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

Melissa N. Jay[1] and Rebecca A. Betensky[2]

[1]Department of Biostatistics, University of Iowa; [2]Department of Biostatistics, NYU School of Global Public Health

Introduction

The Kaplan-Meier estimator is widely used in the medical literature to compare survival distributions between two or more patient groups of interest. It can used to make comparisons between patient groups defined by baseline characteristics but suffers from *guarantee-time bias* if patient groups are defined by time-varying characteristics.[1]

Several valid approaches exist for illustrating the relationship between a time-varying covariate value and patient survival time; however, these methods either require the interpretation that patients always have a single value of the covariate of interest or handle only non-reversible time-varying covariates.

We propose two estimators that can be used to display survival probabilities for hypothetical patient groups defined by clinically relevant sequences of covariate values.

Existing Estimators

Snapinn et al.[2] and Xu et al.[3] proposed estimators to display the relationship between a time-varying covariate value, k, and survival. For patient i at event time j, denote:

- Y_i = follow-up time
- δ_i = event indicator (1 = event, 0 = censored)
- $X_i(t_j)$ = time-varying covariate value
- $\boldsymbol{c_i}(t_j)$ = additional covariate values

The size of the risk set is defined as

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

and the size of the event set is defined as

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

where $w_i(t_j) = 1$ for all i and j in Snapinn et al.'s estimator and

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

in Xu et al.'s estimator. Both estimators have an analogous form to the Kaplan-Meier estimator:

$$\hat{S}_k(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_{jk}}{n_{jk}}\right). \qquad (1)$$

Proposed Estimators

We propose an unweighted estimator and a weighted estimator to display survival for patient groups defined by covariate paths. These estimators directly extend Snapinn et al. and Xu et al.'s estimators.

Let $\boldsymbol{z} = \{z_1, \ldots, z_m\}$ be the set of m values on a covariate path of interest, and let $\boldsymbol{r} = \{r_1, \ldots, r_{m-1}\}$ be the set of $m-1$ transition times at which the covariate value changes. The values on the covariate path can be defined by the function:

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

Let $\underline{Z}(t)$ denote the set of covariate values and transition times on the path up until time t.

The size of the risk set is defined as

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

and the size of the event set is defined as

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

For our proposed unweighted estimator, $w_i(t_j) = 1$ for all i and j. For our proposed weighted estimator, each $w_i(t_j)$ is the stabilized weight

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

where $P(X_i(t_j) = x_i(t_j) | \boldsymbol{c}_i(t_j)))$ can be estimated using an appropriate model specific to each event time.

The estimator is thus

$$\hat{S}_{\underline{Z}(t)}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_{\underline{Z}(t_j)}}{n_{\underline{Z}(t_j)}}\right). \qquad (2)$$

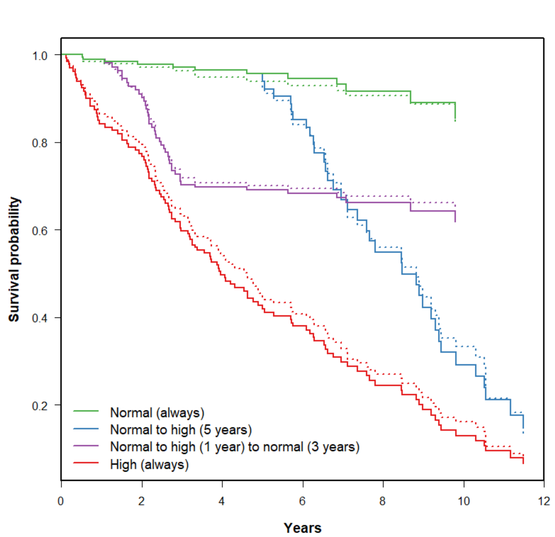
These estimators extend the existing estimators in that they allow for any covariate path, not just a constant covariate path.

Interpretation: *$\hat{S}_{\underline{Z}(t)}(t)$ is the estimated probability of survival beyond time t for a hypothetical group of patients who have the covariate values and transition times defined by $\underline{Z}(t)$.*

Example

Dataset: **pbcseq** from the **survival** package in R. The **pbcseq** dataset includes baseline and follow-up data for 312 patients who were enrolled in a clinical trial at the Mayo Clinic. This trial aimed to assess the efficacy of *D-penicillamine* as a treatment for primary biliary cholangitis (PBC). [4]

Serum bilirubin has been identified as a strong prognostic indicator for PBC, with repeated high levels associated with disease progression.[5] We dichotomize serum bilirubin levels into high and normal categories, defining high levels as greater than 2.0 mg/dL.



The figure above displays the proposed unweighted (solid lines) and weighted (dotted lines) survival estimators for various bilirubin trajectories. The weighted estimators were calculated using patient sex, age, treatment assignment, and PBC stage as covariates in the stabilized weights.

Note that our extended estimators are the purple and blue curves. They provide information not present in the red and green curves and show very different survivor functions for different covariate paths.

Conclusion

- Estimators incorporate clinically relevant milestones on a patient's treatment course to illustrate how covariate changes are associated with survival.
- Estimators have analogous form to the widely-used Kaplan-Meier estimator.
- Careful interpretation is required.
- Weighted estimator becomes unstable or cannot be computed if risk sets get too small.

References

[1] A Giobbie-Hurder, RD Gelber, and MM Regan. Challenges of guarantee-time bias. *J Clin Oncol*, 31(23):2963, 2013.

[2] SM Snapinn, QI Jiang, and B Iglewicz. Illustrating the impact of a time-varying covariate with an extended kaplan-meier estimator. *Am Stat*, 59(4):301–307, 2005.

[3] S Xu, S Shetterly, D Powers, MA Raebel, TT Tsai, PM Ho, and D Magid. Extension of kaplan-meier methods in observational studies with time-varying treatment. *Value Health*, 15(1):167–174, 2012.

[4] T Therneau and P Grambsch. *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer-Verlag, 2000.

[5] WJ Lammers, KV Kowdley, and HR van Buuren. Predicting outcome in primary biliary cirrhosis. *Ann Hepatol*, 13(4):316–326, 2014.

Acknowledgements

- National Science Foundation Graduate Research Fellowship (Grant No. 000390183)
- NIH Training Grant in Quantitative Sciences for Cancer Research (Grant No. T32CA009337)

Contact Information

- Email: melissa-jay@uiowa.edu
- Twitter: @MelissaJay