RESTRICTED MEAN ANALYSIS ACROSS MULTIPLE FOLLOW-UP INTERVALS

by

Nabihah Tayob

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2013

Doctoral Committee:

Associate Professor Susan Murray, Chair
Associate Professor Kevin R. Flaherty
Professor Yi Li
Professor Douglas E. Schaubel

For my parents, Fayroza and Fazul, and my brothers, Rashaad and Amin. Thank you for your unwavering support.

ACKNOWLEDGMENTS

"Education is the most powerful weapon which you can use to change the world."
-Nelson Mandela, former president of South Africa, 1993 Nobel Peace Prize laureate.

I would like to take this opportunity to thank my advisor, Susan Murray, for all her aid and guidance in my years at graduate school. I have thoroughly enjoyed working with you these last five years. I would also like to thank all our collaborators in the pulmonary group at the University of Michigan Hospital: Kevin Flaherty, Meilan Han, Fernando Martinez and Shelley Schmidt.

Graduate school has been a series of challenges and successes. I would like to thank my family and friends for helping me through the difficult times and for celebrating my accomplishments. My parents, Fayroza and Fazul, have been so steadfast in their support. Their joy and pride has made the hard work all worthwhile. My two older brothers, Rashaad and Amin, have been my inspiration since childhood. I am still not sure that I will ever know as much as they do. Thank you also to the rest of my family, Fathima, Sulaimaan, Zakariya and Sarah. I started graduate school at the University of Michigan the year Sulaimaan, my eldest nephew, was born. I am just glad that I finished school before he begins next year!

Thank you also to Laura Fernandes, Naseem Jauhar, Miku Kawakami, Quang Duong and Azarias Reda. We met soon after we all arrived in Ann Arbor and I am so proud and thankful to call you my friends. I would also like to thank my friend

of many, many years Tasneem Bayat. The staff and students in the Biostatistics Department contain so many friends that I am unable to name them all but in particular I would like to thank Ludi Fan, Matthew Flickinger, Matt Zawistowski, Fatma-Zohra Nedjari and Nicole Fenech.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Restricted mean analysis across multiple follow-up intervals

by

Nabihah Tayob

Chair: Susan Murray

The restricted mean survival, first proposed by Irwin (1949), is the expected survival time within a fixed follow-up window. This measure has a meaningful interpretation for both physicians and patients in a clinical setting that motivates its further exploration.

The first paper provides a nonparametric estimate of τ-restricted mean survival that uses additional follow-up information beyond τ, when appropriate, to improve precision. The variance of our estimate must account for correlation between incorporated follow-up windows and we follow an approach by Woodruff (1971) that linearizes random components of the estimate to simplify calculations. Both asymptotic closed form calculations and simulation studies recommend selection of follow-up intervals spaced approximately $\tau/2$ apart.

In the second paper we develop two recurrent events testing procedures. We take advantage of the properties of time-to-first event analyses and use events beyond the

first by combining data across multiple follow-up windows in two different ways. The first pools the data before estimating the τ-restricted mean survival and the second uses the area under the τ-restricted mean residual life function. We consider multiple scenarios of treatment effect in simulation studies and find our testing procedures perform favorably, especially when events are correlated, compared to the robust proportional rates model proposed by Lin et al. (2000) and the nonparametric Ghosh & Lin (2000) test.

A component of the lung allocation score, used to order patients for transplant offers, is the 1-year restricted mean survival on waitlist. In the third paper we develop a restricted mean survival model that combines data from multiple 1-year follow-up windows spaced six months apart to incorporate time-dependent patient risk data, extending work by Xiang et al. (2013) to multiple follow-up intervals. Model parameters are estimated by multiply imputing censored time-to-event data using an inverse transform method; the complete dataset is analyzed using standard methods. The systematic removal of patients from the lung transplant waitlist based on their daily updated LAS results in dependent censoring, which we account for using inverse probability of censoring weights when estimating survival functions. Simulation studies show that our proposed method performs well and incorporating additional follow-up improves efficiency.

CHAPTER I

INTRODUCTION

The restricted mean survival time is the expected survival time within a fixed follow-up interval. It was first proposed by Irwin (1949) since the mean survival time is not estimable in the presence of censoring, which is almost always the case in time to event studies. This measure has a meaningful interpretation for both physicians and patients in a clinical setting. In addition, the restricted mean survival captures information about the immediate future and health economists (Gyrd-Hansen & Sogaard, 1998) have found that patients consider life-years closer to the present more valuable than those in the future. The meaningful and relevant interpretation of the restricted mean survival motivates its further exploration. Statistical methods for estimation, hypothesis testing for treatment effect in a randomized clinical trial and regression models based on the restricted mean survival have been widely studied but research on incorporating follow-up information beyond the first follow-up window has been limited.

The first paper provides a nonparametric estimate of τ-restricted mean survival that uses additional follow-up information beyond τ, when appropriate, to improve precision. The τ-restricted mean residual life function and its associated confidence

bands are a tool to assess the stability of disease prognosis and the validity of combining follow-up intervals for this purpose. The variance of our estimate, the overall τ-restricted mean survival, must account for correlation between incorporated follow-up windows and we follow an approach by Woodruff (1971) that linearizes random components of the estimate to simplify calculations. Both asymptotic closed form calculations and simulation studies recommend selection of follow-up intervals spaced approximately $\tau/2$ apart. In simulations, the variance we propose performs better than the standard sandwich variance estimate. Our analysis approach is illustrated in two settings summarizing prognosis of idiopathic pulmonary fibrosis patients and aspirin treated diabetic retinopathy patients who had deferred photocoagulation.

In the second paper we focus on developing recurrent events testing procedures for two independent samples. Use of a combined endpoint that includes disease progression (recurrent event) as well as mortality (terminating event) improves power for detecting treatment effects. Standard recurrent events analyses require assumptions about the dependence between events and time-to-first event analyses do not use the information beyond the first event. In our approach, we take advantage of the properties of time-to-first event analyses and use events beyond the first by combining data across multiple follow-up windows. The two testing procedures that we develop for evaluating treatment effect combine the multiple follow-up windows in different ways. The first uses the overall τ-restricted mean survival and the second uses the area under the τ-restricted mean residual life function. A simulation study compares our test to the robust proportional rates model proposed by Lin et al. (2000) and the nonparametric Ghosh & Lin (2000) test for recurrent events subject to death. We consider multiple scenarios of treatment effect and find our testing procedures perform favorably, especially when events are correlated. The analysis approach is illustrated

for a randomized trial testing the ability of azithromycin to reduce exacerbations in patients with chronic obstructive pulmonary disease (COPD).

The motivation for the third paper comes from the lung transplant setting. The lung allocation score (LAS) involves the difference between benefit (days of life gained during the next year if a transplant is offered immediately) and urgency (1-year restricted mean while on waitlist), and is used to order patients for organ offers. To date, risk data at listing has been used as one of the primary drivers of model development for the urgency model, and while time-dependent patient risk data has been collected on the waitlist, statistical methods of incorporating this information has been limited. We develop a restricted mean survival model that combines data from multiple 1-year follow-up windows spaced six months apart. This is an extension of the method proposed by Xiang et al. (2013) to multiple follow-up intervals. The model parameters are estimated by multiply imputing censored time-to-event data using an inverse transform method to obtain complete dataset that can be analyzed using standard methods. The systematic removal of patients from the lung transplant waitlist based on their daily updated LAS results in dependent censoring completely captured by LAS. The estimate of the survival function in the multiple imputation procedure adjusts for the bias resulting from dependent censoring using inverse weights. The method will also allow us to describe changes in patient urgency over the waitlist candidate experience. The proposed method is compared to some existing methods for fitting restricted mean survival models in simulation studies. We found that our proposed method performs well and incorporating additional follow-up improves efficiency. A recent release of lung waitlist data was used to implement the proposed methodology and study the patient urgency model and the effect of incorporating multiple follow-up windows.

CHAPTER II

NONPARAMETRIC RESTRICTED MEAN ANALYSIS ACROSS MULTIPLE FOLLOW-UP INTERVALS

2.1 Summary

This research provides a nonparametric estimate of τ-restricted mean survival that uses additional follow-up information beyond τ, when appropriate, to improve precision. The τ-restricted mean residual life function and its associated confidence bands are a tool to assess the stability of disease prognosis and the validity of combining follow-up intervals for this purpose. The variance of our estimate must account for correlation between incorporated follow-up windows and we follow an approach by Woodruff (1971) that linearizes random components of the estimate to simplify calculations. Both asymptotic closed form calculations and simulation studies recommend selection of follow-up intervals spaced approximately $\tau/2$ apart. In simulations, the variance we propose performs better than the standard sandwich variance estimate. Our analysis approach is illustrated in two settings summarizing prognosis of idiopathic pulmonary fibrosis patients and aspirin treated diabetic retinopathy patients who had deferred photocoagulation.

2.2 Introduction

Yearly progression predictions are commonly reported for clinical longitudinal data. For example, Raghu et al. (2011) report that mild to moderate idiopathic pulmonary fibrosis (IPF) patients tend to lose 0.2 liters in forced vital capacity lung function per year. This is a valuable summary statistic, for physicians and patients, that has not been sufficiently explored for censored time to event data. For instance, it would be useful for a physician to be able to report to a patient that IPF patients followed for 10 years were observed to live 91% of each year, on average, given they were alive at the start of the year. This is a concise estimate for the patient about how their disease may affect them in the short term and gives a sense of the stability of their disease when appropriate. In cases like these where yearly progression is reasonably stable, a yearly progression estimate can be made more efficient by combining information from different follow-up periods. Pulmonary researchers and patients are particularly primed to interpret days of life lived in a year since a lung allocation score introduced by Egan et al. (2006) is based on expected days of life lived in the following year without transplant.

In the Early Treatment Diabetic Retinopathy Study (ETDRS), ETDRS Research Group (1991a,b), time to severe vision loss was the primary endpoint. In this case an ophthalmologist might want to report the expected days of good sight per year, given that a patient has not yet reached the severe vision loss endpoint, based on 4-5 years of observed data from this study.

The restricted mean residual life function (RMRL) is the expected days of life per year for those surviving at the beginning of the year and may be used to view trends in these summary statistics over time. Ghorai et al. (1982) proposed an estimator

based on integrated conditional Kaplan-Meier estimates Ghorai & Rejto (1987) and Na & Kim (1999) proposed a smooth-spline estimator for this quantity, among others. Yu (2003) developed confidence bands for restricted mean residual life functions estimated via Nelson-Aalen estimates, Cox model hazard estimates and inverse weighted hazard estimates, calling them expected life prosper functions (ELPF). Stability of these functions suggests an opportunity for producing an overall summary statistic that is more precise.

This paper provides a nonparametric estimate of the expected number of days lived per τ years based on $A > \tau$ years of follow-up, a single statistic describing the cost of the disease to a patient over a stable period of time. In the IPF and diabetic retinopathy settings described above, A is substantially greater than τ. We study several τ-length intervals from the follow-up period to see if follow-up intervals obtained after time zero essentially estimate the same restricted mean and if incorporating these extra intervals provides efficiency gains. In cases where follow-up intervals give non-stable trends for the restricted mean, the overall trend is summarized.

In Section 2.3 we review RMRL estimation, confidence band construction and a simple RMRL smoother. Section 2.4 describes the nonparametric τ-restricted mean survival estimator that combines information across different τ-length intervals of follow-up. Although the data appears clustered by individual, it turns out that the sandwich variance estimation for clustered data, available in many software packages, does not perform well when the cluster is based on many overlapping follow-up intervals. The proposed variance described in Section 2.4 is based on a linearization of random components of the estimator, similar to the approach recommended by Woodruff (1971) and more recently Williams (1995). In Section 2.5 we consider how to choose the number of follow-up intervals useful for obtaining efficiency gains.

A simulation study that assesses the performance of the proposed estimate and its variance against currently available competitors is presented in Section 2.6. Two examples of the proposed analysis approach, pertaining to IPF patients and diabetic retinopathy patients, are given in Section 2.7. A discussion follows in Section 2.8.

2.3 τ-restricted mean residual life

We define notation in Section 2.3.1. Section 2.3.2 describes a nonparametric estimate of RMRL, and Section 2.3.3 gives estimated confidence bands of RMRL, following the style laid out in Yu (2003). Section 2.3.4 provides smoothed RMRL estimates across chosen follow-up intervals as an additional visual tool for assessing whether these follow-up intervals may be combined.

2.3.1 Notation

For each of n patients we define observed event time, $X_i = \min(T_i, C_i)$, with failure indicator $\delta_i = I(T_i \leq C_i)$, based on true failure time, T_i, and censoring time, C_i, $i = 1, \ldots, n$. Calendar time, t, is measured from the start of the study. We define the residual life observed at t as $X_i(t) = (X_i - t)I(X_i \geq t)$ with failure indicator variable $\delta_i(t) = \delta_i I(X_i \geq t)$.

For a τ-length interval starting at calendar time, t, the τ-restricted mean residual lifetime is $\mu(t, \tau) = E\{\min(T - t, \tau)|T > t\} = \int_0^\tau Pr(T > t + u|T > t)du$. Here, u denotes an internal time scale measured from calendar time t. This definition of RMRL, with t as a parameter, allows us to simultaneously discuss values of $\mu(t, \tau)$ measured at different calendar times $t \in \{t_1, ..., t_b\}$. We use the convention of $t_1 = 0$ in all that follows.

Counting process notation includes the two timescales described, t (calendar) and u (internal time measured from t). For individual i at calender time t, the event counting process applied to internal timescale u is $N_i(t,u) = I\{X_i(t) \leq u, \delta_i(t) = 1\}$ with at-risk process $Y_i(t,u) = I\{X_i(t) \geq u\}$. At t, the total number of events occurring no later than u is $N(t,u) = \sum_{i=1}^{n} N_i(t,u)$ and the number at risk at u is $Y(t,u) = \sum_{i=1}^{n} Y_i(t,u)$. We require notation combining counting process quantities across calendar times, $\{t_1, ..., t_b\}$. For individual i and internal time u, $N_i(u) = \sum_{k=1}^{b} N_i(t_k,u)$ and $Y_i(u) = \sum_{k=1}^{b} Y_i(t_k,u)$. Combining information across follow-up intervals and patients gives us $N(u) = \sum_{k=1}^{b} N(t_k,u)$, total number of events occurring no later than u, and $Y(u) = \sum_{k=1}^{b} Y(t_k,u)$, total number at risk at u. Many of the $\{X_i(t_k), \delta_i(t_k)\}$ pairs reflect follow-up times censored at time 0 due to attrition prior to time t_k.

2.3.2 Estimation of τ-restricted mean residual life

The τ-restricted mean residual lifetime function, $\mu(t,\tau)$, tracks subsequent expected lifetime during an interval of length τ given the patient has survived up to time t. Henceforth we call $\mu(t,\tau)$ the RMRL function, submerging τ for brevity. Using notation from the previous section, a consistent nonparametric estimator of the RMRL function is $\hat{\mu}(t,\tau) = \int_0^\tau \hat{P}(t,s)ds$ where $\hat{P}(t,s) = \exp\left\{-\int_0^s \frac{dN(t,u)}{Y(t,u)}\right\}$.

2.3.3 Confidence Bands

We slightly modify work from Yu (2003) for RMRL confidence band calculations applied to times $\{t_1, \ldots, t_b\}$. As opposed to 95% pointwise confidence intervals, designed to cover each separate $\mu(t,\tau)$ value 95% of the time, confidence bands are

designed to cover the entire set of values $\{\mu(t_1,\tau),\ldots,\mu(t_b,\tau)\}$ 95% of the time. Plotting the RMRL function with its corresponding bands is useful in suggesting whether follow-up windows may be combined for estimation or not.

Technical development of confidence bands for $\{\mu(t_1,\tau),\ldots,\mu(t_b,\tau)\}$ is based on the Gaussian process $B(t,\tau) = n^{1/2}\{\hat{\mu}(t,\tau) - \mu(t,\tau)\}$ and the covariance of this process at times t_{j_1} and t_{j_2}, $j_1 = 1,\ldots,b, j_2 = 1,\ldots,b$. The estimated covariance matrix $\hat{\sum}_{b\times b}$, with estimates of $\text{cov}\{B(t_{j_1},\tau), B(t_{j_2},\tau)\}$ as the (j_1,j_2) elements, is

$$n\int_0^\tau\int_0^\tau \hat{P}(t_{j_1},s)\hat{P}(t_{j_2},s')\sum_{i=1}^{n}\int_0^s\int_0^{s'}\frac{dN_i(t_{j_1},u)dN_i(t_{j_2},v)}{Y(t_{j_1},u)Y(t_{j_2},v)}ds\,ds'$$

Following Lin et al. (1994), the asymptotic distribution of $\{B(t_1,\tau),\ldots,B(t_b,\tau)\}$ is approximated by generating a large number of mean zero multivariate Normal samples using the observed covariance structure, $\hat{\sum}_{b\times b}$. Using these samples, we calculate q_α satisfying $Pr\{\max_{t\in\{t_1,\ldots,t_b\}}|\hat{B}(t,\tau)| > q_\alpha\} = \alpha$. Level $100(1-\alpha)\%$ confidence band values surrounding $\mu(t,\tau)$ become $\Big[\hat{\mu}(t,\tau) - n^{-1/2}\times q_\alpha, \hat{\mu}(t,\tau) + n^{-1/2}\times q_\alpha\Big]$, calculated at times $t = \{t_1,\ldots,t_b\}$. In practice, these confidence bands perform well provided that there are at least 25 event times following t_b.

2.3.4 Smoothed RMRL estimated at times $t = \{t_1,\ldots,t_b\}$

Smoothed RMRL values can be useful in visualizing trend from noise across follow-up intervals. For each follow-up interval, $\{t_k, t_k+\tau\}, k = 1,\ldots,b$, adjoining follow-up windows contribute information towards estimation of the smoothed RMRL estimate

at t_k, $\hat{\mu}^*(t_k, \tau)$. In particular,

$$\hat{\mu}^*(t_k,\tau) = \begin{cases} \int_0^\tau \exp\left\{ - \int_0^s \frac{\sum_{j=k}^{k+1} dN(t_j,u)}{\sum_{j=k}^{k+1} Y(t_j,u)} \right\} ds & k=1 \\ \int_0^\tau \exp\left\{ - \int_0^s \frac{\sum_{j=k-1}^{k+1} dN(t_j,u)}{\sum_{j=k-1}^{k+1} Y(t_j,u)} \right\} ds & k=2,\ldots,b-1 \\ \int_0^\tau \exp\left\{ - \int_0^s \frac{\sum_{j=k-1}^{k} dN(t_j,u)}{\sum_{j=k-1}^{k} Y(t_j,u)} \right\} ds & k=b \end{cases}$$

Each smoothed RMRL value at time t_k is a special case of the overall τ-restricted mean survival function developed in the next section. These smoothed RMRL values are superimposed on RMRL plots.

2.4 Overall τ-restricted mean survival

When the RMRL (discussed in Section 2.3.2) and its corresponding confidence bands (discussed in Section 2.3.3) do not indicate a strong trend, we develop a more efficient estimate of the expected number of days lived in the next τ years by pooling appropriate follow-up periods beginning at times $t \in \{t_1, ..., t_b\}$. In Section 2.4.1 we define our proposed estimate of the overall τ-restricted mean survival. The variance of this estimate is developed in Section 2.4.2.

2.4.1 Estimation

The proposed estimate of the overall τ-restricted mean survival is

$$\hat{\mu}^*(\tau) = \int_0^\tau \exp\left\{ - \int_0^s \frac{dN(u)}{Y(u)} \right\} ds. \tag{2.1}$$

Let $\lambda(t_k, u) = \lim_{\Delta u \to 0}[Pr\{u \le X_i(t_k) < u + \Delta u, \delta_i(t_k) = 1 | X_i(t_k) \ge u\}/\Delta u]$ and let $\lambda^W(u) = \sum_{k=1}^b \lambda(t_k, u) Pr\{X_i(t_k) \ge u\} / \sum_{l=1}^b Pr\{X_i(t_l) \ge u\}$. In Appendix

A.1, we show that (2.1) converges in probability to $\mu^*(\tau) = \int_0^\tau \exp\left[-\int_0^s \lambda^W(u)du\right] ds$, which is the mean of the mixture distribution created from combining follow-up times across the different intervals. If the overall dataset reflects a single distribution, that is, $\lambda(t_{k_1}, u) = \lambda(t_{k_2}, u)$ for $0 \leq u \leq \tau$ and $k_1, k_2 \in \{1, ..., b\}$, then $\mu^*(\tau)$ reduces to $\mu(t_1, \tau)$, the usual restricted mean that is typically estimated using a single follow-up period. Variance calculations in the following section acknowledge the potential mixture of hazards that might occur when combining follow-up times across intervals.

2.4.2 Variance of proposed estimate

The proposed variance estimate is calculated via linearizing components of $\sqrt{n}\hat{\mu}^*(\tau)$ via Taylor series approximations. This approach to obtaining variances is an attractive alternative to working through stochastic integrals of martingales for correlated counting process, where appropriate filtrations can be challenging to define. Suppose that in the dataset of combined follow-up times we observe M events at internal times $\{0 < \mathscr{T}_1 < \ldots < \mathscr{T}_M < \tau\}$ where events from the same individual during different follow-up windows are correlated; for convenience, we define $\mathscr{T}_0 = 0$ and $\mathscr{T}_{M+1} = \tau$. Let $F_j\{dN(\mathscr{T}_j), Y(\mathscr{T}_j)\} = dN(\mathscr{T}_j)/Y(\mathscr{T}_j)$. In the following we temporarily submerge arguments of F_j. We define $G_m(F_0, F_1, \ldots, F_m) = \exp(-\sum_{j=0}^m F_j)$. After rewriting $\sqrt{n}\hat{\mu}^*(\tau)$ in terms of $G_m(F_0, F_1, \ldots, F_m)$, the variance of $\sqrt{n}\hat{\mu}^*(\tau)$ becomes

$$Var\left\{\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)G_m(F_0, F_1, \ldots, F_m)\right\}.$$

The non-linear terms $G_0, G_1, \ldots, G_M$ may be made more tractable for variance calculations via linearization based on a Taylor series expansion of $G_m(F_0, F_1, \ldots, F_m)$ about $\lambda^W(\mathscr{T}_j)d\mathscr{T}_j, j = 0, \ldots, m$. The first order partial derivatives of $G_m(F_0, F_1, \ldots, F_m)$

with respect to each F_j are $-\exp\left\{-\sum_{j'=0}^{m} F_{j'}\right\}$, $j = 0, \ldots, m$. The variance of $\sqrt{n}\hat{\mu}^*(\tau)$ is then

$$
\begin{aligned}
& Var\left(\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1}-\mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\right. && \text{(2.2a)}\\
& +\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1}-\mathscr{T}_m)\left[\sum_{j=0}^{m}-\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}F_j\right] && \text{(2.2b)}\\
& -\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1}-\mathscr{T}_m)\left[\sum_{j=0}^{m}-\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\lambda^W(\mathscr{T}_j)d\mathscr{T}_j\right] && \text{(2.2c)}\\
& +\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1}-\mathscr{T}_m)\frac{1}{2!}\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\left[\sum_{j=0}^{m}\{F_j-\lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}\right]^2 && \text{(2.2d)}\\
& \left.+\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1}-\mathscr{T}_m)\,[\text{ higher order terms}]\right). && \text{(2.2e)}
\end{aligned}
$$

Terms (2.2a) and (2.2c) are nonrandom and therefore do not contribute to the variance. The fourth term (2.2d) converges to zero in probability, details given in Appendix A.2. Similarly all higher order terms of the Taylor series linearization (2.2e) converge to zero in probability and the variance reduces to $Var\left[\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1}-\mathscr{T}_m)\right.$ $\left.\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\sum_{j=0}^{m}F_j\right]$, the variance of a linear sum of non-linear components $F_0, F_1, \ldots, F_M$.

Each of the non-linear F_j terms may be made more tractable for variance calculations via further linearization based on a Taylor series expansion of $F_j\{dN(\mathscr{T}_j), Y(\mathscr{T}_j)\}$ about the expected values of $dN(\mathscr{T}_j)$ and $Y(\mathscr{T}_j)$. The expected value of $dN(\mathscr{T}_j)$ is $n\sum_{k=1}^{b}\lambda(t_k,\mathscr{T}_j)Pr\{X_i(t_k)\geq\mathscr{T}_j\}d\mathscr{T}_j$ and the expected value of $Y(\mathscr{T}_j)$ is $n\sum_{k=1}^{b}Pr\{$ $X_i(t_k)\geq\mathscr{T}_j\}$. The ratio of the expectations, $E\{dN(\mathscr{T}_j)\}/E\{Y(\mathscr{T}_j)\}$, simplifies to $\lambda^W(\mathscr{T}_j)$. The first-order partial derivative of F_j with respect to $dN(\mathscr{T}_j)$ is $1/Y(\mathscr{T}_j)$ and the first-order partial derivative of F_j with respect to $Y(\mathscr{T}_j)$ is $-dN(\mathscr{T}_j)/Y(\mathscr{T}_j)^2$.

The variance of $\sqrt{n}\hat{\mu}^*(\tau)$ is then

$$
\begin{aligned}
Var\Bigg[\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1}-\mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\times \\
\left\{\sum_{j=0}^{m}F_j[E\{dN(\mathscr{T}_j)\},E\{Y(\mathscr{T}_j)\}]\right. && \text{(2.3a)}\\
+\sum_{j=0}^{m}\left\{\frac{dN(\mathscr{T}_j)-E\{dN(\mathscr{T}_j)\}}{E\{Y(\mathscr{T}_j)\}}\right\} && \text{(2.3b)}\\
-\sum_{j=0}^{m}\left\{\frac{E\{dN(\mathscr{T}_j)\}\left[Y(\mathscr{T}_j)-E\{Y(\mathscr{T}_j)\}\right]}{E\{Y(\mathscr{T}_j)\}^2}\right\} && \text{(2.3c)}\\
+\sum_{j=0}^{m}\frac{1}{2!}\left\{0-2\frac{[Y(\mathscr{T}_j)-E\{Y(\mathscr{T}_j)\}][dN(\mathscr{T}_j)-E\{dN(\mathscr{T}_j)\}]}{E\{Y(\mathscr{T}_j)\}^2}\right. \\
\left.+\frac{2E\{dN(\mathscr{T}_j)\}}{E\{Y(\mathscr{T}_j)\}^3}[Y(\mathscr{T}_j)-E\{Y(\mathscr{T}_j)\}]^2\right\} && \text{(2.3d)}\\
\left.+\sum_{j=0}^{m}\{\text{ higher order terms}\}\right]. && \text{(2.3e)}
\end{aligned}
$$

Term (2.3a) is a constant and therefore does not contribute to the variance. Terms (2.3b) and (2.3c) simplify to $\sum_{j=0}^{m}\left[dN(\mathscr{T}_j)-Y(\mathscr{T}_j)E\{dN(\mathscr{T}_j)\}/E\{Y(\mathscr{T}_j)\}\right]/E\{Y(\mathscr{T}_j)\}$. The fourth term (2.3d) converges in probability to zero, details given in Appendix A.2. Similarly, all higher order terms of the Taylor series linearization (2.3e) converge to zero in probability. Recall that $dN(\mathscr{T}_j)=\sum_{i=1}^{n}\sum_{k=1}^{b}dN_i(t_k,\mathscr{T}_j)$ and $Y(\mathscr{T}_j)=\sum_{i=1}^{n}\sum_{k=1}^{b}Y_i(t_k,\mathscr{T}_j)$. The expected value of $dN(\mathscr{T}_j)$ is $n\sum_{k=1}^{b}\lambda(t_k,\mathscr{T}_j)Pr\{X_i(t_k)\geq\mathscr{T}_j\}d\mathscr{T}_j$ and the expected value of $Y(\mathscr{T}_j)$ is $n\sum_{k=1}^{b}Pr\{X_i(t_k)\geq\mathscr{T}_j\}$. The variance then reduces to $Var\left[\sqrt{n}\sum_{i=1}^{n}Z_i\{\hat{\mu}^*(\tau)\}/n\right]$ where $Z_i\{\hat{\mu}^*(\tau)\}=\sum_{k=1}^{b}Z_{ik}\{\hat{\mu}^*(\tau)\}$ and

$$
\begin{aligned}
Z_{ik}\{\hat{\mu}^*(\tau)\}=\sum_{m=0}^{M}(\mathscr{T}_{m+1}-\mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\times \\
\sum_{j=0}^{m}\frac{dN_i(t_k,\mathscr{T}_j)-\lambda^W(\mathscr{T}_j)Y_i(t_k,\mathscr{T}_j)d\mathscr{T}_j}{\sum_{l=1}^{b}Pr\{X_i(t_l)\geq\mathscr{T}_j\}}.
\end{aligned}
$$

Whereas event times across overlapping follow-up periods are generally not i.i.d., the $Z_i\{\hat{\mu}^*(\tau)\}$ terms are i.i.d., making empirical variance estimates based on $Z_i\{\hat{\mu}^*(\tau)\}$ appropriate for this setting. In practice, the empirical variance estimator of $\sqrt{n}\hat{\mu}^*(\tau)$ is $\sum_{i=1}^n \left[Z_i\{\hat{\mu}^*(\tau)\} - \bar{Z}\{\hat{\mu}^*(\tau)\}\right]^2/(n-1)$ where $\bar{Z}\{\hat{\mu}^*(\tau)\} = \sum_{i=1}^n Z_i\{\hat{\mu}^*(\tau)\}/n$. The sample estimates of $Z_{ik}\{\hat{\mu}^*(\tau)\}$ are given by

$$z_{ik}\{\hat{\mu}^*(\tau)\} = \sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j=0}^{m}\frac{dN(\mathscr{T}_j)}{Y(\mathscr{T}_j)}\right\}\sum_{j=0}^{m}\frac{dN_i(t_k,\mathscr{T}_j) - \frac{dN(\mathscr{T}_j)}{Y(\mathscr{T}_j)}Y_i(t_k,\mathscr{T}_j)}{Y(\mathscr{T}_j)/n}$$

In understanding design issues discussed in Section 2.5, it is convenient to have the asymptotic closed form variance, $\sigma^2 = Var\left[\sqrt{n}\sum_{i=1}^n\sum_{k=1}^b Z_{ik}\{\hat{\mu}^*(\tau)\}/n\right]$, where

$$Z_{ik}\{\hat{\mu}^*(\tau)\} = \int_0^\tau \exp\left\{-\int_0^s \lambda^W(u)du\right\}\left[\int_0^s \frac{dN_i(t_k,u) - \lambda^W(u)Y_i(t_k,u)du}{\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}}\right]ds$$

in asymptotically equivalent stochastic integral notation. In Appendix A.3 we show that

$$\begin{aligned}
\sigma^2 = &\sum_{k=1}^b\sum_{l=1}^b\int_0^\tau\int_0^\tau \exp\left\{-\int_0^s \lambda^W(u)du\right\}\exp\left\{-\int_0^{s'}\lambda^W(v)dv\right\}\\
&\int_0^s\int_0^{s'}\frac{1}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq u\}\right]\left[\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq v\}\right]}\\
&\Big[\lambda(t_k,u)Pr\{X_i(t_k)\geq u\}I(v = u + t_k - t_l)du\\
&-\lambda^W(u)\lambda(t_l,v)Pr\{X_i(t_l)\geq v\}\{I(u\leq v + t_l - t_k) + I(u=0)I(v < t_k - t_l)\}du\,dv\\
&-\lambda^W(v)\lambda(t_k,u)Pr\{X_i(t_k)\geq u\}\{I(v\leq u + t_k - t_l) + I(v=0)I(u < t_l - t_k)\}du\,dv\\
&+\lambda^W(u)\lambda^W(v)Pr\{X_i(t_k)\geq u, X_i(t_l)\geq v\}du\,dv\\
&-\{\lambda(t_k,u) - \lambda^W(u)\}\{\lambda(t_l,v) - \lambda^W(v)\}Pr\{X_i(t_k)\geq u\}Pr\{X_i(t_l)\geq v\}du\,dv\Big]ds\,ds'.
\end{aligned}$$

2.5 Practical Issues

Number and spacing of follow-up windows should be chosen to increase precision of $\hat{\mu}^*(\tau)$. For this purpose, we examine the estimator's closed form variance σ^2 in the special case where the failure time, T_i, follows an exponential distribution with hazard λ. The censoring time, C_i, is independently sampled from a Uniform$[A - A^*, A]$ distribution, where A is the length of the study with accrual time A^*. In this case, $\lambda(t_k, u) = \lambda^W(u) = \lambda$. Standard probability calculations for $u, v \in (0, \tau]$ give $Pr\{X_i(t_k) \geq u\} = \exp\{-\lambda(u + t_k)\}\{A - \max(A - A^*, u + t_k)\}/A^*$ and $Pr\{X_i(t_k) \geq u, X_i(t_l) \geq v\} = \exp\{-\lambda \max(u + t_k, v + t_l)\}\ \{A - \max(A - A^*, u + t_k, v + t_l)\}/A^*$. Details of these calculations are given in Appendix A.4.

We consider $\tau = 1$ year. The parameter λ was chosen to give a constant 1-year RMRL of 11 months, $A^* = 1$ year, $A = 3$ years and $n = 100$. Figure 2.1 shows the behavior of σ^2/n, the finite sample size variance, for three one-year follow-up windows with $t_1 = 0$, $t_3 = 12$ months and t_2 varying between these values. The one-year windows starting at t_1 and t_3 do not overlap, so the choice of t_2 examines if there is an advantage to adding a 3^{rd} follow-up window that overlaps the other two. The plot suggests that an additional 1-year window starting in the middle of $t_1 = 0$ and $t_3 = 12$ months, i.e. at $t_2 = 6$ months, reduces the variance the most.

Next consider (1) whether additional equally spaced follow-up windows further reduce σ^2/n and (2) the extent to which incorporating an additional year's worth of follow-up information, and corresponding 1-year windows, into construction of $\hat{\mu}^*(\tau)$ reduces its variability. The first entry of Table 2.1 corresponds to the variance obtained if estimating the standard 1-year restricted mean that doesn't use information from additional follow-up intervals. For a given t_b of 12 or 24 months, increasing

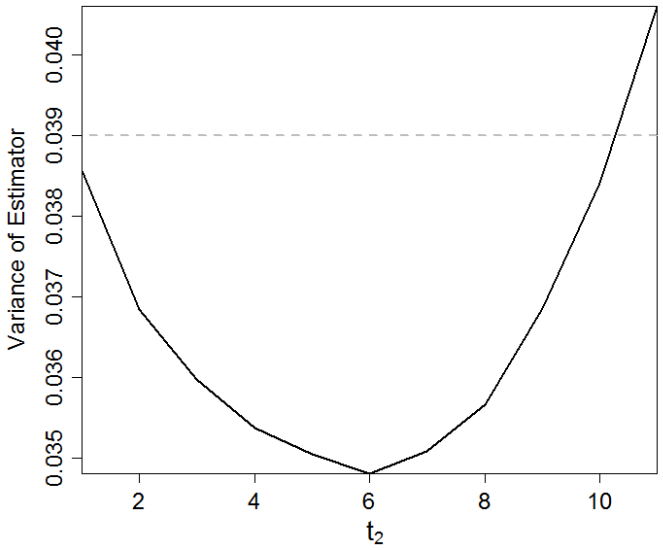


Figure 2.1: Closed form asymptotic variance of $\hat{\mu}^*(\tau)$ for 3 year study with τ=1 year, $t_1 = 0$, $t_3 = 12$ months and varying t_2. Dashed line corresponds to variance of estimator constructed using two follow-up windows $t_1 = 0$ and $t_3 = 12$.

the number of follow-up windows improves the precision of the estimator. However, there are diminishing returns from introducing starting times more frequently than at 6-month intervals.

Table 2.1: Study of follow-up window choices based on calculated variance and Asymptotic Relative Efficiency (ARE) for the special case discussed in Section 4.

	Number of Windows	$\{t_1, \ldots, t_b\}$	σ^2/n	ARE
$t_b = 0$	1	0	0.071	1.00
$t_b = 12$	2	0, 12	0.039	1.82
	3	0, 6, 12	0.035	2.03
	5	0, 3, 6, 9, 12	0.035	2.03
$t_b = 24$	3	0, 12, 24	0.030	2.37
	5	0, 6, 12, 18, 24	0.026	2.73
	9	0, 3, 6, 9, 12, 15, 18, 21, 24	0.025	2.84

It seems clear from exploration of this special case that spacing of $t_1, \ldots, t_b$ should be at 6-month intervals when estimating an overall 1 year restricted mean, with

$t_b = 24$. In general, the recommended number of intervals is based on available follow-up time and we propose using intervals starting from $t_k = (k-1)\frac{\tau}{2}$ for $k = 1, \ldots, b$, with t_b less than $A - \tau$.

2.6 Simulation Study

Simulation experiments were conducted to assess finite sample size performance of $\hat{\mu}^*(\tau)$. We consider whether augmenting the first observation window improved the estimator and the effect of the number of intervals on the performance of the estimator. The performance of our proposed variance estimate is compared to a variance estimate that assumes independence and the sandwich variance (formulae to calculate these are given in Appendix A.5).

The simulation experiment design assumes we wish to estimate 12-month restricted mean survival in a 36-month study. We assume that 30% of the sample ($n = 100$) were recruited at the start of the study and were observed for the full 36 months. The remaining 70% were uniformly accrued over the first 12 months. T_i was simulated from a piecewise Weibull distribution with parameters chosen so that the 12-month restricted mean survival was 11 months at the recommended follow-up times $t_k \in \{0, 6, 12, 18, 24\}, k = 1, \ldots, 5$. The hazard function of the piecewise Weibull distribution is given by $\lambda(x) = \alpha_i \lambda_i x^{\alpha_i - 1}$, where the parameters are constant within a 6 month interval. The parameters are defined as $\alpha = (1.1, 0.9, 1.1, 0.9, 1.1, 0.9)$ and $\lambda = (1.25, 2.01, 1.05, 2.00, 1.26, 0.10) \times 10^{-2}$.

Simulation results in Table 2.2 show that augmenting the data with additional follow-up time improves the precision of the estimator, with ARE ranging from 1.87 to 2.70. The bias is minimal in all cases. As expected based on Section 2.5, the

Table 2.2: Study of follow-up windows and performance of variance estimators in 500 Monte Carlo simulations ($n = 100$) from a piecewise Weibull distribution

$\{t_1, ..., t_b\}$	Emp Mean	Emp Var	Independent Variance		Sandwich Variance		Proposed Variance	
			Emp Mean	Cov 95% CI	Emp Mean	Cov 95% CI	Emp Mean	Cov 95% CI
0	10.986	0.073	0.070	0.938	0.069	0.938	0.070	0.938
0, 12	10.993	0.039	0.039	0.938	0.039	0.934	0.039	0.940
0, 12, 24	10.975	0.030	0.031	0.954	0.031	0.952	0.031	0.952
0, 6, 12, 18, 24	10.980	0.027	0.018	0.878	0.021	0.902	0.028	0.948

The following abbreviations are used: Empirical Mean(Emp Mean);
Empirical Variance (Emp Var);
Coverage of 95% confidence interval (Cov 95% CI)

case with $\{t_1, \ldots, t_5\} = \{0, 6, 12, 18, 24\}$ outperforms other scenarios for, $b = 1, 2, 3$. All three variance methods perform well when $\{t_1, \ldots, t_b\}$ produced disjoint intervals (rows $1-3$ of Table 2.2). With overlapping intervals, the proposed variance gives the best coverage; independent and sandwich variances both underestimate the simulation empirical variance.

2.7 Examples

In a study by Schmidt et al. (2012), which aimed to provide better prognostic information to idiopathic pulmonary fibrosis (IPF) patients, 734 patients were identified through interstitial lung disease databases from three referral centers, the Royal Brompton and Harefield National Health Service Foundation Trust, National Jewish Health and the University of Michigan Health System, from 1981 through 2008. There is currently no effective treatment for IPF, with patients experiencing a steady average decline in lung function per year. A natural question is whether the expected number of days of life in the next year is also stable.

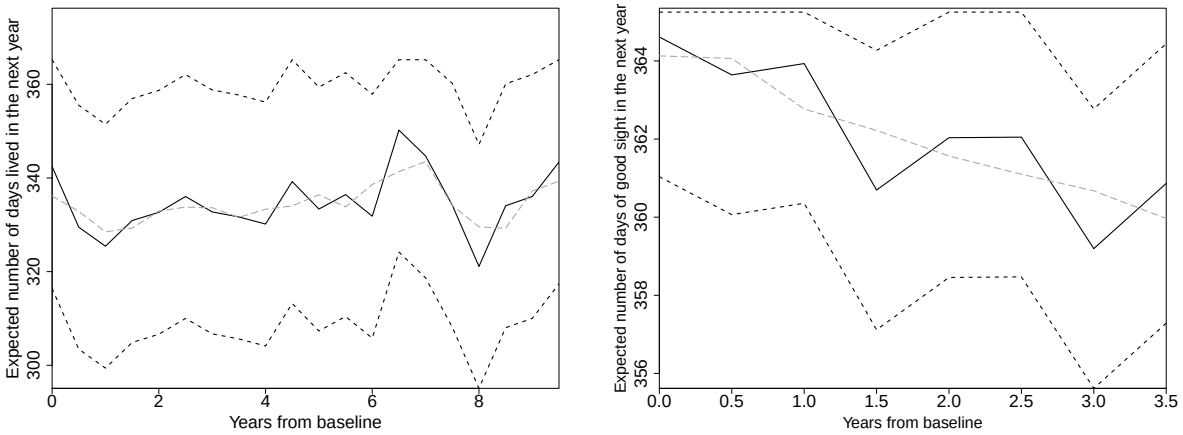
For each patient, calendar time begins at first pulmonary function test at their

referral center within the study period. As recommended, one-year intervals with start times every six months are used. The final interval at 9.5 years, chosen to ensure at least 25 risk set deaths remaining, used follow-up through year 10.5. The smoothed RMRL in Figure 2.2(a) fluctuates between 329 and 343 days with more stability in the first 6 years, where there is more available data. The confidence bands suggest that it is still reasonable to report an overall point estimate that the average number of days of life lived in the next year is 333.6 (95% CI: 330.7-336.4). During the first decade of their disease, IPF patients are expected to live 91% of each year given they were alive at the start of the year.

The Early Treatment Diabetic Retinopathy Study (ETDRS) (ETDRS Research Group, 1991a,b) enrolled patients with severe diabetic retinopathy in both eyes who were taking Aspirin daily. In addition one eye of each patient was randomly assigned to early photocoagulation and the other to deferral of photocoagulation until a later time when high-risk proliferative retinopathy was detected. We focus on the 583 patients who were randomized to the deferred photocoagulation treatment group. The major endpoint of interest was time to severe vision loss, defined as visual acuity less than 5/200 at two consecutive visits.

One-year intervals with start times every six months are used. The final interval at 3.5 years, chosen to ensure at least 25 risk set deaths remaining, used follow-up through year 4.5. The smoothed RMRL in Figure 2.2(b) suggests a slightly declining trend, reflecting somewhat quicker eyesight deterioration over time despite initiation of therapy once patients became especially high risk. Given the narrow width of the confidence bands ($<$ 10 days), one may argue that it is reasonable to report an overall point estimate of average days of sight per year, 362.2 days (95% CI: 361.3-363.1), which falls within the reported confidence bands through the 3.5 year period

considered. In this case, it may also be instructive to include the graphic so that future potential trend may be monitored.



(a) IPF data

(b) ETDRS data

Figure 2.2: One-year restricted mean residual life function (solid) evaluated every six months. Associated CB are given (dashed) as well as the smoothed restricted mean residual life function (long dash).

2.8 Discussion

We have developed a summary statistic of the expected number of days lived in the next τ years that uses additional follow-up information when we observe a stable disease progression. The RMRL plot at each time $t_1, \ldots t_b$ tracks the expected lifetime in the next τ years given the patient has already lived up to time t_k, $k = 1, \ldots, b$. The RMRL plots and associated confidence bands provide a summary of the trend in disease progression. They are a useful diagnostic tool to assist in deciding whether disease progression is stable, i.e. the RMRL is the same at each $t_1, \ldots t_b$, or not. In cases where we observe stable disease progression the additional incorporated follow-up windows give a more precise estimate of the τ-restricted mean survival. An obvious example of stable disease progression is exponential failure times. However, more generally, the RMRL is stable across b windows of follow-up when $\lambda(t_{k_1}, u) =$

$\lambda(t_{k_2}, u)$ for $0 \leq u \leq \tau$ and $k_1, k_2 \in \{1, ..., b\}$. We showed consistency of our overall τ-restricted mean survival in this case. We also suspect that in cases where the integrated survival curves within the follow-up windows are equivalent, our estimate is consistent. A rigorous proof backing up this intuition has eluded us, although special cases of distributions with this property have given $\mu^*(\tau) = \mu(t_1, \tau)$ and simulations also appear to perform well in these settings.

The empirical variance estimate we describe is straightforward to program and performed quite well in finite sample simulations, improving upon sandwich estimation variance results that seem to break down when follow-up intervals overlap. Although our calculations for determining the optimal number of follow-up windows was based on a 1-year follow-up period, these results generalize to any linear transformation of this time-scale and so our results extend to more general cases encountered in clinical trials or observational studies.

Our method pools the data and estimates the τ-restricted mean survival in a combined dataset. We initially considered an alternative method of using additional follow-up information that estimates the τ-restricted mean survival as a weighted average of RMRL estimates across the follow-up intervals. This approach was explored in simulations, for the same Weibull setting we used in Section 2.6, but was found to be less efficient than our recommended method and was not pursued further. Something similar was documented in Murray & Tsiatis (1996, 2001), where weighted averages of integrated survival curves across strata could actually result in reduced efficiency when strata were not prognostically different from one another. That would be the case in our setting where stability of progression across follow-up windows is required before combining data into an overall dataset.

CHAPTER III

NONPARAMETRIC TESTS OF TREATMENT EFFECT FOR A RECURRENT EVENT PROCESS THAT TERMINATES

3.1 Summary

Recurrent and terminal events are common outcomes for studying treatment effects in clinical studies. Existing approaches follow either a time-to-first event analysis approach or a recurrent event modeling approach. Recurrent event analyses are often restricted by independence assumptions on gap-times between events. Although time-to-first event analyses are not subject to this restriction, they discard information that occurs beyond the initial event and are much less powerful for detecting treatment differences. We develop two new approaches for determining treatment effects, motivated by less restrictive assumptions of time-to-first event analyses, that combine information from multiple follow-up intervals. The first testing procedure pools (correlated) short term τ-restricted outcomes from pre-specified intervals starting at times $t_k, k = 1, \ldots, b$, and compares estimated τ-restricted mean survival across treatment groups from this combined dataset. The second procedure calculates conditional τ-restricted means from those at risk at times $t_k, k = 1, \ldots, b$ and compares the area

under a function of these by treatment. Variances calculations, taking into account correlation of short-term outcomes within individuals, linearize random components of the test statistics following Woodruff (1971) and more recently Williams (1995). Simulations compare the finite sample performance of our tests to the robust proportional rates model proposed by Lin et al. (2000) and the Ghosh & Lin (2000) test for recurrent events subject to death. In treatment effect patterns following proportional incidence rates, delayed treatment effect and short duration treatment effect, the proposed methods perform favorably when compared to existing methods. These new analysis approaches also produce correct type I error rates when gap-times between events are correlated. The analysis approach is illustrated in data from a randomized trial of azithromycin in patients with chronic obstructive pulmonary disease (COPD).

3.2 Introduction

Clinical studies, where the outcome of interest is time-to-event, often use a combined endpoint that includes disease progression as well as mortality to improve the power of the study. For example, in idiopathic pulmonary fibrosis (IPF) studies the combined endpoint often used is time to death, lung transplant, acute exacerbation, 10% decline in forced vital capacity (liters) or 15% decline in diffusing capacity of the lung for carbon monoxide (ml/min/mmHg), whichever occurs first. This endpoint is a combination of recurrent and terminating events, where multiple recurrent events are observed in some patients. An analysis to test treatment effect based on time-to-first event will ignore the information contained in later recurrent events and in terminating events observed after a recurrent event. Alternatively, typical recurrent event analyses on gap-times require independent gap-times to avoid bias from de-

pendent censoring, which is sometimes a deterrent from use of these analyses in the clinical trial setting. The aim of our testing procedure is to extract information from both recurrent and terminal events regarding treatment effect while approaching the data from a perspective closer to time-to-first event analysis that avoids independence assumptions and dependent censoring issues.

One advantage of time-to-first event analyses is the immediate applicability to patients about what will happen to them next. For instance, the combined endpoint for IPF studies mentioned above is relevant to the subsequent year of life for the patient. Health economists have noted that patients value life-years closer to the present more than those in the future, Gyrd-Hansen & Sogaard (1998). This motivates a testing procedure that captures short term outcomes and how they evolve over time.

One potential short term summary statistic is τ-restricted mean survival; for the IPF study a 1-year restricted mean would be suitable. However, a standard 1-year restricted mean survival estimate ignores data collected after year one that could contain information on treatment effect. If additional 1-year follow-up windows are available beyond the first year, then they would add information on short term outcomes at different stages of the trial.

We embrace the philosophy that understanding short-term windows of treatment effect can provide an alternative and potentially superior understanding of treatment effect throughout the trial. We propose combining times-to-first-event across multiple follow-up windows of length τ beginning at evenly spaced times $t \in \{t_1, \ldots, t_b\}$. The choice of the spacing between $\{t_1, \ldots, t_b\}$ is based on reducing the variance of the restricted mean. Each starting point t_k of a follow-up interval is chosen as part of a study design, not influenced by observed data, and so avoids complications of dependent censoring that a gap-time analysis would pose for this data structure.

We develop two methods of combining information from follow-up windows in two-sample testing procedures. The first statistic pools data across each of the τ-length follow-up windows and estimates the τ-restricted mean survival in this combined dataset. The variance of this estimate is based on a linearization of random components of the estimator, similar to the approach recommended by Woodruff (1971) and more recently Williams (1995). An alternative statistic combines information across follow-up intervals by first creating a function of conditional τ-restricted means among those at risk at $t_k, k = 1, \ldots, b$ and then calculating the area under this function as a summary statistic for comparison between treatment groups.

There are a number of advantages of our analysis approach for this type of clinical trial. First, we include both recurrent and terminating events by defining time-to-event as a combined endpoint, so treatment differences should emerge with respect to either type of process. Second, we incorporate data collected beyond τ and account for resulting correlation between multiple follow-up windows for each patient. Variance calculations do not require any assumptions about the correlation structure between events observed for each patient. Third, by fixing the follow-up window start times we do not create the problem of dependent censoring commonly encountered when modeling gap-times between recurrent events.

Section 3.3 describes the notation used in the testing procedure. In Section 3.4 we propose two test statistics for comparing treatment effect in a randomized trial. A simulation study is used to compare the proposed two-sample tests to the robust proportional rates model proposed by Lin et al. (2000) and the Ghosh & Lin (2000) test for recurrent events subject to death in finite sample size settings in Section 3.5. The analysis approach is illustrated in Section 3.6 using data from a randomized trial of azithromycin in patients with chronic obstructive pulmonary disease (COPD). A

discussion follows in Section 3.7.

3.3 Notation

Consider the following data structure: D_i is the death time and C_i is the independent censoring time for patient $i = 1, \ldots, n$. For patient i we define $T_{i1} < T_{i2} < \ldots < T_{iJ_i}$ as recurrent event times terminating with death, i.e. $T_{iJ_i} = D_i$. Although we assume C_i is independent of $T_k, k = 1, ..., J_i$, the recurrent event times and the death time may be correlated. The process may not be fully observed due to censoring and we define the observed data for each patient to be $X_{ij} = \min(T_{ij}, C_i)$ with failure indicator variable $\delta_{ij} = I(T_{ij} \leq C_i)$ for $j = 1, 2, \ldots, \tilde{J}_i$, where $\tilde{J}_i \leq J_i$.

Since our analysis approach is to combine time-to-first event outcomes across multiple pre-specified intervals beginning at times $t \in \{t_1, \ldots, t_b\}$, with $t_1 = 0$ in all that follows, we define for patients at risk at time t

$$\eta_i(t) = \min\{j = 1, \ldots, \tilde{J}_i : X_{ij} \geq t\}$$

$$X_i(t) = X_{i\eta_i(t)} - t$$

$$\delta_i(t) = \delta_{i\eta_i(t)}$$

where $X_i(t)$ is the observed time to the next event from t, $\eta_i(t)$ is the corresponding index of the observed event time and $\delta_i(t)$ is the associated failure indicator variable. Otherwise, if a patient is not at risk at time t, we use the convention that $X_i(t) = \delta_i(t) = 0$.

For a follow-up interval starting at t, the time to the next event is censored at time $C_i - t$, which is independent of $X_i(t)$. This would not be the case if we were

considering gap times between recurrent events, except in the special case where gap times are independent of one another. Gap times are traditionally defined as $S_{i1} = T_{i1}$ and $S_{ij} = T_{ij} - T_{i(j-1)}$ for $j = 2, 3, \ldots, J_i$. When gap times S_{ij} are correlated, their corresponding censoring times $C_i - T_{i(j-1)}$ become dependent on earlier gap times as well. Focusing on times to the next event from prespecified times $t_1, \ldots, t_b$ avoids dependent censoring issues faced by gap time analysis approaches, while still taking advantage of additional recurrent event information.

For defining counting process notation, we have two time scales indexed by t and u. The t notation defines the beginning of each pre-specified follow-up interval as measured from the patient's randomization time. The u notation follows continuous time within any particular follow-up window, indexing the statistical processes from 0 to τ. We define the event counting process for each time to first event analysis starting at t as $N_i(t,u) = I\{X_i(t) \leq u, \delta_i(t) = 1\}$ and the at-risk process as $Y_i(t,u) = I\{X_i(t) \geq u\}$. For the interval starting at t, $N(t,u) = \sum_{i=1}^{n} N_i(t,u)$ counts the number of first events from the start of the interval to u and $Y(t,u) = \sum_{i=1}^{n} Y_i(t,u)$ gives the number at-risk at time u for a first event in the interval. The hazard rate within each follow-up interval is defined as $\lambda(t,u) = \lim_{\Delta u \to 0}[Pr\{u \leq X_i(t) < u + \Delta u, \delta_i(t) = 1 | X_i(t) \geq u\}/\Delta u]$. In addition, we define $N(u) = \sum_{i=1}^{n} \sum_{k=1}^{b} N_i(t_k, u)$ and $Y(u) = \sum_{i=1}^{n} \sum_{k=1}^{b} Y_i(t_k, u)$ as the counting and at-risk processes for data that is pooled across follow-up windows. In the remainder of the manuscript, a subscript g is inserted as the first index to indicate treatment group; otherwise notation defined in this section is unchanged.

3.4 Constructing the Two-Sample Tests

3.4.1 Overall τ-restricted mean test

The first proposed summary statistic that we use to compare treatment groups pools data from each of the follow-up windows when estimating a τ-restricted mean time-to-first event. That is, the overall τ-restricted mean within each group is

$$\hat{\mu}_g^*(\tau) = \int_0^\tau \exp\left\{-\int_0^s \frac{dN_g(u)}{Y_g(u)}\right\} ds \text{ where } g = 1, 2.$$

Let $\lambda_g^W(u)$ be defined as $\sum_{k=1}^b \lambda_g(t_k, u) Pr\{X_{gi}(t_k) \geq u\} / \sum_{l=1}^b Pr\{X_{gi}(t_l) \geq u\}$ so that $\mu_g^*(\tau) = \int_0^\tau \exp\left[-\int_0^s \lambda_g^W(u)du\right] ds$ is the mean of the mixture distribution created from combining times-to-first event across the different follow-up intervals. In Appendix A of the Supplementary Materials, we show that $\sqrt{n_g}\{\hat{\mu}_g^*(\tau) - \mu_g^*(\tau)\}$ converges in distribution to a normal random variable with finite variance that is estimated by $\hat{\sigma}_{*g}^2 = \sum_{i=1}^{n_g} \left[z_i\{\hat{\mu}_g^*(\tau)\} - \bar{z}\{\hat{\mu}_g^*(\tau)\}\right]^2 / (n_g - 1)$ where $\bar{z}\{\hat{\mu}_g^*(\tau)\} = \sum_{i=1}^{n_g} z_i\{\hat{\mu}_g^*(\tau)\}/n_g$, $z_i\{\hat{\mu}_g^*(\tau)\} = \sum_{k=1}^b z_{ik}\{\hat{\mu}_g^*(\tau)\}$ and

$$z_{ik}\{\hat{\mu}_g^*(\tau)\} = \int_0^\tau \exp\left\{-\int_0^s \frac{dN_g(u)}{Y_g(u)}\right\} \left\{\int_0^s \frac{dN_{gi}(t_k, u) - \frac{dN_g(u)}{Y_g(u)} Y_{gi}(t_k, u)}{Y_g(u)/n_g}\right\} ds.$$

The variance calculations avoid assumptions about the correlation structure between follow-up intervals within a subject, while taking into account that some of these intervals overlap and some do not. This is a somewhat weaker assumption than what is typically used when calculating robust (sandwich) covariance structures. Sandwich covariance between follow-up intervals [0,12) and [6,18) would assume the same correlation structure for each individual contributing to the analysis, when in

reality, some times-to-first event in these two intervals contain overlapping follow-up segments (e.g., long event times) and some do not (e.g., short event times). In exploratory work, we found that the robust covariance estimated some mixture of these, which affected inference in finite sample sizes.

Let π_g be proportion of individuals in group g, $g = 1, 2$, which can be consistently estimated with $\hat{\pi}_g = n_g/(n_1 + n_2)$. A nonparametric test statistic comparing the overall τ-restricted mean survival in two independent groups of size n_1 and n_2 respectively is

$$\mathscr{T}_* = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left\{ \hat{\mu}_1^*(\tau) - \hat{\mu}_2^*(\tau) \right\}.$$

Under the null hypothesis $H_0 : \mu_1^*(\tau) = \mu_2^*(\tau)$, $\mathscr{T}_*$ has a mean zero normal limiting distribution with variance $\pi_2 \sigma_{*1}^2 + \pi_1 \sigma_{*2}^2$ that can be consistency estimated by $\hat{\pi}_2 \hat{\sigma}_{*1}^2 + \hat{\pi}_1 \hat{\sigma}_{*2}^2$, as shown in more detail in Appendix B of the Supplementary Materials.

3.4.2 Area under the τ-restricted mean residual lifetime function test

An alternative testing procedure is based on a function of conditional τ-restricted means, hereafter called the τ-restricted mean residual lifetime function (RMRL) and denoted by $\mu_g(t, \tau)$. This is the expected time to the next event during an interval of length τ given the patient has survived up to time t in each group $g = 1, 2$. A consistent nonparametric estimator of the τ-RMRL function is $\hat{\mu}_g(t, \tau) = \int_0^\tau \hat{P}_g(t, s) ds$ where $\hat{P}_g(t, s) = \exp\left\{ - \int_0^s dN_g(t, u)/Y_g(t, u) \right\}$. The area under $\mu_g(t, \tau)$ from t_1 to t_b is defined to be

$$\hat{\mu}_g(\cdot, \tau) = \int_{t_1}^{t_b} \hat{\mu}_g(t, \tau) dt.$$

In Appendix C of the Supplementary Materials we show that $\sqrt{n_g}\{\hat{\mu}_g(\cdot,\tau)-\mu_g(\cdot,\tau)\}$ converges in distribution to a normal random variable with finite variance that is estimated by $\hat{\sigma}^2_{R,g}=\sum_{i=1}^{n_g}\left[z_i\{\hat{\mu}_g(\cdot,\tau)\}-\bar{z}\{\hat{\mu}_g(\cdot,\tau)\}\right]^2/\left(n_g-1\right)$ where $\bar{z}\{\hat{\mu}_g(\cdot,\tau)\}=\sum_{i=1}^{n_g}z_i\{\hat{\mu}_g(\cdot,\tau)\}/n_g$ and $z_i\{\hat{\mu}_g(\cdot,\tau)\}=$

$$\int_{t_1}^{t_b}\int_0^{\tau}\exp\left\{-\int_0^s\frac{dN_g(t,u)}{Y_g(t,u)}\right\}\left\{\int_0^s\frac{dN_{gi}(t,u)-\frac{dN_g(t,u)}{Y_g(t,u)}Y_{gi}(t,u)}{Y_g(t,u)/n_g}\right\}ds\,dt.$$

A nonparametric test statistic comparing the integrated τ-RMRL in two independent groups of size n_1 and n_2 respectively is then defined as

$$\mathscr{T}_{\mathscr{R}}=\sqrt{\frac{n_1n_2}{n_1+n_2}}\left\{\hat{\mu}_1(\cdot,\tau)-\hat{\mu}_2(\cdot,\tau)\right\}.$$

Under the null hypothesis of no difference in the area under the τ-RMRL between the two treatment groups, $\mathscr{T}_{\mathscr{R}}$ converges in distribution to a mean zero normal distribution with variance $\pi_2\sigma^2_{R,1}+\pi_1\sigma^2_{R,2}$ that can be estimated with $\hat{\pi}_2\hat{\sigma}^2_{R,1}+\hat{\pi}_1\hat{\sigma}^2_{R,2}$; see Appendix D of the Supplementary Materials for more details.

Although not required for technical results to hold, we have found it convenient to estimate $\hat{\mu}_g(\cdot,\tau)$ using a reduced number of timepoints $t_1,\ldots,t_b$ and trapezoidal rule integration for faster computation. In this case $\hat{\mu}_g(\cdot,\tau)$ becomes $\sum_{k=1}^{b-1}(t_{k+1}-t_k)\left\{\hat{\mu}(t_{k+1},\tau)+\hat{\mu}(t_k,\tau)\right\}/2$ and $z_i\{\hat{\mu}_g(\cdot,\tau)\}=\sum_{k=1}^{b-1}(t_{k+1}-t_k)\left[z_{ik}\{\hat{\mu}(t_{k+1},\tau)\}+z_{ik}\{\hat{\mu}(t_k,\tau)\}\right]/2$ where $z_{ik}\{\hat{\mu}(t_k,\tau)\}=$

$$\int_0^{\tau}\exp\left\{-\int_0^s\frac{dN_g(t,u)}{Y_g(t,u)}\right\}\left\{\int_0^s\frac{dN_{gi}(t,u)-\frac{dN_g(t,u)}{Y_g(t,u)}Y_{gi}(t,u)}{Y_g(t,u)/n_g}\right\}ds\,.$$

3.4.3 Practical issues

The power of $\mathscr{T}_*$ and $\mathscr{T}_R$, are affected by the choice of τ and, for $\mathscr{T}_*$, the number and spacing of the follow-up intervals. Recall that in each follow-up interval of length τ, we only use information up to the first event time. Hence as τ gets larger, the potential loss of information increases as opposed to a gap-time oriented analysis. For this reason, we recommend a follow-up window length that is clinically meaningful and long enough for a treatment difference to emerge, but not excessively long. For example, in our pulmonary setting many study designs are centered on one-year differences, suggesting $\tau = 1$ year as a clinically meaningful window length. Since exacerbation rates average 1-2 per year in this setting, the choice of one year is also reasonable enough to detect treatment differences without ignoring much follow-up information in each window.

We advocate choosing $\{t_1, \ldots, t_b\}$ so that we have the most precise estimate of the statistic $\mu^*(\tau)$. In Appendix E of the Supplementary Materials, we summarize closed form variance calculations and simulations for the special case when a single event is observed for each patient. In short, we observe diminishing returns in precision gains from introducing starting times t_k, more frequently than $\tau/2$ units apart, so that follow-up intervals starting from $t_k = (k-1)\tau/2$ for $k = 1, \ldots, b$ are recommended, where b is chosen so that the final follow-up interval starting at t_b does not exceed the study period.

3.5 Simulations

Simulations based on 500 iterations, with $n_1 = n_2 = 100$, study finite sample properties of the proposed two-sample tests. The proposed 36-month study has a

12-month patient accrual period, where 30% of the sample is recruited at the start of the study and observed the full 36 months; the remaining 70% are uniformly accrued over the first 12 months, i.e., a uniform(24,36) administrative censoring mechanism.

To generate correlated recurrent and terminating events we first simulate mean zero multivariate normal random variables $U_{ij}, j = 1, 2, \ldots, J_i$ and V_i with $\text{Var}(U_{ij}) = \text{Var}(V_i) = 1$ for $j = 1, 2, \ldots, J_i$, $\text{corr}(U_{ij}, U_{ij'}) = \rho_1$ for $j \neq j'$ and $\text{corr}(U_{ij}, V_i) = \rho_2$. The parameter ρ_1 controls the correlation between gap times and the parameter ρ_2 controls the correlation between the gap times and the time to death. Using the probability integral transform method we then convert these multivariate normal random variables $U_{ij}, j = 1, 2, \ldots, J_i$ and V_i to correlated uniformly distributed random variables $U'_{ij}, j = 1, 2, \ldots, J_i$ and V'_i. Finally, the inverse probability integral transform method is used to obtain correlated exponentially distributed gap times $S_{ij}, j = 1, 2, \ldots, J_i$ and time to death D_i. In the simulations that follow we assume either independence ($\rho_1 = \rho_2 = 0$) or dependence ($\rho_1 = 0.5, \rho_2 = 0.3$) of events observed within individual.

Group 1's gap time incidence rate is $\lambda_{S1} = 1/12$ with death hazard $\lambda_{D1} = 1/36$. We consider three scenarios for treatment effect in group 2: an immediate treatment effect, a treatment effect that is delayed by 6 months and a treatment effect that vanishes after 12 months. In each case for group 2 when the treatment effect is active, the gap time incidence rates change to $\lambda_{S2} = \lambda_{S1} * \alpha$ and the death hazard changes to $\lambda_{D2} = \lambda_{D1} * \alpha$ where α takes on values $\{1, 0.9, 0.8, 0.7, 0.6\}$.

The performance of the proposed two-sample tests, $\mathscr{T}_*$ and $\mathscr{T}_R$, are compared to the performance of the robust proportional rates model of Lin et al. (2000) ($\mathscr{T}_{PM}$) applied to the combined event of recurrence or death, the combined statistic of Ghosh & Lin (2000) with equal weights for recurrent and terminal endpoints ($\mathscr{T}_{GL}$) and the

standard time-to-first event analysis based on the logrank test ($\mathscr{T}_{LR}$).

Type I errors of all statistics are presented on first line of Table 3.1 for the case where treatment effect is immediate and carries through the duration of the trial. For each correlation structure, tests maintain type I error rates around the nominal level of 0.05 (within a 95% CI of 0.031-0.069). When the recurrent and terminating events are independent ($\rho_1 = \rho_2 = 0$), the robust proportional rates model is observed to be the most powerful test, with the power of our tests close to that of Ghosh and Lin and larger than the standard logrank time-to-first-event analysis.

When events are correlated ($\rho_1 = 0.5, \rho_2 = 0.3$) , our proposed tests are observed to be more powerful than the alternatives. The power of $\mathscr{T}_{PM}$ and $\mathscr{T}_{LR}$ are comparable and slightly larger than that seen with $\mathscr{T}_{GL}$. Hence, our proposed tests are seen to best handle the correlated event data, while taking into account termination by death in a natural way.

Table 3.1: Power of two-sample hypothesis tests in 500 Monte Carlo simulations for scenario one where we assume proportional incidence rates in the two treatment groups and that treatment is effective immediately.

	$\rho_1 = 0,\ \rho_2 = 0$					$\rho_1 = 0.5,\ \rho_2 = 0.3$				
α	$\mathscr{T}_*$	$\mathscr{T}_R$	$\mathscr{T}_{PM}$	$\mathscr{T}_{GL}$	$\mathscr{T}_{LR}$	$\mathscr{T}_*$	$\mathscr{T}_R$	$\mathscr{T}_{PM}$	$\mathscr{T}_{GL}$	$\mathscr{T}_{LR}$
1.0	0.042	0.033	0.044	0.046	0.048	0.036	0.054	0.050	0.044	0.044
0.9	0.160	0.156	0.190	0.158	0.112	0.116	0.107	0.104	0.092	0.108
0.8	0.520	0.476	0.648	0.560	0.306	0.356	0.340	0.316	0.224	0.302
0.7	0.898	0.851	0.976	0.940	0.684	0.734	0.711	0.636	0.504	0.640
0.6	1.000	0.990	1.000	0.996	0.948	0.960	0.957	0.860	0.784	0.922

Table 3.2 shows results when the treatment effect is delayed by 6 months. The first row, with $\alpha = 1$, again shows that type I error rates are near the nominal level of 0.05, although the logrank test for time-to-first-event is somewhat high. When the events are uncorrelated, the area under the τ-RMRL test and the proportional rates model are most powerful among the alternatives. The power of $\mathscr{T}_*$ is similar to that

of $\mathscr{T}_{GL}$, followed distantly by $\mathscr{T}_{LR}$, which suffers the most from the delayed treatment effect since it uses the time-to-first-event.

In the correlated event case, the area under the τ-RMRL test has the highest power, followed by $\mathscr{T}_*$, while the remaining tests drop in power anywhere from 30-40%. The tests that assume proportional rates suffer somewhat from violation of the assumption in this scenario, in addition to the influence of the correlated data structure on power.

Table 3.2: Power of two-sample hypothesis tests in 500 Monte Carlo simulations for scenario two where we assume proportional hazards in the two treatment groups but that treatment is only effective after a latency period.

	$\rho_1 = 0$, $\rho_2 = 0$					$\rho_1 = 0.5$, $\rho_2 = 0.3$				
α	$\mathscr{T}_*$	$\mathscr{T}_R$	$\mathscr{T}_{PM}$	$\mathscr{T}_{GL}$	$\mathscr{T}_{LR}$	$\mathscr{T}_*$	$\mathscr{T}_R$	$\mathscr{T}_{PM}$	$\mathscr{T}_{GL}$	$\mathscr{T}_{LR}$
1.0	0.042	0.045	0.058	0.042	0.072	0.052	0.056	0.062	0.046	0.056
0.9	0.112	0.142	0.144	0.134	0.068	0.084	0.098	0.072	0.058	0.062
0.8	0.344	0.391	0.396	0.326	0.120	0.212	0.295	0.168	0.114	0.132
0.7	0.688	0.741	0.744	0.618	0.206	0.476	0.563	0.342	0.228	0.270
0.6	0.934	0.966	0.956	0.898	0.342	0.736	0.856	0.544	0.396	0.404

Table 3.3 shows results when the treatment effect vanishes after 12 months. With uncorrelated events, type I error rates are slightly inflated for all tests, although this pattern was not seen in the correlated scenario. In either correlation setting, the time-to-first event analysis has the greatest power compared to all the recurrent event analyses. The recurrent event testing procedures, that use additional follow-up information beyond the first combined-event, add statistical noise without sufficient statistical signal towards the end of follow-up. The area under the τ-RMRL test is especially sensitive to loss of treatment effect over time. This is perhaps the only scenario where the time-to-first event analysis is clearly preferred.

In Appendix B.6, we include plots of the empirical mean 12-month RMRL for group 1 and for each value of alpha for group 2 for each of the scenarios.

Table 3.3: Power of two-sample hypothesis tests in 500 Monte Carlo simulations for scenario 3 where we assume short duration treatment effect.

	$\rho_1 = 0$, $\rho_2 = 0$					$\rho_1 = 0.5$, $\rho_2 = 0.3$				
α	$\mathscr{T}_*$	$\mathscr{T}_R$	$\mathscr{T}_{PM}$	$\mathscr{T}_{GL}$	$\mathscr{T}_{LR}$	$\mathscr{T}_*$	$\mathscr{T}_R$	$\mathscr{T}_{PM}$	$\mathscr{T}_{GL}$	$\mathscr{T}_{LR}$
1.0	0.068	0.083	0.070	0.074	0.064	0.048	0.043	0.058	0.054	0.050
0.9	0.070	0.067	0.082	0.084	0.084	0.058	0.057	0.076	0.052	0.064
0.8	0.152	0.093	0.220	0.218	0.230	0.092	0.068	0.130	0.096	0.186
0.7	0.278	0.153	0.376	0.370	0.414	0.162	0.085	0.172	0.126	0.308
0.6	0.484	0.263	0.608	0.592	0.662	0.334	0.188	0.300	0.230	0.524

3.6 Example

The Azithromycin in COPD Trial (NACT) (Albert et al., 2011) randomized chronic obstructive pulmonary disease (COPD) patients with a history of prior acute exacerbations (AE) to receive either azithromycin or placebo for approximately 12-13 months (380 days) to determine whether azithromycin reduced the frequency of AE. The original study showed a significant benefit in the azithromycin group using traditional methods for recurrent events. To make the example more interesting, we restrict our attention to the first 380 randomized patients ($n_1 = 192$ azithromycin, $n_2 = 188$ placebo), which corresponds to approximately one year of accrual, rather than the full dataset with over 3 years of accrual and 1117 patients. The data includes recurrent AE times as well as information on mortality and loss to follow-up over the study period, with 511 observed total AE and mortality events spread throughout the 380 days of follow-up.

The time to first event analysis gives a hazard ratio of 0.80 (95% CI: 0.62-1.03) with a logrank test p-value of 0.085. The proportional means model recurrent event analysis estimates the intensity rate ratio as 0.78 (95% CI: 0.61-0.99) with a p-value of 0.045. The Ghosh & Lin test for recurrent events subject to death gives a p-value of 0.131.

We apply our testing procedure with $\tau = 250$, where the follow-up window length was chosen so that we could use two follow-up windows starting at days 0 and 130. Limitations in the follow-up data precluded our ability to view several one-year follow-up windows of treatment effect, which we would recommend for use with this statistic in the design phase of a COPD study in the future. Follow-up window start times of 0 and 130 days follow recommendations in Section 3.4.3, with a slight adjustment to ensure the entire follow-up period was used.

The estimated average time to event within the next 250 days across the follow-up windows is 162 days in the placebo group and 179 days in the azithromycin group, with a p-value of 0.030. The RMRL plot in Figure 3.1 indicates there may be a decreasing treatment effect over time but overall there is a significant treatment difference. The p-value of the testing procedure comparing the area under each of these RMRL functions is 0.029.

3.7 Discussion

We have developed two recurrent event testing procedures that are better able to detect treatment effects on a combined-endpoint when there are correlated recurrent and terminating events and the treatment effect continues to manifest in later follow-up periods (first two simulation scenarios). A consequence of correlation in event times within an individual is a significant drop in power that will affect the viability of study designs if not accounted for during the design stage. That is, a study designed assuming independent gap times will be underpowered. A study powered based on our proposed statistics with a reasonable correlation assumption would protect power; if correlation is weaker than that used in the design then power is

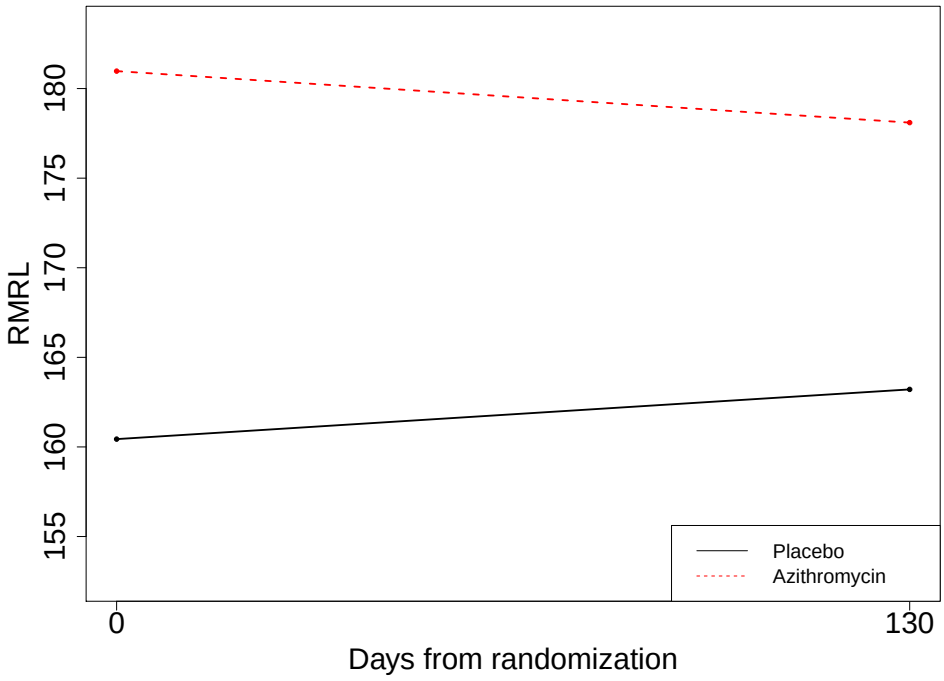


Figure 3.1: Restricted Mean Residual Life function evaluated for follow-up windows of interest beginning at 0 and 130 days from randomization.

better than planned. In the case that events are not correlated and incidence rates are proportional, the proportional means model is hard to beat, with power 10-20% higher than its competitors (first simulation scenario).

Our presentation that analyzes combined death and recurrence endpoints assumes that there is no interest in endpoint-specific treatment effects. When such interest exists, a competing risk testing procedure would be preferred. In the flavor of a competing risks analysis, our method can be applied to recurrent events data where deaths are treated as independent censoring events of the recurrent event process. Alternatively, our method can be applied to deaths without including recurrent events. Ghosh & Lin (2000) consider the treatment effect on the recurrent and terminating event types separately while acknowledging that patients who die cannot experience

any further recurrent events. When analysis of the combined event is sufficient in defining treatment effects, as is standard in the design of pulmonary clinical trials for example, our procedure appears to be more powerful than the Ghosh-Lin method.

CHAPTER IV

STATISTICAL CONSEQUENCES OF A SUCCESSFUL LUNG ALLOCATION SYSTEM – RECOVERING INFORMATION AND REDUCING BIAS IN MODELS FOR URGENCY

4.1 Summary

The national lung allocation system has reduced the number of waitlist deaths by ranking transplant candidates based on a lung allocation score (LAS). The LAS requires estimation of the 1-year restricted mean waitlist survival (urgency) based on current prognosis. Patients are required to update risk factors every 6 months. Fewer waitlist deaths and the systematic removal of candidates from the waitlist for transplantation present statistical challenges that must be addressed when using recent waitlist data. Multiple overlapping 1-year follow-up windows are used in a 1-year restricted mean model that estimates patient urgency based on risk-factors at the start of the window. Censored patients are multiply imputed by sampling from the inverse probability of censoring (IPCW) adjusted survival estimate, within a risk set of patients still at-risk and with similar prognosis to the censored patient. In simulation studies, we found that the multiple imputation procedure was able to produce unbi-

ased parameter estimates with similar efficiency to those obtained if censoring had never occurred. The analysis of 10,740 lung transplant candidates revealed that for most risk factors incorporating additional follow-up windows produced more efficient estimates.

4.2 Introduction

Since 2005, national lung allocation policy for those aged 12 and over has relied on the statistical estimation of a lung transplant candidate's 1-year restricted mean lifetime without opportunity for transplant (urgency) and the number of days to be gained in the next year if a transplant is offered immediately (benefit) (Egan et al., 2006). The United Network for Organ Sharing (UNOS) is charged with collecting and updating patient risk factors so that the lung allocation score (LAS) that determines transplantation priority can change with a candidate's prognosis. In fact, patients are required to update their allocation factors every six months or be penalized with a zero LAS value that effectively puts them at the end of the candidate list (OPTN Policy 3.7; http://optn.transplant.hrsa.gov/PoliciesandBylaws2/policies/pdfs/policy_9.pdf).

One feature that was intentionally designed into the LAS was a lack of influence of waiting time on a patient's score. The allocation method that preceded the LAS was based entirely on waiting time, with those waiting longer given higher priority for transplantation. This influenced listing behavior to the extent that candidates would enter the waitlist before being willing to accept an organ, just to accrue waiting time in the event they needed a transplant later. This also resulted in a high number of deaths among those who were too urgent to accrue the needed waitlist time to get to

the top of the list.

In terms of statistical development and maintenance of the LAS, only risk factors collected at a candidate's entry into the waitlist have been been used to model urgency to date and only the first year of follow-up after listing had been used for restricted mean lifetime estimation. This reflects an unfortunate waste of statistical information in a setting where fewer and fewer waitlist deaths are being observed, due in part to the success of allocation to patients more likely to die. Additional information on 1-year prognosis using windows of follow-up after listing would (1) potentially improve efficiency of estimation from additional events occurring beyond 1 year and (2) potentially expand the knowledge base of measured risk factors that progress beyond the listing stage, increasing the applicability of urgency scores to those on the waitlist beyond one year. This latter feature would be particularly useful since listing recommendations tend to catch patients at a similar state of urgency at the time these listing risk factors are collected, whereas patients progress at quite different rates thereafter. There is also the statistical challenge of dependent censoring of waitlist outcomes that are circumvented by a timely transplant intervention.

Gong & Schaubel (2013) model the distribution of survival time from a set of specified calendar times, conditional on the risk factors measured at each specified calendar time, through Cox-regression models. They account for dependent censoring through inverse probability of censoring weights. An estimate of restricted mean survival can be obtained from a Cox-regression model by integrating the estimated survival curve over the follow-up window of interest. Since our interest lies in estimation of a 1-year restricted mean survival, we have chosen to model the restricted mean survival directly.

In this paper we develop a restricted mean model for transplant urgency that uses

data from multiple follow-up windows during the listing period. Updated risk profile information at the beginning of each follow-up window is used to predict outcomes for the subsequent year. We extend a multiple imputation procedure developed by Xiang et al. (2013) for dependently censored data to address issues of removal from the lung candidate pool based on LAS values involving urgency. This, in turn, allows us to take advantage of generalized estimating equation (GEE) (Liang & Zeger, 1986) software to account for the particular flavor of correlation induced from incorporating (overlapping) follow-up windows from the same patient. In addition, we consider an inverse-weighted pseudo-observation approach, similar to that developed by Xiang & Murray (2012), that also takes advantage of the GEE framework to account for correlated follow-up information.

Methods are summarized in Section 4.3. Notation and the data structure induced by using multiple follow up windows are described in Section 4.3.1. Motivation behind use of GEE methods applied to imputed datasets in this setting are described in Section 4.3.2. It will be convenient to have a working understanding of inverse probability of censoring weighted (IPCW) survival estimation, as described by Robins (1993), as it plays an important role in our methodology. In Section 4.3.3, we briefly review how to construct inverse weighted survival estimates that are consistent for the survival function in the presence of dependent censoring. Our extension of the pseudo-observation restricted mean model, which adjusts for dependent censoring as in Xiang & Murray (2012), is described in Section 4.3.4. Methods for multiple imputation of dependently censored waitlist outcomes are described in Section 4.3.5. Section 4.4 assesses our approach versus alternative approaches for estimating lung candidate urgency via simulation. We then analyze a recent release of lung transplant data collected by UNOS in Section 4.5, providing updated urgency measures in

this cohort as well as evaluating urgency changes over time. A discussion follows in Section 4.6.

4.3 Multiple Imputation Methodology

4.3.1 Notation

Failure and censoring times are denoted by T_i and C_i, respectively, for patient $i = 1, \ldots, n$. The observed event time is $X_i = \min(T_i, C_i)$ with associated failure indicator variable $\delta_i = I(T_i < C_i)$. $\mathbf{V}_i(t)$ and $\mathbf{Z}_i(t)$ are covariates affecting C_i and T_i, respectively. We denote the recorded histories of $\mathbf{V}_i(t)$ and $\mathbf{Z}_i(t)$ up to time t by $\bar{V}_i(t) = \{\mathbf{V}_i(u); 0 \leq u \leq t\}$ and $\bar{Z}_i(t) = \{\mathbf{Z}_i(u); 0 \leq u \leq t\}$, respectively. The event counting process is defined as $N_i(t) = I(X_i \leq t, \delta_i = 1)$ and the at-risk process is defined as $Y_i(t) = I(X_i \geq t)$. We also define the counting process for censoring, $N_{Q_i}(t) = I(X_i \leq t, \delta_i = 0)$.

Our proposed method incorporates information from several follow-up windows of length 1 year, spaced 6 months apart; i.e., windows start at $\{0, 6, 12, ...\}$ months until removal from the candidate list for transplantation. More generally, windows of length τ start at times $\{t_1, \ldots, t_{n_i}\}$, $n_i = 1, \ldots, b$, $i = 1, \ldots, n$. For individuals with available follow-up during a window starting at t, we define $T_i^*(t) = \min(T_i - t, \tau)$ as the restricted time to event from t, where $\tau = 1$ year in the lung allocation setting but otherwise it may be any value where $P(C_i > \tau) > 0$. We will estimate candidate urgency using data pairs $\{T_i^*(t_1), \mathbf{Z}_i(t_1)\}, \{T_i^*(t_2), \mathbf{Z}_i(t_2)\}, \ldots, \{T_i^*(t_{n_i}), \mathbf{Z}_i(t_{n_i})\}$, $i = 1, ..., n$ via the model

$$E[\log\{T_i^*(t_j)\}] = \beta^T \mathbf{Z}_i(t_j). \tag{4.1}$$

In Figure 4.1 we illustrate the relationships between T_i, $T_i^*(t)$ and $\mathbf{Z}_i(t)$ for $t =$ 0, 6, 12 and 18 months. In our first example, patient 1 dies at 20 months post listing ($T_1 = 20$). Hence patient 1 contributes information on one year survival via the data pairs $\{T_1^*(0) = 12 \text{ months}, \mathbf{Z}_1(0)\}$, $\{T_1^*(6) = 12 \text{ months}, \mathbf{Z}_1(6)\}$, $\{T_1^*(12) = 8 \text{ months}, \mathbf{Z}_1(12)\}$, and $\{T_1^*(18) = 2 \text{ months}, \mathbf{Z}_1(18)\}$. Patient 2 is has $T_2 = 7$ months and is therefore observed for two follow-up windows with start times $\{0, 6\}$ and corresponding data pairs $\{T_2^*(0) = 7 \text{ months}, \mathbf{Z}_2(0)\}$ and $\{T_2^*(6) = 1 \text{ month}, \mathbf{Z}_2(6)\}$. Patient 3 was transplanted at $C_3 = 7$ months and is therefore also observed for two follow-up windows with start times $\{0, 6\}$. In Section 4.3.5, we describe a multiple imputation procedure for missing failure times that will be used in our analysis. Hence, if a failure time of 10 months is imputed for patient 3, this patient would contribute data pairs $\{T_3^*(0) = 10 \text{ months}, \mathbf{Z}_3(0)\}$ and $\{T_3^*(6) = 4 \text{ months}, \mathbf{Z}_3(6)\}$ to the analysis for one of the M multiply imputed datasets.

4.3.2 Generalized estimating equations for complete data

Once we have constructed the longitudinal data structure as described in Section 4.3.1, the correlation between observations from different follow-up windows within each patient must be accounted for when fitting model (4.1). GEE provides a framework that easily allows the correlation between overlapping and non-overlapping follow-up windows to differ via the unstructured working correlation matrix. In addition, the robust sandwich variance provides protection against misspecification of working correlation matrix. The wider availability of correlation matrices makes GEE software ideal. Currently available statistical software for correlated censored survival outcomes assumes that correlated event times follow an exchangeable correlation structure between any two pairs of outcomes that doesn't accommodate our

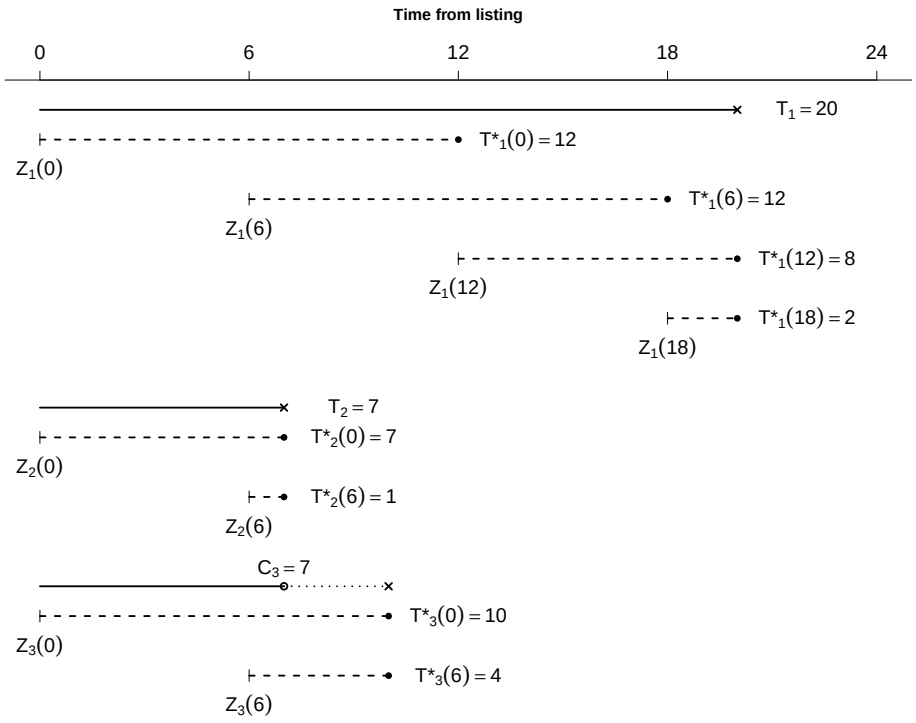


Figure 4.1: Example of how to construct the random variables $T_i^*(t) = \min\{T_i(t), 12\}$ in each of the follow-up windows for which the patients are under observation in three cases. Patient 1 dies at 20 months post-listing; patient 2 dies at 7 months post-listing; and patient 3 is transplanted at 7 months post-listing with one of their M imputed death times equal to 10 months.

data well. For example, the overlapping follow-up windows should have a different correlation structure than the windows that do not overlap.

4.3.3 IPCW Survival Estimation

The method to construct inverse weighted survival estimates, $\hat{S}_T^W(t)$, that are consistent for the survival function, $S_T(t)$, in the presence of dependent censoring proceeds as follows. First we fit a Cox model for censored outcomes via the model,

$$\lambda_Q\{t|\bar{V}(t)\} = \lambda_{Q_0}(t)\exp\{\gamma^T\mathbf{V}(t)\},$$

where $\lambda_Q\{t|\bar{V}(t)\} = \lim_{\Delta t \to 0} P\{t \leq X_i < t + \Delta t, \delta_i = 0 | X_i \geq t, \bar{V}(t)\}/\Delta t$ is the hazard function for the censoring distribution that depends on the recorded history of covariates $\bar{V}(t)$ via a proportional hazards model and $\lambda_{Q_0}(t)$ is an unspecified baseline hazard for the censoring distribution. The parameters from the model, γ, are consistently estimated by $\hat{\gamma}$ and the estimates can be obtained using most standard statistical software.

For each patient $i = 1, \ldots, n$, we define $K_i^{\mathbf{V}}(u) = P\{C_i > u | \bar{V}_i(u)\}$ based on the model above. Then patient i's estimated weight becomes

$$\hat{W}_i(u) = \hat{K}_i^{\mathbf{V}}(u)^{-1} = \exp\left\{ \sum_{j=1}^{n} \int_0^u \frac{e^{\hat{\gamma}^T \mathbf{V}_i(v)} dN_{Q_j}(v)}{\sum_{j'=1}^{n} Y_{j'}(v) e^{\hat{\gamma}^T \mathbf{V}_{j'}(v)}} \right\}.$$

We can then estimate the IPCW cumulative hazard by

$$\hat{\Lambda}_T^W(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)\hat{W}_i(u)}{\sum_{j=1}^{n} Y_j(u)\hat{W}_j(u)}$$

and the IPCW survival estimate adjusted for dependent censoring captured by $\mathbf{V}(t)$ becomes $\hat{S}_T^W(t) = \exp\{-\hat{\Lambda}_T^W(t)\}$.

4.3.4 Restricted Mean Model via Pseudo Observations Adjusted for Dependent Censoring

Andersen et al. (2004) developed a pseudo observation approach to modeling (4.1) that was later extended by Xiang & Murray (2012) to account for dependent censoring through incorporating inverse weighted survival estimates described in Section 4.3.3. Although pseudo observation methods sometimes struggle with intercept bias in small samples, resulting restricted mean estimates are useful in defining risk sets for imputation since these sets are invariant to intercept estimation. These models also

have the advantage of being extremely easy to implement. With the ultimate goal of defining risk sets that will be used in our imputation procedure, we now summarize how to estimate pseudo observations from each follow-up window.

For each patient $i = 1, \ldots, n$ at each time t_j where $j = 1, \ldots, n_i$, the pseudo-observation is defined to be

$$\mathscr{PO}_{ij} = n_j \hat{\delta}_j - (n_j - 1)\hat{\delta}_j^{-i} \tag{4.2}$$

where n_j is the number of patients at-risk at time t_j, and $\hat{\delta}_j$ and $\hat{\delta}_j^{-i}$ are estimates of $E[\log\{T_i^*(t_j)\}]$ based on datasets with and without patient i, respectively. The expectation of $\log\{T_i^*(t_j)\}$ can be written as

$$-\int_0^\tau \log(u) dP(T_i - t_j > u | T_i > t_j) + \log(\tau) P(T_i - t_j > \tau | T_i > t_j).$$

The estimate of $P(T_i - t_j > u | T_i > t_j)$ from the entire dataset is given by

$$\hat{P}(T_i - t_j > u | T_i > t_j) = \frac{\hat{S}_T^W(t_j + u)}{\hat{S}_T^W(t_j)}$$

and the estimate of $P(T_i - t_j > u | T_i > t_j)$ from the dataset without patient i is given by

$$\hat{P}^{(-i)}(T_i - t_j > u | T_i > t_j) = \frac{\hat{S}_T^{W(-i)}(t_j + u)}{\hat{S}_T^{W(-i)}(t_j)} \text{ where}$$
$$\hat{S}_T^{W(-i)}(t) = \exp\left\{ - \sum_{j=1, j\neq i}^{n} \int_0^t \frac{dN_j(u)\hat{W}_j(u)}{\sum_{j'=1, j'\neq i}^{n} Y_{j'}(u)\hat{W}_{j'}(u)} \right\}.$$

We then estimate $\hat{\delta}_j$ and $\hat{\delta}_j^{-i}$ with

$$\hat{\delta}_j = -\int_0^\tau \log(u) d\hat{P}(T_i - t_j > u | T_i > t_j) + \log(\tau)\hat{P}(T_i - t_j > \tau | T_i > t_j), \text{ and}$$
$$\hat{\delta}_j^{-i} = -\int_0^\tau \log(u) d\hat{P}^{(-i)}(T_i - t_j > u | T_i > t_j) + \log(\tau)\hat{P}^{(-i)}(T_i - t_j > \tau | T_i > t_j),$$

inserting these into equation (4.2) to obtain $\mathscr{PO}_{ij}$, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. GEE methodology applied to data pairs $\{\mathscr{PO}_{ij}, \mathbf{Z}_i(t_j)\}$, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$ give us model (4.1) parameter estimates, $\hat{\beta}^{PO^W}$.

4.3.5 Outline of Multiple Imputation Algorithm

Each censored patients' current follow-up window has measured covariates $\mathbf{Z}(t_{n_i})$ and a censored event time $C(t_{n_i}) = \min(C_i - t_{n_i}, \tau)$ that requires multiple imputation for generation of convenient complete datasets for analysis. Several authors have suggested methods for multiple imputation in the presence of dependent censoring including Faucett et al. (2002); Hsu et al. (2006); Liu et al. (2011); Xiang et al. (2013). Our approach extends that used by Xiang et al. (2013) to the setting with multiple follow-up windows, and hence is able to take advantage of time-dependent covariates available at the beginning of each window in model (4.1)'s restricted mean estimation. While Xiang et al. (2013) were able to use time-dependent covariates in forming risk sets similar to what we propose in the following, they were not able to incorporate updated covariate information directly into their restricted mean model.

Inverse Transform Imputation

The simplest case of the inverse transform imputation method is based on a Kaplan-Meier estimate, $\hat{S}_T(t)$, that is consistent for the survival function, $S_T(t)$. In

this case, Taylor et al. (2002) showed that multiple imputation reproduces the the Kaplan-Meier on average. We review this simplest case to avoid notation in delivering the concept. Our proposed multiple imputation method follows the same procedure with a different risk set definition for R_i and a different consistent survival estimate that accounts for dependent censoring within the risk set, $\hat{S}^W_{T_i^*(t_{n_i})}(t|R_i)$, both of which will be given in later in Section 4.3.5.

For a patient censored at C_i, Taylor et al. (2002) generate imputes by sampling from the distribution with survival function $S_T(t|T > C_i)$. Since $S_T(t|T > C_i)$ is not a known function, it is consistently estimated with $\hat{S}_T(t|T > C_i)$ in applying the inverse transform. The impute, t, is sampled by (1) generating a Uniform(0,1) random variable, u, and (2) finding the smallest value t where $\hat{S}_T(t|T > C_i) \leq u$. The risk set, R_i, is defined as the set of patients with comparable risk to the patient censored at C_i and in this simple case it is comprised of patients with $T_j > C_i$ for $j = 1, \ldots, n$. Step (2) can be equivalently expressed as finding the smallest value t where $\hat{S}_T(t|R_i) \leq u$ when using risk set notation.

Hsu et al. (2006) extended this algorithm to more complicated risk set, R_i, comprised of patients with similar survival and censoring hazard estimates. Liu et al. (2011) use restricted mean models in defining risk sets and recommend sampling from the distribution of the residual of the restricted mean model when forming an impute. Xiang et al. (2013) further extended this algorithm to the use of $\hat{S}^W_T(t)$ in the inverse transform method.

In our setting, we impute for an individual censored at $C_i(t_{n_i}) < \tau$ where imputes are sampled from the survival distribution of $T_i^*(t_{n_i})$ within the risk set, R_i. The impute, t, is sampled by (1) generating a Uniform(0,1) random variable, u, and (2)

finding the smallest value t where $\hat{S}^{W}_{T_i^*(t_{n_i})}(t|R_i) \leq u$. Then by (3) identifying the observed event time, $T_k^*(t_{n_i})$, that corresponds to t, we can solve for the associated residual ε using model (4.1) where $\log\{T_k^*(t_{n_i})\} = \hat{\beta}^{PO^W T}\mathbf{Z}_k(t_{n_i}) + \varepsilon$. Lastly, (4) if $t = \tau$ then impute $\tilde{T}_i^*(t_{n_i}) = \tau$, otherwise impute $\tilde{T}_i^*(t_{n_i}) = \exp\{\hat{\beta}^{PO^W T}\mathbf{Z}_i(t_{n_i}) + \varepsilon\}$. If $\tilde{T}_i^*(t_{n_i}) < C_i(t_{n_i})$ then sample another t and repeat steps (3) and (4) until the imputed value is greater than $C_i(t_{n_i})$.

Risk Set Formation

As in the simplest case of risk set formation, the minimal constraint for belonging to R_i is $T_k > C_i$, $k = 1 \ldots, n$. Further restrictions based on information from $\mathbf{Z}_i(t)$ improve similarity of patients in the risk set to the censored individual being imputed. The covariates $\mathbf{Z}_i(t)$ are related to survival and we can improve our imputation by selecting patients with similar urgency at the censoring time C_i based on our linear model (4.1). The second constraint for belonging to risk set R_i is then $|\hat{\beta}^{PO^W T}\mathbf{Z}_k(C_i) - \hat{\beta}^{PO^W T}\mathbf{Z}_i(C_i)| < \epsilon$ where ϵ is the parameter that controls how closely the linear predictors should match at C_i. The choice of epsilon is based on defining a risk set large enough to produce valid multiple imputes but as homogenous as possible with respect to urgency. In addition we require that patient k is in the same diagnosis group as patient i and that $LAS_k(C_i) = LAS_i(C_i)$ so that patients have similar urgency and transplant probability.

Inverse Weighted Survival Estimation Within Risk Set

Within the risk set, the inverse probability of censoring weight for the k^{th} patient is defined as

$$\begin{aligned} W_k^{R_i}(u) &= 1/P\{C_k > u | C_k > C_i, \bar{V}_k(u)\} \\ &= \frac{P\{C_k > C_i | \bar{V}_k(u)\}}{P\{C_k > u | \bar{V}_k(u)\}} \\ &= \frac{K_k^{\mathbf{V}}(C_i)}{K_k^{\mathbf{V}}(u)}. \end{aligned}$$

Then the inverse weighted survival estimate within the risk set is $\hat{S}^W_{T_i^*(t_{n_i})}(t|R_i) = \exp\{-\Lambda^W_{T_i^*(t_{n_i})}(u|R_i)\}$ for $C_i(t_{n_i}) \leq u < \tau$ and $\hat{S}^W_{T_i^*(t_{n_i})}(t|R_i) = 0$ for $u > \tau$ where $\hat{\Lambda}^W_{T_i^*(t_{n_i})}(u|R_i)$ is defined as

$$\sum_{k \in R_i} \int_{C_i}^{u+t_{n_i}} \frac{dN_k(v)\hat{W}_k^{R_i}(v)}{\sum_{j \in R_i} Y_j(v)\hat{W}_j^{R_i}(v)}.$$

Analysis of the M multiply imputed datasets

We repeat the imputation procedure until we obtain M completed datasets. In practice M=10 is usually sufficient to produce valid results. The analysis of the M multiply imputed datasets is given by Little & Rubin (1986). For each complete dataset, we can construct the longitudinal data structure described in Section 4.3.1. Model (4.1) GEE parameter estimates for dataset m are denoted by $\hat{\beta}_m^{MI}$ and their associated variance estimates are denoted by $\widehat{Var}(\hat{\beta}_m^{MI})$, $m = 1, \ldots, M$.

The estimates of β based on the multiple imputation method are $\hat{\beta}^{MI} = \sum_{m=1}^{M} \hat{\beta}_m^{MI}/M$. The associated variance estimate is $\widehat{Var}(\hat{\beta}^{MI}) = W + (1 + M^{-1})B$, where $W = \sum_{m=1}^{M} \widehat{Var}(\hat{\beta}_m^{MI})/M$ is the average within imputation variance and $B = \sum_{m=1}^{M}(\hat{\beta}_m^{MI} -$

$\hat{\beta}^{MI})^2/(M-1)$ is the between imputation variance. The 95% confidence intervals and hypothesis tests for $\hat{\beta}^{MI}$ are constructed based on the asymptotic distribution $(\hat{\beta}^{MI}-\beta)/\sqrt{Var(\hat{\beta}^{MI})} \sim t_\nu$. The degrees of freedom of the t-distribution are given by $\nu = (M-1)[1+W/\{B(M+1)\}]^2$.

4.4 Simulations

In order to better understand finite sample behavior of our methods in relation to other available approaches, we summarize simulation results from 500 Monte Carlo iterations with $n = 300$ patients. Each iteration gives GEE parameter estimates for the model, $E(\log[\min\{T_i(t_j), \tau\}]) = \beta_0 + \beta_1 \times Z_{1i}(t_j) + \beta_2 \times Z_{2i}$, via (a) the IPCW pseudo observation method applied to the first follow-up window as in Xiang & Murray (2012), (b) Xiang, Murray and Liu's multiple imputation method applied to the first follow-up window as in Xiang et al. (2013) and (c) our proposed multiple imputation method that incorporates information from multiple follow up windows. As a benchmark we also present results (d) in the absence of censoring when multiple follow up windows are used in estimation. Each method assumes $\tau = 1$ year. Methods using follow-up beyond year one in estimation incorporate information from 1-year windows starting at $t_1 = 0$, $t_2 = 6$ months and $t_3 = 12$ months. The data are generated as follows.

Step 1: A time-dependent covariate, $Z_{1i}(t_j)$, is simulated from a Uniform(0,1) at $t_1 = 0$, $t_2 = 6$ months and $t_3 = 12$ months. It will be convenient to denote the history of this time-dependent covariate by $\bar{Z}_{1i}(t) = \{Z_{1i}(u); 0 \le u \le t\}$. A time-independent covariate Z_{2i} is simulated from a Uniform(0,0.8).

Step 2: Each failure time T_i is simulated from a piecewise exponential distribution

with hazard equal to λ_{i1} in the interval $[0, 6]$, λ_{i2} in the interval $(6, 12]$ and λ_{i3} in the interval $(12, \infty)$. Then $P\{T_i - t_j > u | T_i > t_j, \bar{Z}_{1i}(u+t_j), Z_{2i}\} = \exp\{-\int_{t_j}^{t_j+u}[\lambda_{i1} I(0 \leq v \leq 6) + \lambda_{i2} I(6 < v \leq 12) + \lambda_{i3} I(12 < v < \infty)]dv\}$. And $E(\log[\min\{T_i(t_j), \tau\}])$ becomes

$$
\begin{aligned}
&-\int_0^{\tau} \log(u) dP\{T_i - t_j > u | T_i > t_j, \bar{Z}_1(u + t_j), Z_2\} \\
&+ \log(\tau) P\{T_i - t_j > \tau | T_i > t_j, \bar{Z}_1(\tau + t_j), Z_2\},
\end{aligned}
$$

where setting the above equal to the restricted mean model, $\beta_0 + \beta_1 Z_{1i}(t_j) + \beta_2 Z_{2i}$ gives us a way to solve for patient specific hazards. In Appendix C.1, we give further details on the algebra involved.

Step 3: The dependent censoring time C_i is generated from the piecewise exponential distribution with hazard $\lambda_i^C(u) = \lambda_0^C(u) \exp\{0.3 Z_{1i}(0) + 0.35 Z_{1i}(6) I(u > 6) + 0.01 Z_{1i}(0) Z_{1i}(6) I(6 < u \leq 12) + 0.4 Z_{1i}(12) I(u > 12) + 0.001 Z_{1i}(0) Z_{1i}(6) Z_{1i}(12) I(u > 12) + 0.1 Z_{2i}\}$ where $\lambda_0^C(u)$ is equal to 0.01 in the interval $[0, 6]$, 0.011 in the interval $(6, 12]$ and 0.012 in the interval $(12, \infty)$, producing approximately 25% censoring prior to 24 months.

Table 4.1 presents the results under the null hypothesis with $\beta_0 = 2.1$, $\beta_1 = 0$ and $\beta_2 = 0$ and when covariates affect survival with $\beta_0 = 2.1$, $\beta_1 = -0.125$ and $\beta_2 = 0.1$. For each of the parameters we present the empirical mean, bias, empirical mean standard error, empirical standard deviation and coverage of the 95% confidence interval for each of the analysis approaches under consideration.

Based on the results in Table 4.1, we observe that the parameter estimates from all methods have minimal bias under both the null hypothesis and for non-zero β's, except for the intercept term in the IPCW pseudo observation method. Several au-

thors have noted issues with intercept bias for pseudo observation methods (Andrei & Murray, 2007; Xiang & Murray, 2012). Low bias for β_1 and β_2 was also seen when pseudo observation regression methods were applied to data from multiple follow-up windows as described in Section 4.3.4 (data not shown), with β_0 underestimated to a similar degree. Hence, pseudo-observation methods seem useful for describing multiplicative effects of risk factors on the restricted mean and forming comparable risk sets of patients, but tend to underestimate restricted means compared to imputation methods with even moderate sample sizes.

Inclusion of follow-up windows starting at $t_2 = 6$ and $t_3 = 12$ months results in more efficient estimates. The asymptotic relative efficiency (ARE) of our proposed method versus the IPCW pseudo observation method is between 3.86 and 6.49 for each of the parameters. The ARE of our proposed method versus the multiple imputation method of Xiang and Murray is between 1.85 and 2.71. The ARE comparing our proposed method with versus without censoring is between 0.95 and 1 indicating that our method effectively handles dependent censoring and produces parameter estimates with nearly the same efficiency as if censoring never occurred.

4.5 Example

The lung waitlist consists of 10,740 transplant candidates aged 12 years and older who were newly listed between September 1, 2006 and March 2, 2012; 7,359 of these patients received a transplant, 884 died while on the waitlist, 1,124 dropped off the waitlist without a transplant and 1,373 were alive on the waitlist on March 2, 2012. Risk factors used to model LAS urgency are given by the OPTN Thoracic Committee (OPTN Policy 3.7) and have been vetted as worthy of inclusion in the algorithm. Most

Table 4.1: Comparison of estimates from model based on one follow-up window the IPCW pseudo-observation (IPCW-PO) and the multiple imputation method (MI); based on three follow-up windows using uncensored observations (Uncensored) and our multiple imputation method (MI*) under two scenarios with 500 Monte Carlo simulations. M=10 in both multiple imputation methods.

Parameter	IPCW-PO	MI	Uncensored	MI*
$\beta_0 = 2.1$	1.753	2.152	2.121	2.101
	[-0.347, 0.182,	[0.052, 0.119,	[0.021, 0.079,	[0.001, 0.081,
	0.170, 0.546]	0.113, 0.940]	0.080, 0.920]	0.080, 0.950]
$\beta_1 = 0$	-0.022	-0.010	-0.000	0.001
	[-0.022, 0.238,	[-0.010, 0.156,	[-0.000, 0.093,	[0.001, 0.097,
	0.237, 0.950]	0.153, 0.950]	0.092, 0.958]	0.093, 0.962]
$\beta_2 = 0$	-0.004	-0.007	-0.000	-0.002
	[-0.004, 0.298,	[-0.007, 0.196,	[-0.000, 0.134,	[-0.002, 0.142,
	0.292, 0.928]	0.199, 0.964]	0.143, 0.940]	0.143, 0.956]
$\beta_0 = 2.1$	1.715	2.150	2.126	2.107
	[-0.385, 0.183,	[0.050, 0.122,	[0.026, 0.081,	[0.007, 0.083,
	0.181, 0.436]	0.121, 0.930]	0.079, 0.940]	0.081, 0.948]
$\beta_1 = -0.125$	-0.142	-0.121	-0.123	-0.117
	[-0.017, 0.240,	[0.004, 0.160,	[0.002, 0.097,	[0.007, 0.100,
	0.245, 0.952]	0.166, 0.928]	0.095, 0.962]	0.097, 0.958]
$\beta_2 = 0.1$	0.164	0.096	0.090	0.085
	[0.064, 0.300,	[-0.004, 0.199,	[-0.001, 0.143,	[-0.015, 0.145,
	0.285, 0.956]	0.197, 0.954]	0.145, 0.946]	0.145, 0.946]

Empirical Mean
[Bias, Empirical Mean Standard Error,
Empirical Standard Deviation, Coverage of 95% Confidence Interval]

have proven historical statistical significance in at least one previous analysis of lung candidate data.

Patients are divided into four overarching diagnosis groups, A through D, by the OPTN Thoracic Committee that are considered to be similar with respect to waitlist and post-transplant survival. The details of the diagnoses that comprise each group are given in OPTN Policy 3.7. In our dataset at listing, of the 3618 patients in Group A, 2924 (81%) were diagnosed with chronic obstructive pulmonary disease; in Group B, 262 (56%) out of 468 were diagnosed with primary pulmonary hypertension; in

Group C, 1284 (99%) out of 1296 were cystic fibrosis patients; and in Group D, 3633 (68%) out of 5358 patients were diagnosed with idiopathic pulmonary fibrosis. A few group A and D diagnoses are allowed to enter the urgency model as their own risk factors. For group A, these are bronchiectasis, lymphangioleiomyomatosis and sarcoidosis with PA mean$\leq$30mm Hg. For group D, these are obliterative bronchiolitis, pulmonary fibrosis other and sarcoidosis with PA mean>30mm Hg. Eisenmenger syndrome, from group B, is also listed as a risk factor in LAS. Most of these smaller diagnosis groups are not statistically different from their larger conglomerate group designation, but having a separate parameter has been important in obtaining public approval of the algorithm.

In Table 4.2, we summarize the risk factors at listing within each of the four diagnoses groups. These factors are age, body mass index (BMI), cardiac index prior to any exercise, central venous pressure (CVP) at rest, whether they were on continuous mechanical ventilation, serum creatinine, whether they were diabetic, percent predicted forced vital capacity (FVC), whether they required assistance with the activities of daily living (ADL), O_2 requirement at rest needed to maintain adequate oxygen saturation, partial pressure of carbon dioxide (PCO_2), pulmonary artery (PA) systolic pressure at rest, and six minute walk distance obtained while receiving supplemental oxygen to maintain oxygen saturation of 88% or greater at rest. Those familiar with the LAS may recall that bilirubin has recently been approved as an urgency risk factor. However, this measure has only recently started being collected by the OPTN and was unavailable in the March 2012 release data used for our analyses.

Patients in group C are on average the youngest transplant candidates (mean age is 29.4 years) and this is expected since the group consists almost entirely of patients with cystic fibrosis, a genetic disorder that results in lung disease from a very young

age. The main diagnoses in the other groups are lung diseases that develop over time so most patients are older when they require a lung transplant. Having a cardiac index less than 2 L/min/m^2 is considered to be an indicator that the heart is not functioning well. Based on this measure, patients in group B tend to be the most severely ill with 17.5% of the patients having a cardiac index < 2 L/min/m^2. The proportion of patients requiring continuous mechanical ventilation is highest in group C (6.3%) and lowest in group A (0.7%). FVC % predicted is a measure of lung function and in terms of this measure the diagnoses groups are ranked C, D, A and B from most severely ill to least severely ill. However the patients in group C are also most likely to need no assistance with activities of daily living (19.2%) and are able to walk much long distances in the six minute walk test compared to the other groups. Group B consists of patients with various hypertensive disorders and as we would expect, patients in this group have the highest average PA systolic pressure at rest (76.7 mm Hg).

As urgent patients receive transplants, they are removed from the waitlist. This systematic removal of patients creates a problem of dependent censoring that we adjust for using inverse probability of censoring weights in the survival estimation procedure discussed in Section 4.3.3. For each patient on the waitlist the probability of being censored is estimated from a time-dependent Cox model, the results of which are presented in Table 4.3. The covariates that influence censoring are gender, race (white, black and other), height, blood type (A, B, O and AB) and time-dependent LAS and listing status (active and inactive). Gender, race and height are all seen to be highly significant characteristics for differentiating which patients will be censored. Blood types A and B were observed to be similar in terms of censoring hazard but patients with blood type O had a lower hazard of censoring compared to blood type A

(hazard ratio=0.96, 95% confidence interval: 0.92-1.00) and patients with blood type AB had a higher hazard of censoring compared to blood type A (hazard ratio=1.12, 95% confidence interval: 1.00-1.25). Patients with an LAS of 0 have the lowest possible score with a very low chance of transplant (censoring) that gets ameliorated a bit by geography when higher risk patients are not in competition for an organ. A one-unit increase in LAS when the $0<\text{LAS}\leq 30$ results in a decreasing hazard which reflects the low probability of being offered a transplant for low LAS scores. The effect of a one unit increase in LAS decreases for higher ranges of LAS scores. This may seem counterintuitive however the probability of being censored is counterbalanced by the probability of surviving until a transplant becomes available.

Figure 4.2 compares the IPCW survival estimate of waitlist survival to the Kaplan-Meier estimate of waitlist survival from listing up to three and a half years post listing. The Kaplan-Meier estimate does not adjust for dependent censoring resulting from transplantation and therefore estimates higher waitlist survival probabilities compared to the IPCW survival.

Using the IPCW survival estimate, we construct IPCW pseudo-observations for 1-year follow-up windows starting at 0, 6, 12, 18, 24 and 30 months following the method described in Section 4.3.4. The start time of the final follow-up window was chosen to ensure we had at least 25 risk set deaths in each follow-up window. The pseudo-observations are used to fit model (4.1) to estimate lung candidate urgency based on LAS risk factors. The parameter estimates are then used in the multiple imputation procedure presented in Section 4.3.5.

In Table 4.4 we present the results of fitting the restricted mean model to estimate lung candidate urgency using three different methods, (a) the IPCW pseudo observation method applied to the first follow-up window as in Xiang & Murray (2012), (b)

Xiang, Murray and Liu's multiple imputation method applied to the first follow-up window as in Xiang et al. (2013) and (c) our proposed multiple imputation method that incorporates information from multiple follow up windows. The exponentiated parameters ($e^{\hat{\beta}}$), 95% confidence intervals and p-values are presented for each risk factor and are therefore interpreted in terms of their multiplicative effect on the number of days lived in the next year.

The intercept estimated from the IPCW pseudo-observation (IPCW-PO) method has a very wide confidence interval indicating that it is not estimated well. All the effect size confidence intervals from the IPCW-PO method are wider than those obtained using the multiple imputation method applied to the first follow-up window. In general, the conclusions reached regarding statistically significant risk factors are the same even if the effect sizes differ between the two methods. The exceptions include having a cardiac index$<$2.0 L/min/m^2 and PA systolic pressure in group B, C or D.

Our proposed multiple imputation procedure estimates the effect of the risk factors across the 6 follow-up windows viewed during the 3.5 years since listing. Hence, our parameter estimates should be comparable to the other methods if time since listing does not play a strong role. Again, the intercept was the most different between our proposed method and the others, approximately 18 days smaller than the multiple imputation method based on only the first follow-up window. Our proposed method gave shorter confidence intervals compared to the other methods indicating increased efficiency resulting from the incorporation of additional follow-up windows. Comparing the two multiple imputation procedures, we observe that time since listing does not play a role and the effect sizes remain similar. The additional information contained in later follow-up windows allows us to confirm the statistical significance

of some of the risk factors including CVP in group B, being a diabetic and PCO_2.

Table 4.2: Summary of LAS urgency covariates at listing, by diagnosis group, in 10,740 lung transplant candidates†.

LAS Covariates	Group A n=3618	Group B n=468	Group C n=1296	Group D n=5358
Age (years)	58.1 (8.3)	45.7 (14.7)	28.4 (10.3)	57.1 (11.0)
BMI (kg/m^2)	24.6 (4.3)	25.1 (4.7)	19.3 (2.9)	27.0 (4.4)
Cardiac Index<2.0 ($L/min/m^2$)	150 (4.1%)	82 (17.5%)	16 (1.2%)	271 (5.1%)
CVP (mm Hg)	7.7 (4.2)	10.8 (6.5)	5.2 (3.8)	5.5 (4.3)
Continuous Mechanical Ventilation	26 (0.7%)	10 (2.1%)	82 (6.3%)	196 (3.7%)
Creatinine (serum mg/dL)	0.8 (0.2)	1.0 (0.6)	0.7 (0.3)	0.9 (0.3)
Diabetes	430 (11.9%)	64 (13.7%)	627 (48.4%)	1249 (23.3%)
FVC (% predicted)	54.0 (17.5)	66.4 (22.7)	40.0 (11.8)	47.6 (17.0)
No assistance with ADL	368 (10.2%)	32 (6.8%)	249 (19.2%)	506 (9.4%)
O_2 requirement at rest (L/min)	3.0 (2.5)	4.0 (4.4)	2.9 (3.5)	4.9 (5.2)
PCO_2 (mm Hg)	49.9 (10.8)	42.4 (6.4)	53.4 (17.4)	44.7 (8.2)
PA systolic (mm Hg)	38.3 (10.8)	76.7 (25.7)	39.0 (10.6)	43.2 (16.2)
Six-min walk distance (feet)	783.4 (347.9)	776.7 (438.4)	970.6 (465.8)	797.2 (464.2)

† For continuous variables, numbers shown are mean (standard deviation)

For binary variables, numbers shown are number (proportions)

Body Mass Index (BMI); Cardiac Index (CI); Central Venous Pressure (CVP)

Activities of Daily Living (ADL); Pulmonary Artery (PA)

4.6 Discussion

LAS implementation has successfully reduced waitlist deaths, which reflects vitally important improvements for end stage lung disease care. This same reduction in deaths results in less power to estimate waitlist survival in current cohorts. It is therefore critical to develop statistical methodology that is able extract as much information as possible from available data.

As we saw in Section 4.5, incorporating additional follow-up windows to estimate transplant urgency resulted greater efficiency, as evidenced by narrower confidence intervals for parameter estimates. We were also able to confirm that low central

Table 4.3: Proportional hazards censoring model for 10,740 lung transplant candidates

	Hazard Ratio	95% Confidence Interval	p-value
Time-independent characteristics			
Female (vs Male)	0.76	(0.72,0.80)	< 0.0001
Black (vs White)	0.77	(0.72,0.83)	< 0.0001
Other (vs White)	0.77	(0.72,0.83)	< 0.0001
Height: $< 5'3''$ (versus $> 5'9''$)	0.63	(0.58,0.68)	< 0.0001
Height: $5'3''$-$5'6''$ (versus $> 5'9''$)	0.78	(0.73,0.83)	< 0.0001
Height: $5'6''$-$5'9''$ (versus $> 5'9''$)	0.87	(0.82,0.92)	< 0.0001
Blood type: B (versus A)	1.00	(0.93,1.07)	0.9285
Blood type: O(versus A)	0.95	(0.91,0.99)	0.0134
Blood type: AB (versus A)	1.12	(1.00,1.25)	0.0473
Time-dependent characteristics			
LAS>0 (versus LAS$=0$)	2224	(22.14, > 3000)	0.0010
Unit increase in LAS: $0<$LAS≤ 30	0.76	(0.65,0.88)	0.0004
Unit increase in LAS: $30<$LAS≤ 35	1.14	(1.11,1.17)	< 0.0001
Unit increase in LAS: $35<$LAS≤ 40	1.11	(1.09,1.12)	< 0.0001
Unit increase in LAS: $40<$LAS≤ 60	1.04	(1.03,1.04)	< 0.0001
Unit increase in LAS: LAS>60	1.03	(1.03,1.04)	< 0.0001
Active vs inactive status	1.79	(1.65,1.95)	< 0.0001

venous pressure remained a statistically significant risk factor for patients in diagnosis group B for the current cohort. Measuring this risk factor is an invasive procedure so it is important to reassure that its collection is useful for ranking patients.

By design, a patient's LAS value does not change based on time-on-waitlist so that there is no advantage to altering listing behavior to game the system and disadvantage other patients. In our analysis we confirm that the effect sizes of most risk factors do not depend on time-on-waitlist.

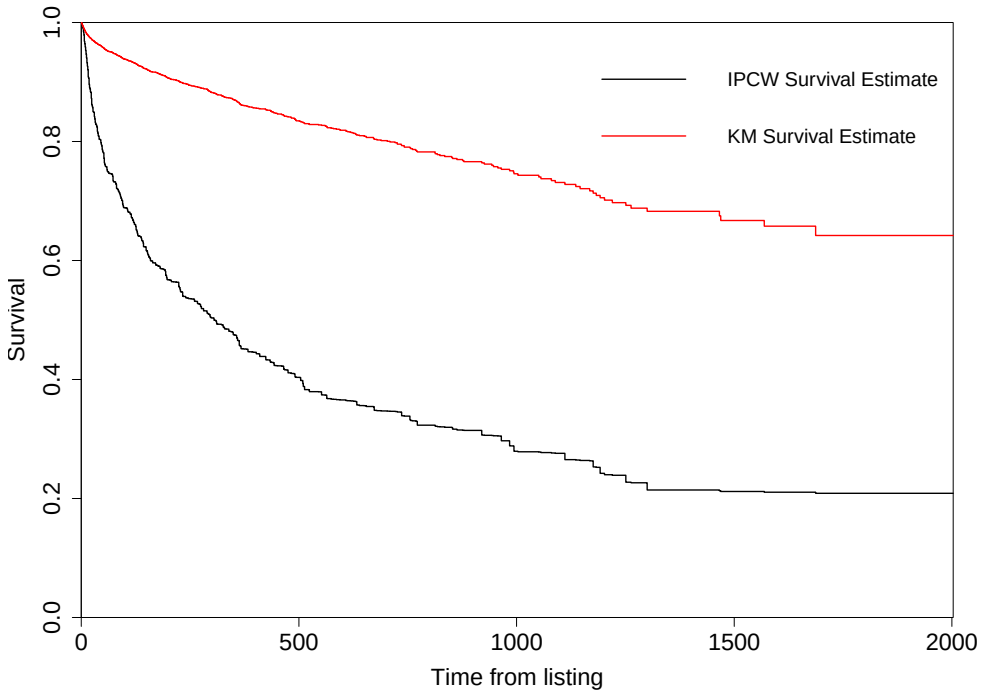


Figure 4.2: Waitlist probability of survival estimated by the Kaplan-Meier method and the IPCW-survival method.

Table 4.4: Results for urgency model based on the first follow-up window using the IPCW pseudo-observation (IPCW-PO) and the multiple imputation method (MI); and based on six follow-up windows using our multiple imputation method (MI*). M=10.

	IPCW-PO			MI			MI*		
	$e^{\hat{\beta}}$	95% CI	p-value	$e^{\hat{\beta}}$	95% CI	p-value	$e^{\hat{\beta}}$	95% CI	p-value
Intercept	1040.44	(274.78,3939.62)	< 0.0001	562.64	(446.18,709.49)	< 0.0001	544.91	(463.16,641.08)	< 0.0001
Age (per 20 year increase)	0.81	(0.66,0.98)	0.0328	0.94	(0.90,0.98)	0.0039	0.93	(0.91,0.96)	0.0002
BMI (per 1 kg/m^2 decrease if BMI<20kg/m^2)	1.02	(0.94,1.11)	0.6295	0.99	(0.97,1.01)	0.3014	0.98	(0.97,1.00)	0.0312
Cardiac Index<2.0 (ref=CI$\geq$ 2)	0.77	(0.47,1.25)	0.2938	0.85	(0.76,0.94)	0.0043	0.88	(0.81,0.96)	0.0062
CVP in Group B (per 10 mm Hg)	0.98	(0.92,1.05)	0.5599	0.84	(0.65,1.07)	0.1483	0.81	(0.68,0.96)	0.00191
Continuous Mechanical Ventilation (ref=other/no ventilation)	0.06	(0.02,0.21)	< 0.0001	0.22	(0.17,0.28)	< 0.0001	0.22	(0.17,0.28)	< 0.0001
Creatinine (Age> 18 years)	1.07	(0.83,1.39)	0.5926	0.95	(0.88,1.02)	0.1424	0.95	(0.90,1.00)	0.0610
Diabetes (ref=no diabetes)	0.90	(0.70,1.16)	0.4352	0.96	(0.90,1.01)	0.1091	0.96	(0.92,1.00)	0.0405
Diagnosis group (ref=group A)									
Group B	1.05	(0.43,2.57)	0.9228	1.12	(0.89,1.42)	0.3126	1.12	(0.94,1.34)	0.02072
Group C	0.73	(0.42,1.27)	0.2673	0.86	(0.77,0.96)	0.0104	0.88	(0.81,0.96)	0.0041
Group D	1.27	(0.72,2.22)	0.4056	0.97	(0.87,1.09)	0.6464	0.99	(0.90,1.08)	0.7972
Bronchiectasis	1.18	(0.89,1.55)	0.2468	0.99	(0.90,1.08)	0.8132	1.01	(0.95,1.08)	0.7189
Eisenmenger Syndrome	1.26	(0.81,1.97)	0.3115	1.14	(0.87,1.49)	0.3108	1.02	(0.74,1.40)	0.8929
Lymphangioleiomyomatosis	1.05	(0.48,2.31)	0.9025	1.05	(0.90,1.22)	0.5229	1.06	(0.96,1.18)	0.2037
Obliterative Bronchiolitis	1.84	(1.32,2.57)	0.0003	1.21	(0.99,1.47)	0.0571	1.16	(1.01,1.35)	0.0411
Pulmonary Fibrosis Other	1.32	(0.88,1.98)	0.1775	1.04	(0.95,1.13)	0.4035	1.05	(0.98,1.13)	0.1733
Sarcoidosis with PA mean>30mm Hg	2.00	(1.15,3.49)	0.0145	1.30	(1.15,1.46)	0.0002	1.27	(1.17,1.38)	< 0.0001
Sarcoidosis with PA mean$\leq$30mm Hg	0.83	(0.52,1.34)	0.4525	0.94	(0.85,1.04)	0.2231	0.93	(0.85,1.01)	0.0713
FVC (per 10% decrease if FVC %predicted<80%) in Group D	0.84	(0.76,0.94)	0.0015	0.96	(0.94,0.98)	0.0014	0.96	(0.95,0.98)	0.0003
No assistance with ADL (ref=some/total assisstance)	0.81	(0.65,1.01)	0.0583	1.00	(0.97,1.04)	0.8877	1.00	(0.97,1.02)	0.7865
O_2 requirement at rest in Group B	0.94	(0.90,0.99)	0.0154	0.93	(0.90,0.97)	0.0008	0.94	(0.91,0.97)	0.0002
O_2 requirement at rest in Groups A, C or D	0.83	(0.78,0.88)	< 0.0001	0.93	(0.92,0.94)	< 0.0001	0.94	(0.93,0.94)	< 0.0001
PCO_2 (per 10mm Hg)	0.91	(0.75,1.12)	0.3825	0.98	(0.95,1.00)	0.0778	0.98	(0.96,1.00)	0.0363
PCO_2 increase of $\geq$ 15% (PCO_2 increase of $<$ 15%)	0.35	(0.04,2.99)	0.3371	0.97	(0.69,1.35)	0.8244	1.01	(0.92,1.10)	0.8767
PA systolic in Group A (per 10 mm Hg increase if PA systolic> 40mm Hg)	1.04	(0.81,1.33)	0.7804	0.99	(0.94,1.03)	0.5164	1.00	(0.96,1.03)	0.7631
PA systolic in Group B, C or D (per 10 mm Hg increase)	0.92	(0.82,1.02)	0.1024	0.96	(0.95,0.98)	0.0005	0.97	(0.95,0.98)	0.0001
Six-min walk distance (per 1000 feet)	1.39	(1.09,1.78)	0.0091	1.13	(1.07,1.19)	0.0003	1.12	(1.08,1.17)	< 0.0001

Effect size, $e^{\hat{\beta}}$, per unit increase in the continuous risk factor unless otherwise stated
Effect size for reference categories of categorical risk factors is 1.
Body Mass Index (BMI); Cardiac Index (CI); Central Venous Pressure (CVP)
Activities of Daily Living (ADL); Pulmonary Artery (PA)

CHAPTER V

CONCLUSION

The objective of this dissertation was to develop statistical methods for estimation, hypothesis testing and regression models for the τ-restricted mean survival that included follow-up information beyond τ. To this end, we chose to use multiple observation windows, of length τ, within the available follow-up period. When estimating the restricted mean survival the additional follow-up windows produced more efficient estimates. In the hypothesis testing procedure we were able to show that both recurrent and terminating events could be included and this resulted in a more powerful test compared to some commonly used approaches when the events are highly correlated. The restricted mean regression models that we develop allow for time-dependent covariates that are updated every 6 months. In each of the statistical methods we have developed we approach the data from a different perspective and incorporate follow-up from multiple observation windows, which allows us to use more of the observed information.

In the Chapter II we introduced the overall τ-restricted mean survival that used follow-up from multiple observation windows to improve precision when estimating the τ-restricted mean survival. The correlated data structure resulting from the mul-

tiple follow-up windows required a variance estimate that was able to better account for the correlation structure than the sandwich estimator. We followed an approach by Woodruff (1971) where the estimator was linearized using Taylor series in order to identify independent sums of random variables using asymptotic arguments. We advocate the use of the empirical variance of these independent sums when estimating the variance of the overall τ-restricted mean survival. The closed form asymptotic variance was also studied, which required understanding the relationship between the event process and the at-risk process across follow-up windows within the same patient. This was necessary in order to study how the number and spacing of the follow-up windows affect the efficiency of the estimator. We found that the using overlapping follow-up windows spaced $\tau/2$ months apart provided the greatest increase in efficiency and that introducing follow-up windows more frequently resulted in greater complexity but only minor gains in efficiency. The finite sample-size performance of our estimator and proposed variance was compared to the sandwich variance and the variance that assumes independence in a simulation study. The estimate of the overall τ-restricted mean survival was unbiased and our proposed variance was better able to account for the correlation structure resulting from multiple overlapping follow-up windows compared to the sandwich variance.

The nonparametric hypothesis tests that we developed in Chapter III incorporated recurrent and terminating events beyond the first by combining time-to-first event analyses in each of the follow-up windows. This approach was an attractive alternative to the usual recurrent event analyses since it did not require any assumptions regarding the correlation between the event times but allowed us to use information beyond the initial event. We studied two methods of combining the follow-up windows for two-sample testing procedures in a randomized trial. The first pooled the

data in the same way as the overall τ-restricted mean survival in the previous chapter and the second used the area under the τ-restricted mean residual life function. For each test statistic we were able to describe the asymptotic distribution under the null using the linearization technique of Woodruff (1971) without making any assumptions about the correlation between events. In simulation studies we found that when events were correlated our methods outperformed the robust proportional rates model of Lin et al. (2000) and the Ghosh & Lin (2000) test for recurrent events subject to death. When events were uncorrelated, the proportional rates model was difficult to beat. We also observed that correlated events resulted in a significant decrease in the power. This observation has consequences for study design since a recurrent events study that does not account for potential correlation between gap-times will be underpowered if there is correlation. Our testing procedures were more powerful when there is correlated data and could be used in study design to ensure the study was sufficiency powered while accounting for potential correlation.

Chapter IV developed regression models for restricted mean survival that used risk-factors measured at the start of each follow-up windows. This allowed us to incorporate information from time-dependent risk factors past baseline. Our aim was to fit the LAS urgency model using time-dependent covariates. We extended work by Xiang et al. (2013) to estimate the parameters from the model. For each censored patient, we multiply imputed a time-to-event using the inverse transform method. The impute was sampled from within a risk-set of patients with similar prognosis to the censored patient but who were still at-risk. In the lung allocation setting, a patient's LAS score is one of the factors used to determine if they are offered a transplant. The transplantation of severely ill patients resulted in dependent censoring that we needed to account for in our methodology. We chose to use inverse probabil-

ity of censoring weights to decrease the bias of the survival function estimate. The complete data was then transformed into a longitudinal data structure. GEE was used to estimate model parameters based on data pairs comprised of 1-year restricted time-to-event from the start of the follow-up window and the risk-factors measured at the start of the follow-up window. The differing correlation between overlapping and non-overlapping follow-up windows was easily handled by GEE using an unstructured working correlation matrix with a robust sandwich variance to protect against misspecification. Simulation studies showed that our method produced unbiased and more efficient parameter estimates compared to methods which only used information from the baseline observation window. The increased efficiency of our method resulted in narrower confidence intervals for parameters in the LAS urgency model and the additional follow-up windows allowed us to study changes in the parameters over time.

APPENDICES

APPENDIX A

Supplementary Materials for Chapter II

A.1 Convergence of estimator

From page 10 in Section 2.4.1, we need to show that

$$\hat{\mu}^*(\tau) = \int_0^\tau \exp\left\{-\int_0^s \frac{dN(u)}{Y(u)}\right\} ds$$

converges in probability to

$$\mu^*(\tau) = \int_0^\tau \exp\left[-\int_0^s \lambda^W(u)du\right] ds,$$

where $\lambda^W(u) = \sum_{k=1}^b \lambda(t_k, u) Pr\{X_i(t_k) \geq u\} / \sum_{l=1}^b Pr\{X_i(t_l) \geq u\}$.

We use the following definitions from Section 2.4.2: $\{0 < \mathscr{T}_1 < \ldots < \mathscr{T}_M < \tau\}$ are observed event times from combined dataset; $\mathscr{T}_0 = 0$ and $\mathscr{T}_{M+1} = \tau$; $F_j\{dN(\mathscr{T}_j), Y(\mathscr{T}_j)\}$ $= dN(\mathscr{T}_j)/Y(\mathscr{T}_j)$; and $G_m(F_0, F_1, \ldots, F_m) = \exp(-\sum_{j=0}^m F_j)$ with arguments of F_j submerged for brevity. We can then rewrite $\hat{\mu}^*(\tau)$ as

$$\hat{\mu}^*(\tau) = \sum_{m=0}^{M} (\mathscr{T}_{m+1} - \mathscr{T}_m) G_m(F_0, F_1, \ldots, F_m).$$

A Taylor series expansion of $G_m(F_0, F_1, \ldots, F_m)$ about $\lambda^W(\mathscr{T}_j)d\mathscr{T}_j$ gives

$$\hat{\mu}^*(\tau) = \sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\} \tag{A.1}$$

$$+ \sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)(-1)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\left[\sum_{j=0}^{m}\{F_j - \lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}\right] \tag{A.2}$$

$$+ \sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\frac{1}{2!}\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\left[\sum_{j=0}^{m}\{F_j - \lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}\right]^2 \tag{A.3}$$

$$+ \sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\left[\text{ higher order terms}\right].$$

Term (A.1) is a non-stochastic numerical approximation of $\mu^*(\tau)$ that converges to $\mu^*(\tau)$ as $n \to \infty$. Our convergence argument is complete upon showing $\sum_{j=0}^{m}\{F_j - \lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}$ in (A.2) and (A.3) converges in probability to zero as $n \to \infty$.

Note that since $\sum_{j=0}^{m}\{F_j - \lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}$ is asymptotically equivalent in distribution to $\int_0^s \left\{dN(u)/Y(u) - \lambda^W(u)du\right\}$ with $s = \mathscr{T}_m$, we may show that $\int_0^s \{dN(u)/Y(u) - \lambda^W(u)du\}$ converges in probability to zero as $n \to \infty$ to complete our convergence argument.

It is convenient to first consider $\sum_{i=1}^{n} Y_i(u)/n$, an empirical measure based on i.i.d random variables. By the Glivenko-Cantelli Theorem, as well as the continuous mapping theorem,

$$\delta_1 = \sup_{u\in[0,\tau]}\left|\frac{Y(u)}{n} - \sum_{k=1}^{b} Pr\{X_i(t_k) \geq u\}\right| \xrightarrow{p} 0 \tag{A.4}$$

$$\text{and } \delta_2 = \sup_{u\in[0,\tau]}\left|\left\{\frac{Y(u)}{n}\right\}^{-1} - \left[\sum_{k=1}^{b} Pr\{X_i(t_k) \geq u\}\right]^{-1}\right| \xrightarrow{p} 0. \tag{A.5}$$

Then $\int_0^s \left\{ dN(u)/Y(u) - \lambda^W(u)du \right\}$ becomes

$$
\begin{aligned}
&\int_0^s \left(\frac{dN(u)}{n} \left[\frac{n}{Y(u)} + \frac{1}{\sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} - \frac{1}{\sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} \right] - \lambda^W(u)du \right) \\
=&\int_0^s \left[\frac{\sum_{k=1}^b dN(t_k,u)}{n \sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} - \frac{\sum_{k=1}^b \lambda(t_k,u) Pr\{X_i(t_k) \geq u\}}{\sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} du \right] \\
&+ \int_0^s \frac{\sum_{k=1}^b dN(t_k,u)}{n} \left[\frac{n}{Y(u)} - \frac{1}{\sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} \right] \\
=&\sum_{k=1}^b \int_0^s \frac{dN(t_k,u) - n\lambda(t_k,u) Pr\{X_i(t_k) \geq u\} du}{n \sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} \\
&+ \sum_{k=1}^b \int_0^s \frac{dN(t_k,u)}{n} \left(\left\{ \frac{Y(u)}{n} \right\}^{-1} - \left[\sum_{l=1}^b Pr\{X_i(t_l) \geq u\} \right]^{-1} \right) \\
\leq&\sum_{k=1}^b \int_0^s \frac{dN(t_k,u) - n\lambda(t_k,u) Pr\{X_i(t_k) \geq u\} du}{n \sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} + \sum_{k=1}^b \int_0^s \frac{dN(t_k,u)}{n} \delta_2. \qquad \text{(A.6)}
\end{aligned}
$$

The first component of term (A.6) is a sum of b mean zero stochastic integrals with respect to martingales. Using standard martingale theory, we derive the asymptotic variance of each of the b stochastic integrals to be the limit of the following expression

$$
\begin{aligned}
&\int_0^s \left[\frac{1}{n \sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} \right]^2 n\lambda(t_k,u) Pr\{X_i(t_k) \geq u\} du \\
=&\int_0^s \frac{\lambda(t_k,u) Pr\{X_i(t_k) \geq u\} du}{n \left[\sum_{l=1}^b Pr\{X_i(t_l) \geq u\} \right]^2}.
\end{aligned}
$$

It is easily seen that the expression converges in probability to zero as $n \to \infty$. The asymptotic covariance between any two of the b stochastic integrals associated with

follow-up windows beginning at t_k and t_l, $k \neq l$, is the limit of

$$\begin{aligned}
&\int_0^s \int_0^s \frac{\sum_{i=1}^n \sum_{j=1}^n cov\{dN_i(t_k,u), dN_j(t_l,v)\}}{n^2[\sum_{l'=1}^b Pr\{X_i(t_l') \geq u\}][\sum_{l'=1}^b Pr\{X_i(t_l') \geq v\}]} \\
=&\int_0^s \int_0^s \frac{n\, cov\{dN_i(t_k,u), dN_i(t_l,v)\}}{n^2[\sum_{l'=1}^b Pr\{X_i(t_l') \geq u\}][\sum_{l'=1}^b Pr\{X_i(t_l') \geq v\}]} \\
=&\int_0^s \int_0^s \frac{E\{dN_i(t_k,u)dN_i(t_l,v)\} - E\{dN_i(t_k,u)\}E\{dN_i(t_l,v)\}}{n[\sum_{l'=1}^b Pr\{X_i(t_l') \geq u\}][\sum_{l'=1}^b Pr\{X_i(t_l') \geq v\}]} \\
=&\int_0^s \frac{\lambda(t_k,u)Pr\{X_i(t_k) \geq u\}I(0 \leq u + t_k - t_l \leq s)du}{n[\sum_{l'=1}^b Pr\{X_i(t_l') \geq u\}][\sum_{l'=1}^b Pr\{X_i(t_l') \geq u + t_k - t_l\}]} \\
&- \int_0^s \int_0^s \frac{\lambda(t_k,u)Pr\{X_i(t_k) \geq u\}\lambda(t_l,v)Pr\{X_i(t_l) \geq v\}du\, dv}{n[\sum_{l'=1}^b Pr\{X_i(t_l') \geq u\}][\sum_{l'=1}^b Pr\{X_i(t_l') \geq v\}]}.
\end{aligned}$$

This expression also converges in probability to zero as $n \to \infty$. Therefore the variance of the sum of the b stochastic integrals converges to zero and the first component of term (A.6) converges in probability to zero. The second component of term (A.6) $\xrightarrow{p} 0$ after applying Lenglart's Inequality and (A.5) to each summand.

Therefore $\int_0^s \left\{dN(u)/Y(u) - \lambda^W(u)du\right\} \xrightarrow{p} 0$ as $n \to \infty$ and we have completed our convergence argument.

A.2 Asymptotic arguments for deriving the variance of the proposed estimate

Continuing from Section 2.4.2, we need to show the following result about the term (2.2d) from page 12:

$$\sqrt{n}\left[\sum_{j=0}^{m}\{F_j - \lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}\right]^2 \xrightarrow{p} 0.$$

This can be shown upon showing that the component asymptotically equivalent

in distribution,

$$\sqrt{n}\left[\int_0^s \left\{dN(u)/Y(u) - \lambda^W(u)du\right\}\right]^2,$$

converges in probability to zero. In Appendix A.1 we have shown that $\int_0^s \{dN(u)/Y(u) - \lambda^W(u)du\} \xrightarrow{p} 0$, hence we need only show that $\sqrt{n}[\int_0^s \{dN(u)/Y(u) - \lambda^W(u)du\}]$ converges to a distribution with finite variance.

Similarly to arguments used to formulate equation (A.6), we have

$$\begin{aligned}
&\sqrt{n}\left[\int_0^s \left\{\frac{dN(u)}{Y(u)} - \lambda^W(u)du\right\}\right] \\
\leq &\sum_{k=1}^{b} \sqrt{n}\int_0^s \frac{dN(t_k,u) - n\lambda(t_k,u)Pr\{X_i(t_k) \geq u\}du}{n\sum_{l=1}^{b} Pr\{X_i(t_l) \geq u\}} + \delta_2 \sum_{k=1}^{b} \sqrt{n}\int_0^s \frac{dN(t_k,u)}{n}. \quad \text{(A.7)}
\end{aligned}$$

The first component of term (A.7) is the sum of b stochastic integrals and by the martingale central limit theorem, each converges to a normal process with mean zero and variance

$$\begin{aligned}
&\int_0^s \left[\frac{\sqrt{n}}{n\sum_{l=1}^{b} Pr\{X_i(t_l) \geq u\}}\right]^2 n\lambda(t_k,u)Pr\{X_i(t_k) \geq u\}du \\
= &\int_0^s \frac{\lambda(t_k,u)Pr\{X_i(t_k) \geq u\}du}{\left[\sum_{l=1}^{b} Pr\{X_i(t_l) \geq u\}\right]^2}.
\end{aligned}$$

The asymptotic covariance between any two of the b stochastic integrals associated

with follow-up windows beginning at t_k and t_l, $k \neq l$, is the limit of

$$\begin{aligned}
&\int_0^s \int_0^s \frac{\sum_{i=1}^n \sum_{j=1}^n cov\{\sqrt{n}\, dN_i(t_k,u), \sqrt{n}\, dN_j(t_l,v)\}}{n^2[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}][\sum_{l'=1}^b Pr\{X_i(t'_l) \geq v\}]} \\
=&\int_0^s \int_0^s \frac{n^2\, cov\{dN_i(t_k,u), dN_i(t_l,v)\}}{n^2[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}][\sum_{l'=1}^b Pr\{X_i(t'_l) \geq v\}]} \\
=&\int_0^s \int_0^s \frac{E\{dN_i(t_k,u)dN_i(t_l,v)\} - E\{dN_i(t_k,u)\}E\{dN_i(t_l,v)\}}{[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}][\sum_{l'=1}^b Pr\{X_i(t'_l) \geq v\}]} \\
=&\int_0^s \frac{\lambda(t_k,u)Pr\{X_i(t_k) \geq u\}I(0 \leq u + t_k - t_l \leq s)du}{[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}][\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u + t_k - t_l\}]} \\
&- \int_0^s \int_0^s \frac{\lambda(t_k,u)Pr\{X_i(t_k) \geq u\}\lambda(t_l,v)Pr\{X_i(t_l) \geq v\}du\, dv}{[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}][\sum_{l'=1}^b Pr\{X_i(t'_l) \geq v\}]}.
\end{aligned}$$

This expression is finite as $n \to \infty$. Therefore the sum of the b stochastic integrals has finite asymptotic variance.

After b applications of the central limit theorem to each term $\sqrt{n}\sum_{i=1}^n \int_0^s dNi(t_k,u)/n$, $k = 1, \ldots, b$, we note that each converges to a normal distribution with finite variance. The covariance between any two terms associated with follow-up windows beginning at t_k and t_l, $k \neq l$ is

$$\begin{aligned}
&\int_0^s \int_0^s \frac{\sum_{i=1}^n \sum_{j=1}^n cov\{\sqrt{n}\, dN_i(t_k,u), \sqrt{n}\, dN_j(t_l,v)\}}{n^2} \\
=&\int_0^s \int_0^s \frac{n^2\, cov\{dN_i(t_k,u), dN_i(t_l,v)\}}{n^2} \\
=&\int_0^s \int_0^s E\{dN_i(t_k,u)dN_i(t_l,v)\} - E\{dN_i(t_k,u)\}E\{dN_i(t_l,v)\} \\
=&\int_0^s \lambda(t_k,u)Pr\{X_i(t_k) \geq u\}I(0 \leq u + t_k - t_l \leq s)du \\
&- \int_0^s \int_0^s \lambda(t_k,u)Pr\{X_i(t_k) \geq u\}\lambda(t_l,v)Pr\{X_i(t_l) \geq v\}du\, dv.
\end{aligned}$$

Therefore $\sum_{k=1}^b \sqrt{n}\sum_{i=1}^n \int_0^s dNi(t_k,u)/n$ converges to a distribution with finite variance. By applying Slutsky's theorem and (A.5), the second component of term (A.7)

converges in probability to zero. Therefore we can conclude that $\sqrt{n}[\int_0^s \{dN(u)/Y(u) - \lambda^W(u)du\}]$ converges to a distribution with finite variance.

Next we show that the term (2.3d) from page 13 converges in probability to zero. Once again we work with asymptotically equivalent stochastic integrals. The result follows after showing each component of the term (2.3d) converges in probability to zero, i.e.

i

$$\sqrt{n}\int_0^s \frac{[Y(u) - E\{Y(u)\}][dN(u) - E\{dN(u)\}]}{E\{Y(u)\}^2} \xrightarrow{p} 0$$

ii

$$\sqrt{n}\int_0^s \frac{[Y(u) - E\{Y(u)\}]^2 E\{dN(u)\}}{E\{Y(u)\}^3} \xrightarrow{p} 0$$

Starting with term (i) and plugging in expectations gives

$$\begin{aligned}
&\sqrt{n}\int_0^s \frac{[Y(u) - n\sum_{k=1}^b Pr\{X_i(t_k) \geq u\}][dN(u) - n\sum_{k=1}^b \lambda(t_k,u)Pr\{X_i(t_k) \geq u\}du]}{n^2\left[\sum_{l=1}^b Pr\{X_i(t_l) \geq u\}\right]^2} \\
=&\sqrt{n}\int_0^s \frac{[Y(u)/n - \sum_{k=1}^b Pr\{X_i(t_k) \geq u\}]\sum_{k=1}^b [dN(t_k,u) - n\lambda(t_k,u)Pr\{X_i(t_k) \geq u\}du]}{n\left[\sum_{l=1}^b Pr\{X_i(t_l) \geq u\}\right]^2} \\
\leq&\delta_1 \sum_{k=1}^b \sqrt{n}\int_0^s \frac{dN(t_k,u) - n\lambda(t_k,u)Pr\{X_i(t_k) \geq u\}du}{n\left[\sum_{l=1}^b Pr\{X_i(t_l) \geq u\}\right]^2}.
\end{aligned}$$

By the martingale central limit theorem, each of the b terms converge to a mean zero

normal process with variance

$$\int_0^s \left(\frac{\sqrt{n}}{n\left[\sum_{l=1}^b Pr\{X_i(t_l) \geq u\}\right]^2} \right)^2 n\lambda(t_k,u) Pr\{X_i(t_k) \geq u\} du$$
$$= \int_0^s \lambda(t_k,u) Pr\{X_i(t_k) \geq u\} / \left[\sum_{l=1}^b Pr\{X_i(t_l) \geq u\} \right]^4 du.$$

The asymptotic covariance between any two of the b stochastic integrals associated with follow-up windows beginning at t_k and t_l, $k \neq l$, is the limit of

$$\begin{aligned}
&\int_0^s \int_0^s \frac{\sum_{i=1}^n \sum_{j=1}^n cov\{\sqrt{n}\, dN_i(t_k,u), \sqrt{n}\, dN_j(t_l,v)\}}{n^2[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}]^2[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq v\}]^2} \\
=&\int_0^s \int_0^s \frac{n^2\, cov\{dN_i(t_k,u), dN_i(t_l,v)\}}{n^2[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}]^2[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq v\}]^2} \\
=&\int_0^s \int_0^s \frac{E\{dN_i(t_k,u) dN_i(t_l,v)\} - E\{dN_i(t_k,u)\} E\{dN_i(t_l,v)\}}{[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}]^2[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq v\}]^2} \\
=&\int_0^s \frac{\lambda(t_k,u) Pr\{X_i(t_k) \geq u\} I(0 \leq u + t_k - t_l \leq s) du}{[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}]^2[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u + t_k - t_l\}]^2} \\
&- \int_0^s \int_0^s \frac{\lambda(t_k,u) Pr\{X_i(t_k) \geq u\} \lambda(t_l,v) Pr\{X_i(t_l) \geq v\} du\, dv}{[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq u\}]^2[\sum_{l'=1}^b Pr\{X_i(t'_l) \geq v\}]^2}.
\end{aligned}$$

This expression is finite as $n \to \infty$ and therefore the sum of the b stochastic integrals has finite asymptotic variance. Applying (A.4) and Slutsky's theorem, it can easily be shown that term (i) converges in probability to zero.

Plugging the expectations into term (ii) gives us

$$\begin{aligned}
&\sqrt{n}\int_0^s [Y(u) - n\sum_{k=1}^{b} Pr\{X_i(t_k) \geq u\}]^2 \frac{\sum_{l=1}^{b} \lambda(t_l,u) Pr\{X_i(t_l) \geq u\}}{n^2 \left[\sum_{l'=1}^{b} Pr\{X_i(t_{l'}) \geq u\}\right]^3} du \\
=&\int_0^s \sqrt{n}\left[\frac{Y(u) - n\sum_{k=1}^{b} Pr\{X_i(t_k) \geq u\}}{n}\right]^2 \frac{\sum_{l=1}^{b} \lambda(t_l,u) Pr\{X_i(t_l) \geq u\}}{\left[\sum_{l'=1}^{b} Pr\{X_i(t_{l'}) \geq u\}\right]^3} du \\
\leq&\delta_1 \int_0^s \sqrt{n}\left[\frac{Y(u) - n\sum_{k=1}^{b} Pr\{X_i(t_k) \geq u\}}{n}\right] \frac{\sum_{l=1}^{b} \lambda(t_l,u) Pr\{X_i(t_l) \geq u\}}{\left[\sum_{l'=1}^{b} Pr\{X_i(t_{l'}) \geq u\}\right]^3} du \\
=&\delta_1 \int_0^s \sqrt{n}\sum_{i=1}^{n}\sum_{k=1}^{b}\left[\frac{Y_i(t_k,u) - Pr\{X_i(t_k) \geq u\}}{n}\right] \frac{\sum_{l=1}^{b} \lambda(t_l,u) Pr\{X_i(t_l) \geq u\}}{\left[\sum_{l'=1}^{b} Pr\{X_i(t_{l'}) \geq u\}\right]^3} du \\
=&\delta_1 \sum_{k=1}^{b}\sqrt{n}\sum_{i=1}^{n}\int_0^s \left[\frac{Y_i(t_k,u) - Pr\{X_i(t_k) \geq u\}}{n}\right] \frac{\sum_{l=1}^{b} \lambda(t_l,u) Pr\{X_i(t_l) \geq u\}}{\left[\sum_{l'=1}^{b} Pr\{X_i(t_{l'}) \geq u\}\right]^3} du.
\end{aligned}$$

By the central limit theorem, each of the b summands converge in distribution to a normal random variable with covariance between any two terms associated with

follow-up windows beginning at t_k and t_l defined as

$$
\begin{aligned}
&\int_0^s \int_0^s \frac{\sum_{i=1}^n \sum_{j=1}^n cov\{\sqrt{n}\, Y_i(t_k,u), \sqrt{n}\, Y_j(t_l,v)\}}{n^2} \times \\
&\left(\frac{\sum_{l=1}^b \lambda(t_l,u) Pr\{X_i(t_l) \geq u\}}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq u\}\right]^3} \right) \left(\frac{\sum_{l=1}^b \lambda(t_l,v) Pr\{X_i(t_l) \geq v\}}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq v\}\right]^3} \right) du\, dv \\
=&\int_0^s \int_0^s \frac{n^2\, cov\{Y_i(t_k,u), Y_i(t_l,v)\}}{n^2} \times \\
&\left(\frac{\sum_{l=1}^b \lambda(t_l,u) Pr\{X_i(t_l) \geq u\}}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq u\}\right]^3} \right) \left(\frac{\sum_{l=1}^b \lambda(t_l,v) Pr\{X_i(t_l) \geq v\}}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq v\}\right]^3} \right) du\, dv \\
=&\int_0^s \int_0^s [E\{Y_i(t_k,u) Y_i(t_l,v)\} - E\{Y_i(t_k,u)\} E\{Y_i(t_l,v)\}] \times \\
&\left(\frac{\sum_{l=1}^b \lambda(t_l,u) Pr\{X_i(t_l) \geq u\}}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq u\}\right]^3} \right) \left(\frac{\sum_{l=1}^b \lambda(t_l,v) Pr\{X_i(t_l) \geq v\}}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq v\}\right]^3} \right) du\, dv \\
=&\int_0^s \int_0^s [Pr\{X_i(t_k) \geq u, X_i(t_l) \geq v\} - Pr\{X_i(t_k) \geq u\} Pr\{X_i(t_l) \geq v\}] \times \\
&\left(\frac{\sum_{l=1}^b \lambda(t_l,u) Pr\{X_i(t_l) \geq u\}}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq u\}\right]^3} \right) \left(\frac{\sum_{l=1}^b \lambda(t_l,v) Pr\{X_i(t_l) \geq v\}}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq v\}\right]^3} \right) du\, dv
\end{aligned}
$$

Since the covariance terms are finite, the sum converges to a distribution with finite variance and it can then be shown that term (ii) converges in probability to zero by applying Slutsky's theorem and (A.4).

A.3 Closed form asymptotic variance

The asymptotic closed form variance is useful for understanding design issues and is defined in Section 2.4.2 on page 14. In this appendix we provide the details for the closed form calculations. We are interested $\sigma^2 = Var\left[\sqrt{n} \sum_{i=1}^n \sum_{k=1}^b Z_{ik}\{\hat{\mu}^*(\tau)\}/n\right]$,

where

$$Z_{ik}\{\hat{\mu}^*(\tau)\}=\int_0^\tau \exp\left\{-\int_0^s \lambda^W(u)du\right\}\left[\int_0^s \frac{dN_i(t_k,u)-\lambda^W(u)Y_i(t_k,u)du}{\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}}\right]ds.$$

Examining the variance more closely, we may simplify and expand terms.

$$\begin{aligned}
\sigma^2 &= Var\left[\sqrt{n}\sum_{i=1}^n\sum_{k=1}^b Z_{ik}\{\hat{\mu}^*(\tau)\}/n\right] \\
&= \sum_{i=1}^n Var\left[\sum_{k=1}^b Z_{ik}\{\hat{\mu}^*(\tau)\}\right]/n \\
&= Var\left[\sum_{k=1}^b Z_{ik}\{\hat{\mu}^*(\tau)\}\right] \\
&= \sum_{k=1}^b\sum_{l=1}^b cov\left[Z_{ik}\{\hat{\mu}^*(\tau)\}, Z_{il}\{\hat{\mu}^*(\tau)\}\right] \\
&= \sum_{k=1}^b\sum_{l=1}^b \int_0^\tau\int_0^\tau exp\left\{-\int_0^s \lambda^W(u)du\right\} exp\left\{-\int_0^{s'} \lambda^W(v)dv\right\} \\
&\quad \int_0^s\int_0^{s'} cov\left[\frac{dN_i(t_k,u)-\lambda^W(u)Y_i(t_k,u)du}{\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq u\}}, \frac{dN_i(t_l,v)-\lambda^W(v)Y_i(t_l,v)dv}{\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq v\}}\right] ds\, ds' \quad \text{(A.8)}
\end{aligned}$$

The covariance that we need to calculate is based on correlated terms from patient i:

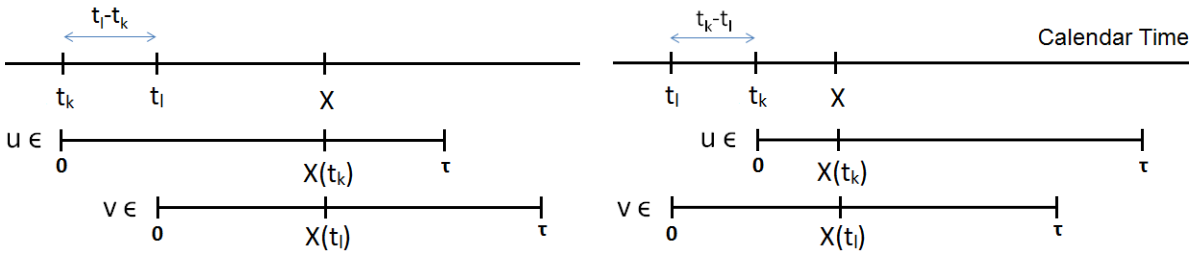
$$\begin{aligned}
&cov\left\{dN_i(t_k,u)-\lambda^W(u)Y_i(t_k,u)du, dN_i(t_l,v)-\lambda^W(v)Y_i(t_l,v)dv\right\} \\
&= E\{dN_i(t_k,u)dN_i(t_l,v)\}-\lambda^W(u)E\{Y_i(t_k,u)dN_i(t_l,v)\}du-\lambda^W(v)E\{Y_i(t_l,v)dN_i(t_k, \\
&u)\}dv+\lambda^W(u)\lambda^W(v)E\{Y_i(t_k,u)Y_i(t_l,v)\}du\,dv-E\left\{dN_i(t_k,u)-\lambda^W(u)Y_i(t_k,u)du\right\} \\
&E\left\{dN_i(t_l,v)-\lambda^W(v)Y_i(t_l,v)dv\right\}.
\end{aligned}$$

$E\{dN_i(t_k,u)dN_i(t_l,v)\}$ involves the failure time of the same subject in different observation windows. From the definitions of $X_i(t_k)$ and $X_i(t_l)$ we can deduce that

the expectation is only non-zero when $u + t_k = v + t_l$. Then

$$
\begin{aligned}
&E\{dN_i(t_k, u)dN_i(t_l, v)\} \\
&= \lim_{\Delta u, \Delta v \to 0} Pr\{u \le X_i(t_k) < u + \Delta u, \delta_i(t_k) = 1, v \le X_i(t_l) < v + \Delta v, \delta_i(t_l) = 1\} \\
&= \lim_{\Delta u \to 0} Pr\{u \le X_i(t_k) < u + \Delta u, \delta_i(t_k) = 1\} I(v = u + t_k - t_l) \\
&= \lim_{\Delta u \to 0} Pr\{u \le X_i(t_k) < u + \Delta u, \delta_i(t_k) = 1 | X_i(t_k) \ge u\} Pr\{X_i(t_k) \ge u\} I(v = u + t_k - t_l) \\
&= \lambda(t_k, u) Pr\{X_i(t_k) \ge u\} I(v = u + t_k - t_l) du,
\end{aligned}
$$

where $du/\Delta u = 1 + o(du)$.



(a) $t_k \le t_l$ (b) $t_k > t_l$

Figure A.1: Visual of relationship between follow-up intervals beginning at t_k and t_l for two possible cases.

There are two cases we need to consider when calculating $E\{Y_i(t_k, u)dN_i(t_l, v)\}$: $t_k \le t_l$ and $t_k > t_l$. The first case is illustrated in Figure A.1(a). If we observe a failure at time $v = X_i(t_l)$ for the observation window beginning at t_l, the subject can only be at risk up to time $v + t_l - t_k$ in the observation window beginning at t_k. Therefore the expectation is only non-zero when $u \le v + t_l - t_k$. The second case is illustrated in Figure A.1(b). The expectation is only non-zero when $u \le v - (t_k - t_l)$. When the failure occurs before calendar time t_k, by definition $X_i(t_k) = 0$. Therefore the expectation is also non-zero when $u = 0$ and $v < t_k - t_l$. Note that if $t_k \le t_l$, the

condition $u = 0$ and $v < t_k - t_l$ is never satisfied. Hence,

$$
\begin{aligned}
&E\{Y_i(t_k, u)dN_i(t_l, v)\} \\
&= \lim_{\Delta v \to 0} Pr\{X_i(t_k) \geq u, v \leq X_i(t_l) < v + \Delta v, \delta_i(t_l) = 1\} \\
&= \lim_{\Delta v \to 0} Pr\{v \leq X_i(t_l) < v + \Delta v, \delta_i(t_l) = 1\}\Big\{I(u \leq v + t_l - t_k) + I(u = 0)I(v < t_k - t_l)\Big\} \\
&= \lambda(t_l, v)Pr\{X_i(t_l) \geq v\}\Big\{I(u \leq v + t_l - t_k) + I(u = 0)I(v < t_k - t_l)\Big\}dv.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&E\{Y_i(t_l, v)dN_i(t_k, u)\} \\
&= \lambda(t_k, u)Pr\{X_i(t_k) \geq u\}\Big\{I(v \leq u + t_k - t_l) + I(v = 0)I(u < t_l - t_k)\Big\}du.
\end{aligned}
$$

Lastly, $E\{Y_i(t_k, u)Y_i(t_l, v)\} = Pr\{X_i(t_k) \geq u, X_i(t_l) \geq v\}$. Substituting these results into equation (A.8) gives us σ^2.

A.4 Standard probability calculations for σ^2 in special case

We provide the details of the calculation of the closed form asymptotic variance, σ^2, in the special case where the failure time, T_i, follows an exponential distribution with hazard λ. The censoring time, C_i, is independently sampled from a Uniform$[A - A^*, A]$ distribution, where A is the length of the study with accrual time A^*. The closed form asymptotic variance for this special case is required in Section 2.5 on page 15.

For $u, v \in (0, \tau]$, we have the following result:

$$
\begin{aligned}
& Pr\{X_i(t_k) \geq u\} \\
=& Pr\{\min(T_i - t_k, C_i - t_k) I(T_i \geq t_k, C_i \geq t_k) \geq u\} \\
=& E[I\{\min(T_i - t_k, C_i - t_k) I(T_i \geq t_k, C_i \geq t_k) \geq u\}] \\
=& E[I\{\min(T_i - t_k, C_i - t_k) \geq u\} I(T_i \geq t_k) I(C_i \geq t_k)] \\
=& E\{I(T_i - t_k \geq u) I(C_i - t_k \geq u) I(T_i \geq t_k) I(C_i \geq t_k)\} \\
=& E\{I(T_i - t_k \geq u) I(T_i \geq t_k)\} E\{I(C_i - t_k \geq u) I(C_i \geq t_k)\} \\
=& Pr(T_i \geq u + t_k) Pr(C_i \geq u + t_k) \\
=& \exp(-\lambda(u + t_k)) \frac{A - \max(A - A^*, u + t_k)}{A^*}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& Pr(X_i(t_k) \geq u, X_i(t_l) \geq v) \\
=& Pr\{\min(T_i - t_k, C_i - t_k) I(T_i \geq t_k, C_i \geq t_k) \geq u, \\
& \min(T_i - t_l, C_i - t_l) I(T_i \geq t_l, C_i \geq t_l) \geq v\} \\
=& E\{I(T_i - t_k \geq u) I(T_i \geq t_k) I(T_i - t_l \geq v) I(T_i \geq t_l)\} \\
& * E\{I(C_i - t_k \geq u) I(C_i \geq t_k) I(C_i - t_l \geq v) I(C_i \geq t_l)\} \\
=& Pr(T_i \geq \max(u + t_k, v + t_l)) Pr(C_i \geq \max(u + t_k, v + t_l)) \\
=& \exp\left(-\lambda \left(\max\left(u + t_k, v + t_l\right)\right)\right) \frac{A - \max(A - A^*, u + t_k, v + t_l)}{A^*}.
\end{aligned}
$$

Since we assume exponential failure times, it is expected that we have constant

hazards in each follow-up window and we show this here.

$$
\begin{aligned}
&\lambda(t_k, u) \\
=&\lim_{\Delta u \to 0} \frac{Pr\{u \le X_i(t_k) < u + \Delta u, \delta_i(t_k) = 1 | X_i(t_k) \ge u\}}{\Delta u} \\
=&\lim_{\Delta u \to 0} \frac{Pr\{u \le X_i(t_k) < u + \Delta u, \delta_i(t_k) = 1\}}{\Delta u Pr\{X_i(t_k) \ge u\}} \\
=&\lim_{\Delta u \to 0} \frac{Pr(u \le T_i - t_k < u + \Delta u) Pr(C_i \ge u + t_k)}{\Delta u Pr(L_i \ge u + t_k) Pr(C_i \ge u + t_k)} \\
=&\lim_{\Delta u \to 0} \frac{\int_{u+t_k}^{u+\Delta u+t_k} \lambda \exp(-\lambda x) dx}{\Delta u \exp\{-\lambda(u + t_k)\}} \\
=&\lim_{\Delta u \to 0} \frac{\exp\{-\lambda(u + t_k)\}\{1 - \exp(-\lambda \Delta u)\}}{\Delta u \exp\{-\lambda(u + t_k)\})} \\
=&\lim_{\Delta u \to 0} \frac{1 - \exp(-\lambda \Delta u)}{\Delta u} \\
=&\lim_{\Delta u \to 0} \frac{\lambda \Delta u - \frac{\lambda^2 (\Delta u)^2}{2!} + \frac{\lambda^3 (\Delta u)^3}{3!} + \dots}{\Delta u} \\
=&\lim_{\Delta u \to 0} \lambda - \frac{\lambda^2 \Delta u}{2!} + \frac{\lambda^3 (\Delta u)^2}{3!} + \dots \\
=&\lambda.
\end{aligned}
$$

Using the above results we then have

$$
\begin{aligned}
&\lambda^W(u) \\
=&\sum_{k=1}^{b} \lambda(t_k, u) \frac{Pr(X_i(t_k) \ge u)}{\sum_{l=1}^{b} Pr(X_i(t_l) \ge u)} \\
=&\sum_{k=1}^{b} \lambda \frac{Pr(X_i(t_k) \ge u)}{\sum_{l=1}^{b} Pr(X_i(t_l) \ge u)} \\
=&\lambda.
\end{aligned}
$$

A.5 Alternative variance estimates

The asymptotic variance of the estimate of $E\{\min(T,\tau)\} = \int_0^\tau \exp\{-\hat{\Lambda}(s)\}ds$, where $\hat{\Lambda}(s)$ is the Nelson-Aalen estimate of the cumulative hazard, is evaluated using the delta method.

$$\begin{aligned} & Var\left[\int_0^\tau \exp\{-\hat{\Lambda}(s)\}ds\right] \\ =& \int_0^\tau \int_0^\tau cov\left[\exp\{-\hat{\Lambda}(s)\}, \exp\{-\hat{\Lambda}(s')\}\right] ds\, ds' \\ =& \int_0^\tau \int_0^\tau \exp\{-\hat{\Lambda}(s)\}\exp\{-\hat{\Lambda}(s')\} Var\left[\hat{\Lambda}\{\min(s,s')\}\right] ds\, ds'. \end{aligned}$$

The estimate of the variance of $\hat{\Lambda}(s)$, assuming independence, is given by $\int_0^s dN(u)/Y(u)^2$.

The robust sandwich estimate of the variance of the cumulative hazard $\hat{\Lambda}(s)$ is given by

$$\sum_{i=1}^n \left\{\int_0^s \frac{dN_i(u)Y(u) - Y_i(u)dN(u)}{Y(u)^2}\right\}^2.$$

This variance estimate is based on fitting an independence working Cox proportional hazards model with covariate $W_i(t_k) = 1$ for $i = 1,\ldots,n$, $k = 1,\ldots,b$ and adjusting the covariance matrix for the association between event times within an individual Lin & Wei (1989).

APPENDIX B

Supplementary Materials for Chapter III

B.1 Asymptotic Distribution of Overall τ-Restricted Mean Survival

Suppose that in the dataset of combined follow-up windows we observe M events at internal times $\{0 = \mathscr{T}_0 < \mathscr{T}_1 < \ldots < \mathscr{T}_M < \mathscr{T}_{M+1} = \tau\}$, where times to event from the same individual are correlated. Let $F_j\{dN(\mathscr{T}_j), Y(\mathscr{T}_j)\} = dN(\mathscr{T}_j)/Y(\mathscr{T}_j)$. For convenience, we submerge arguments of F_j and define $G_m(F_0, F_1, \ldots, F_m) = \exp(-\sum_{j=0}^{m} F_j)$. We can then rewrite $\hat{\mu}^*(\tau)$ as $\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)G_m(F_0, F_1, \ldots, F_m)$.

We then linearize $G_m(F_0, F_1, \ldots, F_m), m = 0, \ldots, M$, via a Taylor series expansion

about $\lambda^W(\mathscr{T}_j)d\mathscr{T}_j$, $j = 0, 1, \ldots, m$, so that

$$\begin{aligned}
&\sqrt{n}\{\hat{\mu}^*(\tau) - \mu^*(\tau)\} \\
=&\sqrt{n}\left[\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\} - \mu^*(\tau)\right] && \text{(B.1)}\\
&+\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}(-1)\left[\sum_{j=0}^{m}\{F_j - \lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}\right] && \text{(B.2)}\\
&+\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\frac{1}{2!}\left[\sum_{j=0}^{m}\{F_j - \lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}\right]^2 && \text{(B.3)}\\
&+\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\{\text{ higher order terms}\}\,. && \text{(B.4)}
\end{aligned}$$

For $s = \mathscr{T}_m$, $\sum_{j=0}^{m}\lambda^W(\mathscr{T}_j)d\mathscr{T}_j$ is a non-stochastic quantity that converges to $\int_0^s \lambda^W(u)du$ as $n \to \infty$. Hence, term (B.1) converges to 0 as $n \to \infty$. And terms (B.3) and (B.4) converge in probability to zero if $\sum_{j=0}^{m}\{F_j - \lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}$, asymptotically equivalent to $\int_0^s \left\{dN(u)/Y(u) - \lambda^W(u)du\right\}$, converges in probability to zero and $\sqrt{n}[\int_0^s\{dN(u)/Y(u) - \lambda^W(u)du\}]$ converges to a distribution with finite variance.

By Glivenko-Cantelli and continuous mapping theorems,

$$\delta_1 = \sup_{u\in[0,\tau]}\left|\frac{Y(u)}{n} - \sum_{k=1}^{b} Pr\{X_i(t_k) \geq u\}\right| \xrightarrow{p} 0 \qquad \text{(B.5)}$$

$$\text{and } \delta_2 = \sup_{u\in[0,\tau]}\left|\left\{\frac{Y(u)}{n}\right\}^{-1} - \left[\sum_{k=1}^{b} Pr\{X_i(t_k) \geq u\}\right]^{-1}\right| \xrightarrow{p} 0. \qquad \text{(B.6)}$$

To verify that the asymptotic variance of $\sqrt{n}[\int_0^s\{dN(u)/Y(u) - \lambda^W(u)du\}]$ is fi-

nite, we rewrite the expression as

$$\begin{aligned}
&\sqrt{n}\int_0^s \left[\frac{dN(u)}{n}\left\{\frac{n}{Y(u)}+\frac{1}{\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}}-\frac{1}{\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}}\right\}\right. \\
&\qquad\qquad \left.-\lambda^W(u)du\right] \\
=&\sqrt{n}\sum_{k=1}^b\int_0^s \frac{dN(t_k,u)-n\lambda(t_k,u)Pr\{X_i(t_k)\geq u\}du}{n\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}} \\
&+\sqrt{n}\sum_{k=1}^b\int_0^s \frac{dN(t_k,u)}{n}\left(\left\{\frac{Y(u)}{n}\right\}^{-1}-\left[\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}\right]^{-1}\right) \\
\leq&\sqrt{n}\sum_{k=1}^b\int_0^s \frac{dN(t_k,u)-n\lambda(t_k,u)Pr\{X_i(t_k)\geq u\}du}{n\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}}+\delta_2\sqrt{n}\sum_{i=1}^n\int_0^s\frac{dN_i(u)}{n} \qquad \text{(B.7)}
\end{aligned}$$

The left term in (B.7) is the sum of b mean zero stochastic integrals of martingales, each with finite asymptotic variance $\int_0^s \lambda(t_k,u)Pr\{X_i(t_k)\geq u\}/\left[\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}\right]^2 du$ via the martingale central limit theorem. The asymptotic covariance of any two of the b stochastic integrals with windows at t_k and t_l, $k\neq l$, is

$$\begin{aligned}
&\int_0^s\int_0^s \frac{1}{[\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq u\}][\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq v\}]}\times \\
&\left[\lim_{\Delta u\to 0,\Delta v\to 0} Pr\{u\leq X_i(t_k)<u+\Delta u,\delta_i(t_k)=1, v\leq X_i(t_l)<v+\Delta v,\delta_i(t_l)=1\}du\,dv\right. \\
&\left.-\lambda(t_k,u)Pr\{X_i(t_k)\geq u\}\lambda(t_l,v)Pr\{X_i(t_l)\geq v\}du\,dv\right],
\end{aligned}$$

which is also finite.

For the right hand term in (B.7), applying the central limit theorem to $\sqrt{n}\sum_{i=1}^n\int_0^s dNi(u)/n$ gives convergence to a normal distribution with finite variance. Applying Slutsky's theorem and (B.6), the right hand term in (B.7) $\xrightarrow{p}0$.

Since $\sqrt{n}[\int_0^s\{dN(u)/Y(u)-\lambda^W(u)du\}]$ converges to a distribution with finite variance,

$\int_0^s \left\{ dN(u)/Y(u) - \lambda^W(u)du \right\}$ converges in probability to zero as $n \to \infty$. This completes the argument that terms (B.3) and (B.4) converge in probability to zero, leaving us with term (B.2).

Therefore $\sqrt{n}\{\hat{\mu}^*(\tau) - \mu^*(\tau)\}$ is asymptotically equivalent in distribution to

$$\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}(-1)\left[\sum_{j=0}^{m}\{F_j - \lambda^W(\mathscr{T}_j)d\mathscr{T}_j\}\right].$$

A Taylor series expansion of $F_j\{dN(\mathscr{T}_j), Y(\mathscr{T}_j)\}$ about the expected values of $dN(\mathscr{T}_j)$ and $Y(\mathscr{T}_j)$, used to make the non-linear F_j terms more tractable for understanding the asymptotic distribution of $\hat{\mu}^*(\tau)$, gives

$$\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\sum_{j=0}^{m}\left[\lambda^W(\mathscr{T}_j)d\mathscr{T}_j - \frac{E\{dN(\mathscr{T}_j)\}}{E\{Y(\mathscr{T}_j)\}}\right] \tag{B.8}$$

$$+\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}(-1)\sum_{j=0}^{m}\left[\frac{dN(\mathscr{T}_j) - \frac{E\{dN(\mathscr{T}_j)\}}{E\{Y(\mathscr{T}_j)\}}Y(\mathscr{T}_j)}{E\{Y(\mathscr{T}_j)\}}\right] \tag{B.9}$$

$$+\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\Bigg($$

$$\sqrt{n}\sum_{j=0}^{m}\frac{[Y(\mathscr{T}_j) - E\{Y(\mathscr{T}_j)\}][dN(\mathscr{T}_j) - E\{dN(\mathscr{T}_j)\}]}{E\{Y(\mathscr{T}_j)\}^2} \tag{B.10}$$

$$-\sqrt{n}\sum_{j=0}^{m}\frac{E\{dN(\mathscr{T}_j)\}}{E\{Y(\mathscr{T}_j)\}^3}[Y(\mathscr{T}_j) - E\{Y(\mathscr{T}_j)\}]^2\Bigg) \tag{B.11}$$

$$+\sqrt{n}\sum_{m=0}^{M}(\mathscr{T}_{m+1} - \mathscr{T}_m)\exp\left\{-\sum_{j'=0}^{m}\lambda^W(\mathscr{T}_{j'})d\mathscr{T}_{j'}\right\}\{\text{ higher order terms}\}. \tag{B.12}$$

Term (B.8) is equal to 0 since $E\{dN(\mathscr{T}_j)\} = n\sum_{k=1}^{b}\lambda(t_k, \mathscr{T}_j)Pr\{X_i(t_k) \geq \mathscr{T}_j\}d\mathscr{T}_j$ and $E\{Y(\mathscr{T}_j)\} = n\sum_{k=1}^{b}Pr\{X_i(t_k) \geq \mathscr{T}_j\}$ giving $E\{dN(\mathscr{T}_j)\}/E\{Y(\mathscr{T}_j)\} = \lambda^W(\mathscr{T}_j)d\mathscr{T}_j$. Also, after plugging in these expectations, terms (B.10) through (B.12) can be shown

to $\xrightarrow{p} 0$ as follows.

Term (B.10) is asymptotically equivalent to

$$\sqrt{n}\int_0^s \frac{[Y(u)-n\sum_{k=1}^b Pr\{X_i(t_k)\geq u\}][dN(u)-n\sum_{k=1}^b \lambda(t_k,u)Pr\{X_i(t_k)\geq u\}du]}{n^2\left[\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}\right]^2}$$
$$\leq \delta_1 \sum_{k=1}^b \sqrt{n}\int_0^s \frac{dN(t_k,u)-n\lambda(t_k,u)Pr\{X_i(t_k)\geq u\}du}{n\left[\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}\right]^2}.$$

By the martingale central limit theorem, each of the b terms converge to a mean zero normal process with finite variance $\int_0^s \lambda(t_k,u)Pr\{X_i(t_k)\geq u\}/\left[\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}\right]^4 du$. The finite asymptotic covariance between any two of the b stochastic integrals with follow-up windows starting at t_k and t_l, $k\neq l$, is

$$\int_0^s\int_0^s \frac{1}{[\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq u\}]^2[\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq v\}]^2}\times$$
$$\Big[\lim_{\Delta u\to 0,\Delta v\to 0} Pr\{u\leq X_i(t_k)<u+\Delta u,\delta_i(t_k)=1,v\leq X_i(t_l)<v+\Delta v,\delta_i(t_l)=1\}du\,dv$$
$$-\lambda(t_k,u)Pr\{X_i(t_k)\geq u\}\lambda(t_l,v)Pr\{X_i(t_l)\geq v\}du\,dv\Big].$$

Hence, applying (B.5) and Slutsky's theorem, term (B.10) $\xrightarrow{p} 0$.

Term (B.11) is asymptotically equivalent to

$$\sqrt{n}\int_0^s [Y(u)-n\sum_{k=1}^b Pr\{X_i(t_k)\geq u\}]^2 \frac{\sum_{l=1}^b \lambda(t_l,u)Pr\{X_i(t_l)\geq u\}}{n^2\left[\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}\right]^3}du$$
$$\leq \delta_1 \sum_{k=1}^b \sqrt{n}\sum_{i=1}^n \int_0^s \left[\frac{Y_i(t_k,u)-Pr\{X_i(t_k)\geq u\}}{n}\right] \frac{\sum_{l=1}^b \lambda(t_l,u)Pr\{X_i(t_l)\geq u\}}{\left[\sum_{l=1}^b Pr\{X_i(t_l)\geq u\}\right]^3}du.$$

By the central limit theorem, each of the b summands converge in distribution to a

normal with finite asymptotic covariance between any two summands equal to

$$\int_0^s \int_0^s \Big[Pr\{X_i(t_k) \geq u, X_i(t_l) \geq v\} - Pr\{X_i(t_k) \geq u\} Pr\{X_i(t_l) \geq v\} \Big] \\ * \frac{\sum_{l'=1}^b \lambda(t_{l'}, u) Pr\{X_i(t_{l'}) \geq u\}}{\Big[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq u\}\Big]^3} \frac{\sum_{l'=1}^b \lambda(t_{l'}, v) Pr\{X_i(t_{l'}) \geq v\}}{\Big[\sum_{l'=1}^b Pr\{X_i(t_{l'}) \geq v\}\Big]^3} du\, dv.$$

Hence an application of Slutsky's theorem and (B.5) have term (B.11) $\xrightarrow{p} 0$. Similar arguments applied to higher order terms show that (B.12) $\xrightarrow{p} 0$, leaving us with term (B.9). Therefore $\sqrt{n}\{\hat{\mu}^*(\tau) - \mu^*(\tau)\}$ is asymptotically equivalent in distribution to

$$\begin{aligned} &\sqrt{n} \int_0^\tau \exp\left\{ - \int_0^s \lambda^W(u) du \right\} (-1) \int_0^s \left[\frac{dN(u) - \lambda^W(u) Y(u) du}{n \sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} \right] ds \\ =&\sqrt{n} \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^b (-1) \int_0^\tau \exp\left\{ - \int_0^s \lambda^W(u) du \right\} \int_0^s \left[\frac{dN_i(t_k, u) - \lambda^W(u) Y_i(t_k, u) du}{\sum_{l=1}^b Pr\{X_i(t_l) \geq u\}} \right] ds \\ =&\sqrt{n} \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^b - Z_{ik}\{\hat{\mu}^*(\tau)\}, \end{aligned}$$

where $Z_{ik}\{\hat{\mu}^*(\tau)\} = \int_0^\tau \exp\left\{ - \int_0^s \lambda^W(u) du \right\} \int_0^s \left[dN_i(t_k, u) - \lambda^W(u) Y_i(t_k, u) du \right] / \Big[\sum_{l=1}^b Pr\{X_i(t_l) \geq u\} \Big] ds$. Hence, the asymptotic distribution of $\sqrt{n}\{\hat{\mu}^*(\tau) - \mu^*(\tau)\}$ boils down to application of the central limit theorem acting on mean zero independent and identically distributed random variables $Z_i\{\hat{\mu}^*(\tau)\} = \sum_{k=1}^b - Z_{ik}\{\hat{\mu}^*(\tau)\}$, $i = 1, \ldots, n$. We define $\bar{Z}\{\hat{\mu}^*(\tau)\} = \sum_{i=1}^n Z_i\{\hat{\mu}^*(\tau)\}/n$. By the central limit theorem $\sqrt{n}[\bar{Z}\{\hat{\mu}^*(\tau)\} - 0]$ has a limiting mean 0 normal distribution with finite variance $\sigma_*^2 = Var[Z_i\{\hat{\mu}^*(\tau)\}]$.

Empirical variance estimates of σ_*^2 are given by $\hat{\sigma}_*^2 = \sum_{i=1}^n \left[z_i\{\hat{\mu}^*(\tau)\} - \bar{z}\{\hat{\mu}^*(\tau)\} \right]^2 / (n-$

1) where $\bar{z}\{\hat{\mu}^*(\tau)\} = \sum_{i=1}^{n} z_i\{\hat{\mu}^*(\tau)\}/n$, $z_i\{\hat{\mu}^*(\tau)\} = \sum_{k=1}^{b} z_{ik}\{\hat{\mu}^*(\tau)\}$ and

$$z_{ik}\{\hat{\mu}^*(\tau)\} = \int_0^\tau \exp\left\{-\int_0^s \frac{dN(u)}{Y(u)}\right\}\left\{\int_0^s \frac{dN_i(t_k,u) - \frac{dN(u)}{Y(u)}Y_i(t_k,u)}{Y(u)/n}\right\} ds.$$

B.2 Asymptotic Distribution of $\mathscr{T}_*$

Under $H_0 : \mu_1^*(\tau) = \mu_2^*(\tau)$,

$$\begin{aligned}\mathscr{T}_* =& \sqrt{\frac{n_1 n_2}{n_1+n_2}}\left\{\hat{\mu}_1^*(\tau) - \hat{\mu}_2^*(\tau)\right\} \\ =& \sqrt{\frac{n_2}{n_1+n_2}}\sqrt{n_1}\left\{\hat{\mu}_1^*(\tau) - \mu_1^*(\tau)\right\} - \sqrt{\frac{n_1}{n_1+n_2}}\sqrt{n_2}\left\{\hat{\mu}_2^*(\tau) - \mu_2^*(\tau)\right\}.\end{aligned}$$

In Appendix B.1 we show that $\sqrt{n_g}\{\hat{\mu}_g^*(\tau) - \mu_g^*(\tau)\}$, $g = 1, 2$, each converge in distribution to a normal with mean 0 and variance σ_{*g}^2. Slutsky's theorem applied to the above expression for $\mathscr{T}_*$ results in the desired mean zero limiting normal distribution with variance $\pi_2\sigma_{*1}^2 + \pi_1\sigma_{*2}^2$.

B.3 Asymptotic Distribution of the Area under $\mu(t,\tau)$ from t_1 to t_b

The integrated τ-RMRL estimator is defined to be

$$\hat{\mu}(\cdot,\tau) = \int_{t_1}^{t_b} \hat{\mu}(t_k,\tau)dt_k.$$

Since $\hat{\mu}(t_k,\tau)$ is a special case of $\hat{\mu}^*(\tau)$ for b=1, results from Appendix B.1 show that $\sqrt{n}\{\hat{\mu}(t_k,\tau) - \mu(t_k,\tau)\}$ is asymptotically equivalent in distribution to

$\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} -Z_{ik}\{\hat{\mu}(t_k,\tau)\}$ where

$$
\begin{aligned}
& Z_{ik}\{\hat{\mu}(t_k,\tau)\} \\
=& \int_0^{\tau} \exp\left\{-\int_0^{s} \lambda(t_k,u)du\right\} \int_0^{s} \left[\frac{dN_i(t_k,u) - \lambda(t_k,u)Y_i(t_k,u)du}{Pr\{X_i(t_k) \geq u\}}\right] ds.
\end{aligned}
$$

Incorporating these results into the integrated τ-RMRL estimator, $\sqrt{n}\{\hat{\mu}(\cdot,\tau) - \mu(\cdot,\tau)\}$ is asymptotically equivalent in distribution to

$$
\begin{aligned}
& \sqrt{n}\int_{t_1}^{t_b} \hat{\mu}(t_k,\tau) - \mu(t_k,\tau)dt_k \\
=& \int_{t_1}^{t_b} \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} -Z_{ik}\{\hat{\mu}(t_k,\tau)\}dt_k \\
=& \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} \int_{t_1}^{t_b} -Z_{ik}\{\hat{\mu}(t_k,\tau)\}dt_k \\
=& \sqrt{n}\frac{1}{n}\sum_{i=1}^{n} Z_i\{\hat{\mu}(\cdot,\tau)\},
\end{aligned}
$$

where $Z_i\{\hat{\mu}(\cdot,\tau)\} = \int_{t_1}^{t_b} -Z_{ik}\{\hat{\mu}(t_k,\tau)\}dt_k$ is a mean zero random variable. Let $\bar{Z}\{\hat{\mu}(\cdot,\tau)\} = \sum_{i=1}^{n} Z_i\{\hat{\mu}(\cdot,\tau)\}/n$. By the central limit theorem $\sqrt{n}[\bar{Z}\{\hat{\mu}(\cdot,\tau)\} - 0]$ has a limiting mean 0 normal distribution with finite variance $\sigma_R^2 = Var[Z_i\{\hat{\mu}(\cdot,\tau)\}]$.

Empirical variance estimates of σ_R^2 are given by $\hat{\sigma}_R^2 = \sum_{i=1}^{n} \left[z_i\{\hat{\mu}(\cdot,\tau)\} - \bar{z}\{\hat{\mu}(\cdot,\tau)\}\right]^2 /(n-1)$ where $\bar{z}\{\hat{\mu}(\cdot,\tau)\} = \sum_{i=1}^{n} z_i\{\hat{\mu}(\cdot,\tau)\}/n$ and

$$
z_i\{\hat{\mu}(\cdot,\tau)\} = \int_{t_1}^{t_b} \int_0^{\tau} \exp\left\{-\int_0^{s} \frac{dN(t_k,u)}{Y(t_k,u)}\right\} \left\{\int_0^{s} \frac{dN_i(t_k,u) - \frac{dN(t_k,u)}{Y(t_k,u)}Y_i(t_k,u)}{Y(t_k,u)/n}\right\} ds\, dt_k.
$$

B.4 Asymptotic Distribution of $\mathscr{T}_{\mathscr{R}}$

Under $H_0 : \mu_1(\cdot, \tau) = \mu_2(\cdot, \tau)$,

$$\begin{aligned}\mathscr{T}_{\mathscr{R}} =& \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left\{\hat{\mu}_1(\cdot, \tau) - \hat{\mu}_2(\cdot, \tau)\right\} \\ =& \sqrt{\frac{n_2}{n_1 + n_2}} \sqrt{n_1} \left\{\hat{\mu}_1(\cdot, \tau) - \mu_1(\cdot, \tau)\right\} - \sqrt{\frac{n_1}{n_1 + n_2}} \sqrt{n_2} \left\{\hat{\mu}_2(\cdot, \tau) - \mu_2(\cdot, \tau)\right\}.\end{aligned}$$

In Appendix B.3 we show that $\sqrt{n_g}\{\hat{\mu}_g(\cdot, \tau) - \mu_g(\cdot, \tau)\}$ each converge in distribution to a mean 0 normal with variance $\sigma_{R,g}$, $g = 1, 2$. By Slutsky's theorem we have that the expression above converges in distribution to a mean zero normal random variable with variance $\pi_2 \sigma^2_{R,1} + \pi_1 \sigma^2_{R,2}$.

B.5 Spacing of Follow-up Windows

The choice of $\{t_1, \ldots, t_b\}$ should be governed by potential to make efficiency gains in estimating $\mu^*(\tau)$. To gain intuition we consider the special case where a single event occurs for each patient. In this case, derivation of an asymptotic closed form variance of $\hat{\mu}^*(\tau)$ is tractable and efficiency of different spacings between $\{t_1, \ldots, t_b\}$

can be evaluated. In particular, the asymptotic variance of $\sqrt{n}\hat{\mu}^*(\tau)$ becomes

$$\sum_{k=1}^{b}\sum_{l=1}^{b}\int_0^\tau\int_0^\tau \exp\left\{-\int_0^s \lambda^W(u)du\right\}\exp\left\{-\int_0^{s'} \lambda^W(v)dv\right\}\times$$
$$\int_0^s\int_0^{s'} \frac{1}{\left[\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq u\}\right]\left[\sum_{l'=1}^b Pr\{X_i(t_{l'})\geq v\}\right]}\times$$
$$\Big[\lambda(t_k,u)Pr\{X_i(t_k)\geq u\}I(v=u+t_k-t_l)du$$
$$-\lambda^W(u)\lambda(t_l,v)Pr\{X_i(t_l)\geq v\}\{I(u\leq v+t_l-t_k)+I(u=0)I(v<t_k-t_l)\}du\,dv$$
$$-\lambda^W(v)\lambda(t_k,u)Pr\{X_i(t_k)\geq u\}\{I(v\leq u+t_k-t_l)+I(v=0)I(u<t_l-t_k)\}du\,dv$$
$$+\lambda^W(u)\lambda^W(v)Pr\{X_i(t_k)\geq u, X_i(t_l)\geq v\}du\,dv$$
$$-\{\lambda(t_k,u)-\lambda^W(u)\}\{\lambda(t_l,v)-\lambda^W(v)\}Pr\{X_i(t_k)\geq u\}Pr\{X_i(t_l)\geq v\}du\,dv\Big]ds\,ds'.$$

In the special case where the event time follows an exponential distribution and the censoring mechanism is uniform, $\lambda(t_k,u)=\lambda^W(u)=\lambda$. Also, standard probability calculations for $u,v\in(0,\tau]$ give $Pr\{X_i(t_k)\geq u\}=\exp\{-\lambda(u+t_k)\}\{A-\max(A-A^*,u+t_k)\}/A^*$ and $Pr\{X_i(t_k)\geq u, X_i(t_l)\geq v\}=\exp\{-\lambda\max(u+t_k,v+t_l)\}\{A-\max(A-A^*,u+t_k,v+t_l)\}/A^*$, where A is the length of the study with accrual time, A^*. In the results that follow we assume $n=100$, an exponential hazard corresponding to $\mu^*(12)=11$ months ($\tau=12$ months) and a uniform(24,36) administrative censoring mechanism.

Figure B.1 evaluates the potential for an overlapping follow-up window to contribute efficiency in estimation of $\hat{\mu}^*(\tau)$. Fixed (non-overlapping) follow-up windows at $t_1=0$ and $t_3=12$ are included in estimation and the vertical axis shows values of the closed form asymptotic variance as an additional follow-up window starting at $t_2\in(0,12)$ is used in estimation. We observe that an additional 12-month window starting at 6 months optimizes the available precision from including a single overlap-

ping follow-up window. Since the time-scale used in these calculations is arbitrary, Figure B.1 suggests that a window evenly spaced between two existing windows has the best potential to improve efficiency.

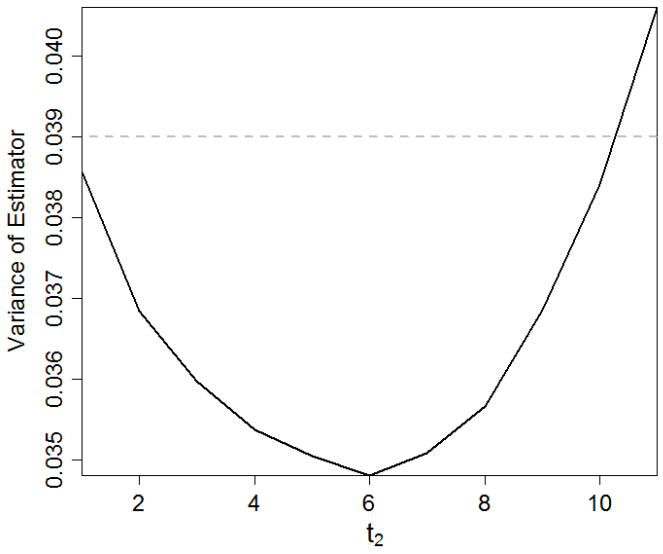


Figure B.1: Finite sample ($n = 100$) closed form asymptotic variance of $\hat{\mu}^*(\tau)$ assuming a single exponential event time for a 36 month study with an administrative censoring mechanism when τ=12 months. $t_1 = 0$, $t_3 = 12$ months and we vary t_2. Dashed line corresponds to variance of estimator constructed using two follow-up windows $t_1 = 0$ and $t_3 = 12$.

Table B.1 explores the potential for efficiency gain with additional equally spaced follow-up windows. The optimal variance identified in Figure B.1 is shown for the case with $t_b = 12$ with 3 windows; the dashed line in this figure corresponds to $t_b = 12$ with 2 windows. Asymptotic relative efficiencies relative to use of a single window at $t_b = 0$ are given in the final column. Results indicate (a) additional follow-up time, as indicated by increasing t_b, provides the highest improvement in efficiency and (b) introducing windows more frequently that at 6-month intervals results in diminishing returns. Gains provided by (a) may require additional resources in the implementation of a clinical trial, whereas available gains provided by (b) are cost-free.

Based on this special case, we recommend using intervals starting from $t_k = (k-1)\tau/2$ for $k = 1, \ldots, b$, where b is chosen so that the final follow-up interval starting at t_b does not exceed the study period. Simulation methods are recommended to explore the merits of more complex designs than given here.

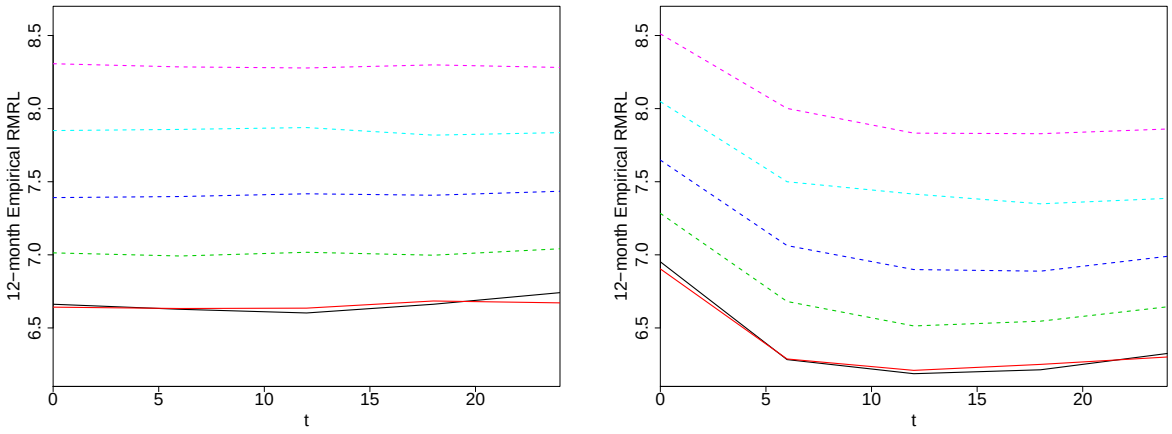
Table B.1: Study of follow-up window choices based on finite sample ($n = 100$) closed form variance (σ^2/n) and Asymptotic Relative Efficiency (ARE) for the special case were we assume a single exponential event time per patient in a 36 month study with an administrative censoring mechanism.

	Number of Windows	$\{t_1, \ldots, t_b\}$	σ^2/n	ARE
$t_b = 0$	1	0	0.071	1.00
$t_b = 12$	2	0, 12	0.039	1.82
	3	0, 6, 12	0.035	2.03
	5	0, 3, 6, 9, 12	0.035	2.03
$t_b = 24$	3	0, 12, 24	0.030	2.37
	5	0, 6, 12, 18, 24	0.026	2.73
	9	0, 3, 6, 9, 12, 15, 18, 21, 24	0.025	2.84

B.6 Simulation Study: Empirical Mean 12-month RMRL

The plot of the empirical mean 12-month RMRL for group 1 and for each value of alpha for group 2 is given in Figure B.2, B.3 and B.4 for scenarios of immediate treatment effect, delayed treatment effect and short duration treatment effect respectively.

In these plots we observe that assuming proportional hazard results in constant parallel trajectories for the τ-RMRL when the events are uncorrelated. When the events are correlated, we no longer have constant τ-RMRL functions, but rather the τ-RMRL are initially decreasing then constant. We were initially perplexed by why this would be the case. In order to better understand the dependent data structure we considered the simpler scenario where we only have recurrent events with no



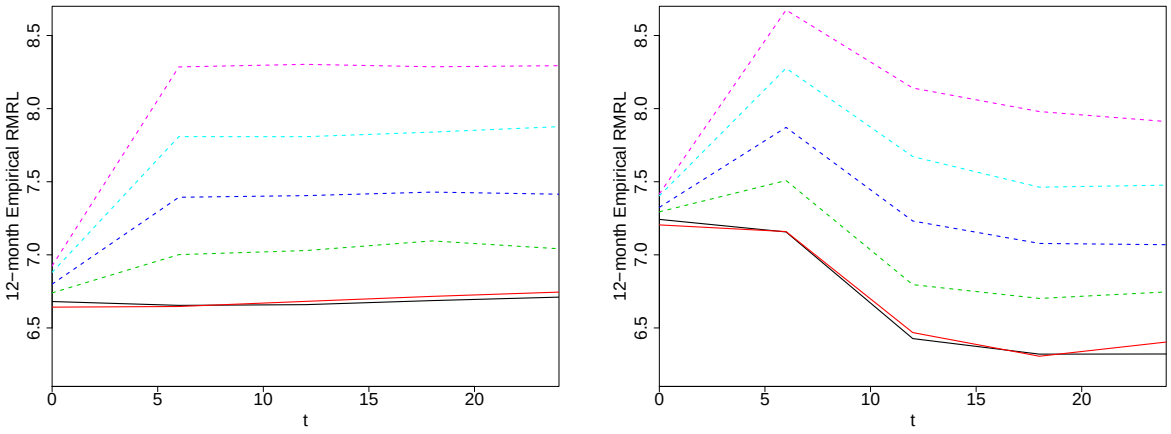
(a) Uncorrelated events

(b) Strongly correlated events

Figure B.2: Plot of the empirical mean 12-month RMRL for group 1 (black) and for each value of alpha for group 2 ($\alpha = 1$: red, $\alpha = 0.9$: green, $\alpha = 0.8$: dark blue, $\alpha = 0.7$: light blue, $\alpha = 0.6$: pink) when we assume proportional hazards.

terminating events. In this case the simulations still showed decreasing τ-RMRL functions with greater slopes for stronger positive correlations. To explain the τ-RMRL function at 6 months, we need to understand the distribution of the next event from 6 months. T_1 and T_2 are the times to recurrent event from time 0. They are defined as $T_1 = S_1$ and $T_2 = S_1 + S_2$ in terms of gap times S_1 and S_2. In our simulations, we assume gap times are exponentially distributed with parameter λ_S.

If $T_1 > 6$ then the next event at 6 months will follow the same exponential distribution as time 0, i.e $\sim$ Exponential(λ_S). If $T_1 < 6$ and $T_2 > 6$ then the next event from 6 months $\sim T_2|T_1 < 6, T_2 > 6$. The survival function of this distribution is $Pr(T_2 > t, T_1 < 6)/Pr(T_2 > 6, T_1 < 6)$ for $t > 6$. The probability $Pr(T_2 > t, T_1 < 6)$



(a) Uncorrelated events

(b) Strongly correlated events

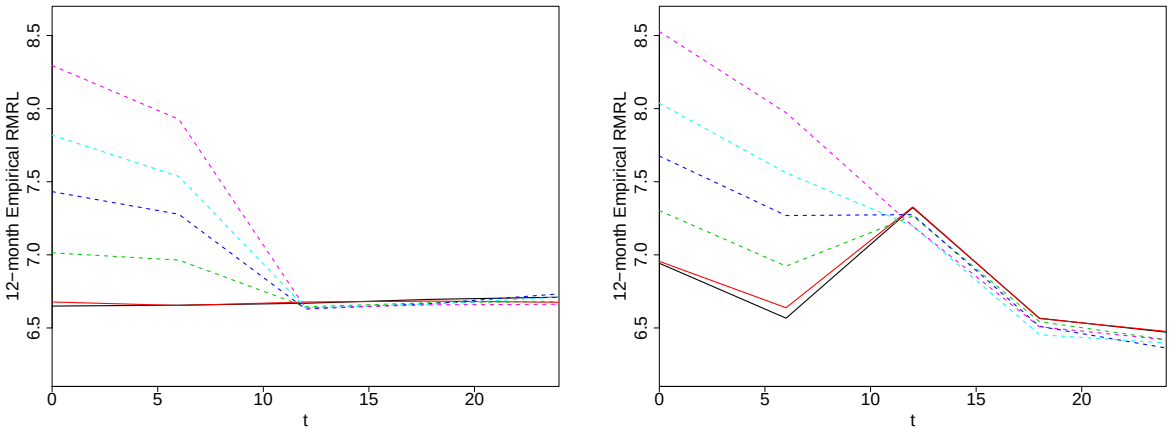
Figure B.3: Plot of the empirical 12-month RMRL for group 1 (black) and for each value of alpha for group 2 ($\alpha = 1$: red, $\alpha = 0.9$: green, $\alpha = 0.8$: dark blue, $\alpha = 0.7$: light blue, $\alpha = 0.6$: pink) when we assume delayed proportional treatment effect.

can be written in terms of the random variables S_1 and S_2,

$$\begin{aligned} & Pr(T_2 > t, T_1 < 6) \\ =& Pr(S_1 + S_2 > t, S_1 < 6) \\ =& \int_0^6 Pr(S_2 > t - s | S_1 = s) f_{S_1}(s) ds. \end{aligned}$$

If S_1 and S_2 are independent, the survival function of $T_2|T_1 < 6, T_2 > 6$ is $\exp\{-\lambda_S(t-6)\}$ for $t > 6$, i.e $\sim$ Exponential(λ_S). If S_1 and S_2 are positively correlated $Pr(S_2 > t - s|S_1 = s) < Pr(S_2 > t - s)$, where $s \in [0, 6)$ and $t > 6$, since S_1 is small and therefore S_2 is then more likely to be small. We can define $k(t)$ to be a function such that

$$\begin{aligned} & Pr(T_2 > t, T_1 < 6) \\ =& k(t) 6 \lambda_S \exp(-\lambda_S t), \end{aligned}$$



(a) Uncorrelated events

(b) Strongly correlated events

Figure B.4: Plot of the empirical 12-month RMRL for group 1 (black) and for each value of alpha for group 2 ($\alpha = 1$: red, $\alpha = 0.9$: green, $\alpha = 0.8$: dark blue, $\alpha = 0.7$: light blue, $\alpha = 0.6$: pink) when we assume short duration treatment effect.

where $0 < k(t) < 1$ for $t > 6$ and is a decreasing function of t. Intuitively $k(t)$ must be a decreasing function because when S_1 and S_2 are positively correlated the conditions on S_2 are more restrictive and it is less likely that $S_2 > t - s$ for any $s \in [0, 6)$. Therefore the survival function of $T_2|T_1 < 6, T_2 > 6$ is $k(t)\exp\{-\lambda_S(t-6)\}/k(6)$ where $k(t)/k(6) < 1$ for $t > 6$. Therefore the area under the survival function up to τ is smaller for positively correlated recurrent events.

APPENDIX C

Supplementary Materials for Chapter IV

C.1 Patient specific hazards

To simulate the failure time T_i for $i = 1, \ldots, n$, we require patient specific hazards for the survival function $P\{T_i - t_j > u | T_i > t_j, \bar{Z}_{1i}(u+t_j), Z_{2i}\} = \exp\{-\int_{t_j}^{t_j+u}[\lambda_{i1} I(0 \leq v \leq 6) + \lambda_{i2} I(6 < v \leq 12) + \lambda_{i3} I(12 < v < \infty)]dv\}$, that obey the restricted mean model $E(\log[\min\{T_i(t_j), \tau\}]) = \beta_0 + \beta_1 Z_{1i}(t_j) + \beta_2 Z_{2i}$.

Therefore we have

$$
\begin{aligned}
&\beta_0 + \beta_1 Z_{1i}(t_j) + \beta_2 Z_{2i} \\
&= -\int_0^\tau \log(u) dP\{T_i - t_j > u | T_i > t_j, \bar{Z}_1(u + t_j), Z_2\} \\
&+ \log(\tau) P\{T_i - t_j > \tau | T_i > t_j, \bar{Z}_1(\tau + t_j), Z_2\},
\end{aligned}
$$

for each window t_j, $j = 1, 2, 3$. In this case we have three nonlinear equations with three unknown parameters that we need to solve for: λ_{i1}, λ_{i2} and λ_{i3}. In each case the solution is obtained using numerical algorithms for nonlinear equations.

Starting with $j = 3$, we have

$$\begin{aligned}
&\beta_0 + \beta_1 Z_{1i}(12) + \beta_2 Z_{2i} \\
&= \int_0^{12} \log(u)\lambda_{i3} \exp\{-u\lambda_{i3}\} du + \log(12) \exp\{-12\lambda_{i3}\},
\end{aligned}$$

where λ_{i3} is the only unknown quantity.

For $j = 2$, we have

$$\begin{aligned}
&\beta_0 + \beta_1 Z_{1i}(6) + \beta_2 Z_{2i} \\
&= \int_0^{6} \log(u)\lambda_{i2} \exp\{-u\lambda_{i2}\} du + \int_6^{12} \log(u)\lambda_{i3} \exp\{-6\lambda_{i2} - u\lambda_{i3}\} \\
&+ \log(12) \exp\{-6\lambda_{i2} - 6\lambda_{i3}\},
\end{aligned}$$

where λ_{i2} is the only unknown quantity.

For $j = 1$, we have

$$\begin{aligned}
&\beta_0 + \beta_1 Z_{1i}(0) + \beta_2 Z_{2i} \\
&= \int_0^{6} \log(u)\lambda_{i1} \exp\{-u\lambda_{i1}\} du \\
&+ \int_6^{12} \log(u)\lambda_{i2} \exp\{-6\lambda_{i1} - u\lambda_{i2}\} + \log(12) \exp\{-6\lambda_{i1} - 6\lambda_{i2}\},
\end{aligned}$$

where λ_{i1} is the only unknown quantity.

BIBLIOGRAPHY

BIBLIOGRAPHY

Albert, R. K.; Connett, J.; Bailey, W. C.; Casaburi, R.; Cooper, J. A. D.; Criner, G. J.; Curtis, J. L.; Dransfield, M. T.; Han, M. K.; Lazarus, S. C.; Make, B.; Marchetti, N.; Martinez, F. J.; Madinger, N. E.; McEvoy, C.; Niewoehner, D. E.; Porsasz, J.; Price, C. S.; Reilly, J.; Scanlon, P. D.; Sciurba, F. C.; Scharf, S. M.; Washko, G. R.; Woodruff, P. G. & Anthonisen, N. R. (2011): Azithromycin for Prevention of Exacerbations of COPD. *New England Journal of Medicine* **365**(8):689–698. PMID: 21864166.

Andersen, P.; Hansen, M. & Klein, J. (2004): Regression Analysis of Restricted Mean Survival Time Based on Pseudo-Observations. *Lifetime Data Analysis* **10**(4):335–350.

Andrei, A.-C. & Murray, S. (2007): Regression models for the mean of the quality-of-life-adjusted restricted survival time using pseudo-observations. *Biometrics* **63**:398–404.

Egan, T. M.; Murray, S.; Bustami, R. T.; Shearon, T. H.; McCullogh, K. P.; Edwards, L. B.; Coke, M. A.; Garrity, E. R.; Sweet, S. C.; Heiney, D. A. & Grover, F. L. (2006): Development of the New Lung Allocation System in the United States. *American Journal of Transplantation* **6**(2):1212–1227.

ETDRS Research Group (1991a): Early phtocoagulation for diabetic retinopathy: ETDRS report numer 9. *Opthalmology* **98**:766–785.

ETDRS Research Group (1991b): Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics: ETDRS report numer 7. *Opthalmology* **98**:741–756.

Faucett, C. L.; Schenker, N. & Taylor, J. M. G. (2002): Survival Analysis Using Auxiliary Variables Via Multiple Imputation, with Application to AIDS Clinical Trial Data. *Biometrics* **58**(1):pp. 37–47.

Ghorai, J.; Susarla, A.; Susarla, V.; & Van-Ryzin, J. (1982): Nonparametric Estimation of Mean Residual Life Time with Censored Data. *Nonparametric Statistical Inference*, volume 1 of *Colloquia Mathematica Societatis, 32*, pages 269–291. North-Holland, Amsterdam-New York.

Ghorai, J. K. & Rejto, L. (1987): Estimation of mean residual life with censorded data under the proportional hazard model. *Communications in Statistics- Theory and Methods* **16**:2097–2114.

Ghosh, D. & Lin, D. Y. (2000): Nonparametric Analysis of Recurrent Events and Death. *Biometrics* **56**(2):554–562.

Gong, Q. & Schaubel, D. E. (2013): Partly Conditional Estimation of the Effect of

a Time-Dependent Factor in the Presence of Dependent Censoring. *Biometrics* **69**(2):338–347.

Gyrd-Hansen, D. & Sogaard, J. (1998): Discounting life years: Whither time preference? *Health Economics* **7**:121–127.

Hsu, C.-H.; Taylor, J. M. G.; Murray, S. & Commenges, D. (2006): Survival analysis using auxiliary variables via non-parametric multiple imputation. *Statistics in Medicine* **25**(20):3503–3517.

Irwin, J. O. (1949): The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *Journal of Hygiene* **47**:188–189.

Liang, K. Y. & Zeger, S. L. (1986): Longitudinal Data Analysis using Generalized Linear Models. *Biometrika* **73**:13–22.

Lin, D. J. & Wei, L. J. (1989): The Robust Inference for the Cox Proportional Hazards Model. *Journal of the American Statistical Association* **84**:1074–1078.

Lin, D. Y.; Fleming, T. R. & Wei, L. J. (1994): Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**(1):73–81.

Lin, D. Y.; Wei, L. J.; Yang, I. & Ying, Z. (2000): Semiparametric Regression for the Mean and Rate Functions of Recurrent Events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(4):711–730.

Little, R. J. A. & Rubin, D. (1986): *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc., New York, NY, USA.

Liu, L. X.; Murray, S. & Tsodikov, A. (2011): Multiple imputation based on restricted mean model for censored data. *Statistics in Medicine* **30**(12):1339–1350.

Murray, S. & Tsiatis, A. A. (1996): Nonparametric survival estimation using prognostic longitudinal covariates. *Biometrics* **52**(1):137–151.

Murray, S. & Tsiatis, A. A. (2001): Using auxiliary time-dependent covariates to recover information in nonparametric testing with censored data. *Lifetime Data Analysis* **7**(2):125–141.

Na, M. H. & Kim, J. J. (1999): On inference for mean residual life. *Communications in Statistics- Theory and Methods* **28**:2917–2833.

Raghu, G.; Collard, H. R.; Egan, J. J.; Martinez, F. J.; Behr, J.; Brown, K. K.; Colby, T. V.; Cordier, J. F.; Flaherty, K. R.; Lasky, J. A.; Lynch, D. A.; Ryu, J. H.; Swigris, J. J.; Wells, A. U.; Ancochea, J.; Bouros, D.; Carvalho, C.; Costabel, U.; Ebina, M.; Hansell, D. M.; Johkoh, T.; Kim, D. S.; King, T. E.; Kondoh, Y.; Myers, J.; Müller, N. L.; Nicholson, A. G.; Richeldi, L.; Selman, M.; Dudden, R. F.; Griss, B. S.; Protzko, S. L. & Schünemann, H. J. (2011): An Official ATS/ERS/JRS/ALAT Statement: Idiopathic Pulmonary Fibrosis: Evidence-based Guidelines for Diagnosis and Management. *American Journal of Respiratory and Critical Care Medicine* **183**:788–823.

Robins, J. M. (1993): Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, pages 24–33. Alexandria, Virginia.

Schmidt, S. L.; Tayob, N.; Han, M. K.; Zappala, C.; Kervitsky, D.; Murray, S.; Toews, G. B.; Wells, A. U.; Brown, K. K.; Martinez, F. J. & Flaherty, K. R. (2012): Idiopathic Pulmonary Fibrosis: Predicting Future Disease Course from Past Trends in Pulmonary Function. *Unpublished* .

Taylor, J. M. G.; Murray, S. & Hsu, C.-H. (2002): Survival estimation and testing via multiple imputation. *Statistics & Probability Letters* **58**(3):221 – 232.

Williams, R. L. (1995): Product-Limit Survival Funtions with Correlated Survival Times. *Lifetime Data Analysis* **1**:171–186.

Woodruff, D. (1971): A Simple Method for Approximating the Variance of a Complicated Estimate. *Journal of the American Statistical Association* **66**:411–414.

Xiang, F. & Murray, S. (2012): Restricted mean models for transplant benefit and urgency. *Statistics in Medicine* **31**(6):561–576.

Xiang, F.; Murray, S. & Liu, X. (2013): Analysis of Transplant Urgency and Benefit via Multiple Imputation. *Submitted* .

Yu, J. (2003): *Prosper function analysis for organ allocation: A counting process and martingale approach.* Ph.D. thesis, University of Michigan Department of Biostatistics, Ann Arbor, MI.