependent coefficient to accommodate certain situations where the functional form of the time-varying Cox model is not appropriate. The estimated survival curves produced by this method are interpreted on an individual patient level, so a fixed set of baseline patient characteristics must be specified to obtain survival probabilities. This method is convenient for qualitatively comparing subjects with the same characteristics, except with regard to the intervening event of interest at a pre-determined time point, which some have and some do not have. It is limited, however, by its inability to incorporate other time-varying covariate values into the estimation process and its assumptions about the functional form of the time-varying covariate's effect.

In survival analysis, it is important to make the distinction between internal and external time-varying covariates, as the type of covariate dictates the appropriate analysis methods and interpretations. Internal time-varying covariates, which frequently arise in the medical setting, are covariates whose values are affected by the underlying event process. Two examples of internal covariates include tumor response status and systolic blood pressure measurements. A concern with internal time-varying covariates is the possible distortion of a treatment effect. If the internal time-varying covariate lies on the causal pathway between a baseline treatment and the survival outcome, the Cox model and related nonparametric analyses could fail to capture the true treatment effect, as it may be largely reflected in the effect of the time-varying covariate.[13] In contrast, external time-varying covariates are not affected by the underlying event process and do not fall along relevant causal pathways. The aforementioned survival estimators are useful for displaying associations between time-varying covariate values and survival but should be interpreted carefully as measures of association when used with internal covariates.

None of the aforementioned methods display survival for patient groups that are defined by one or more covariate value changes at clinically important intermediate time points, while requiring no functional assumptions about the time-varying covariate's effect; we fill this significant gap in this article. We propose two extensions of the Kaplan-Meier estimator to display survival of patient groups defined by *any* covariate path, including those that are non-constant and reversible during follow-up. This is useful for visualizing the impact of evolving measures on survival. In Section 2, we introduce notation and summarize existing estimators. In Section 3, we propose extensions of Snapinn et al's[5] and Xu et al's[6] estimators, providing crude and confounder-adjusted estimators for patient groups defined by covariate paths. In Section 4, we apply our proposed estimators to a lung cancer dataset and a primary biliary cholangitis dataset. We assess the properties of our proposed estimators via a simulation study in Section 5, and conclude with a discussion of advantages and limitations of our proposed estimators in Section 6.

2 | NOTATION AND EXISTING ESTIMATORS

2.1 | Notation

To facilitate straightforward comparison with Snapinn et al,[5] we adopt their notation. For subject i, let T_i denote the event time, C_i denote the censoring time, and $Y_i = \min(T_i, C_i)$ denote the follow-up time. Let δ_i denote the event indicator, with $\delta_i = 1$ if the the event is observed, that is, $T_i < C_i$, and $\delta_i = 0$ otherwise. We denote the ordered set of J unique event times as $\boldsymbol{t} = \left\{t_{(j)} : j \in 1, \ldots, J\right\}$ and define $X_i(t)$ to be the value of the time-varying covariate for individual i at time t.

2.2 | Survival estimator of Snapinn et al

Snapinn et al's[5] extension of the Kaplan-Meier estimator is based on constructing risk sets that are specific to the possible values of the time-varying covariate. In particular, the size of the risk set at time t_j for the subjects with covariate value k at t_j is defined as

$$n_{jk} = \sum_i I(X_i(t_j) = k) I(Y_i \geq t_j).$$

Analogously, the size of the event set at time t_j for subjects with covariate value k at t_j is defined as

$$d_{jk} = \sum_i I(X_i(t_j) = k) I(\delta_i = 1) I(Y_i = t_j).$$

Their extended Kaplan-Meier estimator for a subject with constant covariate value k simply aggregates these counts across event times and is given by

$$\tilde{S}_k(t) = P(T > t | X(s) = k,\ s \leq t) = \prod_{j : t_j \leq t} \left(1 - \frac{d_{jk}}{n_{jk}}\right). \tag{1}$$

An important underlying assumption of this estimator is that the risk of death at any given time is dependent only on a subject's current value of the covariate, and not on any past values. If $X_i(t_j)$ were constant over time for each subject,

then Snapinn et al's[5] extended estimator would be equivalent to the Kaplan-Meier estimator. However, because it is not, subjects contribute to risks sets and event sets only at those times at which their covariate is equal to the value of interest.

Because the covariate is time-varying, the sizes of the risk sets are not necessarily a decreasing function of time, as they would be with a baseline covariate. This feature of Snapinn et al's[5] method can be a limitation in some cases. If a risk set contains too few individuals at a given time point, the survival estimate may suffer from poor precision. Furthermore, if there is a brief period in which no individuals are in a risk set, the estimated survival probability will become zero and remain zero, even if the size of the risk set later increases. Snapinn et al[5] suggest two ways to attempt to mitigate this problem. If the time-varying covariate of interest has been discretized, they suggest defining bins with different cut points that result in larger risk sets. Alternatively, they indicate that the use of more granular time units, such as weeks instead of days, may eliminate this problem. This problem is not unfamiliar in survival analysis as it can arise in survival estimation from data with left truncation (ie, delayed entry).

Despite this limitation, the similarity of this estimator to the Kaplan-Meier estimator and its simplicity make it a popular and proper choice for displaying the relationship between survival probability and a time-varying covariate value of interest for many datasets. The graphical display proposed by Snapinn et al[5] has a close correspondence to the effect estimate from a time-varying Cox model when its assumptions hold and for the special case of the covariate remaining constant over time. It is also highly accessible, as it can be implemented using the `survfit` function in the `survival` package in R after putting the data in an interval format. For additional details on utilizing this estimator in R, see Snapinn et al[5] or the R code in the Supplementary Materials. This estimator falls short, however, in that it does not provide a survival estimate for patients with non-constant covariate histories.

2.3 | Survival estimator of Xu et al

Xu et al[6] developed a similar extension of the Kaplan-Meier estimator but included stabilized weights to adjust for confounding variables in observational studies. Their aim was to display the relationship between a time-varying treatment and the time to a pre-specified medical event or complication. In the setting of an observational study, several variables influence whether a patient is currently on or off a given treatment. Snapinn et al's[5] extended estimator, or another crude estimator that does not account for these patient characteristics, could be misleading for observational study, which includes a clinical trial in which there is subject-specific non-compliance. Thus, Xu et al[6] defined the weighted risk set size at time t_j for the group with covariate value k as

$$n_{jk}^{w} = \sum_{i} I(X_i(t_j) = k) I(Y_i \geq t_j) w_i(t_j)$$

and the weighted event set size at time t_j for the group with covariate value k as

$$d_{jk}^{w} = \sum_{i} I(X_i(t_j) = k) I(\delta_i = 1) I(Y_i = t_j) w_i(t_j).$$

The stabilized weight $w_i(t_j)$ for patient i at time t_j is given by

$$w_i(t_j) = \frac{P(X_i(t_j) = x_i(t_j))}{P(X_i(t_j) = x_i(t_j) | \boldsymbol{c}_i(t_j))},$$

where the denominator is patient i's propensity score at time t_j, conditional on the additional measured covariate values for patient i at time t_j, $\boldsymbol{c}_i(t_j)$. Xu et al[6] presented their estimator in the context of a binary time-varying treatment covariate and thus estimated $P(X_i(t_j) = x_i(t_j) | \boldsymbol{c}_i(t_j))$ for each patient at each event time using logistic regression models. Note, though, that Xu et al's[6] estimator can also accommodate time-varying covariates with more than two levels. A multinomial regression or alternative modeling approach may be employed to compute the stabilized weights in this case. Their extension of the Kaplan-Meier estimator is thus

$$\tilde{S}_k^{w}(t) = \prod_{j: t_j \leq t} \left(1 - \frac{d_{jk}^{w}}{n_{jk}^{w}}\right). \tag{2}$$

Similar to Snappin et al's[5] estimator, Xu et al's[6] estimator can be interpreted as the survival function for a group of patients who always has covariate value k, after adjusting for the variables that influence treatment selection. Like Snapinn et al's[5] extended estimator, Xu et al's[6] estimator assumes that the risk of an event is dependent only on the current value of the covariate and it features non-monotonic risk sets. It offers an important extension to Snapinn et al's[5] estimator, as it adjusts for time-varying confounding. We provide R code to implement Xu et al's[6] estimator in the Supplementary Materials.

3 | PROPOSED ESTIMATORS

In this section, we present unadjusted and covariate-adjusted extensions of the Kaplan-Meier estimator for displaying a survival distribution that incorporates any covariate path. This is highly relevant in clinical practice, where patients and clinicians want to know the time-to-event distributions for several possible covariate trajectories, and in particular, trajectories that are truly time-varying and non-constant. In the oncology setting, it might be of clinical interest to compare the survival curves for patients who achieve tumor response 30 days after initiating treatment compared to 60 days after initiating treatment. However, such displays are not provided by the Snappin et al[5] or Xu et al[6] estimators.

To define a covariate path of interest, we let $\boldsymbol{z} = \{z_1, \dots, z_m\}$ denote the set of m covariate values on the path, and let $\boldsymbol{r} = \{r_1, \dots, r_{m-1}\}$ be the set of $m-1$ transition times at which the covariate value changes. The covariate path is then defined by the function

$$z(t) = \begin{cases} z_1, & 0 \le t \le r_1 \\ z_2, & r_1 < t \le r_2 \\ \vdots & \vdots \\ z_m, & r_{m-1} < t. \end{cases}$$

We let $\underline{Z}(t)$ denote the set of covariate values and transition times on the path up until time t. We derive estimators for $P\big(T > t|\underline{Z}(t)\big)$.

3.1 | Unadjusted estimator

Our unadjusted estimator extends Snapinn et al's[5] estimator by replacing the indicator in the risk and event sets, $I(X_i(t_j) = k)$, with the alternative indicator, $I(X_i(t_j) = z(t_j))$. Instead of extracting data for a covariate value k that is constant over time, we now incorporate an indicator to count the number of patients who have the discrete covariate value equal to the current path value, $z(t_j)$. This can be a different value from one event time to the next, depending on how the covariate path is defined. This substitution results in a risk set size at time t_j for patients who have covariate value $z(t_j)$ of

$$n_{\underline{Z}(t_j)} = \sum_{i=1}^{n} I(Y_i \ge t_j) I(X_i(t_j) = z(t_j)),$$

and a corresponding event set size of

$$d_{\underline{Z}(t_j)} = \sum_{i=1}^{n} I(Y_i = t_j) I(X_i(t_j) = z(t_j)) I(\delta_i = 1).$$

With these updated risk and event sets, we arrive at our unadjusted estimator

$$\hat{S}_{\underline{Z}(t)}(t) = \prod_{j:\, t_j \le t} \left(1 - \frac{d_{\underline{Z}(t_j)}}{n_{\underline{Z}(t_j)}}\right). \tag{3}$$

$\hat{S}_{\underline{Z}(t)}(t)$ is the estimated probability of survival beyond time t for a hypothetical group of patients who have the covariate values and transition times $\underline{Z}(t)$. This estimator takes into account the current covariate path value and any past covariate path values, as well as their transition times. In the case of tumor response at day 30, the estimator would have different interpretations at days 20 and 60. $\hat{S}_{\underline{Z}(20)}(20)$ represents the probability of surviving beyond day 20 for a hypothetical group of patients who have not had a tumor response by day 20, and $\hat{S}_{\underline{Z}(60)}(60)$ is the probability of surviving beyond day 60 for a hypothetical group of patients who achieved tumor response at day 30.

This estimator, like Snapinn et al's[5] estimator, is constructed under the assumption that hazard of the event depends on the patient's current, and not past, time-varying covariate value. With this assumption, $\hat{S}_{\underline{Z}(t)}(t)$ is consistent, and a Greenwood-type variance can be derived. See the Supplementary Materials for corresponding proofs and derivations. The Greenwood 95% confidence interval for $\hat{S}_{\underline{Z}(t)}(t)$ is

$$\hat{S}_{\underline{Z}(t)}(t) \pm 1.96\sqrt{\widehat{\mathrm{Var}}\left(\hat{S}_{\underline{Z}(t)}(t)\right)}, \tag{4}$$

where

$$\widehat{\mathrm{Var}}\left(\hat{S}_{\underline{Z}(t)}(t)\right) = \left(\hat{S}_{\underline{Z}(t)}(t)\right)^2 \sum_{j:t_j \le t} \frac{d_{\underline{Z}(t_j)}}{n_{\underline{Z}(t_j)}\left(n_{\underline{Z}(t_j)} - d_{\underline{Z}(t_j)}\right)}.$$

Note that the bounds of these confidence intervals are not restricted to [0, 1], and use of a log(-log) confidence interval may be preferred in this setting. Confidence intervals based on this transformation or other transformations can be obtained using the Delta Method. We provide R code to implement our proposed unadjusted estimator in the Supplementary Materials.

3.2 | Adjusted estimator

In observational studies, a confounder-adjusted Kaplan-Meier curve may be more useful for displaying the relationship between a covariate and survival. We extend Xu et al's[6] weighted estimator to handle covariate paths by changing the indicator in the risk and event sets from $I(X_i(t_j) = k)$ to $I(X_i(t_j) = z(t_j))$. The risk and event sets used to compute the adjusted estimator are identical to those used to compute the unadjusted estimator, but include $w_i(t_j)$, a weight for patient i at event time t_j. We compute this estimator using stabilized weights, calculated as

$$w_i(t_j) = \frac{P(X_i(t_j) = x(t_j))}{P(X_i(t_j) = x(t_j)|\boldsymbol{c}_i(t_j))}.$$

However, inverse probability weights or another choice of weights could be implemented as an alternative. The adjusted estimator's weighted risk set size at time t_j is thus

$$n^w_{\underline{Z}(t_j)} = \sum_{i=1}^{n} I(Y_i \ge t_j) I(X_i(t_j) = z(t_j)) w_i(t_j),$$

and its weighted event set size at time t_j is

$$d^w_{\underline{Z}(t_j)} = \sum_{i=1}^{n} I(Y_i = t_j) I(X_i(t_j) = z(t_j)) I(\delta_i = 1) w_i(t_j).$$

With these updated sets, the proposed adjusted estimator is

$$\hat{S}^w_{\underline{Z}(t)}(t) = \prod_{j:t_j \le t} \left(1 - \frac{d^w_{\underline{Z}(t_j)}}{n^w_{\underline{Z}(t_j)}}\right). \tag{5}$$

The interpretation of $\hat{S}^w_{\underline{Z}(t)}(t)$ remains similar to the that of the unadjusted estimator. $\hat{S}^w_{\underline{Z}(t)}(t)$ is the estimated probability of survival beyond time t for a patient who has covariate values and transition times defined by $\underline{Z}(t)$, after adjusting for the covariates included in matrix $\boldsymbol{C}$. $\hat{S}^w_{\underline{Z}(t)}(t)$ is also a consistent estimator (see the Supplementary Materials). Its analogous Greenwood variance estimator is:

$$\widehat{\text{Var}}\left(\hat{S}^w_{\underline{Z}(t)}(t)\right) = \left[\hat{S}^w_{\underline{Z}(t)}(t)\right]^2 \sum_{j:t_j \le t} \frac{d^w_{\underline{Z}(t_j)}}{n^w_{\underline{Z}(t_j)}\left(n^w_{\underline{Z}(t_j)} - d^w_{\underline{Z}(t_j)}\right)}. \tag{6}$$

This variance estimator is derived assuming that the weights are known, rather than estimated, and so it may underestimate the true variability when the weights are estimated. Bootstrapping may be used for a more robust estimate of the weighted estimator's variance and is most appropriate in the case of small risk sets. See the Supplementary Materials for a complete derivation. R code is also available in the Supplementary Materials to facilitate the implementation of our proposed adjusted estimator.

4 | EXAMPLES

In this section, we illustrate the utility of the proposed estimators with a lung cancer study and a clinical trial of primary biliary cholangitis. In addition to applying the proposed estimators, we compare the resulting survival estimates to those from Snapinn et al[5] and Xu et al's[6] methods, and to survival estimates obtained from a Cox proportional hazards models with time-varying covariates. Smith et al[12] describe how one can obtain fitted survival probabilities based on a covariate path from a time-varying Cox model in R, and we apply their approach. All analyses are performed using the R statistical software.[14]

4.1 | Lung cancer and tumor response

We first apply the unadjusted estimator to a dataset of 48 patients with extensive small cell carcinoma of the lung found in Simon and Makuch.[4] The lung cancer patients in this dataset were identified as either tumor responders or tumor non-responders at each recorded death time, and thus tumor response is a time-varying covariate. A covariate value of 0 represented no tumor response and a covariate value of 1 represented tumor response; as such, the covariate was nonreversible, meaning those who had tumor response could never return to the tumor non-response state. Three survival curves are displayed using solid lines in Figure 1. The black curve is the estimated survival function for a patient who had tumor response at the start of follow-up and remained a responder for the duration of the follow-up (ie, transitioning from covariate value 0 to 1 on day 0). This curve is equivalent to Snapinn et al's[5] estimator for patients who were always tumor responders. The light gray curve is equivalent to Snapinn et al's[5] estimator for never responders, where each patient's time at risk spent as a non-responder is used in the estimation. The medium gray curve is one example of our proposed estimator, which depicts survival for patients who had a tumor response at 30 days. This covariate path is represented as $z(t) = 0$ for $t < 30$ and $z(t) = 1$ for $t \ge 30$. This estimate illustrates that a delay of 30 days in tumor response is associated with a decrease in the probability of survival early on, but beyond 400 days, the survival experiences of patients who respond at time 0 versus time 30 are nearly equivalent.

We also obtain survival estimates from the time-varying Cox model, $\lambda_i(t) = \lambda_0(t)\exp\{\beta z_i(t)\}$, where $z_i(t)$ represents tumor response status at time t. These estimates are displayed using dotted lines in Figure 1. The black dotted curve corresponds to survival estimates from this model when setting $z_i(t) = 1$ for all t, and the light gray dotted curve corresponds to the survival estimates when setting $z_i(t) = 0$ for all t. The dark gray dotted line corresponds to the survival estimates from the Cox model when setting $z_i(t) = 0$ for $t < 30$ and $z_i(t) = 1$ afterward. Since the always responder and never responder curves cross when applying Snapinn et al's estimator,[5] the functional form of the time-varying Cox model may not be appropriate for this dataset. The underlying functional assumption of this time-varying Cox model is that $\lambda_i(t; z_i(t) = 0) = \lambda_0(t)$ for the never responders and $\lambda_i(t; z_i(t) = 1) = \lambda_0(t)\exp\{\beta\}$ for the always responders. This functional form requires one of these curves to always fall above the other, depending on the sign of β. The tumor responder curve produced by the proposed unadjusted estimator closely matches the tumor responder curve from the Cox model; however, the tumor non-responder curves are very different, as evidenced by the solid light gray and dotted light

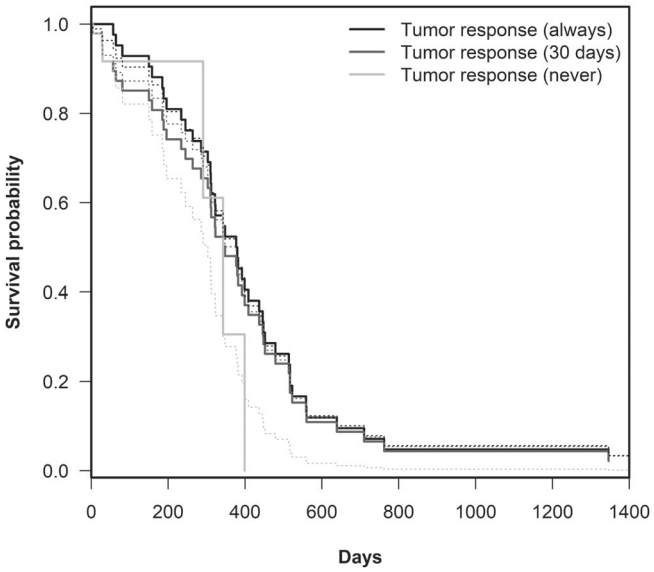

FIGURE 1 Comparison of estimators as applied to the lung cancer dataset from Simon and Makuch.[4] The solid black and light gray curves depict the survival estimates for patients who always have a tumor response and who never have a tumor response using the Snapinn et al[5] method. The solid gray curve displays the proposed unadjusted survival estimator for patients who achieve tumor response at 30 days. The dotted curves display the estimates from a time-varying Cox model

gray curves, which are far apart. The curve depicting the hypothetical patient group with a tumor response at 30 days aligns relatively closely with the comparable curve produced by the Cox model, since the assumed functional form of the Cox model appears to be more reasonable at the earlier time points. This example illustrates the robustness of our proposed unadjusted estimator when the assumptions of the time-varying Cox model do not hold for a dataset; our proposed unadjusted estimator does not make assumptions on the form of the relationship between a covariate path and survival.

4.2 | Primary biliary cholangitis and bilirubin level

Our second example illustrates the difference in prognosis for patients with primary biliary cholangitis (PBC) who had normal versus elevated bilirubin levels. Serum bilirubin has been identified as a strong prognostic indicator for PBC, with repeated high levels associated with disease progression.[15] It is of clinical interest to visualize the survival function not just for patients who have consistently high bilirubin levels but also for patients whose bilirubin levels increase or decrease during the course of treatment. We apply both our adjusted and unadjusted survival estimators for illustration.

The dataset used in this example includes baseline and follow-up data for 312 patients enrolled in a clinical trial at the Mayo Clinic, which aimed to assess the efficacy of D-penicillamine as a treatment for PBC. Along with treatment information, laboratory measurements were collected at follow-up visits, and patient survival status was also recorded. There were 1945 patient visits recorded over the course of 14 years. These data are available in the `survival` package in R,[16] with baseline measurements found in the `pbc` dataset and measurements from sequential visits found in the `pbcseq` dataset. Additional details on these datasets can be found in Murtaugh et al[17] and Therneau and Grambsch.[18]

The data were prepared by first dichotomizing the continuous bilirubin variable into high and normal categories. High levels were defined as bilirubin measurements of >2.0 mg/dL, consistent with the fact that repeated measurements at or above this threshold have been associated with overall progression.[15] Consistent with previous studies, at each time, we grouped those who received a liver transplant with those who were lost to follow-up, since survival without transplant was the primary outcome of interest. We also considered the baseline characteristics of sex, age, and treatment assignment, along with time-varying stage of PBC, as potential confounders.

Figure 2 provides a comparison between our proposed estimators and those of Snapinn et al[5] and Xu et al.[6] The solid lines depict the unadjusted estimators and the dotted lines depict the corresponding adjusted estimators. The stabilized weights were constructed using the potential confounders. The close proximity of the adjusted and unadjusted estimators suggests that the potential confounders that were considered are not contributing substantially to confounding, meaning that they may not be strongly associated with bilirubin and with survival. The large gap between the curves representing always normal and always high bilirubin levels provides clear evidence of the poor prognosis associated with high bilirubin. However, these curves fail to take the fluctuating nature of bilirubin levels into consideration, and thus are not applicable to many patients. Alongside these curves, we display our proposed survival estimates for patients defined by

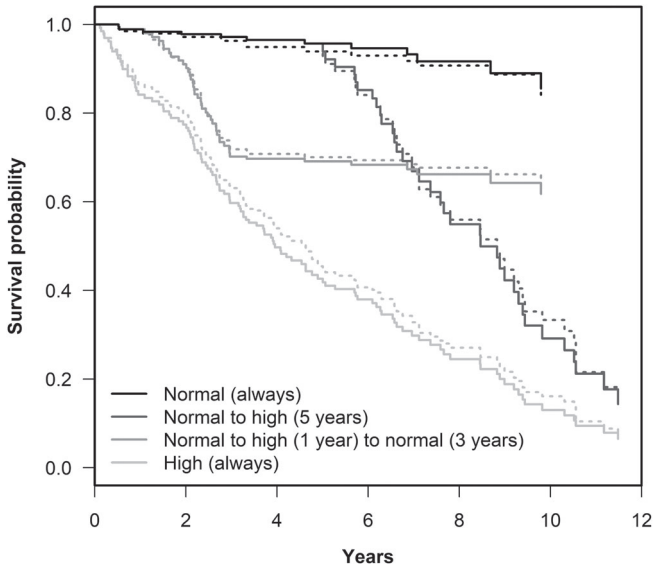

FIGURE 2 Comparison of survival estimators for primary biliary cholangitis patients. The solid lines depict unadjusted estimators and the dotted lines depict adjusted estimators. The middle curves depict two examples of the proposed estimators in this article, while the outer curves correspond to Snapinn et al[5] (solid line) and Xu et al[6] (dotted line) estimators

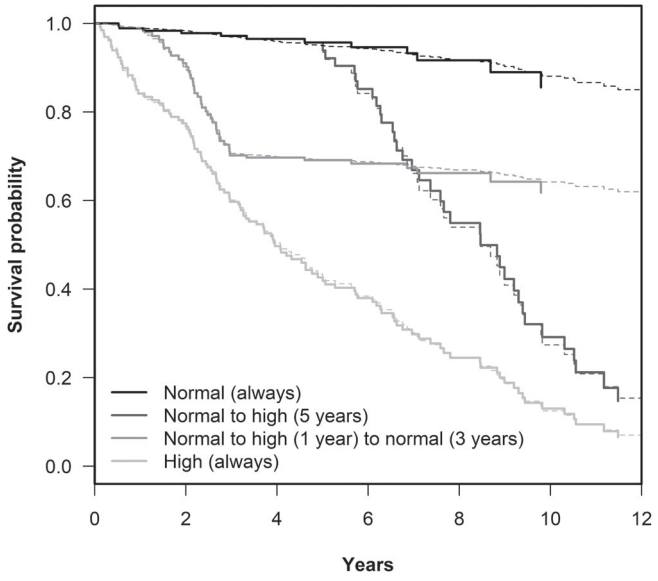

FIGURE 3 Comparison of the proposed unadjusted survival estimator to estimates obtained from a time-varying Cox model for primary biliary cholangitis patients. The solid curves correspond to the proposed estimators and the dotted curves correspond to the time-varying Cox model estimates

two potential covariate paths: in one, a patient has onset of high bilirubin at year five and remains high thereafter, and in the second, a patient has onset of high bilirubin at year one, and returns to normal at year three, and remains normal thereafter. As expected, the survival estimates associated with these intermediate bilirubin trajectories are intermediate to the always normal and always high estimates. They provide important insight into the sensitivity of survival to the timing of the bilirubin changes and could justify further study of aggressive treatment to lower bilirubin and to maintain it at a low level, if possible.

We finally examine how closely the estimates from a Cox model with a time-varying covariate align with the estimates produced by our proposed unadjusted estimator. We fit a Cox model with time-varying bilirubin level as the only covariate. We perform a test on the Schoenfeld residuals and obtain a *P*-value of 0.85. Thus, there is not evidence that the survival curves corresponding to patients with always normal vs. always high levels of bilirubin cross. We next obtain fitted survival probabilities from the time-varying Cox model and compare them to the survival curves produced by our proposed unadjusted estimator. The plot in Figure 3 displays our proposed unadjusted estimators in solid lines and the survival curve from the Cox model in the dotted lines. This time, there is a near perfect correspondence between estimates in this example. The assumed functional form of the time-varying Cox model appears to be quite reasonable in this example.

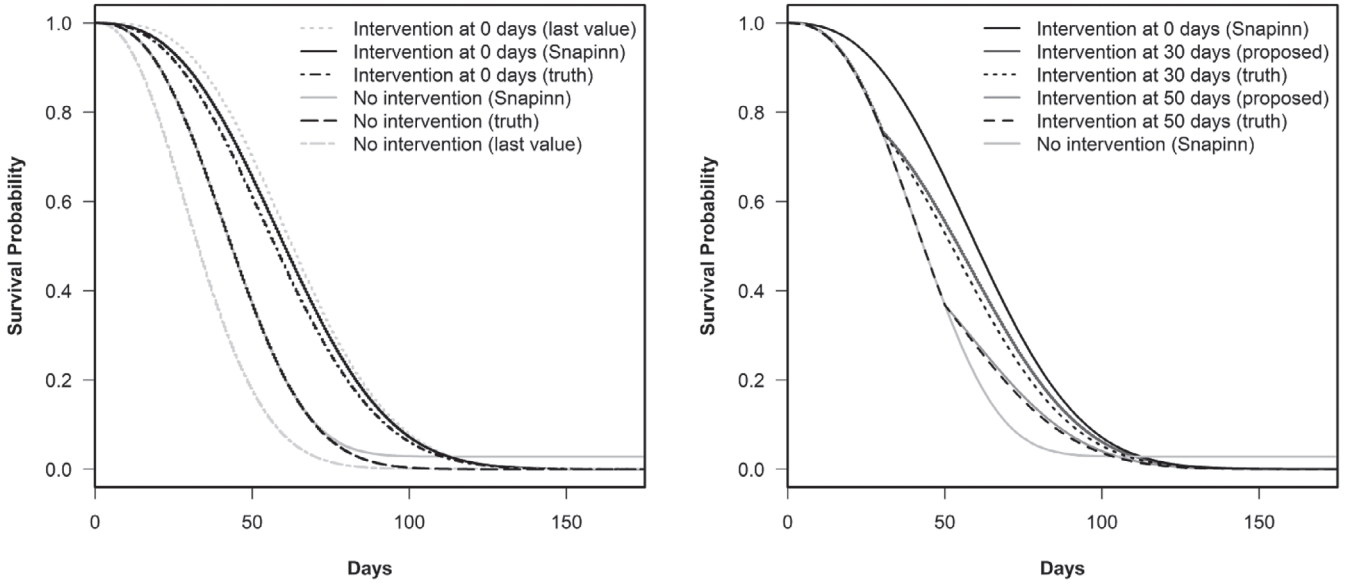

FIGURE 4 Comparison of various survival estimators to the true survival functions using 10 000 simulated datasets. The left plot displays the Snapinn et al[5] estimators, the true survival curves representing never having the intervention and always having the intervention, and the biased last covariate value estimators. The right plot displays the proposed estimators with covariate paths defined by interventions at 30 or 50 days, the true survival curves corresponding to these paths, and the Snapinn et al[5] estimators for comparison

5 | SIMULATION STUDY

We conducted a simulation study to evaluate the properties of our proposed unadjusted estimator. In this simulation, we provide a comparison of our proposed estimator, the Snapinn et al[5] estimator, and the last covariate value approach; we illustrate the consistency of our proposed unadjusted estimator and examine how the distribution of the time-varying covariate value affects how close the estimated survival functions are to the underlying truth in a finite sample. We generated survival times from a proportional hazards model, $\lambda_i(t) = \lambda_0(t)\exp\{\beta z_i(t)\}$, which incorporated a single time-varying covariate $z_i(t)$. This covariate encodes whether or not a patient was given a certain intervention by time t, where $z_i(t) = 0$ if the patient had not received the intervention by time t and $z_i(t) = 1$ if the patient received the intervention by time t. A β value of -0.7 was selected so that the true hazard ratio comparing those who always received the intervention to those who never received the intervention was 0.5. The baseline hazard was taken to be that for a Weibull distribution with parameter values $\lambda = 0.02$ and $\gamma = 2.5$ under the parametrization $\lambda_0(t) = \lambda\gamma(t\lambda)^{\gamma-1}$. A censoring time was generated for each patient from an exponential distribution with a rate parameter of 0.0025. An intervention time was generated for each patient from an exponential distribution with a rate parameter of 0.033, so that the average intervention time was approximately 30 days into follow-up. This set-up resulted in some patients with observed intervention times, and other patients with unobserved intervention times due to their time of death or censoring preceding their simulated intervention time. We simulated 10 000 datasets with 150 patients each to evaluate the proposed estimator. For each dataset, the proposed unadjusted estimator was calculated for two covariate paths: (a) time of intervention of 30 and (b) time of intervention of 50. We also calculated the Snapinn et al[5] estimates for having the intervention at time 0 and never having the intervention. For comparison, we also applied the last covariate value method to each simulated dataset.

The average values of the estimates are depicted in Figure 4, along with the true values. The left plot in Figure 4 reveals the substantial magnitude of the guarantee-time bias associated with the last covariate value approach. The Snapinn et al[5] estimate for never having an intervention almost perfectly approximates the true underlying survival function, while the Snapinn et al[5] estimate for having the intervention from time 0 exhibits a small bias. This can be attributed to the fact that the generated time-varying covariate is non-reversible with an average intervention time of 30 days. Had the dataset been generated with an earlier average intervention time, we would expect the Snapinn et al[5] estimator to align more closely for the intervention group.

The right plot in Figure 4 displays the Snapinn et al[5] curves, the proposed estimators for interventions at time 30 or 50, and the true survival functions corresponding to these intervention times. This plot demonstrates the consistency of

	Time (days)	Empirical SD	Average Greenwood SE
$\hat{S}_{\underline{Z}(25)}(25)$	25	0.042	0.042
$\hat{S}_{\underline{Z}(50)}(50)$	50	0.054	0.053
$\hat{S}_{\underline{Z}(75)}(75)$	75	0.042	0.041
$\hat{S}_{\underline{Z}(100)}(100)$	100	0.023	*0.022

Note: *13 observations were omitted from this average since the Greenwood variances were undefined.

TABLE 1 Comparison of the empirical standard deviation and average analytical Greenwood standard error of the proposed estimator with intervention time of 30 days across 10 000 simulated datasets

the proposed unadjusted estimators, and their advantage over the Snapinn et al[5] estimators, when the target estimand is the survival distribution associated with a change in covariate value at a given intermediate time point.

Table 1 compares the empirical standard deviations of the proposed estimator for intervention at time 30 with the Greenwood-type analytical estimator (4). The analytical standard deviation closely aligns with the empirical standard deviation at 25, 50, 75, and 100 days of patient follow-up.

6 | DISCUSSION

We have proposed weighted and unweighted extensions of the Kaplan-Meier estimator for the purpose of displaying the survival distributions for patients defined by any covariate paths. These estimators build upon the work of Snapinn et al[5] and Xu et al,[6] and allow for visualization of survival in the presence of any covariate path, and not just the extreme cases in which the covariate is constant-valued. Our adjusted estimator offers the opportunity to incorporate additional clinical features to avoid confounding bias. These additional features may be baseline features or time-varying features, since weights are calculated at each event time with the most current information. In addition, our estimators maintain a simplistic form while providing an unbiased representation of survival experience with time-varying covariates.

Like Snapinn et al's[5] and Xu et al's[6] estimators, our proposed extensions integrate each patient's time-varying information to display the association between repeatedly measured covariates and survival. We have demonstrated each estimator's utility with examples involving both reversible and non-reversible binary covariates. While binary time-varying covariates are common in the medical literature and thus the focus of our examples, the proposed estimators can be used for any time-varying covariates, as long as they are either discrete or appropriately discretized.

Our proposed estimators are constructed under a key underlying assumption that requires careful consideration. We assume that the hazard of the event depends only on the current value of the time-varying covariate, and not on its history. In the case of a non-reversible binary covariate such as onset of intervention, this implies that the risk of the event at time t for a patient who initiated the intervention at time t^- is the same as for a patient who initiated the intervention at time 0. While this may be appropriate in some settings, it may not be in others. Note that in the case of very large datasets, it may be possible to relax this assumption. One could estimate the hazard using a group of subjects with the specified patterns over the past defined time intervals, rather than relying on only the current covariate value for hazard estimation.

The second example illustrates that the effect of the covariate path on survival displayed by the proposed unadjusted estimator closely corresponds to the effect estimate in the time-varying Cox model, when the Cox model's assumed functional form is reasonable. This follows from the correspondence between Snapinn et al's estimator[5] and the effect in the time-varying Cox model. Our methods are applicable regardless of whether this assumption is satisfied, but this correspondence is a strength of our method.

There are also several practical considerations for implementing these estimators. One must ensure that the specified covariate paths are clinically reasonable to avoid misleading conclusions. It is also important to check that the specified covariate paths result in sufficiently large risk set sizes at each event time. Otherwise, the survival estimates will be subject to considerable variability. Our simulations illustrate that covariate paths with transition times that are close to a central measure of the time-varying covariate's distribution, such as the mean, may yield better results in finite samples. Like the estimators of Snapinn et al[5] and Xu et al,[6] it is also possible for the size of a risk set to decrease to zero and later increase. A similar issue can arise when estimating survival probabilities with left-truncated data. We described suggestions from Snapinn et al[5] for handling this situation, but we acknowledge that this is also a limitation of our proposed estimators.

When using the adjusted survival estimator, a propensity score is calculated at each event time to construct the stabilized weights. These calculations can be problematic at late event times if the sample size is small, as convergence issues can arise. Violations of the positivity assumption can occur when either a covariate group is no longer represented, or when a level of an explanatory variable is no longer represented in either group.[19] Furthermore, our adjusted estimator requires risk sets that are large enough so that complete separation of dependent variables does not occur. Even when complete separation or the positivity assumption does not pose a problem, stabilized weights will be subject to considerable variability with small sample sizes. For this reason, the proposed adjusted estimator may perform best in studies with large sample sizes and in studies where many of the patient characteristics are continuous, rather than categorical.

While we primarily focus on providing visual evidence of survival function differences using our unadjusted and adjusted estimators, hypothesis testing may be of interest. Xu et al[6] describe modified log-rank and Wilcoxon tests for testing the differences in survival functions between two hypothetical patient groups: those who are always on and those who are always off a treatment. These methods would not apply to the proposed estimators, since the patient groups may overlap between cohorts of interest. To test for significant differences in the survival probability between two hypothetical cohorts, we suggest using a permutation test to determine whether the difference between the survival probabilities at a time point of interest is significant. For example, after 2 months of follow-up, it might be of interest to determine whether those who are always on the treatment have a statistically significant difference in survival probability compared to a hypothetical group of patients who stop receiving the treatment after 1 month. One could hold each patient's covariate values fixed and permute (Y, δ) across subjects. Then, the observed difference in survival estimates at the fixed time point of interest could be compared to the differences obtained under each permutation. This would be best applied to external time-varying covariates.

In conclusion, our proposed extensions of the Kaplan-Meier estimator have the benefit of displaying the survival function for individual patients defined by clinically relevant covariate paths. These nonparametric estimators are attractive for their simplicity and familiarity due to their similarity to the commonly used Kaplan-Meier estimator, but their interpretation must be carefully stated. Additionally, we have derived analytical variance estimators, which enable confidence interval construction. These estimators greatly improve upon widely used incorrect approach of treating time-varying status as a fixed baseline covariate.

ACKNOWLEDGEMENTS

We are grateful to Dr. Bella Vakulenko-Lagun and to the members of Harvard's Biostatistics Cancer Working Group for their helpful suggestions and feedback on this research. This research was supported by the National Science Foundation Graduate Research Fellowship under Grant No. 000390183 (MJ) and NIH Training Grant in Quantitative Sciences for Cancer Research under Grant No. T32CA009337 (MJ) and R01NS094610 (RAB). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are available in the `survival` package in R and in the article by Simon and Makuch.[4] The R code for the analysis and simulation study is provided as a Supplementary Material.

ORCID

Melissa Jay https://orcid.org/0000-0002-8740-4803
Rebecca A. Betensky https://orcid.org/0000-0002-3793-1437

REFERENCES

1. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.
2. D'Agostino RB. Beyond baseline data: the use of time-varying covariates. *J Hypertens*. 2008;26(4):639-640.
3. Beyersmann J, Gastmeier P, Wolkewitz M, Schumacher M. An easy mathematical proof showed that time-dependent bias inevitably leads to biased effect estimation. *J Clin Epidemiol*. 2008;61(12):1216-1221.
4. Simon R, Makuch RW. A non-parametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias. *Stat Med*. 1984;3(1):35-44.
5. Snapinn SM, Jiang QI, Iglewicz B. Illustrating the impact of a time-varying covariate with an extended Kaplan-Meier estimator. *Am Stat*. 2005;59(4):301-307.
6. Xu S, Shetterly S, Powers D, et al. Extension of Kaplan-Meier methods in observational studies with time-varying treatment. *Value Health*. 2012;15(1):167-174.

7. Dubin JA, Müller H, Wang J. Event history graphs for censored survival data. *Stat Med*. 2001;20(19):2951-2964.
8. Feuer EJ, Hankey BF, Gaynor JJ, Wesley MN, Baker SG, Meyer JS. Graphical representation of survival curves associated with a binary non-reversible time dependent covariate. *Stat Med*. 1992;11(4):455-474.
9. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*. 1983;1(11):710-719.
10. Li Y, Schaubel DE, He K. Matching methods for obtaining survival functions to estimate the effect of a time-dependent treatment. *Stat Biosci*. 2014;6(1):105-126.
11. Giobbie-Hurder A, Gelber RD, Regan MM. Challenges of guarantee-time bias. *J Clin Oncol*. 2013;31(23):2963.
12. Smith AR, Goodrich NP, Beil CA, et al. Graphical representation of survival curves in the presence of time-dependent categorical covariates with application to liver transplantation. *J Appl Stat*. 2019;46(9):1702-1713.
13. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. Hoboken, NJ: John Wiley & Sons; 2011.
14. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020. https://www.R-project.org/.
15. Lammers WJ, Kowdley KV, Buuren HR. Predicting outcome in primary biliary cirrhosis. *Ann Hepatol*. 2014;13(4):316-326.
16. Therneau T. *A Package for Survival Analysis in R*. R package. Version 3.2-7; 2000.
17. Murtaugh PA, Dickson ER, Van Dam GM, et al. Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology*. 1994;20(1):126-134.
18. Therneau T, Grambsch P. *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer; 2000.
19. Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol*. 2010;171(6):674-677.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Jay M, Betensky RA. Displaying survival of patient groups defined by covariate paths: Extensions of the Kaplan-Meier estimator. *Statistics in Medicine*. 2021;40:2024–2036. https://doi.org/10.1002/sim.8888