DEVELOPING PSEUDO-OBSERVATION AND MULTIPLE IMPUTATION APPROACHES FOR ANALYSIS OF DEPENDENTLY CENSORED SURVIVAL AND QUALITY-ADJUSTED SURVIVAL DATA

by

Fang Xiang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2012

Doctoral Committee:

Associate Professor Susan Murray, Chair
Associate Professor Douglas Schaubel
Associate Research Professor Cathie Spino
Assistant Professor Laura Rozek

To my parents and sister

ACKNOWLEDGEMENTS

I am heartily thankful to my advisor, Dr. Susan Murray, for her inspiration, guidance, supervision, and support throughout my dissertation work. To me, she is not only a supervisor, but also a friend. Her sense of humor have made my Ph.D. years very pleasant. Without her, this work would not have been possible. It is to her that I owe my deepest gratitude.

I would like to thank Dr. Cathie Spino and Dr. Michael Boehnke who have provided financial support to me throughout the years. Dr. Boehnke was also my advisor for two years when I first came to the department. Special thanks go to Dr. Spino, who has always given me generous help, profound understanding and warm encouragement. I have benefited much from her expertise in clinical trials.

My thanks go to Dr. Douglas Schaubel and Laura Rozek for serving on my committee and providing insightful comments and suggestions. It has been my pleasure to work with Dr. Schaubel. I am also grateful to Tempie Shearon, Kathryn Meyer, and Ying Qian, formerly of the SRTR for providing the valuable lung transplant data.

I must acknowledge as well the many friends, colleagues, and professors from the University of Michigan who have helped or accompanied me over the years.

Last, I would like to thank my parents and my sister, whose unconditional love and support have made possible every single accomplishment of mine.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

INTRODUCTION

Dependent censoring is a common issue in survival and quality-adjusted survival analysis. This thesis develops pseudo-observation and multiple imputation approaches for analysis of these types of data, and illustrates them in both simulations and data examples. Our motivating survival analysis example takes place in the lung allocation setting, where more urgent patients are removed for transplant. Our motivating quality-adjusted survival analysis example takes place when a quality-of-life function is applied to follow-up in breast cancer patients, inducing dependent censoring on the quality-of-life timescale.

In both examples we consider the restricted mean event time as the expected event time when restricted to τ, a time window of interest. When censoring is non-informative, the restricted mean can be estimated nonparametrically by the area under the Kaplan-Meier survival curve. In cases when additional covariate information is available, one can model the hazards using a Cox proportional hazards model, and estimate each individual's restricted mean by integrating their Cox survival curves. A more direct model links the restricted mean to a linear function of covariates. One advantage of this type of model is that needed assumptions are on the scale of the mean rather than the hazard of the distribution, which is what

we are accustomed to in non-censored regression settings. However, in both of our examples, dependent censoring needs to be addressed.

We consider various approaches for estimation of restricted means when censoring is dependent: (1) an inverse-weighted pseudo-observation (PO) approach described in Chapter II; (2) a multiple imputation approach for dependently censored outcomes, described in Chapter III for survival and in Chapter IV for quality-adjusted survival; (3) integration of inverse-weighted Cox survival curves, where this last existing approach provides a basis of comparison when survival is modeled on the hazard scale.

Appropriate inverse weights vary according to whether one is modeling survival or quality-adjusted survival outcomes. In the survival setting, inverse weights have been studied by Robins and Rotinitzky (1992), Robins (1993), Robins and Finkelstein (2000), Satten, Data, and Robins (2001), Scharfstein et al. (2001) and others in various settings. In Chapter II we summarize how these weights may be calculated for use in either Cox or restricted mean models. When quality-adjusted survival is being modeled, Zhao and Tsiatis (1997) modified the concept of inverse weights on the quality-adjusted time scale. We summarize their inverse weighting approach in Chapter IV as an integral component of our multiple imputation strategy in that setting.

When censoring is not dependent, Andersen, Hansen, and Klein (2004) defined a PO for individual i, as $n\int_0^\tau \hat{S}(t)dt - (n-1)\int_0^\tau \hat{S}^{-i}(t)dt$, where $\hat{S}(t)$ is the Kaplan-Meier estimate of the survival probability at time t, $\hat{S}^{-i}(t)$ is its leave-one-out version for subject $i, i = 1, \ldots, n$, and n is the total number of subjects. They model the τ-restricted mean $E[min(\tau, T)]$ as a function of covariates using these pseudo-observations as the dependent variables, since each PO has the same conditional

expectation as the uncensored event time being modeled. Using this approach, restricted mean regression parameters can be obtained using standard software without going through complicated variance calculations. Estimated restricted mean structures can be described to practitioners so that they may calculate and interpret them without having to understand what a baseline hazard is. However Kaplan-Meier estimates used in Andersen et al.'s pseudo-observations are subject to bias due to dependent censoring. In Chapter II we develop a PO approach on the log scale that adjusts for dependent censoring bias via inverse weights, which is more appropriate for modeling restricted means in the lung allocation setting. We compare weighted and unweighted PO approaches and assess performance of these with respect to the lung transplant data.

In contrast to PO methods, which are mainly about modeling the restricted mean, multiple imputation (MI) allows creation of complete data sets that can be used for a variety of standard analyses. In the missing data literature, censoring has been viewed as missing data that can be multiply imputed. Several authors have based imputation methods on the inverse transform relationship, $T = S^{-1}(U)$, where U is a uniform random variable (Taylor, Murray and Hsu (2002), Hsu, Taylor, Murray, and Commenges (2006), Liu, Murray, and Tsodikov (2011)). The key to sampling either survival or quality-adjusted survival outcomes, based on this inverse transform approach, is obtaining a relevant estimate of $S(t)$ within an appropriate risk set. We provide more details on imputation of survival outcomes using inverse weighted survival estimates in Chapter III. Quality-adjusted survival estimates are used as part of the imputation procedure in Chapter IV. Inverse weighted imputation approaches are compared with inverse weighed PO approaches throughout the thesis via simulations and examples mentioned earlier.

CHAPTER II

PSEUDO OBSERVATIONS FOR DEPENDENTLY CENSORED SURVIVAL DATA

2.1 Introduction

To appreciate the statistical aspects of lung transplant candidate data, some background is required. To get a lung transplant in the United States, candidates register with the Organ Procurement and Transplantation Network (OPTN) to obtain placement on a lung waiting list. When these transplants were infrequent, a first come, first served policy seemed equitable to those waiting for transplant. But as the demand increased, so did the average waiting time to transplant, and an increasing number of end-stage lung disease patients died while waiting for an organ offer. Published in 1998 and enacted in 2000, a Final Rule, crafted by the Health Resources and Services Administration of the U.S. Department of Health and Human Services, dictated, among other things, that a more equitable organ allocation algorithm needed to be created and maintained based on objective medical data [8].

In the case of patients waiting for a lung transplant, a statistical algorithm (lung allocation score or LAS) for ranking patients was implemented on May 4, 2005 [9]. The LAS includes measures of the net benefit of the transplant to the candidate as well as the candidate's clinical urgency over the upcoming year. The measure for

net transplant benefit is calculated by subtracting the patient's estimated number of days lived on the waiting list without a transplant over the next year (i.e. transplant urgency) from the estimated number of days lived during the first year following transplantation (i.e. post-transplant survival measure). This is an individual measure of transplant benefit rather than a collective measure of transplant benefit that's sometimes obtained through use of a time dependent covariate for transplant, as in analyses done for the original Stanford heart transplant study [7].

Figure 2.1 shows estimated patient specific urgency by anticipated transplant benefit for a group of lung candidates actively listed between 9/1/2006 and 9/30/2008. It was recognized that ordering patients based on urgency alone, i.e., from left to right in Figure 2.1, might prioritize patients with little or no transplant benefit. Whereas ordering patients solely based on higher benefit, i.e., from top to bottom of Figure 2.1, would likely result in many deaths of urgent patients who would not live until an organ offer. In the end a compromise was reached so that both benefit and urgency were taken into consideration, i.e., allocation according to the diagonal line moving from top left to bottom right of Figure 2.1. The LAS takes the difference between the net transplant benefit and the transplant urgency, with the final score normalized to produce a range from 0 to 100.

Estimates for both urgency and benefit depend on accurate estimation of waitlist days lived during the year following listing. Patients' risk factors measured at listing include diagnosis, age, body mass index (BMI), diabetes, assistance with activities of daily living (ADL), six-minute walk distance (6MWD), forced vital capacity (FVC), oxygen (O_2) requirement at rest, pulmonary artery (PA) systolic pressure, partial pressure of carbon dioxide in the blood (PCO_2), continuous mechanical ventilation, creatinine, and cardiac index. For estimating days lived in the year following trans-

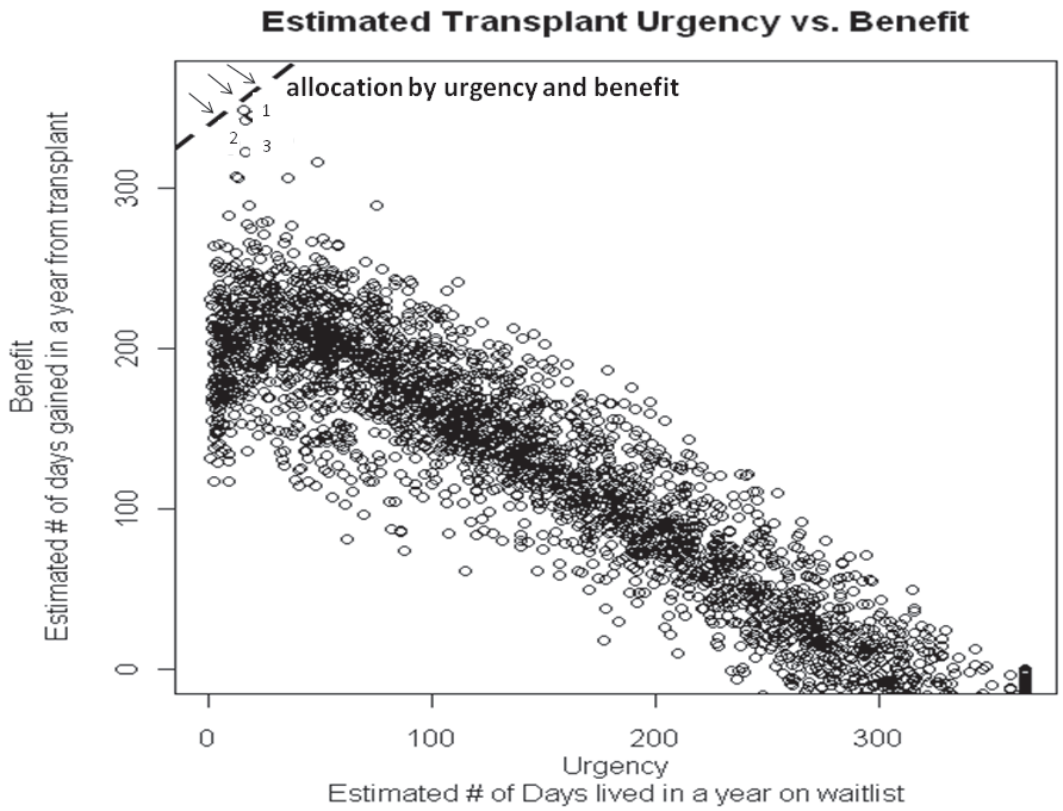


Figure 2.1: Scatterplots of estimated days lived in 1-year on the waiting list versus estimated 1-year transplant benefit (transplant benefit = estimated post-transplant days lived in next year minus estimated waiting list days lived in a year without transplant) at time of listing for n=3701 patients. Allocation follows according to the movement of the diagonal line from the top left to the bottom right. Patients marked as 1, 2, and 3 are the first patients to be offered a lung transplant.

plant, risk factors used in the LAS include diagnosis, age, assistance with ADL, 6MWD, FVC, continuous mechanical ventilation, cardiac index, O_2 requirement at rest, and creatinine. More details on statistical methodology will be given shortly.

The LAS has been largely successful since its implementation. The number of deaths on the waiting list and waiting time for transplant have decreased. As opposed to 512 waitlist deaths in 2004, there were only 266 deaths in 2008, in spite of more urgent patients being listed in 2008 [38]. Listing behavior of end-stage lung patients has changed dramatically since LAS implementation. With no advantage to accruing

waiting time in the new allocation score, the number of patients actively listed for transplant decreased from 2,163 candidates at the end of 2004 to 1,089 patients at the end of 2008 [38]. That is, patients not yet ready to accept an organ offer began to remove themselves from the active candidate pool and delay entering the pool until further progression of disease. As a consequence, the median waiting time has dropped from 792 days in 2004 to 200 days or less after the LAS was used [38]. As successful as the LAS has been, national policy dictates that the algorithm must be continually updated to reflect more recent cohorts of patients, and this is occurring right now in a post-LAS implementation cohort.

The estimated number of days lived during a year in the calculation of LAS is sometimes called the restricted mean life. When estimated nonparametrically, under independent censoring, it is typically defined using the area under a Kaplan-Meier survival curve [21] for the time period of interest (0 to 1 year). In the original development of the LAS, a Cox proportional hazards model [6] was used to estimate each individual's survival curve, and the area under the first year of the survival curve was used to estimate the restricted mean life. We feel that a more appropriate model would target the restricted mean more directly, rather than modeling the hazard ratio. Not only would regression parameters be more directly linked to the restricted mean of interest, but increased transparency of how scores for the LAS are produced would be welcomed by patients and physicians following allocation scores.

One modeling strategy for the restricted mean was introduced by Andersen, Hansen, and Klein [1]. First pseudo observations are generated that have the same conditional mean of interest for regression modeling as the original individual level data. The advantage of creating pseudo observations in the first modeling step is that they can be modeled using traditional uncensored linear models. Pseudo obser-

vations (PO) for mean restricted life, as defined by Andersen, Hansen, and Klein, are created using marginal estimates of restricted mean life, i.e., $\hat{\alpha}_0 = \hat{E}[min(\tau, T)] = \int_0^\tau \hat{P}(T > t)dt$, where T denotes the failure time, τ is the upper limit of a time window of interest and $\hat{P}(T > t)$ is the KM estimate. Then the pseudo observation for each individual, also known from jackknife methodology, is calculated as

$$n\hat{\alpha}_0 - (n-1)\hat{\alpha}_0^{-i}, \tag{2.1}$$

where $\hat{\alpha}_0^{-i} = \int_0^\tau \hat{P}^{-i}(T > t)dt$ with $\hat{P}^{-i}(T > t)$ the KM estimate based on data leaving out patient i.

The intuition behind pseudo observations given in (2.1) is that any nonparametric estimator of $\alpha_0 = E[min(\tau, T)]$ is also implicitly an estimator of

$$E_Z[E[min(\tau, T)|Z]], \tag{2.2}$$

where Z is a vector of covariates, and the inner expectation is of interest in regression modeling. In the case where the outermost expectation is viewed with respect to the empirical distribution of Z, with $\tilde{\alpha}_0 = \frac{1}{n}\sum_{i=1}^n E[min(\tau, T)|Z_i]$, pseudo observations take the form $n\tilde{\alpha}_0 - (n-1)\tilde{\alpha}_0^{-i} = n[\frac{1}{n}\sum_{i=1}^n E[min(\tau, T)|Z_i]] - (n-1)[\frac{1}{n-1}\sum_{j=1, j\neq i}^n E[min(\tau, T)|Z_j]] = E[min(\tau, T)|Z_i]$, the quantity of interest in regression modeling. Andersen, Hansen, and Klein make the case that $\hat{\alpha}_0$ and $\tilde{\alpha}_0$ are both consistent for α_0, hence pseudo observations based on (2.1), which are estimable from censored data, can be used to estimate regression parameters predicting $E[min(\tau, T)|Z_i]$ using readily available linear models. That is, models based on individual specific pseudo observations in (2.1), $i = 1, ..., n$ will have regression parameters similar to a model fit to $min(\tau, T)$ values, $i = 1, ..., n$, if these values were available (uncensored). Graw et al. [11] formalize this argument and verify appropriate asymptotics of parametric estimates.

Any modeling strategy for estimating restricted means requires taking into account an especially interesting dependent censoring issue when updating the LAS to a more current cohort of patients. By removing more urgent patients from the waiting list to get transplanted (via the LAS), our resulting analysis dataset is dependently censored in direct relationship to daily changing LAS of individual patients. In considering a restricted mean model approach, KM estimates used in creating pseudo observations are especially subject to dependent censoring bias. Inverse probability of censoring weighted (IPCW) methods, such as those discussed by Robins and Finkelstein [27], Robins [26], Robins and Rotnitzky [28], Satten, Datta, and Robins [30], Scharfstein, Robins, Eddings, and Rotnitzky [32] among others have been successful in counteracting this type of bias and can be used to consistently estimate $S_T(t) = P(T > t)$, cumulative hazard functions and other quantities of interest.

In this chapter we propose to estimate transplant urgency and benefit by using a pseudo observation approach to estimate one year restricted mean life separately in waitlist and post-transplant cohorts. Our approach will modify each waitlist restricted mean pseudo observation by including IPCW-based survival estimates in place of KM estimates to account for dependent censoring linked to time-dependent LAS of patients. One-year transplant benefit will be estimated for each patient using a restricted mean model estimate of days lived in a year following transplant minus a separate restricted mean model estimate of days lived in a year following listing without transplant.

The rest of this chapter is structured as follows: in Section 2.2, we formally describe the mean structure for restricted life given covariates, an appropriate pseudo observation approach to fit this mean structure, and the IPCW implementation of pseudo observations required to account for censoring via time-dependent LAS. Sec-

tion 2.3 briefly shows simulation studies that ensure our overall analysis approach in the presence of dependent censoring is sound in finite sample populations. In Section 2.4 we present a restricted mean model for lung waitlist candidates and separately for post-transplant recipients to be used in constructing an LAS for each patient. Results on estimated days of life without transplant in a 1-year period (urgency) are given as well as estimated days gained from transplant over the following 1-year period (benefit). Estimated LAS are also displayed for lung transplant candidates using the new methodology. Results are also given using estimates of restricted means for lung candidates based on integrating both traditional and IPCW-adjusted Cox PH model survival curves. Discussion follows in Section 2.5.

2.2 Estimating restricted mean life using IPCW PO

2.2.1 *Mean structure for restricted mean life*

The mean structure for the restricted mean life is

$$E[\log\{min(\tau, T)\}] = \beta^T Z, \tag{2.3}$$

where τ is fixed and within the range of the observed data. When there is no censoring, uncensored data applied to model (2.3) becomes a standard linear model on $\log\{min(\tau, T)\}$. However when censoring is present and informative, using observed data will lead to biased results.

Andersen, Hansen, and Klein [1] formulated pseudo observations using (2.1) and then fit model (2.3) to the resulting pseudo observations using standard linear models. An equally appropriate approach would be to fit the model of the restricted mean using a log link. We have found that the intercept estimator is somewhat improved upon creating pseudo observations based on the transformed random variable $\log\{min(\tau, T)\}$. That is, instead of creating pseudo observations based on

$\hat{\alpha}_0 = \hat{E}[min(\tau, T)] = \int_0^\tau \hat{S}(t)dt$ and log transforming the pseudo values, we will create pseudo observations based on marginal estimates of $\hat{\delta}_0 = \hat{E}[\log\{min(\tau, T)\}]$. Let $Y = \log\{min(\tau, T)\}$, ranging from $-\infty$ to $\log \tau$. Assume for the moment the simplest form of model (2.3), where $E[Y] = \delta_0$, i.e., the marginal mean of $log\{min(\tau, T)\}$ that doesn't depend on any covariates. We may derive the mean of Y as follows.

$$\begin{aligned} E(Y) = \delta_0 = E[\log\{min(\tau, T)\}] &= \int_0^\infty \log[min(\tau, T)]dF_T(t) \\ &= \int_0^\tau \log t \; dF_T(t) + \int_\tau^\infty \log \tau \; dF_T(t) \\ &= \int_0^\tau \log t \; d(1 - S_T(t)) + \log \tau(1 - F_T(\tau)) \\ &= -\int_0^\tau \log t \; dS_T(t) + \log \tau \cdot S_T(\tau). \end{aligned} \tag{2.4}$$

2.2.2 *Pseudo observation approach*

From Section 2.2.1 equation (2.4), the marginal mean of $\log\{min(\tau, T)\}$ can be estimated via

$$\hat{\delta}_0 = \hat{E}(Y) = -\int_0^\tau \log t \; d\hat{P}(T > t) + \log \tau \cdot \hat{P}(T > \tau) \tag{2.5}$$

where $\hat{P}(T > t)$ is some marginal survival estimate on the original time scale. In the context of dependent censoring, we will describe a consistent estimate for $S_T(t)$ in Section 2.2.3, $\hat{S}_T^W(t)$, that uses an inverse weight approach and show that its use in (2.5) gives a consistent estimate of δ_0.

Arguments justifying the use of pseudo observations in fitting (2.3) proceed similarly to the original justification made for the pseudo observation approach. That is, any estimator of δ_0 is implicitly an estimator of $E_Z[E[\log\{min(\tau, T)\}|Z]]$. When the outermost expectation is viewed with respect to the empirical distribution of Z with $\tilde{\delta}_0 = \frac{1}{n}\sum_{j=1}^n E[\log\{min(\tau, T)\}|Z_j]$ and $\tilde{\delta}_0^{-i} = \frac{1}{n-1}\sum_{j=1, j\neq i}^n E[\log\{min(\tau, T)\}|Z_j]$, pseudo observations, $n\tilde{\delta}_0 - (n-1)\tilde{\delta}_0^{-i}$ reduce to $E[\log\{min(\tau, T)\}|Z_i]$, which matches

in expectation the quantity we wish to model. Although both $\tilde{\delta}_0$ and $\hat{\delta}_0$ are consistent for δ_0, the latter gives the most useful form for estimating δ_0 based on censored survival data. So similar to the strategy employed on the scale of α_0, we base our inference on pseudo observations, $n\hat{\delta}_0 - (n-1)\hat{\delta}_0^{-i}$, where $\hat{\delta}_0^{-i}$ is estimated from (2.5) leaving out individual i.

Once pseudo observations, $\mathcal{PO} = (\mathcal{PO}_1, \mathcal{PO}_2, ..., \mathcal{PO}_n)$ are obtained, the regression model (2.3) can be estimated using $\mathcal{PO}$ as the response. Our parameter estimates become $\hat{\beta} = (Z^T Z)^{-1} Z^T \mathcal{PO}$ with estimated covariance matrix $\hat{V}(\hat{\beta}) = \hat{\sigma}^2 (Z^T Z)^{-1}$, where $\hat{\sigma}^2$ is computed in the usual way as $(\mathcal{PO} - \mathbf{Z}\beta)^T (\mathcal{PO} - \mathbf{Z}\beta)/(n-p)$ for p parameters in the model. These results can be estimated from nearly any statistical software package once $\mathcal{PO}_i, i = 1, ..., n$ are obtained.

2.2.3 *IPCW estimates of survival probability*

In the case of dependent censoring care needs to be taken in estimating $S_T(t)$. Of all the potential methods for consistently modeling marginal survival in the presence of dependent censoring, methods based on inverse probability censoring weights used by Robins and his coauthors are perhaps the easiest to apply when there are many time-dependent measures over time, so we selected that approach for estimating $S(t)$ for the lung allocation data.

First one estimates the censoring survival function at any fixed time t, denoted by $\hat{K}_i^{\mathbf{V}}(t) = P(C_i > t | \bar{\mathbf{V}}_i(t))$, where C_i is the censoring time for patient i, and $\bar{\mathbf{V}}_i(t) = \{\mathbf{V}_i(u); 0 \leq u \leq t\}$ is the patient's recorded history up to time t of a vector of possibly time dependent covariates, $\mathbf{V}_i$, that predict the censoring time C_i. In the case of the lung allocation data, C_i is the time a patient is removed from the lung waitlist for transplant, and $\bar{\mathbf{V}}_i(t)$ consists of patient lung allocation scores used

to rank patients for transplant from time 0 (listing time) to time t as well as a few additional predictors including race, gender, blood type, height, and active waiting status. In calculating the contribution of a subject at risk at time t, the subject is given a weight inversely proportional to his estimated probability of remaining uncensored until time t with a history of $\bar{V}_i(t)$, i.e.

$$\hat{W}_i(t) = 1/\hat{K}_i^{\mathbf{V}}(t).$$

The Cox model for censoring survival is often used in inverse weighting approaches because of its flexibility in modeling time-dependent covariates. Since time-dependent LAS is an issue in our case, this is the approach we adopt as well. A Cox model for the censoring hazard is given by

$$\lambda_Q\{t|\bar{\mathbf{V}}(t)\} = \lambda_{Q0}(t)exp\{\gamma'\mathbf{V}(t)\}, \tag{2.6}$$

In the case of the lung waitlist data $\gamma'\mathbf{V}(t)$ becomes γ_1 LAS(t)+γ_2 race+γ_3 gender+γ_4 blood type +γ_5 height + γ_6 I(active waiting status). Then a consistent estimate of the probability that subject i gets censored after time t, $K_i^V(t)$, becomes

$$\hat{K}_i^V(t) = exp\{-\sum_{k=1}^{n}\int_0^t \frac{e^{\hat{\gamma}'\mathbf{V}_i(u)}dN_{Q_k}(u)}{\sum_{j=1}^n Y_j(u)e^{\hat{\gamma}'\mathbf{V}_j(u)}}\},$$

where $N_{Q_i} = I(X_i \leq u, \delta_i = 0)$ is the observable counting process for censoring (transplant), with X_i the observed event time and δ_i the censoring indicator, and $Y_i(u) = I(X_i \geq u)$ is the risk indicator for subject i at time u. The subject specific weight then becomes

$$\hat{W}_i(t) = 1/\hat{K}_i^V(t) = exp\{\sum_{k=1}^{n}\int_0^t \frac{e^{\hat{\gamma}'\mathbf{V}_i(u)}dN_{Q_k}(u)}{\sum_{j=1}^n Y_j(u)e^{\hat{\gamma}'\mathbf{V}_j(u)}}\}.$$

An IPCW version of Nelson-Aalen estimator for cumulative hazard, $\Lambda(t)$, is calculated using

$$\hat{\Lambda}^W(t) = \sum_{i=1}^{n}\int_0^t \frac{dN_{T_i}(u)\cdot\hat{W}_i(u)}{\sum_{j=1}^n Y_j(u)\cdot\hat{W}_j(u)},$$

where $N_{T_i}(u) = I(X_i \leq u, \delta_i = 1)$ is the observable counting process for death. Then the survival probability is estimated with $\hat{S}_T^W(t) = exp\{-\hat{\Lambda}^W(t)\}$. The adjusted pseudo observations described in Section 2.2.2 use $\hat{P}(T > t) = \hat{S}_T^W(t)$ in equation 2.5. Product-integral versions of inverse weighted survival functions such as those described by Satten and Datta [29] would also be appropriate for use as an alternative to $\hat{S}_T^W(t)$.

Proof of consistency of $\hat{S}_T^W(t) = \exp(-\hat{\Lambda}^W(t))$ for $S_T(t)$ proceeds from consistency of $\hat{\Lambda}^W(t)$ for $\Lambda(t)$, a property that was studied extensively by Robins [26] and Robins and Finkelstein [27]. Conditions required for this consistency to hold are that (a) $\lambda_Q(t|\bar{V}(t))$ follows the form given in equation (2.6) and that (b) $\lambda_Q(t|\bar{V}(t), T, T > t) = \lambda_Q(t|\bar{V}(t), T > t)$. Consistency of $\hat{E}(Y)$, used in creating pseudo observations in this chapter, follows from noting that $\int_0^\tau \log t \; d\hat{S}_T^W(t)$ in (2.5) can be written as

$$\lim_{m\to\infty} \sum_{j=1}^{m} \log t_j \; \triangle \; \hat{S}_T^W(t_j) \xrightarrow{p} \lim_{m\to\infty} \sum_{j=1}^{m} \log t_j \; \triangle \; S_T(t_j) = \int_0^\tau \log t \; dS_T(t).$$

2.3 Simulation Study

To validate the method used in analyzing the lung candidate waitlist data subjected to dependent censoring, we conducted a simulation study comparing parameter estimates of model (2.3) using linear regression when (a) $\log[min(\tau, T)]$ is uncensored; (b) $\log[min(\tau, T)]$ is subject to censoring and is replaced by log-transformed pseudo observations defined by (2.1); (c) $\log[min(\tau, T)]$ is subject to censoring and is replaced by IPCW adjusted pseudo observations.

In each simulation, we perform the following procedures:

Step 1: We simulate Z_0 from a Bernoulli(0.5) distribution, Z_1 from Bernoulli(0.5), and Z_2 from Uniform(0,1), where Z_0 is a binary covariate measured at time 0, Z_1 is the time-dependent covariate measured at time $t_1 = 0.2$, and Z_2 is a continuous

time-independent covariate.

Step 2: Failure times, T_i, are simulated from piecewise exponential distributions, i.e. $T_i \sim exp(\lambda_{z_0})$ before time t_1, and $T_i \sim exp(\lambda_{z_0 z_1})$, after time t_1, where $\lambda_0 = 0.3$ and $\lambda_1 = 0.2$ are fixed, $\lambda_{00}, \lambda_{01}, \lambda_{10}$ and λ_{11} are solved so that the mean structure $E[\log\{min(\tau, T)\}] = \beta_0 + \beta_1 Z_0 + \beta_2 Z_2$ is satisfied for a pre-specified $\beta = (\beta_0, \beta_1, \beta_2)$. That is, although T_i is influenced by the time dependent covariate, Z_1, the restricted mean of interest is captured by baseline predictors Z_0 and Z_2. More details for these calculations are given in the appendix.

Step 3: Dependent censoring times C_i are also generated from piecewise exponential distributions. The hazard rates are obtained based on the Cox model $\lambda^C(t|\bar{\mathbf{Z}}(\mathbf{t})) = \lambda_0^C(t) \exp\{\gamma_0 Z_0 + \gamma_1 I[Z_0 = 0, Z_1 = 1, t > t_1] + \gamma_2 I[Z_0 = 1, Z_1 = 0, t > t_1] + \gamma_3 I[Z_0 = 1, Z_1 = 1, t > t_1] + \gamma_4 Z_2\}$, where $\lambda_0^C(t) = 0.15$ for $t \leq t_1$ and $\lambda_0^C(t) = 0.4$ for $t > t_1$, $\gamma_0 = 0.3, \gamma_1 = -1.4, \gamma_2 = 0.5, \gamma_3 = -1.5$, and $\gamma_4 = 1$. So Z_0, Z_2 and time-dependent Z_1 influence censoring.

For each scenario of β, 1000 simulations are run with $\tau = 5\ yrs$ and either $n = 150$ or $n = 300$ patients. Results for the scenario with $\beta_0 = 0.8, \beta_1 = \beta_2 = 0$ are located in part 1) of Table 2.1. In this case the true baseline covariate effects on survival are zero, but dependent censoring is being driven by the time dependent covariate. The unadjusted PO method gives more biased estimates for all the parameters, especially for β_0 and β_1. The IPCW adjusted PO method reduces bias substantially and also has smaller standard error after adjusting for dependent censoring. When both baseline covariates are equal to 0.5, the PO method underestimates the time lived during the 5-year period by 14 months on average while the IPCW PO method is off by only 2 months over the 5-year period. Empirical standard deviations were comparable to standard errors averaged across simulations with the exception of the intercept term

for the traditional PO method. Coverage for the traditional PO intercept was 66.7%, in spite of its much wider confidence interval width, due to the increased bias and underestimated variability for that term.

Simulation results for the scenario with $\beta_0 = 1, \beta_1 = -0.8, \beta_2 = -0.5$, i.e., non-zero baseline covariate effects, are located in part 2) of Table 2.1. Again, bias is higher for β_0 and β_1 using the traditional PO, but in this case the value for β_2 is largely unaffected by the dependent censoring. The adjusted PO method has both smaller bias and smaller standard error for β_0 and β_1. The overall degree of bias for estimating the time lived during the 5 year period was smaller in this scenario, with the traditional PO method off by approximately 4-5 months of life lived and the IPCW PO method off by 1 month over the 5 year period for a patient with $z_0 = z_2 = 0.5$.

Parameter estimates were similar for cases with $n = 150$ and $n = 300$. We were unable to explore larger sample sizes due to limitations in computing speed, so it is not clear at what sample size remaining bias with the IPCW PO method vanishes. Across a grid of possible covariate values for these two scenarios, the bias for the IPCW PO method did not exceed 3.6 months over the 5 years of follow-up. But bias as high as 16 months was seen using the unadjusted PO method.

2.4 Example

This section is organized into 3 components. Section 2.4.1 summarizes analyses for the lung waitlist candidates, section 2.4.2 summarizes analyses for the post-transplant cohort and section 2.4.3 interprets these analyses in terms of urgency, benefit, and lung allocation scores calculated for the waitlist patient cohort. Results are typically reported by the 4 defined diagnosis groups A, B, C, and D. Diagnosis

Table 2.1: Comparison of estimates from model (2.3) using uncensored observations (Uncensored), the traditional pseudo observation (PO) approach and the IPCW PO approach under two scenarios, 1000 iterations.

Parameter	Uncensored	PO	IPCW PO	Uncensored SE [a]	PO SE	IPCW PO SE	Uncensored ESD[b]	PO ESD	IPCW ESD
1) **Covariate effects are zero, sample size = 150**									
$\beta_0 = 0.8$	0.808	0.105	0.702	0.210	0.425	0.243	0.220	0.576	0.235
$\beta_1 = 0$	0.012	-0.266	0.059	0.187	0.380	0.217	0.195	0.400	0.215
$\beta_2 = 0$	-0.013	0.154	-0.022	0.325	0.658	0.377	0.328	0.705	0.372
Covariate effects are zero, sample size = 300									
$\beta_0 = 0.8$	0.798	0.137	0.697	0.148	0.294	0.172	0.153	0.460	0.163
$\beta_1 = 0$	0.001	-0.296	0.049	0.132	0.263	0.154	0.134	0.281	0.155
$\beta_2 = 0$	0.008	0.170	0.000	0.230	0.456	0.267	0.234	0.472	0.262
2) **Covariate effects are non-zero, sample size = 150**									
$\beta_0 = 1$	0.998	0.759	0.891	0.202	0.309	0.239	0.205	0.324	0.220
$\beta_1 = -0.8$	-0.799	-1.025	-0.796	0.181	0.276	0.214	0.180	0.326	0.194
$\beta_2 = -0.5$	-0.497	-0.443	-0.426	0.314	0.479	0.372	0.315	0.496	0.351
Covariate effects are non-zero, sample size = 300									
$\beta_0 = 1$	0.992	0.772	0.883	0.142	0.206	0.170	0.146	0.220	0.158
$\beta_1 = -0.8$	-0.799	-0.971	-0.792	0.128	0.184	0.152	0.127	0.225	0.138
$\beta_2 = -0.5$	-0.486	-0.413	-0.417	0.221	0.319	0.263	0.224	0.325	0.247

[a]SE is the average of estimated standard errors across 1000 iterations.
[b]ESD is the empirical standard deviation of 1000 parameter estimates.

group A is obstructive lung disease, primarily chronic obstructive pulmonary disease (COPD). Group B consists of pulmonary vascular diseases, primarily idiopathic pulmonary arterial hypertension (iPAH). Group C consists of cystic fibrosis (CF), as well as immunodeficiency disorders. Group D is restrictive lung disease, primarily interstitial pulmonary fibrosis (IPF). All lung waitlist patients are classified into one of these 4 diagnosis groups by the OPTN Thoracic Committee for the purpose of estimating diagnosis group influence on urgency and benefit and to model interactions across diagnosis groups. A few diagnoses, such as Bronchiectasis, are given a parameter to distinguish their estimated days of life from that of their overall diagnosis group. These parameters have historically not been significantly different from those of their overall diagnosis group, and yet patient advocates have actively pursued the ability to estimate urgency and benefit more specifically for their patients to the extent that enough data are available to do so.

2.4.1 *Lung candidate analysis*

The waitlist candidate data contain 3701 lung candidates aged 12+ who were newly listed on the lung waitlist during 9/1/2006-9/30/2008. Censoring within 1-year of listing only occurs when a candidate is transplanted, which was the case with 2698 (73%) of the candidates. By diagnosis group, 923 (70%) of 1317 group A candidates, 67 (58%) of 116 group B candidates, 294 (69%) of 428 group C candidates, and 1414 (77%) of 1840 group D candidates were transplanted. Historically of the four groups, diagnosis group D has had the poorest waitlist survival and, with the LAS based in part on urgency, this group also currently experiences the shortest time to transplant. The median time to transplant for group D is only 71 days, as opposed to 170 days for group A, 221 days for group B and 126 days for group C. Baseline characteristics

by primary diagnosis group are shown in Table 2.2. At listing, group D patients typically have very high severity and poor physiologic reserve. In contrast, group A patients have historically had much lower urgency for transplant as measured by survival. These patients are often seeking a transplant based on improving quality of life as opposed to lengthening life.

The P(censoring occurs after time t| candidate's history up to and including t) is used to calculate an inverse weight used in consistent estimation of survival curves and adjusted POs as described in Sections 2.2.2 and 2.2.3. In particular, a time-dependent Cox model for time to censoring is used that includes patients' daily updated LAS, gender, race, blood type, status (active, inactive, offlist), and height, as given in equation (2.6). Parameter estimates from the Cox model on the censoring hazard are summarized in Table 2.3. Probability of transplant is strongly influenced by current LAS values. Although one might expect the probability of transplant to increase monotonically as current LAS increases, in fact the higher transplant priority is tempered by a lower chance of surviving until an organ becomes available for those with the very highest LAS values. This feature is reflected in the parameter estimates shown.

To estimate lung candidate urgency, we fit model (2.3) using both IPCW adjusted pseudo observations and traditional pseudo observations. For comparison, restricted means for lung candidates were estimated by integrating traditional and IPCW-adjusted Cox-model based survival curves [6]. Predictors included in all lung candidate models are the same as those proposed by the OPTN Thoracic Committee in modeling waitlist survival for this cohort [25]. All predictors have been vetted extensively by the Thoracic Committee as being worthy of inclusion in the algorithm based on statistical and/or clinical validity based on either the current

Table 2.2: Baseline characteristics by diagnosis group for 3701 lung waitlist patients.

Characteristics	Group A (primarily COPD) N=1317	Group B (primarily iPAH) N=116	Group C (primarily CF) N=428	Group D (primarily IPF) N=1840
Physiologic Reserve				
Age (yrs)	57.6 ± 8.1 [a]	45.9 ± 15.1	27.9 ± 10.3	55.9 ± 11.6
BMI (kg/m^2)	24.6 ± 4.3	24.8 ± 4.3	19.1 ± 2.9	26.7 ± 4.6
Diabetes	11.0% [b]	10.3%	48.4%	22.9%
No assistance with ADL [c]	10.4%	9.4%	20.9%	9.6%
Six-min walk distance ($feet$)	770.5 ± 356.0	733.3 ± 480.8	902.2 ± 508.0	755.8 ± 482.0
Severity				
FVC (% predicted)	52.4 ± 17.2	69.3 ± 23.1	38.1 ± 11.1	47.0 ± 16.7
O_2 requirement at rest (L/min)	3.1 ± 2.5	4.6 ± 5.1	3.4 ± 4.1	4.6 ± 4.7
PA systolic (mm/Hg)	38.0 ± 10.8	78.5 ± 24.0	38.8 ± 10.6	42.9 ± 17.0
PCO_2 (mmHg)	50.3 ± 11.7	43.5 ± 7.0	56.2 ± 20.0	45.9 ± 11.6
Continuous Mechanical Ventilation	1.1%	2.6%	7.5%	6.1%
Serum Creatinine (mg/dl)	0.82 ± 0.2	0.94 ± 0.3	0.67 ± 0.2	0.91 ± 0.3
Cardiac Index$<$ 2.0 ($L/min/min^2$)	5.5%	35.6%	1.6%	7.8%

[a]For continuous variables, the numbers shown are mean±standard deviation.
[b]For binary variables, the numbers shown are proportions.
[c]ADL: activities of daily living.

Table 2.3: Proportional hazards censoring model; 3701 candidates. Inverse weights based on this model are capped at 20.

Parameter	Hazard Ratio	95% CI	p-value
Characteristic at Listing			
Female (vs. Male)	0.72	$(0.63, 0.82)$	$< .0001$
Race: Black (vs. White)	0.81	$(0.68, 0.95)$	0.0116
Race: Other (vs. White)	0.91	$(0.77, 1.08)$	0.3016
Height: $< 5'3''$ (vs. $> 5'9''$)	0.54	$(0.45, 0.65)$	$< .0001$
Height: $5'3''$ to $5'6''$ (vs. $> 5'9''$)	0.73	$(0.62, 0.86)$	0.0001
Height: $5'6''$ to $5'9''$ (vs. $> 5'9''$)	0.80	$(0.71, 0.90)$	0.0001
Blood type: B (vs. A)	1.06	$(0.91, 1.23)$	0.4801
Blood type: O (vs. A)	0.92	$(0.84, 1.02)$	0.1118
Blood type: AB (vs. A)	1.07	$(0.85, 1.33)$	0.5669
Time Dependent Patient Condition and Listing Status			
LAS=0 (vs. LAS$>$ 0)	0.16	$(0.02, 1.18)$	0.0728
LAS: linear spline for 30+	1.12^a	$(1.06, 1.19)$	$< .0001$
LAS: linear spline for 35+	0.98^b	$(0.91, 1.06)$	0.6779
LAS: linear spline for 40+	0.95^c	$(0.91, 0.99)$	0.0070
LAS: linear spline for 60+	0.97^d	$(0.95, 0.98)$	$< .0001$
Inactive Status (vs. Active)	0.00	$(0, > 1000)$	0.8780
Off the Waitlist (vs. Active)	0.00	$(0, > 1000)$	0.9410

[a]HR corresponding to one unit increase for LAS 30+ relative to those with $0 < LAS < 30$

[b]HR corresponding to spline term for LAS 35+, giving HR due to one unit increase in LAS in the range $35 \leq LAS < 40$ of $1.12 * 0.98 = 1.10$ relative to $0 < LAS < 30$.

[c]HR corresponding to spline term for LAS 40+, giving HR due to one unit increase in LAS in the range $40 \leq LAS < 60$ of $1.12 * 0.98 * 0.95 = 1.04$ relative to $0 < LAS < 30$.

[d]HR corresponding to spline term for LAS 60+, giving HR due to one unit increase in LAS in the range $LAS \geq 60$ of $1.12 * 0.98 * 0.95 * 0.97 = 1.01$ relative to $0 < LAS < 30$.

or a prior waitlist cohort studied. In some cases, statistically insignificant parameters are maintained as placeholders with the expectation that statistical significance will reassert itself in future cohorts; age, assistance with ADL, PA systolic, PCO_2 and creatinine fall into this category. With many fewer waitlist deaths available for modeling purposes after LAS implementation, loss of statistical power has also been cited as an argument for maintaining a predictor in the LAS that has previously been shown to be statistically significant.

Parameter estimates for fitting model (2.3) to patients awaiting transplant are shown in Table 2.4. Risk factors, $e^{\hat{\beta}}$ act multiplicatively on the number of days lived in a year. For instance the estimated number of days lived is $102.15 \times 0.11 \times 1.07^{25} \times 0.49 \times 1.07^{3} \times 1.25^{5.5} \times 0.87^{2} \times 0.60^{0.8} = 63$ days based on IPCW PO and $301.71 \times 0.40 \times 1.03^{25} \times 0.92 \times 1.04^{3} \times 1.06^{5.5} \times 0.92^{2} \times 0.91^{0.8} = 283$ days based on traditional PO for a 55 year old diagnosis group D patient, who has a BMI of 25, has diabetes, requires assistance with ADL, walks 300 feet in six minutes, has 55% predicted FVC, requires $2L/min$ of O_2 at rest, is not on a ventilator, has a stable creatinine of $0.8mg/dl$, has no partial pressure of CO_2 in their blood and has a cardiac index $> 2L/min/min^2$. So the unadjusted PO overestimates the number of days lived in the next year without transplant by 220 days for this very urgent group D patient. Estimated KM and IPCW survival curves used in PO calculations are shown in Figure 2.2; overestimation of days lived using the traditional PO model stems from the overestimation of survival by the KM method.

Traditional and IPCW-adjusted Cox model hazard ratios estimated from the lung candidate data are shown in Table 2.5. The difference in area under the baseline survival curves over a year for the two methods was 11 days, with a more favorable survival profile without adjustment. Integrating a survival curve estimated from a traditional Cox proportional hazards model for this patient yields an estimated number of days lived of 330, which exceeds the estimate based on the unadjusted PO model. The IPCW-adjusted Cox model estimates a restricted mean of 301 for the same patient. Since bias due to dependent censoring has been accounted for in both IPCW PO method and the IPCW Cox method, different modeling restrictions account for the observed differences in estimation. Additional information on urgency estimates based on different modeling paradigms is given in Section 2.4.3 below.

Table 2.4: Lung waitlist results for model (2.3) using IPCW PO and Traditional PO methods for 3701 lung candidates.

	IPCW PO			Traditional PO		
	$e^{\hat{\beta}}$ [a]	95% CI	p-value	$e^{\hat{\beta}}$	95% CI	p-value
(Intercept)	102.15	(15.96, 653.55)	< 0.0001	301.71	(162.16, 561.37)	< 0.0001
Diagnosis Group (ref=Group A, primarily COPD)						
Group B (primarily iPAH)	0.19	(0.05, 0.73)	0.0158	0.25	(0.16, 0.39)	< 0.0001
Group C (primarily CF)	0.57	(0.17, 1.90)	0.3627	0.56	(0.38, 0.84)	0.0047
Group D (primarily IPF)	0.11	(0.03, 0.38)	0.0004	0.40	(0.27, 0.60)	< 0.0001
Diagnosis [b]						
Bronchiectasis	0.76	(0.21, 2.72)	0.6694	0.98	(0.64, 1.50)	0.9230
Lymphangioleiomyomatosis	1.25	(0.10, 16.41)	0.8639	0.93	(0.39, 2.20)	0.8711
Obliterativebronchiolitis	0.25	(0.03, 1.86)	0.1744	1.61	(0.82, 3.17)	0.1667
Pulmonary Fibrosis other	0.77	(0.37, 1.59)	0.4741	1.04	(0.82, 1.33)	0.7445
Sarcoidosis and PA mean $> 30mm/Hg$	0.46	(0.16, 1.37)	0.1641	1.37	(0.95, 1.97)	0.0910
Sarcoidosis and PA mean $\leq 30mm/Hg$	0.84	(0.22, 3.13)	0.7893	0.64	(0.41, 1.00)	0.0480
Physiologic Reserve						
Age (yrs)	1.00	(0.98, 1.02)	0.9919	1.00	(1.00, 1.01)	0.8456
BMI (kg/m^2)	1.07	(1.03, 1.11)	0.0011	1.03	(1.01, 1.04)	0.0001
Diabetes	0.49	(0.33, 0.73)	0.0005	0.92	(0.80, 1.05)	0.2131
No Assistance with ADL [c]	1.08	(0.65, 1.80)	0.7609	1.00	(0.85, 1.19)	0.9642
Six minute walk (per $100ft$)	1.07	(1.03, 1.12)	0.0012	1.04	(1.02, 1.05)	< 0.0001
Severity						
FVC for Group D (per 10% predicted)	1.25	(1.08, 1.43)	0.0019	1.06	(1.01, 1.11)	0.0243
O_2 requirement for Group A,C,D (L/min)	0.87	(0.83, 0.91)	< 0.0001	0.92	(0.90, 0.93)	< 0.0001
PA systolic (per $10mm/Hg$) for Group A	0.91	(0.71, 1.16)	0.4331	0.94	(0.87, 1.02)	0.1613
PCO_2 increase of $\geq 15\%$	1.03	(0.41, 2.59)	0.9449	1.00	(0.73, 1.36)	0.9889
PCO_2 ($mmHg$)	1.00	(0.99, 1.02)	0.6560	0.99	(0.99, 1.00)	0.0068
Continuous Mechanical Ventilation	0.07	(0.03, 0.20)	< 0.0001	0.13	(0.10, 0.18)	< 0.0001
Creatinine (mg/dl)	0.60	(0.32, 1.12)	0.1111	0.91	(0.74, 1.13)	0.3999
Cardiac Index $< 2.0(L/min/min^2)$	0.48	(0.25, 0.94)	0.0315	0.62	(0.50, 0.77)	< 0.0001

[a] For risk factors, $e^{\hat{\beta}}$ acts multiplicatively on the number of days lived in a year.

[b] These diagnoses were grouped into larger diagnosis groups (A, B, C, D) by the OPTN Thoracic Committee for purpose of modeling risk factors that may vary by diagnosis group. Bronchiectasis, Lymphangioleiomyomatosis, and Sarcoidosid and PA mean $\leq 30mm/Hg$ share risk factor parameters with diagnosis group A, Eisenmenger with group B, Obliterativebronchiolitis, Pulmonary Fibrosis other, and Sarcoidosis and PA mean $> 30mm/Hg$ with group D.

[c] ADL: activities of daily living; ref=Some/total assistance with ADL.

Table 2.5: Lung waitlist survival model using IPCW Cox and Traditional Cox methods for 3701 lung candidates.

	Hazard Ratio	IPCW Cox 95% CI	p-value	Hazard Ratio	Traditional Cox 95% CI	p-value
Diagnosis Group (ref=Group A, primarily COPD)						
Group B (primarily iPAH)	2.75	$(1.70, 4.45)$	$< .0001$	2.47	$(1.78, 3.44)$	$< .0001$
Group C (primarily CF)	1.65	$(1.05, 2.59)$	0.0299	1.46	$(1.09, 1.95)$	0.0105
Group D (primarily IPF)	2.58	$(1.56, 4.27)$	0.0002	2.39	(1.75, 3.25)	$< .0001$
Diagnosis [a]						
Bronchiectasis	1.27	$(0.78, 2.10)$	0.3381	1.19	(0.86, 1.65)	0.3027
Lymphangioleiomyomatosis	1.10	$(0.40, 2.98)$	0.8529	0.93	(0.44, 1.93)	0.8358
Obliterativebronchiolitis	0.61	$(0.19, 1.90)$	0.3917	1.39	(0.89, 2.16)	0.1429
Pulmonary Fibrosis other	1.07	$(0.78, 1.49)$	0.6673	0.80	(0.67, 0.96)	0.0180
Sarcoidosis and PA mean $> 30mm/Hg$	0.72	$(0.49, 1.07)$	0.1063	0.89	(0.73, 1.10)	0.2803
Sarcoidosis and PA mean $\leq 30mm/Hg$	1.25	$(0.80, 1.96)$	0.3292	1.03	(0.76, 1.39)	0.8513
Physiologic Reserve						
Age (yrs)	1.00	$(1.00, 1.01)$	0.6546	1.00	(1.00, 1.01)	0.4194
BMI (kg/m^2)	0.98	$(0.97, 1.00)$	0.0428	0.97	(0.96, 0.98)	$< .0001$
Diabetes	1.23	$(1.04, 1.44)$	0.0144	1.38	(1.25, 1.52)	$< .0001$
No Assistance with ADL [b]	1.00	$(0.82, 1.21)$	0.9987	1.00	(0.88, 1.14)	0.9882
Six minute walk (per $100ft$)	0.97	$(0.95, 0.98)$	0.0001	0.97	(0.96, 0.98)	$< .0001$
Severity						
FVC for Group D (per 10% predicted)	0.93	$(0.87, 0.99)$	0.0161	0.94	(0.90, 0.97)	0.0005
O_2 requirement for Group A,C,D (L/min)	1.11	$(1.09, 1.13)$	$< .0001$	1.10	(1.09, 1.11)	$< .0001$
PA systolic (per $10mm/Hg$) for Group A	1.04	$(0.94, 1.14)$	0.4304	1.05	(0.99, 1.12)	0.0924
PCO_2 increase of $\geq 15\%$	1.13	$(0.79, 1.62)$	0.5023	1.38	(1.08, 1.75)	0.0095
PCO_2 $(mmHg)$	1.01	$(1.01, 1.02)$	$< .0001$	1.01	(1.01, 1.01)	$< .0001$
Continuous Mechanical Ventilation	5.59	$(4.08, 7.66)$	$< .0001$	4.29	(3.49, 5.28)	$< .0001$
Creatinine (mg/dl)	1.57	$(1.23, 2.01)$	0.0003	1.68	(1.43, 1.98)	$< .0001$
Cardiac Index $< 2.0(L/min/min^2)$	1.36	$(1.07, 1.73)$	0.0128	1.30	(1.11, 1.52)	0.0010

[a]These diagnoses were grouped into larger diagnosis groups (A, B, C, D) by the OPTN Thoracic Committee for purpose of modeling risk factors that may vary by diagnosis group. Bronchiectasis, Lymphangioleiomyomatosis, and Sarcoidosid and PA mean $\leq 30mm/Hg$ share risk factor parameters with diagnosis group A, Eisenmenger with group B, Obliterativebronchiolitis, Pulmonary Fibrosis other, and Sarcoidosis and PA mean $> 30mm/Hg$ with group D.

[b]ADL: activities of daily living; ref=Some/total assistance with ADL.

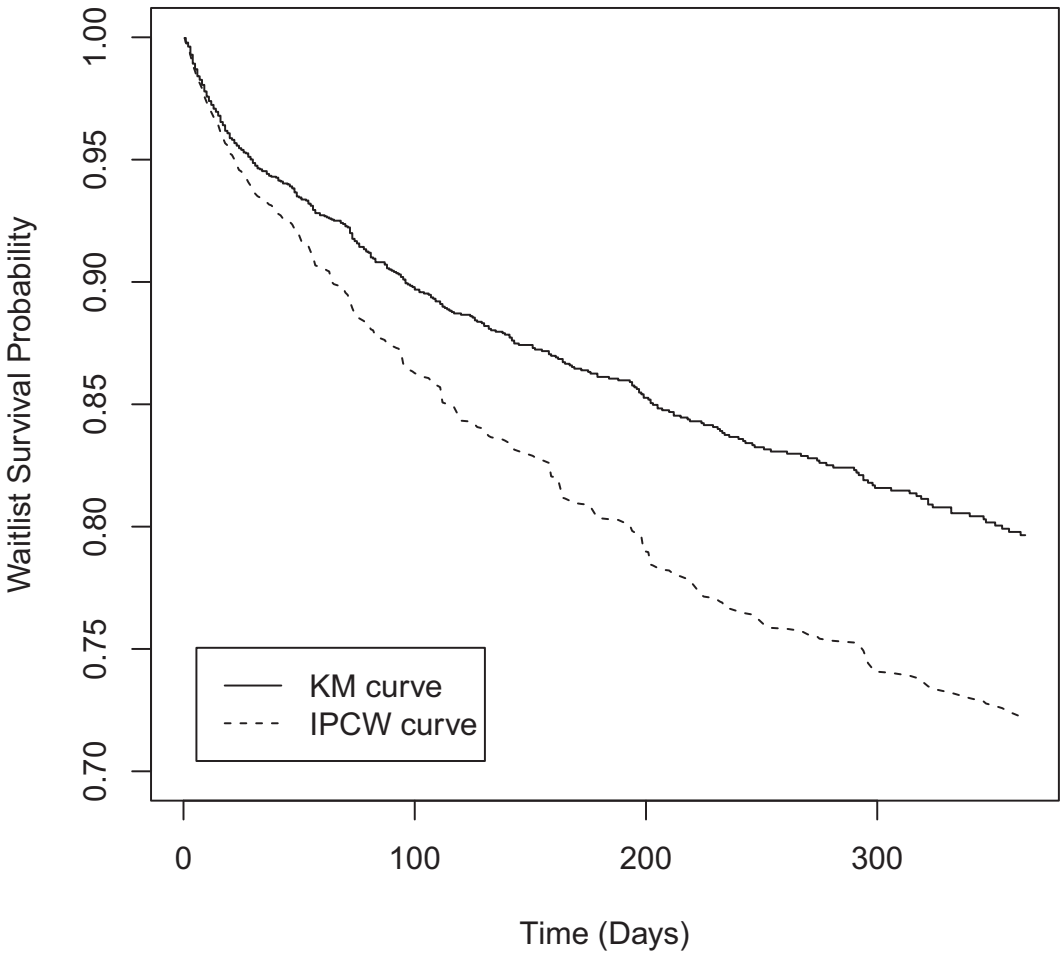


Figure 2.2: Waitlist survival probabilities estimated using KM and IPCW.

2.4.2 *Lung recipient analysis*

The post-transplant data contain 4784 patients aged 12+ who received a lung transplant between 5/4/2005 and 9/3/2008. One-year event rates from the time of transplant are perfectly known, i.e., no censoring, with 816 (17%) deaths within that first year. Results from fitting model (2.3) in the uncensored case are shown in Table 2.6. For the same group D waitlist patient described in section 2.4.1, the

estimated days lived in the first year following transplant is 254 days based on the model in Table 2.6. Recall that the estimated days gained in the first year following transplant is calculated using the estimated days lived one year post-transplant as in Section 2.4.2 minus the estimated days lived one year without transplant as in Section 2.4.1. So the IPCW PO method estimates $254-63 = 191$ days gained during the first year after a transplant for this patient. All other methods indicate days of life lost if transplanted. The IPCW Cox method, the traditional PO method, and the unadjusted Cox method give $254 - 301 = -47$ days, $254 - 283 = -29$ days, and $254 - 330 = -76$ days, respectively.

2.4.3 *Urgency, Benefit, and Lung Allocation Scores*

Figure 2.3 shows boxplots of estimated transplant urgency for the 3701 waitlisted patients by diagnosis group, the two PO modeling paradigms laid out in Table 2.4 and the two Cox modeling paradigms in Table 2.5. By adjusting for dependent censoring, pseudo observations based on IPCW weighted survival curves estimated a higher urgency in Group D patients, which better matches group D survival experience seen before organ allocation took urgency into account (i.e., before LAS-induced dependent censoring was introduced). The integrated IPCW Cox survival curves also show more urgency than the integrated Cox survival curves that don't take into account dependent censoring. The interquartile range for urgency estimates based on integrated Cox PH survival curves is decidedly more narrow than the range of estimated restricted means based on model 2.3 for either the PO or the IPCW PO methods. In addition the Cox modeling approaches tend to estimate many more days lived without transplant compared to the PO methods.

Transplant benefit calculations similar to those done for the hypothetical group

Table 2.6: Lung post-transplant results for model (2.3) for 4784 transplant recipients (no censored data).

	$e^{\hat{\beta}}$ [a]	95% CI	p-value
(Intercept)	344.42	(303.36, 391.03)	< 0.0001
Diagnosis Group (ref=Group A, primarily COPD)			
Group B (primarily iPAH)	0.65	(0.52, 0.81)	0.0002
Group C (primarily CF)	0.92	(0.82, 1.04)	0.1713
Group D (primarily IPF)	0.84	(0.73, 0.96)	0.0107
Diagnosis [b]			
Bronchiectasis	0.96	(0.78, 1.17)	0.6796
Eisenmenger	0.32	(0.11, 0.91)	0.0331
Lymphangioleiomyomatosis	1.24	(0.89, 1.74)	0.2060
Obliterativebronchiolitis	1.25	(0.93, 1.69)	0.1437
Pulmonary Fibrosis other	1.01	(0.89, 1.15)	0.8734
Sarcoidosis and PA mean $> 30mm/Hg$	0.90	(0.74, 1.08)	0.2561
Sarcoidosis and PA mean $\leq 30mm/Hg$	1.00	(0.79, 1.26)	0.9927
Physiologic Reserve			
Age > 45 spline [c] (yrs)	0.99	(0.99, 1.00)	0.0139
No Assistance with ADL [d]	1.02	(0.94, 1.11)	0.6648
Six minute walk (per $100ft$)	1.01	(1.01, 1.02)	0.0002
Severity			
Creatinine at transplant (mg/dl)	0.89	(0.83, 0.96)	0.0017
FVC for Dgn Groups B, D (per 10% predicted)	1.01	(0.99, 1.03)	0.4567
Continuous Mechanical Ventilation at transplant	0.72	(0.63, 0.83)	< 0.0001
Cardiac Index $< 2.0(L/min/min^2)$	0.86	(0.74, 1.00)	0.0496
O_2 at rest for Dgn Group A (L/min)	0.97	(0.96, 0.99)	0.0063
O_2 at rest for Dgn Group B, C, D (L/min)	0.99	(0.98, 1.00)	0.2129
Change in Creatinine $\geq 150\%$	0.78	(0.65, 0.95)	0.0132

[a] For risk factors, $e^{\hat{\beta}}$ acts multiplicatively on the number of days lived in a year.

[b] These diagnoses were grouped into larger diagnosis groups (A, B, C, D) by the OPTN Thoracic Committee for purpose of modeling risk factors that may vary by diagnosis group. Bronchiectasis, Lymphangioleiomyomatosis, and Sarcoidosid and PA mean $\leq 30mm/Hg$ share risk factor parameters with diagnosis group A, Eisenmenger with group B, Obliterativebron-chiolitis, Pulmonary Fibrosis other, and Sarcoidosis and PA mean $> 30mm/Hg$ with group D.

[c] age$>$ 45 spline: the maximum of 0 and age-45.

[d] ADL: activities of daily living.

D waitlist patient described in sections 2.4.1 and 2.4.2 were done for all 3701 patients in our waitlist cohort. Figure 2.4 shows boxplots of estimated transplant benefit by diagnosis group using IPCW PO, traditional PO, integrated IPCW Cox PH sur-

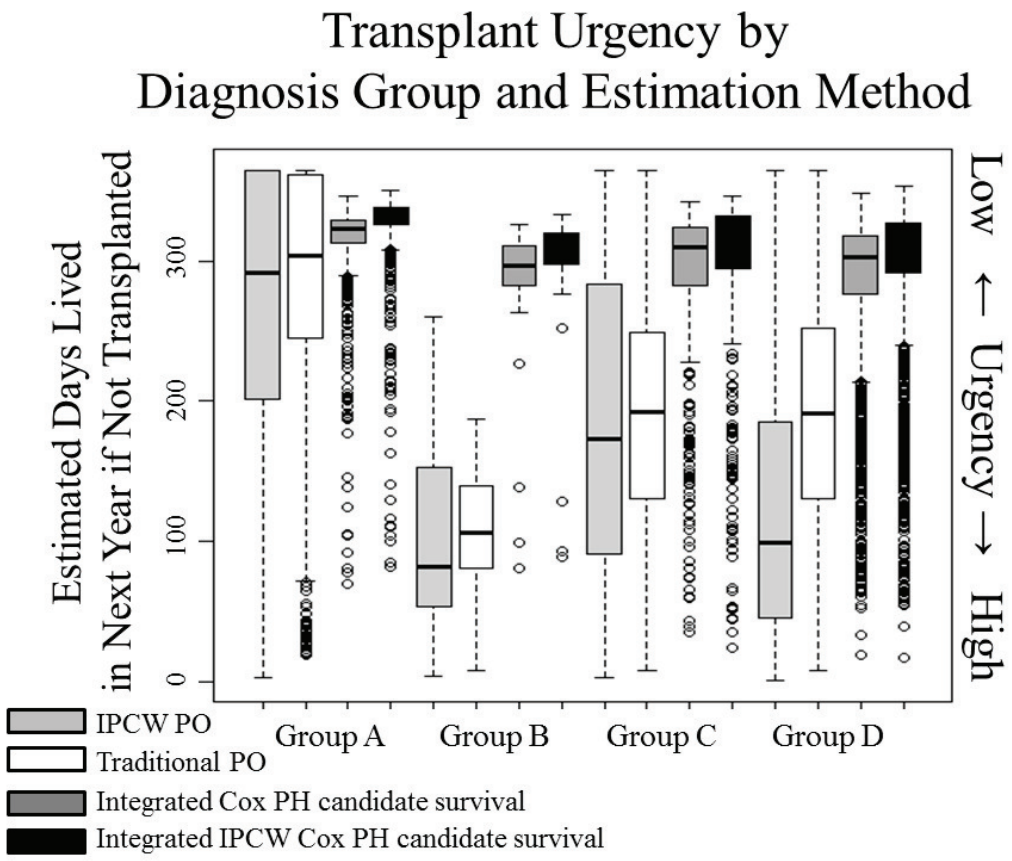


Figure 2.3: Urgency by Diagnosis Group.

vival and integrated Cox PH survival for this cohort. In each case the lung recipient model used model (2.3) applied to the original (perfectly observed) data for this setting. Patient transplant benefit calculations incorporating the IPCW PO model for urgency identified more benefit in Group D patients than when using any other modeling paradigm. Use of integrated IPCW Cox survival curves also exhibit more estimated benefits than use of integrated Cox survival curves not adjusting for dependent censoring. However, benefit estimates remain low, with tight interquartile ranges, when compared to either of the PO methods.

Figure 2.5a shows scatter plots of LAS calculated using IPCW PO versus traditional PO by diagnosis group, with a 45° line superimposed on the plot. For each diagnosis group, the LAS scores change substantially when taking into account

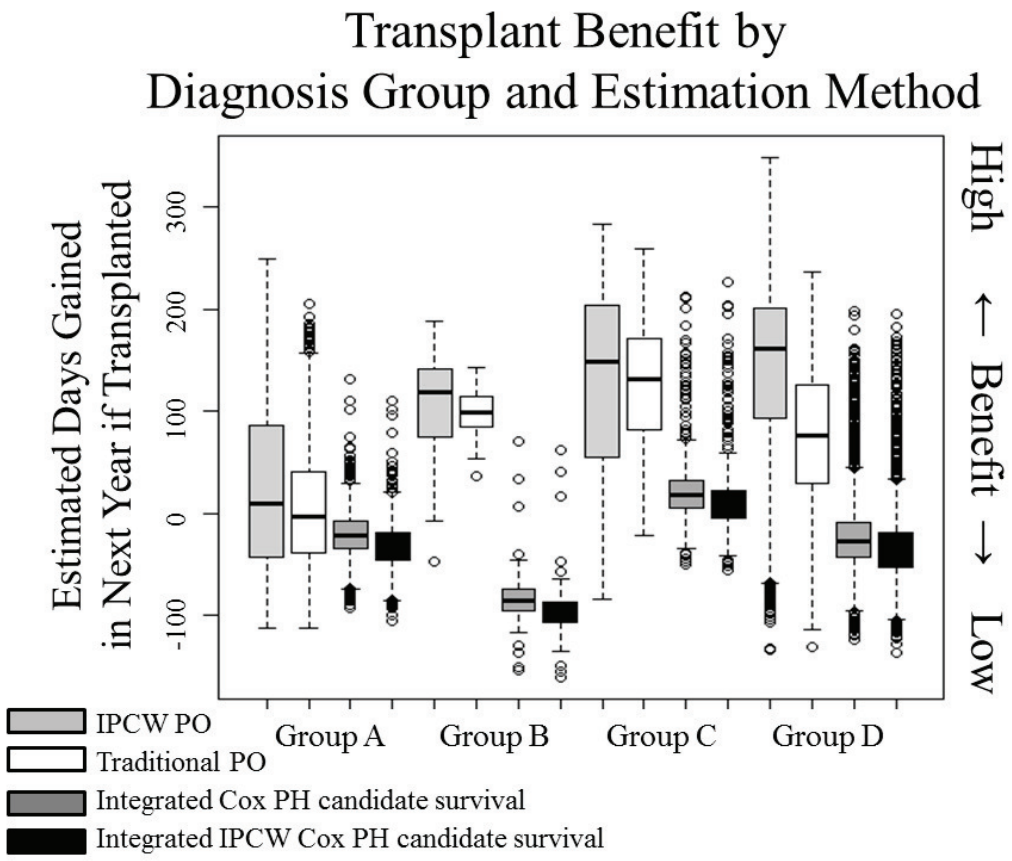


Figure 2.4: Estimated transplant benefit at time of listing by diagnosis group and PO estimation method for 3701 lung transplant candidates.

dependent censoring. When looking at the top 100 ranked patients based on their IPCW PO derived LAS, their scores estimated using traditional PO methods dropped by approximately 16 points on average (0.8 standard deviations of the estimated LAS distribution) when not adjusting for dependent censoring. Similarly scores dropped by approximately 36 points on average using the traditional Cox model integrated waitlist survival curves. Model paradigm selection and adjustment for dependent censoring have a serious impact on time to transplant for those top priority candidates identified using IPCW PO methodology. Figure 2.5b shows a scatter plot of LAS values when calculated using the IPCW PO method vs. using integrated IPCW Cox survival curves for waitlist urgency. Circled values represent patients who would

move from having a low allocation priority using proportional hazards assumptions to a very high allocation priority using model (2.3). Only 30 patients are ranked in the top 100 scores regardless of the IPCW PO method used, IPCOW PO or IPCW Cox integrated survival curves.

2.5 Discussion

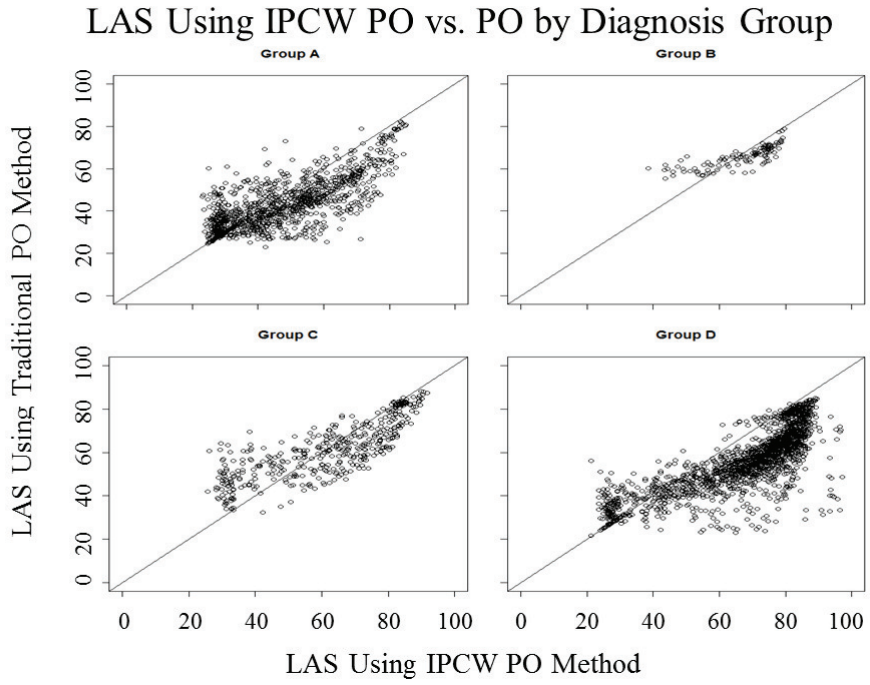
We present new methodology for estimating restricted means in the presence of dependent censoring captured by longitudinal covariates. Upon estimation of $S_T(t)$ using inverse weight methodology, remaining inference becomes very straight forward using our suggested approach. In particular, it is not necessary to program complicated variances of inverse weighted estimates, since the pseudo observation approach merges nicely into use of more standard software package for regression in evaluating parameter estimates. Hence this approach can realistically be implemented by statistical practitioners.

Statistical input into properly modeling components of the LAS in the US has high impact on perhaps 1000 patients at any given time, as the rate of new listings and waitlist removals seems to balance at that level. The success of devising and maintaining an intelligent and practical allocation system for urgent patients introduces a uniquely interesting set of statistical issues. Defining 1-year transplant urgency and benefit at the individual level can be achieved successfully using our described methodology, even when subjected to dependent censoring by transplant for more urgent patients. Hence the LAS can be updated now and in the future using the most recent cohort of patients with minimal bias.

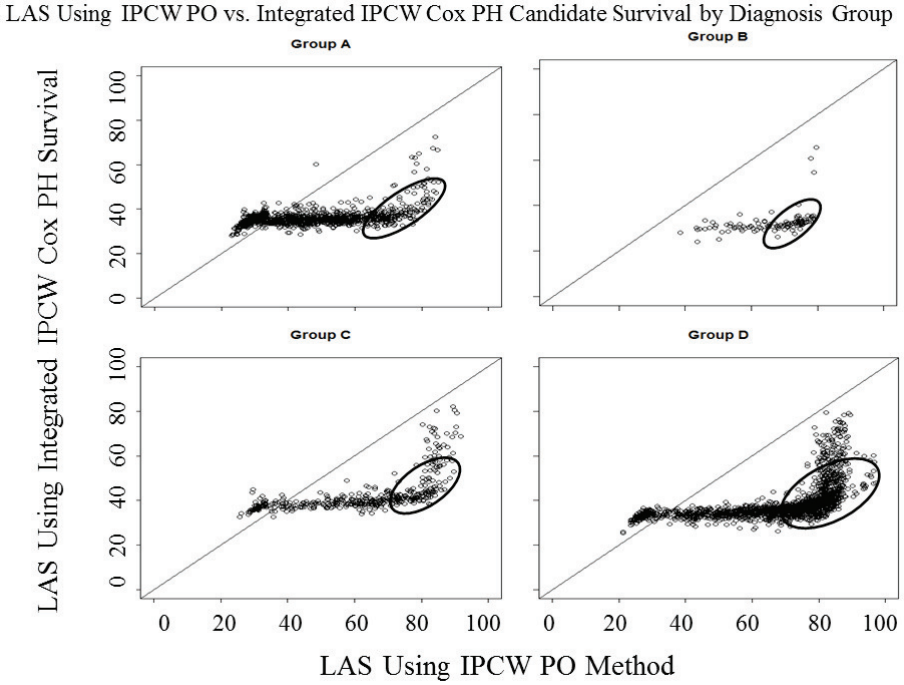
When applied to the lung candidate data, the IPCW PO method gives a broader range of urgency estimates than when estimating urgency based on integrated IPCW

Cox model survival curves. This in turn leads to a broader range of LAS values with which to prioritize the candidates. Our feeling is that parameterization on the scale of the restricted mean leads to more appropriate urgency estimates than parameterization based on constant hazard ratios over time, particularly after viewing the range of scores from using different modeling paradigms in Figure 2.5b.

Availability of this methodology also opens up the important possibility of adding new predictors to the LAS, as these are identified as relevant by the transplant community and collected on OPTN lung transplant candidates. The LAS is the first organ allocation system to explicitly order patients by both estimated urgency and transplant benefit, although liver allocation introduced an urgency score in prioritizing patients around the same time the LAS was developed. No OPTN organ allocation committees have yet updated their algorithms with adjustment for dependent censoring in more recent cohorts of patients. Hence this type of analysis could be applied to other allocation settings with similar dependent censoring issues as well.



(a)



(b)

Figure 2.5: LAS at listing calculated using IPCW PO, PO and Cox PH model by diagnosis group for 3701 lung candidates.

CHAPTER III

MULTIPLE IMPUTATION FOR DEPENDENTLY CENSORED SURVIVAL DATA

3.1 Introduction

In Chapter II, we use the IPCW PO method to estimate transplant urgency and benefit, one of the most important goals of the lung allocation setting. However, compared to the adjusted PO method, MI offers additional advantages. First, instead of being restricted to a single analysis, MI lends itself to survival estimation, two-sample testing, and estimation of restricted means, all within a framework that allows quick and accurate estimation of variability terms. In addition, work by several authors, including Hsu et al. (2002), Liu et al. (2010), and Faucett et al. (2002), has indicated that efficiency and bias reduction can be achieved using MI methods. In particular, Liu et al. (2010) found efficiency gains using an MI-based restricted mean model when compared to a PO-based restricted mean model. In this chapter we continue our analysis of the lung allocation data by developing an MI approach based on restricted mean models that takes into account time-dependent lung allocation scores that induce dependent censoring.

In survival analysis censored observations are one type of missing data, and can thus be imputed via MI. Several authors have suggested approaches for imputation

of such outcomes. Wei and Tanner (1991) applied MI to the analysis of censored regression data, using two general algorithms for the analysis of missing-data problems: the iterative Poor Man's Data Augmentation algorithm and the Asymptotic Data Augmentation algorithm. Faucett, Schenker, and Taylor (2002) used auxiliary variables to recover information from censored observations based on MI with a joint model of a hierarchical change-point model and a time-dependent proportional hazards model. Taylor, Murray and Hsu (2002) proposed nonparametric multiple-imputation methods to handle missing event times for censored observations. Hsu et al. (2006) then proposed an MI method to estimate that uses their nonparametric imputation strategy within risk sets based on similarity of hazards for event times and censoring times. Liu et al. (2010) adopted an MI strategy where risk sets used to select imputes come from a restricted mean model. However their algorithm only accounts for dependent censoring via predictors known at time zero. In the lung transplant setting, the LAS changes across time as a patient progresses, and is highly linked to both a patient's survival times and censoring time. Therefore, it is important to adjust for not only the baseline LAS, but also the changing LAS across time.

In this chapter we propose a multiple imputation algorithm that adjusts for dependent censoring when imputing the censored observations. This algorithm uses a restricted mean model described in Section 2.2.1 when building risk sets. We check the performance of the algorithm through simulation studies. The setting of the simulations is exactly the same as described in Section 2.3 so that comparisons can be made between the IPCW adjusted PO method and our IPCW adjusted MI method. We will also apply the IPCW MI method to the lung transplant patient data, i.e. estimate transplant urgency and benefit by using a multiple imputation approach.

The rest of the chapter is structured as follows: in Section 3.2, we mainly describe the MI method. Section 3.3 shows the performance of our IPCW MI method in simulation studies under the same settings as those conducted in Section 2.3. Section 3.4 provides an analysis of the lung transplant data using MI methods and Section 3.5 provides a brief discussion of the chapter.

3.2 Multiple Imputation Methodology

3.2.1 *Background and Notation*

Let T denote the failure time, C be the corresponding censoring time, $T^* = min(\tau, T)$, $\boldsymbol{Z}$ be some covariates affecting T^*, and $\bar{\boldsymbol{V}}(t) = \{\boldsymbol{V}(u); 0 \leq u \leq t\}$ be the recorded history up to time t of a vector of possibly time dependent covariates $\boldsymbol{V}$ that predict the censoring time C. The choice of $\tau = 1$ year is of interest in the lung allocation example; otherwise it may be any value where $P(C > \tau) > 0$. If C is less than T^*, then we would not be able to observe the restricted failure time. Let $X = min(T^*, C)$ denote the observable random variable, and $\Delta = I(T^* \leq C)$ be the failure indicator variable. In cases where $\boldsymbol{V}$ influences T^* through C, we have dependent censoring and traditional survival analyses are no longer unbiased. In the remainder of the chapter, use of the subscript, i, denotes the previously defined random variable associated with individual i, $i = 1, 2, ..., n$.

Imputation approaches often select imputes from an appropriate model (parametric) or risk set (semi-parametric). When the risk set is selected solely based on being at risk at time C_i, Taylor et al. (2002) give an imputation strategy that corresponds to the Kaplan-Meier estimate in expectation. The idea mimics the inverse transform method often used to simulate outcomes from a particular distribution. Recall the

inverse transform method result that for $U \sim$ Uniform(0,1), $S_T^{-1}(U)$ will follow the distribution of T when $S_T(t)$ is the survival function for T. For a person censored at C_i, Taylor et al. sample $T|T > C_i$ based on an estimate of $S^{-1}_{T|T>C_i}(U)$ that employs the conditional Kaplan-Meier estimate among those at risk at C_i, $\hat{S}_{T|T>C_i}(t)$. In particular, the sampled T is the smallest time that satisfies $\hat{S}_{T|T>C_i}(T) \leq U$, which always corresponds to an observed failure time occurring beyond C_i. The probability a particular T is sampled is equal to the size of the drop in the conditional Kaplan-Meier curve at T.

In the case where censoring is dependent this imputation strategy is biased, since the Kaplan-Meier estimate does not adequately stand in for $S_{T|T>C_i}(t)$ in the inverse transform relationship. Selecting a risk set for a patient censored at C_i based upon $T > C_i$ and additional covariate information can substantially reduce bias. In a follow-up paper Hsu et al. (2006) introduced further constraints, R_i, on the risk set members, requiring similar Cox model hazards to the patient censored at C_i so that the impute is based upon an estimate of $S^{-1}_{T|T>C_i,R_i}(U)$.

We introduce several useful modifications to the inverse transform strategy for imputation to accommodate features of the lung candidate data. First, instead of using a Kaplan-Meier based estimate, we estimate $S^{-1}_{T^*|T^*>C_i,R_i}(U)$ using an IPCW strategy that incorporates time-dependent LAS values after C_i, adjusting for additional dependent censoring within the risk set. We review inverse weight methodology for this purpose in section 3.2.2. In section 3.2.3, we describe how our risk set is chosen. Part of this selection is based on a τ-restricted mean model that decomposes $\log T^*$ into a piece depending on a linear predictor $\boldsymbol{\beta}^T\boldsymbol{Z}$ and a residual term ϵ. This model conveniently provides residuals for each observed failure time that we use subsequently in the imputation procedure described in section 3.2.4. That is,

once T^* is sampled from the modified inverse transform imputation approach, we use the restricted mean model residual corresponding to the sampled T^* and the linear predictor $\boldsymbol{\beta}^T \boldsymbol{Z}_i$ from the censored individual to create the final impute, rather than using T^* directly. A similar strategy of selecting residuals from a linear model was considered by by Schenker and Taylor (1996) within a complete case risk set as well as by Liu et al. (2011). The goal of sampling residuals is to base the variability of the impute on the selected failure time while using a mean failure time attributed to the patient censored at C_i, i.e., the mean predicted from the censored patient's particular risk factors.

3.2.2 *IPCW estimate of survival*

A popular method to account for dependent censoring is to estimate $S(t)$ using $\hat{S}^W(t) = e^{-\hat{\Lambda}^W(t)}$, where $\hat{\Lambda}^W(t)$ is as described by Robins and Finkelstein (2000) and obtained as follows.

Each subject at time t is given a weight, $W_i(t)$, inversely proportional to his or her probability of getting censored after time t, $K_i^{\boldsymbol{V}}(t)$. That is,

$$W_i(t) = 1/K_i^{\boldsymbol{V}}(t) = 1/P(C_i > t|\bar{\boldsymbol{V}}_i(t)).$$

A common strategy, which we also employ, is to estimate the censoring probabilities using a Cox model with time-dependent covariates,

$$\lambda_Q\{t|\bar{\boldsymbol{V}}(t)\} = \lambda_{Q_0}(t)exp\{\gamma'\boldsymbol{V}(t)\}. \tag{3.1}$$

In the case of the lung candidate data, time-dependent covariates in $\boldsymbol{V}(t)$ include the LAS at time t as well as active or inactive waiting status at t. Time independent predictors in $\boldsymbol{V}(t)$ include race, gender, blood type and height at listing. Let $N_{Q_i}(u) = I(X_i \leq u, \Delta_i = 0)$ be the observable counting process for censoring,

$N_{T_i^*}(u) = I(X_i \leq u, \Delta_i = 1)$ be the observable counting process for death, and $Y_i(u) = I(X_i \geq u)$ be the at risk indicator for subject i at time u. Then a consistent estimate of $K_i^V(t)$ is

$$\hat{K}_i^V(t) = exp\{-\sum_{k=1}^{n} \int_0^t \frac{e^{\hat{\gamma}' \boldsymbol{V}_i(u)} dN_{Q_k}(u)}{\sum_{j=1}^n Y_j(u) e^{\hat{\gamma}' \boldsymbol{V}_j(u)}}\},$$

the subject specific weight becomes

$$\hat{W}_i(t) = 1/\hat{K}_i^V(t) = exp\{\sum_{k=1}^{n} \int_0^t \frac{e^{\hat{\gamma}' \boldsymbol{V}_i(u)} dN_{Q_k}(u)}{\sum_{j=1}^n Y_j(u) e^{\hat{\gamma}' \boldsymbol{V}_j(u)}}\},$$

and an IPCW estimator for $\Lambda(t)$ is calculated using

$$\hat{\Lambda}^W(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_{T_i^*}(u) \cdot \hat{W}_i(u)}{\sum_{j=1}^n Y_j(u) \cdot \hat{W}_j(u)}.$$

Although we estimate $S(t)$ using $\hat{S}^W(t) = e^{-\hat{\Lambda}^W(t)}$, alternative inverse weight survival estimates developed by Satten et al. (2001) would also be appropriate. A conditional survival estimate among those in a risk set restricted to those with $T^* > C_i$ who also satisfy constraint, R_i, is given by

$$\hat{S}^W_{T^*|T^*>Ci,R_i}(t) = \frac{\hat{S}^W_{T^*|R_i}(t)}{\hat{S}^W_{T^*|R_i}(C_i)} = \frac{e^{-\hat{\Lambda}^W_{T^*|R_i}(t)}}{e^{-\hat{\Lambda}^W_{T^*|R_i}(C_i)}},$$

where $\hat{S}^W_{T^*|R_i}(t)$ and $\hat{\Lambda}^W_{T^*|R_i}(t)$ are IPCW estimates calculated within the risk set. Additional details of risk set selection are in the following section.

3.2.3 *Risk set selection for individual i censored at $C_i < \tau$*

The first risk set requirement is that members have $T^* > C_i$ so that imputes don't contradict previously observed survival for patient i up to C_i. The majority of this section describes how to define an additional constraint based on incorporating information from covariates, $\boldsymbol{Z}_i$ as well as longitudinal information pertaining to survival in $\boldsymbol{V}_i(t)$. Although it is often possible to select individuals based on a

limited number of risk factors, the presence of many covariates related to survival makes grouping based on a model with a linear predictor more attractive. In the censored data setting, with typically finite follow-up, a common regression model is the restricted mean model

$$E[\log T^*] = \boldsymbol{\beta}^T \boldsymbol{Z}, \tag{3.2}$$

where $\boldsymbol{Z}$ is a vector of time-independent predictors. In the case of the lung candidate data, an individual's LAS urgency component may be estimated according to their risk factors when $T^* = min(1, T)$.

When there is no censoring, standard linear model software is available to fit model (3.2). When censoring is dependent, Xiang and Murray (2012) developed an IPCW modified pseudo observation approach for obtaining estimates $\hat{\boldsymbol{\beta}}^{POW}$ for model (3.2).

In addition to requiring $T^* > C_i$ for those in lung candidate i's risk set, we employ the constraint, R_i, that (a) $\hat{\boldsymbol{\beta}}^{POW\,T}\boldsymbol{Z}$ be within a window of $\hat{\boldsymbol{\beta}}^{POW\,T}\boldsymbol{Z}_i$, where $\boldsymbol{Z}_i$ is the vector of predictors for the individual censored at C_i. This ensures a risk set with similar urgency to individual i at listing. We further impose the constraint that (b) members of the risk set share the same diagnosis group (categorical component of $\boldsymbol{Z}_i$) with (c) similar LAS values at time C_i to individual i (longitudinal component of $\boldsymbol{V}_i(t)$ pertaining to urgency as well as transplant).

3.2.4 *Multiple imputation (MI) of censored observations*

The detailed steps of the algorithm is as follows.

Step 1: For each individual i censored at C_i, we select an appropriate risk set as described in section 3.2.3.

Step 2: Within this risk set, we then calculate the conditional survival probabilities $\hat{S}^W_{T^*|T^*>C_i,R_i}$ as described in section 3.2.2.

Step 3: Generate U, a Uniform(0,1) random variable, and identify the observed restricted failure time T^*_j, where T^*_j is the smallest restricted failure time T^* that satisfies $\hat{S}^W_{T^*|T^*>C_i,R_i}(T^*) \leq U$.

Step 4: Identify the residual ϵ_j from the model $\log T^*_j = \hat{\boldsymbol{\beta}}^{PO^W T}\boldsymbol{Z}_j + \epsilon_j$ corresponding to the T^*_j value selected in Step 3.

Step 5: If $T^*_j = \tau$, we impute T^*_i by τ; otherwise we add the residual of the jth subject, ϵ_j, to $\hat{E}[log(T^*_i)] = \hat{\boldsymbol{\beta}}^{PO^W T}\boldsymbol{Z}_i$, and use this as the imputed value for $log(T^*_i)$. If the imputed $\log(T^*_i) < \log(C_i)$ then repeat from step 3 until the imputed value is greater than $\log(C_i)$.

Step 6: Repeat steps 1-5 until all the censored observations from the observed data set are imputed.

Step 7: Repeat steps 1-6 M times so that we have M completed versions of the observed dataset.

Once M completed data sets are obtained, we may perform analyses using the formulaic approach given by Rubin and Little (1987). We summarize two analyses below that are used in sections 3.2.5 and 3.2.6.

3.2.5 *Restricted mean model analysis on completed datasets*

For each complete dataset, fit model (3.2) with respect to the covariates $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_k)$ to get the parameter estimates $\hat{\boldsymbol{\beta}}_{\boldsymbol{m}} = (\hat{\beta}_{m0}, \hat{\beta}_{m1}, \ldots, \hat{\beta}_{mk})$ with associated variance matrix $\hat{\boldsymbol{W}}_m$, $m = 1, 2, ..., M$.

Our final vector of estimated MI coefficients for the restricted mean model that adjusts for dependent censoring is $\hat{\boldsymbol{\beta}}^{MI^W} = \sum_{m=1}^{M} \hat{\boldsymbol{\beta}}_{\boldsymbol{m}}/M$. The variability associ-

ated with the parameter estimates are composed of two parts, the average within-imputation variance $\bar{\boldsymbol{W}}_M = \sum_{m=1}^{M} \hat{\boldsymbol{W}}_m/M$, and the between-imputation variance $\boldsymbol{B}_M = \sum(\hat{\boldsymbol{\beta}}_m - \hat{\boldsymbol{\beta}}^{MI^W})^2/(M-1)$. The variance of $\hat{\boldsymbol{\beta}}^{MI^W}$ is then $\boldsymbol{V} = \bar{\boldsymbol{W}}_M + (1 - M^{-1})\boldsymbol{B}_M$.

3.2.6 *Marginal survival analysis on completed datasets*

For each complete dataset, we calculate the sample proportions corresponding to the survival estimates $\hat{S}_1(t), \hat{S}_2(t), ..., \hat{S}_M(t)$, and their sample proportion variances $\hat{V}_1(t), \hat{V}_2(t), ..., \hat{V}_M(t)$, where $\hat{V}_m(t) = \hat{S}_m(t)(1 - \hat{S}_m(t))/n$, $m = 1, 2, ..., M$. Our final MI survival estimate that adjusts for dependent censoring is $\hat{S}^{MI^W}(t) = \sum_{m=1}^{M} \hat{S}_m(t)/M$ with estimated variance

$$\hat{V}^{MI^W}(t) = M^{-1} \sum_{m=1}^{M} \hat{V}_m(t) + (1 + M^{-1}) \sum_{m=1}^{M} [\hat{S}_m(t) - \hat{S}^{MI^W}(t)]^2/(M-1).$$

3.3 Simulation Study

To study our multiple imputation method in finite sample sizes, we conducted a simulation study where a time-dependent variable influences censoring and survival and the mean structure follows (3.2) with $T^* = min(5, T)$. Parameter estimates for (3.2) are calculated in cases when (a) $\log T^*$ is uncensored; (b) $\log T^*$ is subject to censoring and is replaced by log-transformed pseudo observations defined by Andersen, Hansen and Klein (2004); (c) $\log T^*$ is subject to censoring and is replaced by IPCW pseudo observations as described by Xiang and Murray (2012) and (d) $\log T^*$ is subject to censoring and is imputed as described in Section 3.2.4.

The multiple imputation approach allows for more possible analyses than merely fitting model (3.2), hence as an example of an additional analysis of interest, the strategy in (d) for computing point estimates for survival is compared to the Kaplan-

Meier estimate.

In each simulation, we perform the following:

Step 1: We generate Z_0 from a Bernoulli(0.5), Z_1 from a Bernoulli(0.5), and Z_2 from a Uniform(0,1), where Z_0 and Z_2 are measured at time 0 and Z_1 is a time-dependent covariate measured at time $t_1 = 0.2$.

Step 2: Failure times, T, are generated from piecewise exponential distributions, i.e. T has a constant hazard λ_{z_0} before time t_1 that changes to $\lambda_{z_0 z_1}$ after time t_1, where $\lambda_0 = 0.3$, $\lambda_1 = 0.2$, $\lambda_{01} = 0.1$, $\lambda_{11} = 0.5$ are fixed and λ_{00} and λ_{10} are solved so that the mean structure $E[\log T^*] = \beta_0 + \beta_1 Z_0 + \beta_2 Z_2$ is satisfied for a pre-specified $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$. That is, although T^* is influenced by the time dependent covariate, Z_1, the restricted mean of interest is captured by baseline predictors Z_0 and Z_2. Further details on solving for parameters that satisfy the mean structure can be found in the appendix of Xiang and Murray (2012), where a similar simulation strategy is used.

Step 3: Piecewise constant hazards leading to dependent censoring times, C, are based on the Cox model $\lambda^C(t|\bar{\boldsymbol{Z}}(\boldsymbol{t})) = \lambda_0^C(t) \exp\{\gamma_0 Z_0 + \gamma_1 I[Z_0 = 0, Z_1 = 1, t > t_1] + \gamma_2 I[Z_0 = 1, Z_1 = 0, t > t_1] + \gamma_3 I[Z_0 = 1, Z_1 = 1, t > t_1] + \gamma_4 Z_2\}$, where $\lambda_0^C(t) = 0.15$ for $t \leq t_1$ and $\lambda_0^C(t) = 0.4$ for $t > t_1$, $\gamma_0 = 0.3, \gamma_1 = -1.4, \gamma_2 = 0.5, \gamma_3 = -1.5$, and $\gamma_4 = 1$. This causes censoring to be influenced by both the first and the second covariates in the mean structure, Z_0 and Z_2. And Z_1, while not directly influencing the form of the mean structure, is very much tied to both the time-to-event and censoring mechanisms.

In interpreting results, it is useful to consider the degree of bias. We consider two scenarios: $\boldsymbol{\beta} = (\beta_0 = 0.8, \beta_1 = \beta_2 = 0)$ (i.e., true baseline covariate effects are zero) and $\boldsymbol{\beta} = (\beta_0 = 1, \beta_1 = -0.8, \beta_2 = -0.5)$ (i.e., non-zero baseline covariate effects). In

the first scenario modest bias is produced overall with censored individuals, approximately 50%, tending to have longer times-to-event. In the second scenario, overall bias is reduced from scenario 1 with censored individuals, approximately 36%, tending to have shorter times-to-event. Methods that ignore potential bias based on any covariates (i.e., unadjusted PO method, Kaplan-Meier) will suffer in both scenarios 1 and 2. For each scenario, 1000 simulations are run with $n = 150$ subjects.

We begin by summarizing results under scenario 1. Results for fitting model (3.2) under scenario 1 ($\beta_0 = 0.8, \beta_1 = \beta_2 = 0$) are located in part (A) of Table 3.1. The PO method that doesn't adjust for dependent censoring gives biased estimates for all parameters, with bias$(\hat{\beta}_0^{PO}) >$ bias$(\hat{\beta}_1^{PO}) >$ bias$(\hat{\beta}_2^{PO})$. Parameter estimates, $\hat{\boldsymbol{\beta}}^{PO^W}$, using the IPCW-adjusted PO method, labeled PO^W, have comparably small bias, with a slight tendency to underestimate β_0. The IPCW adjusted MI method, labeled MI^W, overall has the best track record of removing bias, particularly for the intercept term.

Recall that in setting (1), censored individuals tend to have longer times-to-event so that methods struggling with bias will tend to underestimate restricted lifetimes. In the case with no censoring, i.e. no bias, the proportion of times model (3.2) estimates a lower restricted lifetime than the true time-to-event is 31%. This percentage increases as bias from estimation of model (3.2) increases. The proportion of times that estimated restricted means undershoot the observed times-to-event is highest for PO (55%), followed by PO^W, and MI^W (34%, and 31% respectively).

For an individual with $Z_0 = 1$ and $Z_2 = 0.5$, the PO and PO^W methods underestimate the time lived during the 5-year period by 16 months and 1.5 months on average while the MI^W method is off by less than 1 day over the 5-year period.

Empirical standard deviations are comparable to standard errors averaged across

Table 3.1: Comparison of estimates using uncensored observations (Uncensored), unadjusted pseudo observation (PO), IPCW-adjusted PO (PO^W), and IPCW-adjusted MI (MI^W) under two scenarios.

Parameters	Uncensored	PO	PO^W	MI^W
(A) Covariate effects are zero				
$\beta_0 = 0.8$	0.816 [0.016, 0.210, 0.216] [a]	0.149 [-0.651, 0.421, 0.569]	0.709 [-0.091, 0.245, 0.229]	0.839 [0.039, 0.215, 0.233]
$\beta_1 = 0$	-0.001 [-0.001, 0.188, 0.183]	-0.291 [-0.291, 0.376, 0.374]	0.043 [0.043, 0.219, 0.208]	-0.018 [-0.018, 0.194, 0.205]
$\beta_2 = 0$	-0.015 [-0.015, 0.326, 0.325]	0.135 [0.135, 0.652, 0.681]	-0.019 [-0.019, 0.380, 0.364]	-0.045 [-0.045, 0.335, 0.366]
(B) Covariate effects are non-zero				
$\beta_0 = 1$	0.996 [-0.004, 0.201, 0.196]	0.745 [-0.255, 0.316, 0.313]	0.885 [-0.115, 0.238, 0.215]	0.991 [-0.009, 0.206, 0.215]
$\beta_1 = -0.8$	-0.803 [-0.003, 0.180, 0.180]	-1.037 [-0.237, 0.283, 0.349]	-0.794 [0.006, 0.214, 0.197]	-0.801 [-0.001, 0.186, 0.197]
$\beta_2 = -0.5$	-0.490 [0.010, 0.312, 0.304]	-0.451 [0.049, 0.490, 0.508]	-0.421 [0.079, 0.370, 0.345]	-0.472 [0.028, 0.321, 0.354]

[a] [bias, average standard error based on individual analysis, empirical standard deviation of parameter estimates across 1000 simulations]

simulations with the exception of the intercept term for the PO method. We also investigated a version of the MI^W method that incorporated a bootstrap step, as this sometimes improves variance estimation in MI procedures. However, no tangible improvement in variance estimation was observed from adding this extra step to the procedure (data not shown).

In addition to fitting model (3.2), the MI^W procedure can perform other traditional analyses of interest. For instance, the MI^W approach gives nearly unbiased estimates for survival for all t shown in Table 3.2 part (A) for scenario 1. This table gives marginal survival probability estimates at years 1, 2, 3, 4 and 5. An analysis based on an uncensored version of the data is given as well as analysis for the censored version of the data using Kaplan-Meier survival. At 5 years, the KM underestimates 5-year survival by 7%.

Recall that in scenario (2), $(\beta_0 = 1, \beta_1 = -0.8, \beta_2 = -0.5)$, censored patients tended to have shorter times-to-event. The only method that does not properly adjust for dependent censoring captured by baseline covariates, PO, once again gives strong bias for β_0 and β_1 (see Table 3.1 (B)). For the remaining procedures (PO^W, MI^W), the magnitude of bias is minimal. The overall degree of bias for estimating the time lived during the 5 year period is approximately 4 months for the PO method, 1 month for the PO^W method, and 1 day for the MI^W method over the 5 year period for a patient with $Z_0 = 1, Z_2 = 0.5$. Table 3.2(B) shows that bias in survival estimates based on KM as opposed to MI^W, does not exceed a roughly 5% overestimate of $S(t)$ in this case.

Table 3.2: Comparison of survival estimates using uncensored observations ($\hat{S}^{Uncen}(t)$), Kaplan Meier ($\hat{S}^{KM}(t)$) and IPCW-adjusted MI ($\hat{S}^{MI^W}(t)$) under two scenarios.

t	$S(t)$	$\hat{S}^{Uncen}(t)$	$\hat{S}^{KM}(t)$	$\hat{S}^{MI^W}(t)$
(A) Covariate effects are zero				
1	0.806	0.806 [0.000, 0.032, 0.033] [a]	0.798 [-0.008, 0.035,0.037]	0.805 [-0.001, 0.033, 0.036]
2	0.664	0.671 [0.007, 0.038, 0.039]	0.645 [-0.019, 0.046, 0.047]	0.672 [0.008, 0.041, 0.046]
3	0.569	0.574 [0.005, 0.040, 0.040]	0.531 [-0.038, 0.052, 0.053]	0.576 [0.007, 0.043, 0.054]
4	0.501	0.501 [0.000, 0.041, 0.040]	0.442 [-0.059, 0.056, 0.055]	0.501 [0.000, 0.044, 0.065]
5	0.450	0.445 [-0.005, 0.040, 0.040]	0.377 [-0.073, 0.058, 0.058]	0.446 [-0.004, 0.043, 0.073]
(B) Covariate effects are non-zero				
1	0.648	0.645 [-0.003, 0.039, 0.037]	0.667 [0.019, 0.042, 0.041]	0.647 [-0.001, 0.041, 0.043]
2	0.458	0.456 [-0.002, 0.041, 0.039]	0.495 [0.037, 0.047, 0.045]	0.457 [-0.001, 0.043, 0.048]
3	0.350	0.346 [-0.004, 0.039, 0.038]	0.391 [0.041, 0.048, 0.047]	0.346 [-0.004, 0.041, 0.047]
4	0.279	0.275 [-0.003, 0.036, 0.036]	0.322 [0.043, 0.048, 0.048]	0.279 [0.000, 0.039, 0.048]
5	0.230	0.227 [-0.003, 0.034, 0.034]	0.273 [0.043, 0.048, 0.049]	0.234 [0.004, 0.037, 0.049]

[a] [bias, average standard error based on individual analysis, empirical standard deviation of parameter estimates across 1000 simulations]

3.4 Example

We now return to the OPTN lung transplant setting. Several analyses are of interest. For instance, physicians treating patients typically want to know the survival distribution of listed lung candidates as they await transplantation. Transplant urgency and benefit by individual risk factors are also of interest, and lung allocation scores based on these.

3.4.1 *Lung candidate analysis and urgency*

Our lung waitlist cohort consists of 3701 candidates aged 12 or older. During the first year after listing, the censoring percentage is 73% overall, with differing rates by diagnosis. Group D, made up of interstitial pulmonary fibrosis (IPF) and other restrictive lung disease, has the highest censoring percentage (77%), followed by Group A (primarily obstructive pulmonary disease, 70% censoring), Group C (cystic fibrosis, 69% censoring) and Group B (primarily idiopathic pulmonary arterial hypertension, 58% censoring).

Inverse weights used in our MI procedure are based on a time-dependent Cox model for time to transplant including patients' daily updated LAS, sex, race, blood type, status (active, inactive, offlist), and height, as given in equation (3.1). Parameter estimates are displayed in 2.3. Although one might suppose that the probability of transplant should increase monotonically with higher LAS value, the parameter estimates from Table 2.2 suggest otherwise, likely because patients with high LAS values are also more likely to die before an organ offer manifests.

The imputation risk set for the i^{th} censored candidate is found by choosing candidates at risk at time C_i who have a similar LAS value at C_i and are in the same diagnosis group as the censored patient i. The risk set is further restricted to hav-

ing a similar urgency estimate based on $\hat{\boldsymbol{\beta}}^{PO^W}$ as shown in the second column of Table 3.3. The predictors used in this model precisely match those proposed by the OPTN Thoracic Committee (OPTN Thoracic Organ Transplantation meeting minutes March 23, 2010); some predictors are included based on significant association with survival seen in previously studied cohorts.

Ten imputed datasets were built from the MI procedure outlined in section 3.2.4. Estimates, $\hat{\boldsymbol{\beta}}^{MI^W}$, based on the imputed datasets are shown in the rightmost column of Table 3.3. Differences seen in parameter estimates using the PO^W and MI^W methods are typically minor with the exception of the intercept term, which is lower using the PO^W method than the MI^W method. This pattern was also observed to some degree in the simulation section, where the MI^W method was seen to estimate the true intercept with less bias.

Parameters in the restricted mean model act multiplicatively on the number of days lived in a year. For instance the estimated number of days lived is $467.34 \times 0.78 \times 1.01^{25} \times 1.01 \times 1.02^3 \times 0.92^2 \times 0.99^{50} \times 0.86^{0.8} = 227$ days based on MI^W, and $102.15 \times 0.57 \times 1.07^{25} \times 1.08 \times 1.07^3 \times 0.87^2 \times 0.60^{0.8} = 210$ days based on PO^W for a 55 year old diagnosis group C patient, who has a BMI of 25, has no diabetes, requires no assistance with ADL, walks 300 feet in six minutes, requires $2L/min$ of O_2 at rest, has PCO_2 of $50mmHg$, is not on a ventilator, has a stable creatinine of $0.8mg/dl$ and has a cardiac index $> 2L/min/min^2$.

The PO^W method seems to estimate less waitlist days lived in a year compared to the MI^W method. For comparison we also include urgency estimates built from an integrated IPCW Cox model (Robins and Rotnitzky (1992), Robins (1993)), where

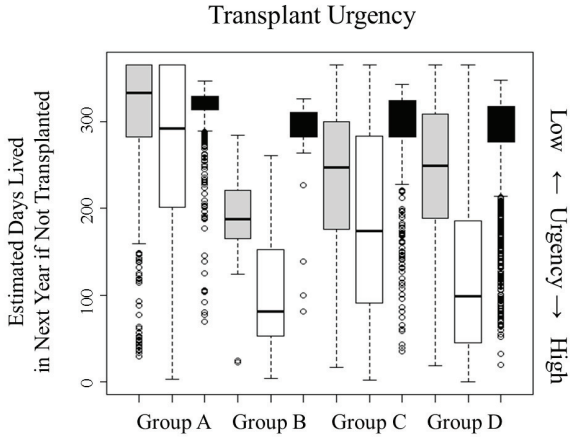
the partial likelihood score function incorporating censoring weights becomes

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} W_i(X_i)\Delta_i\{\boldsymbol{Z}_i(X_i) - \frac{\sum_{i=1}^{n} W_i(X_i)Y_i(X_i)e^{\boldsymbol{\beta}\boldsymbol{Z}_i(X_i)}}{\sum_{i=1}^{n} W_i(X_i)Y_i(X_i)e^{\boldsymbol{\beta}\boldsymbol{Z}_i(X_i)}\boldsymbol{Z}_i(X_i)}\}.$$

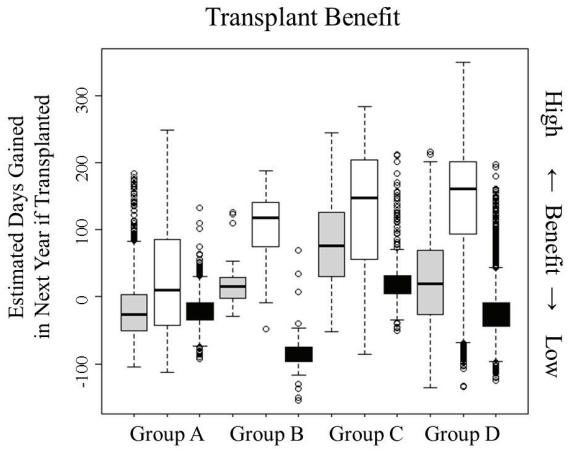
Parameter estimates for the IPCW Cox model are located in Table 2.5. Boxplots of patient urgency by diagnosis group are located in Figure 3.1a. Although all three methods account for dependent censoring, differences in urgency are seen based on method used. Methods based on restricted mean models seem to give a broader range of estimates than urgency estimated via integrated IPCW Cox-model hazards. The IPCW Cox urgency estimates also tend to give the highest number of days lived during first year of listing.

A few model diagnostics were performed to assess the value of the different methods used. For instance, for censored candidates we would hope to predict a larger number of days lived than the observed censoring time for an individual. PO^W urgency estimates undershoot their observed censoring times 35% of the time, while the proportions are 20% and 16% for the IPCW Cox and MI^W methods, respectively. Among observed failures, the sum of squared residuals between the observed (restricted) failure times (urgencies) and the predicted days lived during the first year of listing is smallest (10,501,085) for the MI^W based estimates, followed by the IPCW Cox urgency estimates (15,373,491) and the PO^W estimates (17,921,077). We also look at the concordance index (Harrell et al. 1982; Harrell, Lee, and Mark 1996), which is the percent of times that a model predicts two pairs of data $(X_i, \Delta_i), (X_j, \Delta_j)$ in the correct order, where the correct ordering is observable. The numbers are $80\%, 77\%$, and 79% for the MI^W, PO^W and IPCW Cox methods, respectively.

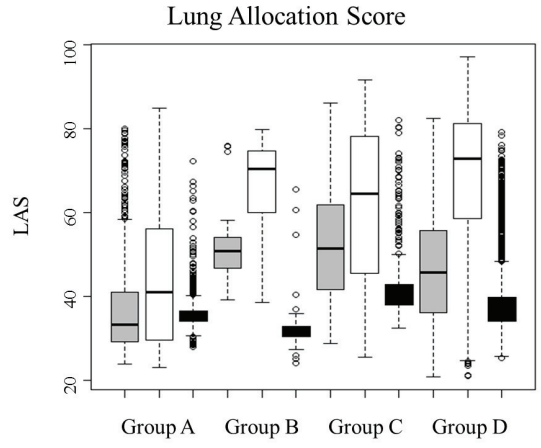
Marginal survival analysis shows lower survival curve estimates in group D using MI^W versus the KM (Figure 3.2). This group also experiences the shortest average



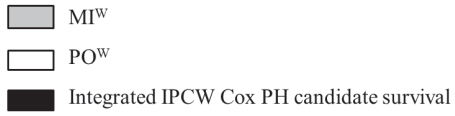
(a)



(b)



(c)



(d) Legend

Figure 3.1: Estimated transplant urgency, benefit, and LAS at time of listing by diagnosis group and estimation method for 3701 lung transplant candidates. High LAS values get earliest lung offers.

times to transplant (71 days for group D, versus 126 days for group C, 170 days for group A, and 221 days for group B (Yusen et al. 2010)).

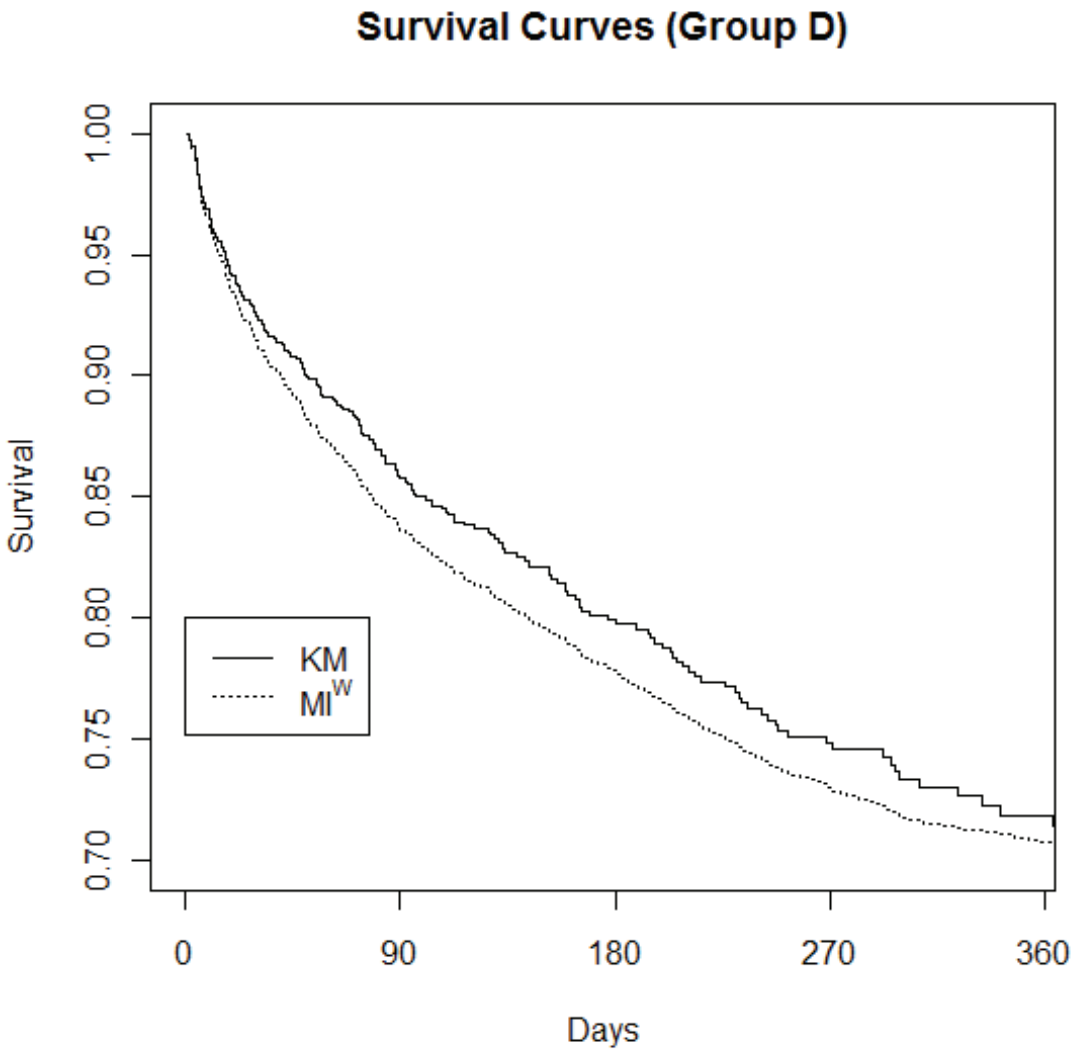


Figure 3.2: Survival Curves for Group D patients using KM and MI^W.

3.4.2 *Lung recipient analysis, transplant benefit, and LAS*

Recall that transplant benefit is calculated by subtracting the estimated days lived without transplant in a year (candidate urgency) from the estimated days lived in a year following transplant. Our lung transplant cohort consists of 4784 patients

aged 12 or older. All patients were followed for at least one year from the time of transplant, i.e., no censoring of the 1-year restricted mean, with 816 (17%) deaths within that first year. Results from fitting model (3.2) in this uncensored case are shown in Table 2.6. For the same group C waitlist patient described in section 4.1, the estimated days lived in the first year following transplant is 266 days based on the model in Table 2.6. So the PO^W method estimates 266 - 210 = 56 days gained during the first year after a transplant for this patient, while the MI^W method estimates a gain of of 266 - 227 = 39 days during that year.

Figure 3.1b shows boxplots of estimated transplant benefit for all 3701 patients in our waitlist cohort by diagnosis group using MI^W, PO^W and IPCW Cox methods. Patient transplant benefit calculations incorporating the PO^W method tend to estimate higher benefit, going along with their lower estimates of waitlist days lived without transplant. LAS values using the MI^W, PO^W and IPCW Cox methods are shown in Figure 3.1c.

3.5 Discussion

The LAS system has been successful in prioritizing candidates for a lung transplant, however, the fact that more urgent patients are more likely to be selected for transplant has led to a statistical issue of dependent censoring. Hence all analyses based on the waitlist cohort must have some adjustment to avoid bias. We provide a useful approach for creating multiply imputed datasets that adjust for dependent censoring. Availability of these completed datasets, along with standard ways of combining results from complete case analyses, allows convenient and quick additional analyses to be conducted as they arise.

Simulations indicate that marginal estimates of survival and estimates of re-

stricted means, both very commonly explored in the lung candidate cohort, perform well using the imputation strategy proposed. Our example also indicates improved model prediction over both an existing PO^W approach and an IPCW Cox approach.

Our analyses indicate that the choice of modeling paradigm strongly influences LAS values and therefore a patient's chances of receiving an organ offer in time. Use of the IPCW Cox method gives a more narrow distribution of LAS values, while MI^W and PO^W approaches give a broader range of scores. The PO^W approach seems to inflate LAS scores somewhat compared to the others.

The LAS affects roughly 1000 lung candidates at any one time. The ability to update the score appropriately is critical in providing fairness to candidates as concomitant care continues to evolve and new markers are discovered that are worthy of inclusion in the LAS algorithm. Our MI^W approach can successfully navigate dependent censoring issues on the lung waitlist and is also appropriate in other settings where longitudinal factors influence dropout as well as survival over time.

Table 3.3: Lung waitlist models using 2 different methods for 3701 lung candidates.

	PO^W	MI^W
	$e^{\hat{\beta}}$ [a] (95% CI), p-value	$e^{\hat{\beta}}$ (95% CI, p-value)
(Intercept)	102.15 (15.96, 653.55), < 0.0001	467.34 (326.04, 669.89), < 0.0001
Diagnosis Group (ref=Group A, primarily COPD)		
Group B (primarily iPAH)	0.19 (0.05, 0.73), 0.0158	0.54 (0.42,0.70), < 0.0001
Group C (primarily CF)	0.57 (0.17, 1.90), 0.3267	0.78 (0.62, 0.99), 0.0418
Group D (primarily IPF)	0.11 (0.03, 0.38), 0.0004	0.57 (0.45, 0.72), < 0.0001
Diagnosis [b]		
Bronchiectasis	0.76 (0.21, 2.72), 0.6694	0.94 (0.75, 1.17), 0.5730
Lymphangioleiomyomatosis	1.25 (0.10, 16.41), 0.8639	1.03 (0.69, 1.55), 0.8729
Obliterativebronchiolitis	0.25 (0.03, 1.86), 0.1744	1.14 (0.79, 1.65), 0.4892
Pulmonary Fibrosis other	0.77 (0.37, 1.59), 0.4741	0.97 (0.85, 1.11), 0.6570
Sarcoidosis and PA mean $> 30mm/Hg$	0.46 (0.16, 1.37), 0.1641	1.18 (0.95, 1.46), 0.1280
Sarcoidosis and PA mean $\leq 30mm/Hg$	0.84 (0.22, 3.13), 0.7893	0.85 (0.69, 1.06), 0.1454
Physiologic Reserve		
Age (yrs)	1.00 (0.98, 1.02), 0.9919	1.00 (0.99, 1.00), 0.3583
BMI (kg/m^2)	1.07 (1.03, 1.11), 0.0011	1.01 (1.01, 1.02), 0.0010
Diabetes	0.49 (0.33, 0.73), 0.0005	0.90 (0.84, 0.97), 0.0035
No Assistance with ADL [c]	1.08 (0.65, 1.80), 0.7609	1.01 (0.93, 1.10), 0.8762
Six minute walk (per $100ft$)	1.07 (1.03, 1.12), 0.0012	1.02 (1.01, 1.03), < 0.0001
Severity		
FVC for Group D (per 10% predicted)	1.25 (1.08, 1.43), 0.0019	1.05 (1.02, 1.08), 0.0003
O_2 requirement for Group A,C,D (L/min)	0.87 (0.83, 0.91), < 0.0001	0.92 (0.91, 0.93), < 0.0001
PA systolic (per $10mm/Hg$) for Group A	0.91 (0.71, 1.16), 0.4331	0.99 (0.95, 1.04), 0.6922
PCO_2 increase of $\geq 15\%$	1.03 (0.41, 2.59), 0.9449	1.13 (0.95, 1.35), 0.1733
PCO_2 ($mmHg$)	1.00 (0.99, 1.02), 0.6560	0.99 (0.99, 1.00), 0.0010
Ventilator	0.07 (0.03, 0.20), < 0.0001	0.16 (0.13, 0.20), < 0.0001
Creatinine (mg/dl)	0.60 (0.32, 1.12), 0.1111	0.86 (0.76, 0.96), 0.0078
Cardiac Index $< 2.0(L/min/min^2)$	0.48 (0.25, 0.94), 0.0315	0.81 (0.72, 0.92), 0.0007

[a]For risk factors, $e^{\hat{\beta}}$ acts multiplicatively on the number of days lived in a year.

[b]These diagnoses were grouped into larger diagnosis groups (A, B, C, D) by the OPTN Thoracic Committee for purpose of modeling risk factors that may vary by diagnosis group. Bronchiectasis, Lymphangioleiomyomatosis, and Sarcoidosid and PA mean $\leq 30mm/Hg$ share risk factor parameters with diagnosis group A, Eisenmenger with group B, Obliterativebronchiolitis, Pulmonary Fibrosis other, and Sarcoidosis and PA mean $> 30mm/Hg$ with group D.

[c]ADL: activities of daily living; ref= Some/total assistance with ADL

CHAPTER IV

MULTIPLE IMPUTATION FOR QUALITY-ADJUSTED SURVIVAL DATA

4.1 Introduction

Quality-adjusted lifetime (QAL) analysis is of interest in settings where there is a trade-off between survival benefit and QAL benefit in recommending a therapy. A common example occurs in breast cancer research, where initial investments in radiation and/or chemotherapy early on can reap rewards of longer life down the line. Patients first experience treatment toxicity (TOX), then go through a disease-free period known as time without symptoms or toxicity (TWiST) and finally experience a period of cancer relapse (REL), if not precluded by death. Such is the case in the IBCSG Ludwig Trial V clinical trial of adjuvant chemoendocrine therapy for stage II breast cancer patients. The study compares a short regimen of perioperative systemic treatment versus a prolonged adjuvant therapy regimen. Longer therapy was hypothesized to prolong lifetime, but at the cost of a longer TOX period. A natural, patient-oriented, question is whether the gains in survival are worth the extra time spent enduring the unpleasant side effects of treatment.

Quality-adjusted survival analysis takes patients' quality of life (QOL) into consideration when assessing a time-to-event. Utility scores ranging from 0 to 1 assign

partial credit for time spent in different health states. Smaller utility scores imply less preference for a health state, with 0 credit being the harshest penalty for quality of life in a state; indicating either a death state or a life state worth giving up entirely rather than enduring for a moment. When data are uncensored, a quality-adjusted survival time is constructed by summing the utility scores times the length of their associated health states. The resulting QAL, QT, is always less than or equal to the unadjusted lifetime, T. In this way both quantity and quality of life are considered in an analysis of the combined endpoint.

When there is no censoring, one may estimate the distribution of QT using sample proportions and $E(QT)$ using a standard sample mean. Multivariate regression on QT is also available using standard software for generalized linear models. However in the more common case where follow-up is incomplete, Gelber, Gelman and Goldhirsch (1989) documented dependent censoring problems associated with shrinking the time scale via QOL utilities as described above. A very nice summary of this type of bias is given by Glasziou, Simes and Gelber (1990). Briefly, those with longer follow-up times on the standard lifetime scale accumulate quality-adjusted survival more easily on the shifted quality-adjusted lifetime scale. For example, it is easier to observe an individual accumulating a single QAL year when followed for 10 years as opposed to one year. Overestimation of restricted means and other QAL summary measures often occurs when no adjustment is made.

Numerous authors have developed QAL analyses, appropriate in the presence of censoring, that pertain to the distribution of QAL, its restricted mean or regression models linked to parameters of its distribution. Many of the earliest contributions to this area built upon restricted mean QAL as envisioned by Gelber and Goldhirsch with their various coauthors. They introduce the quality adjusted end-

points, TWiST (1986, 1987, 1989), and later Q-TWiST (Quality-Adjusted Time Without Symptoms or Toxicity, 1989). When estimating restricted mean QAL using a Q-TWiST oriented method, one first estimates restricted means for lifetime spent within partitioned health states (unadjusted for quality-of-life). When health transitions are based on ordered times-to-event, such as time to end of toxicity, T_{TOX}, time to relapse, T_{REL} and time to death, T_{OS}, one may estimate health state specific restricted means by taking differences between estimated survival curves based on these events. Marginal restricted mean QAL estimates are then obtained by summing state-specific restricted mean lifetimes weighted by their associated utility scores. Glasziou, Simes, and Gelber (1990) constructed Q-TWiST styled estimators using differences between integrated Kaplan-Meier curves to estimate restricted mean lifetimes within partitioned health states and then performed two-sample treatment difference tests for different choices of utility scores. Cole, Gelber, and Goldhirsch (1993) used a Q-TWiST approach where restricted mean lifetimes in each state were estimated via integrated multivariate Cox survival curves, allowing for covariate influence in estimation. Cole, Gelber and Anderson (1994) suggested a semi-Markov stochastic process representation of transitions between partitioned health states and estimated restricted mean lifetimes within each health state using a parametric accelerated failure time regression approach. Gelber, Cole, Gelber and Goldhirsch (1995) give a nice review of many of the Q-TWiST oriented approaches mentioned above. Many additional authors over the years have applied and broadened Q-TWiST flavored, utility weighted combinations of restricted mean lifetimes within partitioned health states and this approach to estimating restricted mean QAL continues to be popular.

An important contribution to quality-of-life adjusted censored survival method-

ology literature was Zhao and Tsiatis' method for consistently estimating the distribution of QT, where an inverse weight approach modified for the censored QAL setting is used to account for the dependent censoring bias observed by previous authors (Zhao and Tsiatis 1997, 1999). In addition to being able to graphically display and make inference on estimated survival functions for QT, estimation of restricted mean QAL reduced to integrating Zhao and Tsiatis' estimate rather than estimating, weighting and summing restricted mean lifetimes separately for each partitioned health state as had been the standard practice using variations of Q-TWiST. In follow-up papers, Zhao and Tsiatis proposed a broader class of estimators for restricted mean QAL (2000) and developed two-sample tests (2001), again working on the QT timescale. Since their initial work, more authors have moved toward using similar inverse weight approaches that allow estimation, testing and regression directly on censored QT outcomes rather than estimation, testing and regression within separate health states. Among authors working along these lines, Bang and Tsiatis (2002) developed regression models for median QAL, Wang and Zhao (2007) developed regression models for restricted mean QAL using weighted estimating equations and Andrei and Murray (2007) proposed regression models for restricted mean QAL using a pseudo observation (PO) approach.

When a practitioner desires a particular type of QAL analysis, a literature search can often turn up something appropriate. An email to the authors may result in obtaining useful software, or software close to what is required with modification. In cases where a literature search does not provide a satisfactory answer, however, one is left to develop needed methodology and software without much assistance. An alternative to obtaining or developing software for each specific type of desired analysis in the QAL setting is to create uncensored versions of the QAL outcomes,

QT, via a multiple imputation approach and take advantage of the extensive library of available software that may be applied to uncensored outcomes.

In the censored survival setting, multiple imputation approaches have been successfully developed by many authors including Wei and Tanner (1991), Taylor, Murray and Hsu (2002), Faucett Schenker, and Taylor (2002), Hsu, Taylor, and Murray (2006) and Liu, Murray, and Tsodikov (2011). But to date, we are not aware of any methods that extend the advantages of the multiple imputation approach to the censored QAL setting.

In this chapter we describe a multiple imputation approach appropriate for use with censored QAL outcomes. We incorporate two strategies that have been successful when censoring is dependent in the censored survival setting: (1) definition of a risk set of patients similar to the censored patient being imputed and (2) estimation of an appropriate inverse-transform relationship for sampling particular failure times from the risk set when imputing the censored outcome. Zhao and Tsiatis' estimates of the distribution of QT within these risk sets are useful toward this goal. We also employ a restricted mean QAL model that defines QT failure times in terms of an estimated restricted mean and a residual as part of a technique to calibrate an impute based on the observed covariates of the censored individual.

The rest of the chapter is structured as follows: in section 2, we describe our MI method in more detail. Section 3 gives finite sample results of analyses based on multiply imputed data sets. An analysis of data from the International Breast Cancer Study Group Ludwig Trial V appears in section 4. Discussion follows in section 5.

4.2 Multiple Imputation Methodology

In section 4.2.1, notation and definitions related to quality-adjusted lifetimes are given. An inverse transform method of selecting imputes from a particular distribution is then reviewed in section 4.2.2. In the context of quality adjusted survival analysis, this inverse transform imputation method will require Zhao and Tsiatis' consistent estimator for the distribution of the quality-adjusted life, which we review in section 4.2.3. Risk set selection constraining imputes to those with similar quality-adjusted restricted means is summarized in section 4.2.4. Steps for building complete datasets via multiple imputation are listed in section 4.2.5 using tools from earlier sections. Finally, a review of how to summarize inference based on the completed datasets, now in the context of quality-adjusted survival analysis, is given in sections 4.2.6 and 4.2.7.

4.2.1 *Background and Notation*

We borrow notation for quality-adjusted outcomes from Zhao and Tsiatis (1997), as well as working assumptions. Assume that continuous lifetime, T, is subject to independent right-censoring by continuously-distributed variable, C. The observed follow-up time is $X = \min\{T, C\}$, with censoring indicator, $\Delta = I(T \leq C)$. Also assume a constant $L > 0$ such that the support of T is included in $[0, L]$ and $P(C > L) > 0$. Individual health history is captured by a continuous-time stochastic process $V(\cdot)$ with states $0, 1, ..., S$, where a deterministic utility function $Q(\cdot)$ assigns to each health state a value between 0 and 1. A score of 0 on this scale corresponds to a quality of life equivalent to (or worse than) death and a score of 1 is interpreted as QOL in the absence of disease. In addition, assume a p-dimensional baseline covariate vector $\boldsymbol{Z}$ collected on every individual. Throughout the chapter, the subscript, i,

attached to a previously defined quantity indicates that it corresponds to the i^{th} individual in the sample. Altogether, the observed data are $\mathcal{D} = [X_i, \Delta_i, \{V_i(t); 0 \leq t \leq X_i\}, \boldsymbol{Z}_i]$, for $i \in \{1, ..., n\}$.

QOL-adjusted lifetime is defined as $QT = \int_0^T Q\{V(t)\}dt$ so that, instead of the full accumulated length of life in each state, a person receives different levels of partial credit for enduring states of life of less than perfect quality. Similarly, the censoring random variable on the quality-adjusted lifetime scale is $QC = \int_0^C Q\{V(t)\}dt$, which measures the amount of quality-adjusted lifetime accrued by time C. The observed quality-adjusted follow-up time is $QX = \int_0^X Q\{V(t)\}dt = \min\{QT, QC\}$.

In many clinical trials, a pre-specified timeframe, say $[0, \tau]$, is of interest where $\tau \leq L$. When the number of quality-adjusted days lived during a study period is of interest we may choose $\tau \approx L$; in other contexts a $\tau = 1$-year or $\tau = 5$-year summary measure may be desired. When QT is restricted by τ, we define $QT^* = \int_0^{T \wedge \tau} Q\{V(t)\}dt$, and $QX^* = \int_0^{X \wedge \tau} Q\{V(t)\}dt$. In general, the τ-restricted mean QOL-adjusted lifetime is defined as $\mu_Q(\tau) = \int_0^\tau P(QT^* > q)dq$.

4.2.2 *Inverse transform imputation strategy*

The following inverse transform approach for selecting imputes has been used in traditional censored survival analysis literature (Taylor, Murray and Hsu (2002), Hsu et al. (2006), Liu, Murray and Tsodikov (2011)) but has never been discussed in the context of quality-adjusted survival. We build from the result that for any random variable $U \sim \text{Uniform}(0, 1)$, $S^{-1}_{QT^*}(U)$ will follow the distribution of QT^*, where $S_{QT^*}(t) = P(QT^* > t)$.

For an individual censored at $C_i < \tau$, we wish to sample from the distribution of QT^* given $QT^* > QC_i$. Additional constraints, R_i, involving $\boldsymbol{Z}_i$ are discussed

later in section 4.2.4, so that in total, we wish to sample from the distribution of $QT^*|QT^* > QC_i, R_i$.

The inverse transform result described above suggests sampling QT^* based on an estimate of $S^{-1}_{QT^*|QT^*>QC_i,R_i}(U)$. In the following section we provide details on how to construct a Zhao and Tsiatis consistent estimate of $S_{QT^*|QT^*>QC_i,R_i}(t)$, $\hat{S}_{QT^*|QT^*>QC_i,R_i}(t)$, within the subset of individuals satisfying the constraint, R_i. The sampled QT^* based on the inverse transform approach is the smallest restricted quality-adjusted failure time that satisfies $\hat{S}_{QT^*|QT^*>QC_i,R_i}(QT^*) \leq U$. The sampled QT^* is always one of the observed QT^* values occurring after QC_i that also satisfies the constraint, R_i. The probability of QT^* being sampled corresponds to the size of the drop in $\hat{S}_{QT^*|QT^*>QC_i,R_i}(t)$ at QT^*.

4.2.3 *A consistent estimator for the distribution of quality-adjusted survival time*

In developing an unbiased estimator for the distribution of quality-adjusted lifetime, Zhao and Tsiatis (1997) defined $D_i(x)$ as the time it would take on the traditional time scale to know whether or not the i^{th} individual accumulates at least x units of quality- adjusted lifetime on the quality-adjusted timescale. That is, $D_i(x) = inf[s : \int_0^s Q\{V_i(t)\}dt \geq x] \wedge min(T_i, \tau)$. Use of $D_i(x)$ is convenient for relating outcomes on the traditional timescale, such as C_i, to outcomes on the quality-adjusted timescale, such as QT^* or QX^*. An indicator function for being able to observe whether t units of quality-adjusted lifetime are accumulated on the quality-adjusted timescale becomes $I\{C_i > D_i(t)\}$. Zhao and Tsiatis then define a consistent (inverse-weighted) estimator for the survival function of QT^* as

$$\hat{S}_{QT^*}(t) = n^{-1} \sum_{i=1}^{n} \frac{I\{C_i > D_i(t)\}}{\hat{G}\{D_i(t)\}} I\{QX_i^* > t\}, \tag{4.1}$$

where $\hat{G}(\cdot)$ is the Kaplan-Meier estimator for the censoring time C, based on data $\{(X_i, \Delta_i), i = 1, ..., n\}$. Using members of the risk set defined for individual i censored at C_i, as described in the following section, we estimate $\hat{S}_{QT^*|QT^*>QC_i,R_i}(t)$ by applying the same approach in equation (4.1) to this group of individuals.

4.2.4 *Risk set selection for an individual censored at* $C_i < \tau$

The idea behind risk sets in imputation literature is that within such a set, failure times are more homogeneous, improving the quality of imputes when selected from members of the set. First, we require members of the risk set to have $QT^* > QC_i$, so that the imputed quality-adjusted failure time will be larger than individual i's already accumulated quality-adjusted lifetime at C_i. We also propose using covariate information to gather individuals with similar estimated quality-adjusted survival via a restricted mean model

$$E[log(QT^*)|Z] = \boldsymbol{\beta}^T \boldsymbol{Z}, \tag{4.2}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ is a p-dimensional parameter. That is, in addition to requiring $QT^* > QC_i$, an additional constraint, R_i, for belonging to the risk set is to require $\hat{\boldsymbol{\beta}}^T \boldsymbol{Z}$ to be similar to $\hat{\boldsymbol{\beta}}^T \boldsymbol{Z}_i$ for the patient censored at C_i. This is the primary component of the constraint, R_i, although for a particularly predictive covariate within (4.2), it may be attractive to impose a further restriction that links membership in the risk set to having this characteristic in common with individual i. One may also consider constraining a risk set membership to those with similar $V(C_i)$ to the patient censored at C_i, provided that the sample size of the study is large enough to support this additional restriction. In the Appendix we briefly describe how to obtain estimates of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^{PO}$, from a quality-adjusted restricted mean model (4.2) using pseudo-observations similar in flavor to those used by Andrei and Murray (2007).

We modify their pseudo-observations so that they are constructed on the $\log QT^*$ scale rather than the QT^* scale. Once $\hat{\boldsymbol{\beta}}^{PO}$ estimates are obtained, we use $\hat{\boldsymbol{\beta}}^{PO}\boldsymbol{Z}_i$ to define a constraint, R_i, used in imputing QT^* for individual i.

4.2.5 *MI algorithm*

Step 1: For each individual censored at C_i, we select an appropriate risk set as described in section 4.2.4.

Step 2: Within this risk set, we then calculate $\hat{S}_{QT^*|QT^*>QC_i,R_i}(t)$ as described in section 4.2.3.

Step 3: Generate a Uniform$(0,1)$ random variable, U, and identify the value, QT_j^*, that is the smallest quality-adjusted failure time satisfying $\hat{S}_{QT^*|QT^*>QC_i,R_i}(QT_j^*) \leq U$.

Step 4: Identify the residual ϵ_j from the model $\log(QT_j^*) = \hat{\boldsymbol{\beta}}^{PO^T}\boldsymbol{Z}_j + \epsilon_j$ corresponding to the observed QT_j^* value selected in Step 3.

Step 5: If $QT_j^* = \tau$, we impute QT_i^* by τ; otherwise we add the residual of the jth subject, ϵ_j, to $\hat{E}[log(QT_i^*)] = \hat{\boldsymbol{\beta}}^{PO^T}\boldsymbol{Z}_i$, and use this as the imputed value for $log(QT_i^*)$. If the imputed $\log(QT_i^*) < \log(QC_i)$ then repeat from step 3 until the imputed value is greater than $\log(QC_i)$.

Step 6: Repeat steps 1-5 until all the censored observations from the observed data set are imputed.

Step 7: Repeat steps 1-6 M times so that we have M completed versions of the observed dataset.

Once M completed data sets are obtained, we may perform analyses using the formulaic approach given by Rubin and Little (1987). We summarize two analyses below that are used in sections 4.2.6 and 4.2.7.

4.2.6 *Quality-adjusted restricted mean model analysis on completed datasets*

For each complete dataset, $m = 1, 2, ..., M$, obtain the parameter estimates $\hat{\boldsymbol{\beta}}_{\boldsymbol{m}} = (\hat{\beta}_{m0}, \hat{\beta}_{m1}, \ldots, \hat{\beta}_{mp})$ with associated estimated variance matrix $\hat{\boldsymbol{W}}_m$ for the covariates $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_p)$ based on model (4.2).

The final MI estimates combining the M vectors of coefficient estimates for the quality-adjusted restricted mean model is $\hat{\boldsymbol{\beta}}^{MI} = \sum_{m=1}^{M} \hat{\boldsymbol{\beta}}_{\boldsymbol{m}}/M$. The estimated variability associated with the parameter estimates are composed of two parts, the average within-imputation variance estimate $\bar{\boldsymbol{W}}_M = \sum_{m=1}^{M} \hat{\boldsymbol{W}}_m/M$, and the between-imputation variance estimate $\boldsymbol{B}_M = \sum(\hat{\boldsymbol{\beta}}_m - \hat{\boldsymbol{\beta}}^{MI})^2/(M-1)$. The estimated variance of $\hat{\boldsymbol{\beta}}^{MI}$ is then $\boldsymbol{V} = \bar{\boldsymbol{W}}_M + (1 - M^{-1})\boldsymbol{B}_M$.

4.2.7 *Marginal quality-adjusted survival analysis on completed datasets*

For each complete dataset, $m = 1, 2, ..., M$, we calculate the quality-adjusted survival estimate $\hat{S}_{QT(m)}(t)$ as the sample proportion of individuals in data set m with $QT^* > t$. Its estimated variance becomes $\hat{V}_m(t) = \hat{S}_{QT(m)}(t)(1 - \hat{S}_{QT(m)}(t))/n$. Our final MI survival estimate that adjusts for dependent censoring is $\hat{S}^{MI}_{QT}(t) = \sum_{m=1}^{M} \hat{S}_{QT(m)}(t)/M$ with estimated variance $\hat{V}^{MI}(t) = M^{-1} \sum_{m=1}^{M} \hat{V}_m(t) + (1 + M^{-1}) \sum_{m=1}^{M} [\hat{S}_{QT(m)}(t) - \hat{S}^{MI}_{QT}(t)]^2/(M-1)$.

4.3 Simulation Studies

To access finite sample performance of our MI method in estimating the marginal distribution of QT^* and parameters from multivariate restricted mean QAL models as in equation (4.2), we have conducted simulations under three scenarios with $n = 150$ where accrual of QAL over $\tau = 2$ years is of interest. The utility function $Q(s) = s/100$, where $s \in S = 0, 1, ..., 100$, is used in all settings. We summarize

results based on 1000 simulations per scenario. In each simulation, we perform the following:

Step 1: Independently generate Z_1 from a Bernoulli(0.5) distribution and Z_2 from a Uniform(0,1) distribution.

Step 2: Generate failure times, T_i, from Exponential(λ_i) distributions, $i = 1, \ldots, 150$ where the parameters λ_i depend on individual i's covariates so that the mean structure $E[log(QT^*)] = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2$ is satisfied. Algebraic details for calculation of λ_i are given in the Appendix.

Step 3: Generate censoring times, C, from an Exponential(1) distribution for scenarios 1 and 2. In scenario 3, we generate C_i based on individual i's Z_{2i} covariate using an Exponential$\{2(1-Z_{2i})\}$. The resulting censoring percentages range between $40 - 50\%$ for the 3 settings.

Step 4: Generate the health state process, $V(t)$, for the 3 different settings as follows.

In the first scenario, the health process remains in the best state for all patients, i.e., $V(t) = 100$ for all t. Therefore, QT^* reduces to min(T, 2 years). In the second scenario, patient i's health process depends on their death time T_i via

$$V_i(t) = \begin{cases} 100 & \text{if } t \in [0, \frac{T_i}{4}] \\ 90 & \text{if } t \in (\frac{T_i}{4}, \frac{T_i}{2}] \\ 80 & \text{if } t \in (\frac{T_i}{2}, \frac{3T_i}{4}] \\ 70 & \text{if } t \in (\frac{3T_i}{4}, T_i]. \end{cases}$$

In scenario 3, patient i's health state process depends on both T_i and their co-

variates:

$$V_i(t) = \begin{cases} 100 & if\ t \in [0, \frac{T_i}{4}], Z_{1i} = 1, \text{and } Z_{2i} > 0.5 \\ 95 & if\ t \in [0, \frac{T_i}{4}], Z_{1i} = 1, \text{and } Z_{2i} \leq 0.5 \\ 90 & if\ t \in [0, \frac{T_i}{4}], \text{and } Z_{1i} = 0 \\ 90 & if\ t \in (\frac{T_i}{4}, \frac{T_i}{2}], Z_{1i} = 1, \text{and } Z_{2i} > 0.5 \\ 85 & if\ t \in (\frac{T_i}{4}, \frac{T_i}{2}], Z_{1i} = 1, \text{and } Z_{2i} \leq 0.5 \\ 80 & if\ t \in (\frac{T_i}{4}, \frac{T_i}{2}], \text{and } Z_{1i} = 0 \\ 80 & if\ t \in (\frac{T_i}{2}, \frac{3T_i}{4}], Z_{1i} = 1, \text{and } Z_{2i} > 0.5 \\ 75 & if\ t \in (\frac{T_i}{2}, \frac{3T_i}{4}], Z_{1i} = 1, \text{and } Z_{2i} \leq 0.5 \\ 70 & if\ t \in (\frac{T_i}{2}, \frac{3T_i}{4}], Z_{1i} = 0, \text{and } Z_{2i} > 0.5 \\ 65 & if\ t \in (\frac{T_i}{2}, \frac{3T_i}{4}], Z_{1i} = 0, \text{and } Z_{2i} \leq 0.5 \\ 70 & if\ t \in (\frac{3T_i}{4}, T_i], Z_{1i} = 1, \text{and } Z_{2i} > 0.5. \\ 65 & if\ t \in (\frac{3T_i}{4}, T_i], Z_{1i} = 1, \text{and } Z_{2i} \leq 0.5. \\ 60 & if\ t \in (\frac{3T_i}{4}, T_i], Z_{1i} = 0, \text{and } Z_{2i} > 0.5. \\ 55 & if\ t \in (\frac{3T_i}{4}, T_i], Z_{1i} = 0, \text{and } Z_{2i} \leq 0.5. \end{cases}$$

In each scenario, we compare parameter estimates from fitting model (4.2) using (i) QAL pseudo-observations as defined in the Appendix, and (ii) our MI approach. The results for scenario 1 are shown in Table 4.1(A), where all individuals have perfect health and are assigned utility score of 1 all the time. There is, therefore, no induced informative censoring involved. The two methods, PO and MI, have similar performance in terms of bias and standard errors. In the second scenario {Table 4.1(B)}, results using PO and MI are similar except for the intercept term, where the MI method is less biased than the PO method. Recall that in the third scenario, censoring depends on Z_2 so that larger Z_2 values allow larger follow-up times. The PO estimate for β_2 seems slightly underestimated in this case. Although the PO

method adjusts for dependent censoring caused by shifting the the QAL timescale, it does not adjust for dependent censoring associated with covariate Z_2. Since the MI method uses information on estimated restricted means when defining risk sets, the MI estimate for β_2 is less affected by the censoring pattern.

Marginal quality-adjusted survival estimates, calculated using MI as described in section 4.2.7, are given in Table 4.2. For comparison we also give Kaplan-Meier estimates for the distribution of QT that are susceptible to overestimating quality-adjusted survival. In scenario 1, the Kaplan-Meier method does not have to contend with dependent censoring and remains the nonparametric maximum likelihood estimator in this case. Point estimates based on the MI method are reasonably close to the correct probabilities, but the Kaplan-Meier clearly remains the preferred estimate in this case. In scenario 2, Kaplan-Meier estimates are known to have positive bias, and moderate overestimates of quality-adjusted survival are seen. The MI estimates in this case outperform the Kaplan-Meier. Interestingly, in scenario 3 the direction of bias caused by Z_2 (censored patients have longer QAL) seems to counteract the Kaplan-Meier's tendency to overestimate QAL survival, so both MI and Kaplan-Meier estimates perform well in this case, although this behavior for the Kaplan Meier does not generally hold. We also calculated Zhao and Tsiatis's survival estimates and found good performance in all but the final scenario, where point estimates for quality-adjusted survival slightly underestimate the truth due to dependent censoring from Z_2.

4.4 IBCSG Ludwig Trial V Example

We apply our imputation method to a subset of 715 premenopausal breast cancer patients from the IBCSG Ludwig Trial V who were randomized to one course of

Table 4.1: Comparison of estimates fitting model (4.2) using pseudo observation (PO), and multiple imputations (MI) under three scenarios.

Parameters	PO	MI
(A) Scenario 1: 40.3% censoring		
$\beta_0 = -1$	-0.957 [0.043, 0.231, 0.241] [a]	-1.005 [-0.005, 0.234, 0.264]
$\beta_1 = 1$	0.973 [-0.027, 0.207, 0.204]	1.003 [0.003, 0.209, 0.223]
$\beta_2 = -1$	-0.962 [0.038, 0.359, 0.366]	-0.994 [0.006, 0.363, 0.413]
(B) Scenario 2: 48.8% censoring		
$\beta_0 = -1$	-0.821 [0.179, 0.235, 0.290]	-1.005 [-0.005, 0.231, 0.252]
$\beta_1 = 1$	1.037 [0.037, 0.209, 0.207]	1.032 [0.032, 0.206, 0.233]
$\beta_2 = -1$	-1.055 [-0.055, 0.364, 0.376]	-0.985 [0.015, 0.356, 0.399]
(C) Scenario 3: 43.1% censoring		
$\beta_0 = -1$	-0.990 [0.010, 0.228, 0.246]	-1.005 [-0.005, 0.236, 0.274]
$\beta_1 = 1$	0.966 [0.034, 0.203, 0.196]	1.027 [0.027, 0.208, 0.219]
$\beta_2 = -1$	-0.904 [0.096, 0.353, 0.368]	-0.969 [0.031, 0.359, 0.404]

[a] [bias, average standard error based on individual analysis, empirical standard deviation of parameter estimates across 1000 simulations]

Table 4.2: Comparison of quality-adjusted survival estimates using, Kaplan Meier ($\hat{S}_{QT}^{KM}(t)$) and multiple imputation ($\hat{\bar{S}}_{QT}^{MI^W}(t)$) under three scenarios.

t	$S_{QT}(t)$	$\hat{S}_{QT}^{KM}(t)$	$\hat{S}_{QT}^{MI}(t)$
(A) Scenario 1			
0.5	0.470	0.471 [0.001, 0.047,0.049] [a]	0.470 [0.000, 0.044, 0.049]
1	0.263	0.264 [0.001, 0.048, 0.047]	0.262 [-0.001, 0.040, 0.048]
1.5	0.162	0.163 [0.001, 0.048, 0.048]	0.161 [-0.001, 0.034, 0.049]
2^-	0.105	0.108 [0.003, 0.047, 0.047]	0 [-0.105, 0, 0]
(B) Scenario 2			
0.5	0.490	0.518 [0.028, 0.047,0.050]	0.483 [-0.007, 0.044, 0.055]
1	0.290	0.343 [0.053, 0.054, 0.056]	0.269 [-0.021, 0.041, 0.055]
1.5	0.172	0.236 [0.064, 0.062, 0.064]	0.159 [-0.013, 0.034, 0.053]
2^-	0.022	0.055 [0.033, 0.058, 0.043]	0 [-0.022, 0, 0]
(C) Scenario 3			
0.5	0.480	0.477 [-0.003, 0.048,0.049]	0.483 [0.003, 0.045, 0.051]
1	0.277	0.266 [-0.011, 0.051, 0.053]	0.276 [-0.001, 0.042, 0.059]
1.5	0.172	0.157 [-0.015, 0.050, 0.054]	0.168 [-0.004, 0.034, 0.062]
2^-	0.000	0.025 [0.025, 0.039, 0.024]	0 [0.000, 0, 0]

[a][bias, average standard error based on individual analysis, empirical standard deviation of parameter estimates across 1000 simulations]

perioperative systemic treatment (short duration) or 6 to 7 courses of prolonged adjuvant therapy (long duration). As mentioned in the introduction, the course of breast cancer may be summarized by three states: TOX, TWiST, and REL. Let T_{TOX} be the duration of TOX, which occurs from the beginning of study until the end of the treatment, T_{TWiST} be the length of TWiST, and T_{REL} be the time a person lives after disease recurrence (REL). The quality-adjusted lifetime is then $QT = \mu_{TOX}T_{TOX} + T_{TWiST} + \mu_{REL}T_{REL}$, where μ_{TOX} and μ_{REL} are utility scores for the toxicity and the relapse periods, ranging between 0 and 1. Here we use $\mu_{TOX} = 0.5$ as employed by Gelber et al. (1991).

A thorough analysis of QAL includes estimation of QAL survival in the two treatment groups and multivariate analysis studying the influence of patient risk profile on restricted mean QAL. We evaluate QAL that would be accumulated over a 5 year period and its relationship with age, estrogen-receptor (ER) status (positive or negative/unknown), tumor size measured in centimeters (greater than 2cm or not), and node group. Sensitivity analyses are done to address the effect of μ_{REL}.

In the case where μ_{REL} is taken to be 0.5, the PO component of Table 4.3 was used to construct risk sets for imputation as described in section 4.2.4. The resulting MI algorithm, summarized in section 4.2.5, produced the long and short duration treatment quality-adjusted survival curve estimates displayed in Figure 4.1.

Quality-adjusted survival curves can be a bit tricky to interpret when one is used to interpreting unadjusted survival curves. Each point on the quality-adjusted survival curve needs to be interpreted in terms of the amount of QAL accrued over 5 years. For instance, the overlap in quality-adjusted survival curves at month 22 indicates that slightly over 80% of patients in either treatment group is able to accumulate at least 22 months of quality-adjusted survival time over the five year period.

However it is not possible to know from the curves, without further information, how much follow-up time it took the short duration treatment group to accrue 22 months of QAL compared to the long duration group. In fact, since the long duration treatment was exposed to toxicity for 5-6 months longer it is likely that patients on that treatment arm had to be followed for another 5-6 more months before accruing those 22 months of QAL.

The QAL benefit associated with the long duration treatment is shown when larger percentages of those patients are able to accrue the higher QAL times between 22 and about 54-55 months. This happens because the time to relapse is delayed in these patients. Only 27.6% of long-duration patients have relapsed at 5 years as opposed to 38.9% of the short duration therapy patients. However, since the long duration patients had roughly 6-7 total courses of toxic therapy, that final 5 months or so of QAL between 55 and 60 months is very difficult to accrue, causing the quality-adjusted survival curve to drop precipitously at around 55 months. Hence this final drop in quality-adjusted survival is not due to a sudden disadvantage in either QOL or survival at 55 months, but is caused by the extra months of toxicity experienced at the beginning of the trial.

Our analysis continues with MI results for the restricted mean QAL analysis in Table 4.3, again assuming $\mu_{REL} = 0.5$. Parameter estimates are very similar using both the MI and the PO methods. To estimate a restricted mean based on individual risk factors from the table, one multiplies the exponentiated parameter estimates raised to the power of the covariate value. For instance, using the MI model results, a 40 year old premenopausal woman with negative ER status, 10 positive nodes, a 3 cm diameter tumor in the long-duration treatment therapy will accrue $32.39 * 1.07^4 * 0.84 * 0.61 * 1.1 = 24$ QAL months over the 5 year period, whereas

on the short duration therapy this same patient would accrue 22 QAL months over the 5 year period. So the extra QAL gained for an extremely ill patient like this is merely 2 months when selecting the long-duration treatment and the extra time on toxicity. However, a patient projected to live 54 QAL months on the short duration treatment would accrue 54*1.10=59.4 QAL months on the long duration treatment, a gain of 5.4 QAL months that might be worth the extra time in toxicity.

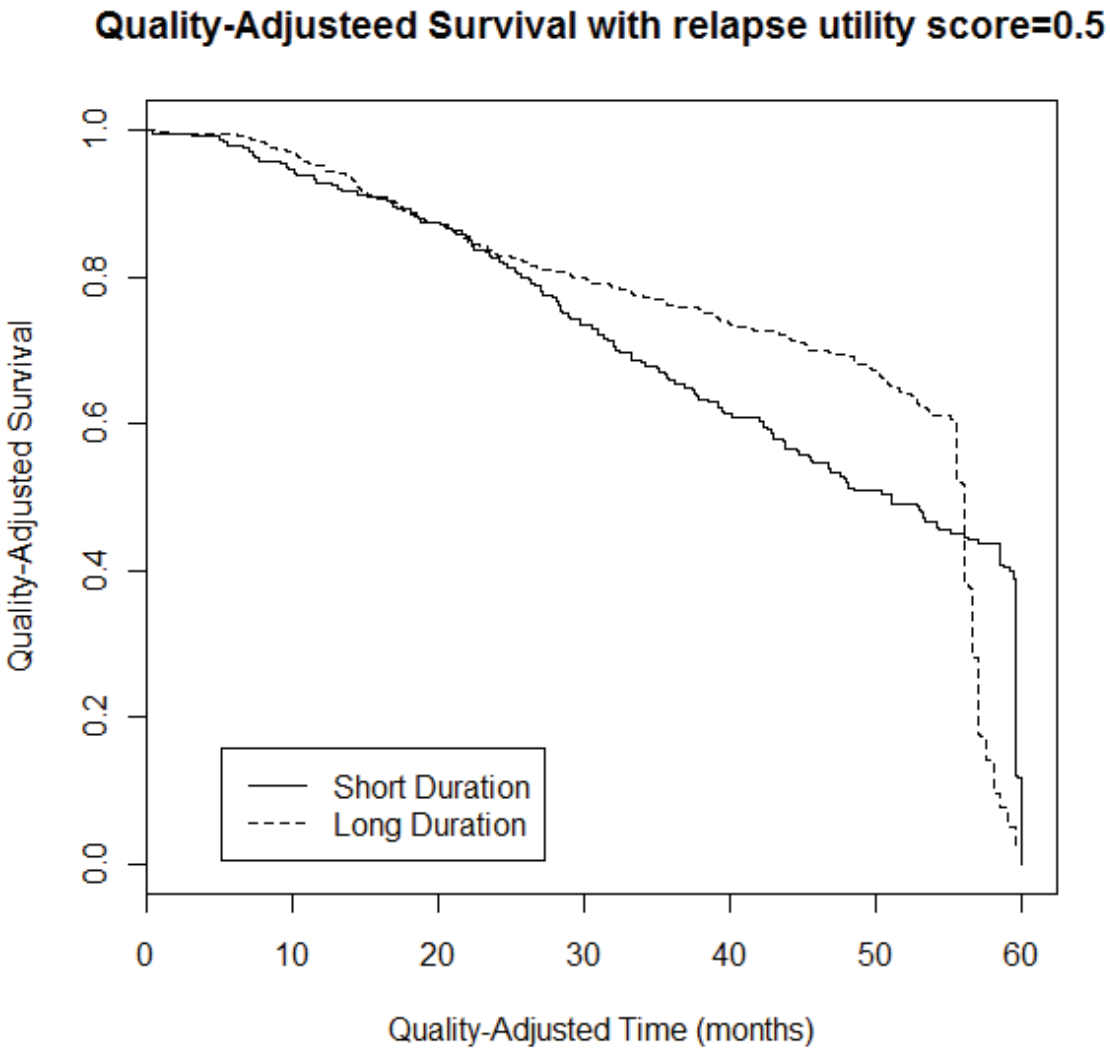
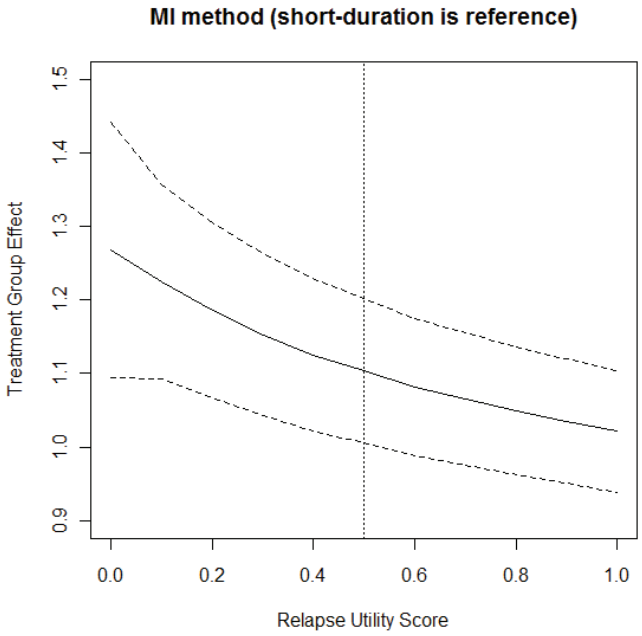


Figure 4.1: Quality-adjusted survival by treatment group for 715 breast cancer patients, with $\mu_{REL} = 0.5$, and $\mu_{TOX} = 0.5$.

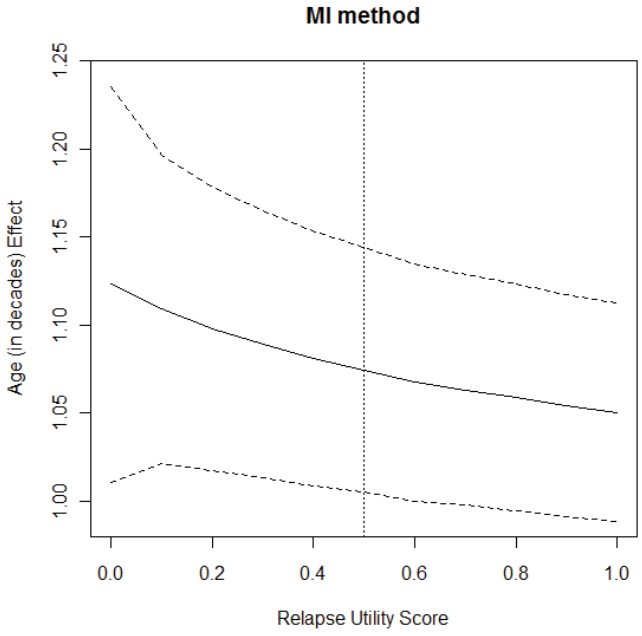
The choice of treatment regimen is also affected by the choice of the μ_{REL} utility score for the relapse period, which shifts the QAL values for the study and alters the fitted restricted mean model parameters. Figure 4.2a shows how the treatment effect parameter changes as a function of μ_{REL}, where the vertical dashed line at $\mu_{REL} = 0.5$ corresponds to the estimated treatment parameter seen in Table 4.3. So for those who have little tolerance for the QOL during relapse, small μ_{REL}, the argument for enduring the long duration therapy becomes slightly more persuasive. And for those who weight time spent in relapse with a higher μ_{REL}, the treatment effect loses statistical significance and treatment choices become more ambivalent in terms of QAL gains. Figures 4.2b, 4.3a, 4.3b, 4.3c and 4.3d show changes in estimated parameters for age, number of positive nodes, tumor size and ER status as functions of μ_{REL}.

4.5 Discussion

Quality-adjusted survival estimates play a key role in computation for both the PO method and the MI method discussed in this work. However, interpretation of point estimates for quality-adjusted survival on the quality-adjusted time scale can be difficult when making comparisons. For instance when calculating 1-year quality-adjusted survival, a subgroup experiencing perfect health with no mortality will accrue 1-year of QAL over a single year of traditional follow-up time while a subgroup consistently experiencing 50% QOL with no mortality will require 2-years of traditional follow-up time to accrue 1 year of QAL. Without any deaths and with at least 2 years of traditional follow-up, both subgroups will have $S_{QT^*}(1 \text{ year})=1$, looking comparable at this quality-adjusted time point even though the experience of the subgroups differ. Summary statistics computed over a common traditional follow-



(a)



(b)

Figure 4.2: Treatment and age coefficient estimates adjusting for other covariates using MI.

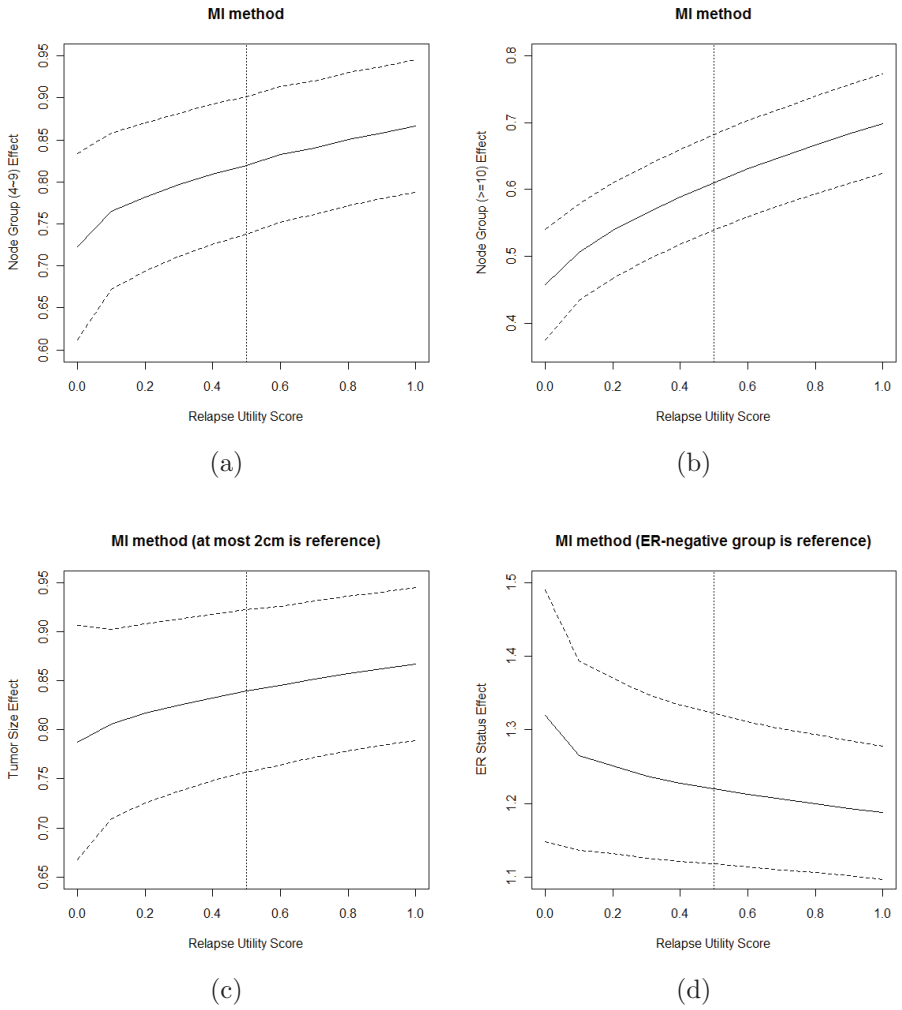


Figure 4.3: Node group, tumor size, and estrogen receptor status coefficient estimates adjusting for other covariates using MI.

up time, such as restricted mean QAL or quantiles of restricted QAL, continue to be the most interpretable ways of comparing QAL experiences of groups. Techniques requiring proportional hazards have not risen in popularity for QAL settings since, as in the breast cancer example, hazards for treatment regimens with different lengths of toxicity tend to eventually cross.

We studied our MI approach for producing restricted mean parameters for model (4.2) as well as quality-adjusted survival estimates. However, these datasets may also be used for other analyses as well. One need only locate software in the more common case where there is no censoring, obtain desired estimates and their variances, and combine results using standard approaches as in sections 4.2.6 and 4.2.7.

In our work, we have observed that PO methods for QAL data have difficulty in estimating intercepts in the restricted mean framework. This problem seems to be somewhat improved by constructing pseudo observations on the log scale, but remains an issue to watch for when using those methods, particularly when there is a large proportion of censoring. Liu et al. (2011) and Xiang and Murray (2012) also reported a slight underestimation of the intercept when using pseudo observations and inverse weighted pseudo observations in restricted mean models. Our MI approach takes advantage of the simplicity of PO regression when grouping individuals into risk sets; risk set selection is essentially unaffected by intercept issues. Our MI procedure has an improved fit for the intercept compared to the PO method, giving better restricted mean estimates. In addition, our simulation results indicate that dependent censoring caused by a baseline covariate is accounted for well using the MI approach. This is consistent to what Liu et al. (2011) noted when using a MI method for restricted mean analysis on the traditional time scale.

Table 4.3: Quality-adjusted restricted mean model for 715 breast cancer patients using PO and MI methods with $\mu_{TOX} = 0.5$, and $\mu_{REL} = 0.5$.

	PO			MI		
	$e^{\hat{\beta}}$	95% CI	p-value	$e^{\hat{\beta}}$	95% CI	p-value
(intercept)	32.38	(22.81, 41.94)	< 0.0001	32.39	(22.41, 42.37)	< 0.0001
Age (in decades)	1.08	(1.01, 1.15)	0.0175	1.07	(1.00, 1.14)	0.0297
ER Positive	1.21	(1.12, 1.31)	< 0.0001	1.22	(1.12, 1.32)	< 0.0001
Tumor Size ($\geq$ 2 cm)	0.84	(0.76, 0.92)	0.0004	0.84	(0.76, 0.92)	0.0005
4-9 Positive Nodes	0.83	(0.75, 0.91)	0.0001	0.82	(0.74, 0.90)	< 0.0001
10+ Positive Nodes	0.62	(0.55, 0.69)	< 0.0001	0.61	(0.54, 0.68)	< 0.0001
Long Duration Treatment	1.09	(0.99, 1.18)	0.0544	1.10	(1.01, 1.20)	0.0300

APPENDICES

APPENDIX A

Appendix for Chapter II

In Section 2.3, we simulate settings when dependent censoring affects estimation of (2.3) unless adjustments are made via the IPCW PO method. The survival function for piecewise exponential failure times used in Section 2.3 is

$$S_T(t) = \begin{cases} e^{-\lambda_{Z_0} t} & 0 \le t \le t_1 \\ e^{-\lambda_{Z_0} t_1} e^{-\lambda_{Z_0 Z_1}(t-t_1)} & t > t_1 \end{cases},$$

with pdf

$$f_T(t) = \begin{cases} \lambda_{Z_0} \cdot e^{-\lambda_{Z_0} t} & 0 \le t \le t_1 \\ \lambda_{Z_0 Z_1} \cdot e^{-(\lambda_{Z_0} - \lambda_{Z_0 Z_1}) t_1} \cdot e^{-\lambda_{Z_0 Z_1} t} & t > t_1 \end{cases}.$$

In Section 2.3 step 2, parameters $\lambda_0 = 0.3, \lambda_1 = 0.2, \lambda_{01} = 0.1$ and $\lambda_{11} = 0.5$ are fixed. The remaining parameters λ_{00} and λ_{10} are chosen to satisfy (2.3) as described below. Recall that Z_1 is measured at time $t_1 = 0.2$ and that $\tau = 5$. Also Z_0 and Z_1 are generated from independent Bernoulli(0.5).

$$\begin{aligned} & E[\log\{min(\tau, T)\}] \\ =& E\big[E[\log\{min(\tau, T)\}|Z_1]\big] \\ =& \sum_{z_1} P(Z_1 = z_1) E[\log\{min(\tau, T)\}|Z_1 = z_1] \\ =& 0.5E[\log\{min(\tau, T)\}|Z_1 = 0] + 0.5E[\log\{min(\tau, T)\}|Z_1 = 1] \end{aligned}$$

When $Z_0 = 0$, we have

$$
\begin{aligned}
&E[\log\{min(\tau, T)\}] \\
=&0.5[(\int_0^{t_1} \log t \cdot \lambda_0 e^{-\lambda_0 t} dt + \int_{t_1}^{\tau} \log t \cdot \lambda_{00} \cdot e^{-(\lambda_0 - \lambda_{00})t_1} \cdot e^{-\lambda_{00} t} dt \\
&+ \log \tau \cdot e^{-\lambda_0 t_1} \cdot e^{-\lambda_{00}(\tau - t_1)}] + 0.5[(\int_0^{t_1} \log t \cdot \lambda_0 e^{-\lambda_0 t} dt \\
&+ \int_{t_1}^{\tau} \log t \cdot \lambda_{01} \cdot e^{-(\lambda_0 - \lambda_{01})t_1} \cdot e^{-\lambda_{01} t} dt + \log \tau \cdot e^{-\lambda_0 t_1} \cdot e^{-\lambda_{01}(\tau - t_1)}] \\
=&\beta_0 + \beta_2 Z_2
\end{aligned}
$$

We solve the above equation for λ_{00}. Similarly when $Z_0 = 1$ we solve the following equation for λ_{10}.

$$
\begin{aligned}
&E[\log\{min(\tau, T)\}] \\
=&0.5[(\int_0^{t_1} \log t \cdot \lambda_1 e^{-\lambda_1 t} dt + \int_{t_1}^{\tau} \log t \cdot \lambda_{10} \cdot e^{-(\lambda_1 - \lambda_{10})t_1} \cdot e^{-\lambda_{10} t} dt \\
&+ \log \tau \cdot e^{-\lambda_1 t_1} \cdot e^{-\lambda_{10}(\tau - t_1)}] + 0.5[(\int_0^{t_1} \log t \cdot \lambda_1 e^{-\lambda_1 t} dt \\
&+ \int_{t_1}^{\tau} \log t \cdot \lambda_{11} \cdot e^{-(\lambda_1 - \lambda_{11})t_1} \cdot e^{-\lambda_{11} t} dt + \log \tau \cdot e^{-\lambda_1 t_1} \cdot e^{-\lambda_{11}(\tau - t_1)}]) \\
=&\beta_0 + \beta_1 + \beta_2 Z_2
\end{aligned}
$$

The resulting parameters λ_{00} and λ_{10} vary according to values of Z_2 for patient i, $i = 1, ..., n$.

APPENDIX B

Appendix for Chapter IV

In section 4.3 we simulate failure times T from exponential distributions with hazard rates λ based on the quality-adjusted restricted mean model (4.2). λ is solved to satisfy the following equation, where $V_j = V(t), t \in [(j-1)t/4, jt/4]$.

$$
\begin{aligned}
& E[\log QT^*] \\
=& E[\log\{min(\tau, T) \cdot \sum_{j=1}^{4} Q\{V_j\}/4\}] \\
=& \int_0^\tau \log t \cdot \lambda e^{-\lambda t} dt + \int_\tau^\infty \log \tau \cdot \lambda e^{-\lambda t} dt + \log \sum_{j=1}^{4} Q\{V_j\}/4 \\
=& \int_0^\tau \log t \cdot \lambda e^{-\lambda t} dt + \log \tau \cdot e^{-\lambda \tau} + \log \sum_{j=1}^{4} Q\{V_j\}/4 \\
=& \beta_0 + \beta_1 Z_1 + \beta_2 Z_2
\end{aligned}
$$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Andersen, P.K., Hansen, M. and Klein, J.P. (2004). Regression analysis of restricted mean survival time based on pseudo observations. *Life Data Analysis* **10,** 335-350.

[2] Andrei, A. C. and Murray, S. (2007). Regression models for the mean of quality-of-life adjusted restricted survival time using pseudo-observations. *Biometrics* **63,** 398404.

[3] Bang H, and Tsiatis AA. Median regression with censored cost data. (2002). *Biometrics* **58,** 643-649.

[4] Cole, B. F., Gelber, R. D., and Goldhirsch, A. (1993). Cox regression models for quality-adjusted survival analysis. *Statistics in Medicine* **12,** 975-987.

[5] Cole, B. F., Gelber, R. D. and Anderson K. M. (1994) Parametric Approaches to Quality-Adjusted Survival Analysis. *Biometrics* **50,** 621-631.

[6] Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187-200.

[7] Crowley, J. and Hu M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* **72,** 27-36.

[8] Department of Health and Human Services. Organ Procurement and Transplantation Network; Final Rule. In: Federal Register 42 CFR, Part 121, October 20, 1999: 56649-56661.

[9] Egan T.M., Murray S., Bustami R.T., Shearon T.H., McCullough 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. *Am J Transplant* **6 (Part 2),** 1212-1227.

[10] Faucett, C.L., Schenker, N. and Taylor, J.M.G. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics* **58,** 37-47.

[11] Graw, F., Gerds, T. and Schumacher, M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Analysis* 2009; **15**: 241-255.

[12] Gelber, R. D., Cole, B. F., Gelber, S., and Goldhirsch, A. (1995). Comparing treatments using quality-adjusted survival: The Q-TWIST method. *The American Statistician* **49,** 161-169.

[13] Gelber, R. D., Gelman, R. S. and Goldhirsch, A. (1989). A quality of life oriented end point for comparing therapies. *Biometrics* **45,** 781-796.

[14] Gelber, R. D. and Goldhirsch, A. (1986). A new endpoint for the assessment of adjuvant therapy in postmenopausal women with operable breast cancer. *Journal of Clinical Oncology* **4,** 1772-1779.

[15] Gelber, R. D., Goldhirsch, A., Castiglione, M., Price, K., Isley, M. and Coates, A. (1987). For the Ludwig Breast Cancer Study Group. Time without symptoms and toxicity (TWIST): a quality-of-life-oriented endpoint to evaluate adjuvant therapy. *Adjuvant Therapy of Cancer* **V,** Salmon, S. E. (ed.), 455-465. Orlando, Florida: Grune & Stratton.

[16] Glasziou, P. P., Simes, R. J., and Gelber, R. D. (1990). Quality Adjusted Survival Analysis. *Statistics in Medicine* **9,** 1259-1276.

[17] Goldhirsch, A., Gelber, R. D., Simes, R. J., Glasziou, P. P. and Coates, A. (1989). Costs and benefits of adjuvant therapy in breast cancer: a quality adjusted survival analysis. *Journal of Clinical Oncology* **7,** 36-44.

[18] Harrel, F. E., Califf, R. M., Pryor, D. B., Lee, K. L. and Rosati, R. A. (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association* **247,** 2543-2546.

[19] Harrell, F. E., Lee, K. L. and Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* **15,** 361387.

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

[21] Kaplan, E.L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *Journal of American Statistics Association* **53,** 457-481, 562-563.

[22] Little, R.J.A. and Rubin, D. (1986). *Statistical analysis with missing data*, 1st Edition. New York: Wiley.

[23] Liu, L. X., Murray, S. and Tsodikov, A. (2011). Multiple imputation based on restricted mean model for censored data. *Statistics in Medicine* **30,** 1339-1350.

[24] Mantel, N. (1966). Evaluation of survival data and two new rank-order statistics arising in its consideration. *Cancer Chemotherapy Reports* **50,** 163-170.

[25] OPTN Thoracic Organ Transplantation meeting minutes March 23, 2010

[26] Robins, J. M. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association* 1993; 24-33. Alexandria, Virginia: American Statistical Association.

[27] Robins, J.M., and Finkelstein, D.M. (2000).Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* **56,** 779-788.

[28] Robins, J. M. and Rotinitzky, A. Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology-Methodological Issues* 1992; N.Jewell, K. Dietz, and B. Farewell (eds): 297-331. Boston: Birkhäuser.

[29] Satten G. A. and Datta S. The Kaplan-Meier Estimator as an Inverse-Probability-of-Censoring Weighted Average. *The American Statistician* 2001; **55**: 207-210.

[30] Satten G.A., Datta S. and Robins J.M. (2001). Estimating the marginal survival function in the presence of time dependent covariates. *Statist. Probab. Lett.* **54,** 397403.

[31] Savage, I. R. (1956). Contributions to the theory of rank-order statistics – the two sample case. *Annals of Mathematical Statistics* **27,** 590-615.

[32] Scharfstein, D., Robins, J.M., Eddings, W., and Rotnitzky,A. (2001). Inference in randomized studies with informative censoring and discrete time-to-event endpoints. *Biometrics* **57,** 404-413.

[33] Schenker N., Taylor J. M. G. (1996). Partially parametric techniques for multiple imputation. *Computational Statistics and Data Analysis* **22,** 425-446.

[34] Taylor, J.M.G., Murray S., and Hsu C.H. (2002). Survival estimation and testing via multiple imputation. *Statistics and Probability Letters* **58,** 221-232.

[35] Wang, H. and Zhao, H. (2007). Regression Analysis of Mean Quality-Adjusted Lifetime with Censored Data. *Biostatistics* **8,** 368-382.

[36] Wei, G.C.G., and Tanner, M.A. (1991). Application of Multiple Imputation to the Analysis of Censored Regression Data. *Biometrics* **47,** 1297-1309.

[37] Xiang, F. and Murray S. (2012). Restricted mean models for transplant benefit and urgency. *Statistics in Medicine* **31,** 561-576

[38] Yusen R.D., Shearon T.H., Qian Y., Kotloff R., Barr M.L., Sweet S., Dyke D.B., Murray S. (2010). Lung transplantation in the United States, 1999-2008. *Am J Transplant* **10 (Part 2),** 1047-1068.

[39] Zhao, H. and Tsiatis, A. A. (1997). A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* **84,** 339-348.

[40] Zhao, H. and Tsiatis, A. A. (1999). Efficient estimation of the distribution of quality adjusted survival time. *Biometrics* **55,** 1101-1107.

[41] Zhao, H. and Tsiatis, A. A. (2000). Estimating mean quality adjusted lifetime with censored data. *Sankhyar, Series B* **62,** 175-188.

[42] Zhao, H., and Tsiatis, A.A. (2001). Testing Equality of Survival Functions of Quality Adjusted Lifetime. *Biometrics* **57,** 861-867.