

Analysis of transplant urgency and benefit via multiple imputation

Fang Xiang,^a Susan Murray^{b*†} and Xiaohong Liu^c

Missing (censored) death times for lung candidates in urgent need of transplant are a signpost of success for allocation policy makers. However, statisticians analyzing these data must properly account for dependent censoring as the sickest patients are removed from the waitlist. Multiple imputation allows the creation of complete data sets that can be used for a variety of standard analyses in this setting. We propose an approach to multiply impute lung candidate outcomes that incorporates (i) time-varying factors predicting removal from the waitlist and (ii) estimates of transplant urgency via restricted mean models. The measures of transplant urgency and benefit for individual patient profiles are discussed in the context of lung allocation score modeling in the USA. Marginal survival estimates in the event that a transplant does not occur are also provided. Simulations suggest that the proposed imputation method gives attractive results when compared with existing methods. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: dependent censoring; multiple imputation; restricted mean life; survival; transplant benefit; transplant urgency

1. Introduction

In the USA, lung transplant candidates are placed on a waiting list until a donor lung becomes available. Donor lungs are allocated using a lung allocation score (LAS), with the highest scored candidate getting the next available lung [1]. The LAS is based upon measures of transplant benefit as well as clinical urgency over the next year. In particular, clinical urgency is defined as the expected number of days that a person would live without a transplant during the following year (1-year restricted mean). Transplant benefit is defined as the days in life gained during the next year if a transplant is offered immediately (difference in 1-year restricted means). LAS values are calculated by subtracting urgency days from benefit days and normalizing the score to produce a range from 0 to 100. The scoring system rewards those who have limited days of life without transplant and a high transplant benefit, while avoiding futile transplants where urgency and benefit are essentially the same.

With the LAS prioritizing lung offers, urgent patients who may benefit from transplant are more likely to be removed from the waiting list for transplant. Their death times without transplant are not observed and, in fact, are very dependently censored by the allocation algorithm in place. Traditional survival analysis methods [2–5] are subject to bias in such cases. To counteract this type of bias, authors including Robins and Finkelstein [6], Robins [7], Robins and Rotnitzky [8], Satten *et al.* [9], and Scharfstein *et al.* [10], among others, have developed inverse probability of censoring weighted (IPCW) methods for many traditional analyses. Xiang and Murray [11] used an IPCW-adjusted method for modeling the restricted mean via pseudo-observations (POs). As successful as IPCW methods have been in correcting useful survival analyses for dependent censoring issues, one drawback to this approach is the need to develop inverse weighted analysis methods to address each separate statistical analysis of interest. As new statistical analyses become of interest for a particular data set, the appropriate IPCW-adjusted method must be identified from the statistical literature or newly developed if not currently available.

^aNovartis Institute of Biomedical Research, Cambridge, MA 02139, U.S.A.

^bDepartment of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, U.S.A.

^cAmgen, South San Francisco, CA 94080, U.S.A.

*Correspondence to: Susan Murray, Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, U.S.A.

†E-mail: skmurray@umich.edu

More and more, censoring is being viewed in the context of missing data literature as something that can be multiply imputed. Taylor *et al.* [12] linked a nonparametric multiple imputation (MI) procedure to results using traditional Kaplan–Meier (KM) and logrank analyses. Hsu *et al.* [13] incorporated auxiliary variables into MI of censored observations. Faucett *et al.* [14] 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. Wei and Tanner [15] 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. Liu *et al.* [16] proposed an MI method based on semiparametric selection of residuals from a restricted mean model. The advantage of MI approaches to the analysis of censored data is that widely available complete case analyses may then be used on the imputed data sets. This advantage is especially attractive in the context of dependent censoring where IPCW methods may have limited availability in mainstream statistical software.

The MI approaches based on selecting failures from an appropriate risk set are attractive in the case of dependent censoring, the idea being that survival within the risk set is homogeneous enough to provide an unbiased impute. In the case where lung allocation causes removal from the waitlist, candidates’ time-dependent LAS values are available to group surviving candidates with a similar score for impute selection. Heterogeneity within this risk set may further be reduced by grouping on restricted mean estimates of survival as in Liu *et al.* [16]. However, even within this selected risk set, daily LAS changes give differing probabilities of transplant over time linked to patient progression that may cause bias. To account for this, our MI approach incorporates an IPCW weight based on the probability of transplant in the final stage of imputation from the risk set.

The rest of the manuscript is structured as follows. In Section 2, we briefly describe our proposed MI method that accounts for dependent censoring linked to daily changing allocation scores. Section 3 gives simulation studies of analyses based on the multiply imputed data sets, in particular, parameter estimates from a restricted mean model used to estimate patients urgency and point estimates of marginal survival. In Section 4, we conduct an analysis of the lung transplant data, providing transplant urgency and benefit estimates that would be appropriate for building a revised LAS based on this more recent cohort of data. Discussion follows in Section 5.

2. Estimating restricted mean life via multiple imputation

2.1. Background and notation

For subject i , $i = 1, \dots, n$, let T_i denote the true failure time, C_i be the corresponding censoring time, \mathbf{Z}_i be a vector of covariates affecting T_i , and $\bar{\mathbf{V}}_i(t) = \{\mathbf{V}_i(u); 0 \leq u \leq t\}$ be the subject’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 . If C_i is less than T_i , then we would not be able to observe the failure time for subject i . Let $X_i = \min(T_i, C_i)$ denote the observable random variable and $\Delta_i = I(T_i \leq C_i)$ be the failure indicator variable. In cases where \mathbf{V}_i influences T_i through C_i , we have dependent censoring, and traditional survival analyses are no longer unbiased.

Imputation approaches often select imputes from an appropriate model (parametric) or risk set (semiparametric). When the risk set is selected solely based on being at risk at time C_i , Taylor *et al.* [12] give an imputation strategy that corresponds to the KM 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 $U \sim \text{Uniform}(0,1)$, $S_T^{-1}(U)$ 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 $T|T > C_i$ from $\hat{S}_{T|T>C_i}^{-1}(U_i)$, where $\hat{S}_{T|T>C_i}^{-1}(t)$ is the conditional KM estimate among those at risk at C_i and U_i is a random $\text{Uniform}(0, 1)$ value. The sampled T is always one of the observed failures occurring beyond C_i with probability of being sampled equal to the size of the drop in the conditional KM curve at T .

In the case where censoring is dependent, this imputation strategy is biased, because the KM estimate does not adequately stand in for $S_{T|T>C_i}^{-1}(t)$ 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.* [13] further restrict the risk set, R_i , to those with similar hazards estimated from covariates measured up to C_i , so that the impute is sampled from $\hat{S}_{T|T>C_i, R_i}^{-1}(U_i)$.

We introduce several useful modifications to the inverse transform strategy for imputation to accommodate the features of the lung candidate data. First, instead of estimating survival with the KM estimate,

we estimate $\hat{S}_{T|T>C_i, R_i}^{-1}(U_i)$ 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 2.2. In Section 2.3, we describe how our risk set, R_i , is chosen. Part of this selection is based on a restricted mean model that decomposes $\log(T)$ into a piece depending on a linear predictor $\beta^T \mathbf{Z}$ 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 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 $\beta^T \mathbf{Z}_i$ 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 Schenker and Taylor [17] within a complete case risk set as well as by Liu *et al.* [16]. 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 , that is, the mean predicted from the censored patient's particular risk factors.

2.2. Inverse probability of censoring weighted 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 [6] 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 , $K_i^V(t)$. That is,

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

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

$$\lambda_Q \{t | \bar{\mathbf{V}}(t)\} = \lambda_{Q_0}(t) \exp\{\gamma' \mathbf{V}(t)\}. \tag{1}$$

In the case of the lung candidate data, time-dependent covariates in $\mathbf{V}(t)$ include the LAS at time t as well as active or inactive waiting status at t . Time-independent predictors in $\mathbf{V}(t)$ 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

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

the subject specific weight becomes

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

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.* [9] would also be appropriate. A conditional survival estimate among those in the risk set, R_i , at time C_i is given by

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

where $\hat{S}_{T|R_i}^W(t)$ and $\hat{\Lambda}_{T|R_i}^W(t)$ are IPCW estimates calculated within the risk set, R_i .

2.3. Risk set selection

Selection for the risk set used to estimate the appropriate failure time distribution typically incorporates information from covariates, \mathbf{Z}_i as well as longitudinal information pertaining to survival in $\mathbf{V}_i(t)$. 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

$$E[\log\{\min(\tau, T)\}] = \beta^T \mathbf{Z}, \quad (2)$$

where τ is a time window of interest and \mathbf{Z} is a vector of time-independent predictors. In the case of the lung candidate data, τ is 1 year, and this model allows for estimation of the days lived in a year based on an individual's risk factors.

When there is no censoring, standard linear model software is available to fit model (2). When censoring is dependent, Xiang and Murray [11] developed an IPCW-modified PO approach for obtaining estimates $\hat{\beta}^{PO^W}$ for model (2) that we use in constructing a risk set, R_i , for a candidate censored at C_i . In particular, the risk set is composed of those with $T > C_i$ and $\hat{\beta}^{PO^WT} \mathbf{Z}$ within a window of $\hat{\beta}^{PO^WT} \mathbf{Z}_i$, where \mathbf{Z}_i is the vector of predictors for the individual censored at C_i . In the lung candidate data, we further restrict the risk set R_i to those in the same diagnosis group (categorical component of \mathbf{Z}_i) with similar LAS values at time C_i to the candidate censored at C_i (longitudinal component of $\mathbf{V}_i(t)$ pertaining to urgency as well as transplant). When there are many individuals surviving beyond τ , it is rare to obtain a risk set where the final observation for $\min(\tau, T)$ is a censored value. In the few cases where this occurs, we recommend expanding the window size about $\hat{\beta}^{PO^WT} \mathbf{Z}_i$ until the last observation for $\min(\tau, T)$ is observed.

2.4. Multiple imputation of censored observations

The detailed steps of the algorithm are as follows.

- Step 1:* For each censored observation C_i , we select an appropriate risk set R_i as described in Section 2.3.
- Step 2:* Within this risk set, we then calculate the conditional survival probabilities $\hat{S}_{T|T>C_i, R_i}^W$ as described in Section 2.2.
- Step 3:* Generate a Uniform(0, 1) random variable, and identify the smallest observed failure time T_j , where $\hat{S}_{T|T>C_i, R_i}^W(T_j) \leq U$.
- Step 4:* Identify the residual ϵ_j from the model $\log\{\min(T_j, \tau)\} = \hat{\beta}^{PO^WT} \mathbf{Z}_j + \epsilon_j$ 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 j th subject, ϵ_j , to $\hat{E}[\log(T_i)] = \hat{\beta}^{PO^WT} \mathbf{Z}_i$, 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 data set.

Once M completed data sets are obtained, we may perform analyses using the formulaic approach given by Little and Rubin [18]. We summarize two analyses of interest in Sections 2.5 and 2.6.

2.5. Restricted mean model analysis on completed data sets

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

Our final vector of estimated MI coefficients for the restricted mean model that adjusts for dependent censoring is $\hat{\beta}^{MI^W} = \sum_{m=1}^M \hat{\beta}_m / M$. The variability associated with the parameter estimates is composed of two parts: the average within-imputation variance $\bar{W}_M = \sum_{m=1}^M \hat{W}_m / M$ and the between-imputation variance $\mathbf{B}_M = \sum (\hat{\beta}_m - \hat{\beta}^{MI^W})^2 / (M - 1)$. The variance of $\hat{\beta}^{MI^W}$ is then $\mathbf{V} = \bar{W}_M + (1 - M^{-1})\mathbf{B}_M$.

2.6. Marginal survival analysis on completed data sets

For each complete data set, we calculate the sample proportions corresponding to the survival estimates $\hat{S}_1(t), \hat{S}_2(t), \dots, \hat{S}_M(t)$, and their sample proportion variances $\hat{V}_1(t), \hat{V}_2(t), \dots, \hat{V}_M(t)$, where $\hat{V}_m(t) = \hat{S}_m(t)(1 - \hat{S}_m(t))/n$, $m = 1, 2, \dots, M$. 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. Simulation study

To study our MI method in finite sample sizes, we conducted a simulation study where a time-dependent variable influences censoring and survival and the mean structure follows (2). Parameter estimates for (2) are calculated in cases when (i) $\log[\min(\tau, T)]$ is uncensored; (ii) $\log[\min(\tau, T)]$ is subject to censoring and is replaced by log-transformed POs defined by Andersen *et al.* [19]; (iii) $\log[\min(\tau, T)]$ is subject to censoring and is imputed using only baseline covariates as in Liu *et al.* [16]; (iv) $\log[\min(\tau, T)]$ is subject to censoring and is replaced by IPCW POs as described by Xiang and Murray [11]; and (v) $\log[\min(\tau, T)]$ is subject to censoring and is imputed as described in Section 2.4. Imputation strategies in (iii) and (v) are also conducted with an additional bootstrap step, where each of the M imputed data sets is produced from a different bootstrap sample from the original observed data. Incorporating a bootstrap step in this manner is consistent with the recommendations from Rubin and Schenker [20], Heitjan and Little [21], Taylor *et al.* [12], and others in approximating a parameter draw from a posterior distribution as part of a ‘proper’ imputation procedure.

The MI approach allows for more possible analyses than merely fitting model (2); hence, as an example of an additional analysis of interest, the various MI strategies for computing point estimates for survival are compared with the KM estimate.

In each simulation, we perform the following:

- Step 1:* We generate Z_0 from a Bernoulli(0.5) distribution, 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_i , are generated from piecewise exponential distributions; that is, T_i has a constant hazard λ_{z_0} before time t_1 that changes to $\lambda_{z_0 z_1}$ after time t_1 , where $\lambda_0 = 0.3$, $\lambda_1 = 0.2$, $\lambda_{01} = 0.1$, and $\lambda_{11} = 0.5$ are fixed and λ_{00} and λ_{10} 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 . Further details on solving for parameters that satisfy the mean structure can be found in the appendix of Xiang and Murray [11], where a similar simulation strategy is used.
- Step 3:* Piecewise constant hazards leading to dependent censoring times, C_i , are based on the Cox model $\lambda^C(t|\mathbf{Z}(t)) = \lambda_0^C(t) \exp\{\gamma_0 Z_0 + \gamma_1 I[Z_0 = 0, Z_1 = 1, t > t_1] + \gamma_2 I[Z_0 = 1, Z_1 = 0, t > t_1] + \gamma_3 I[Z_0 = 1, Z_1 = 1, t > t_1] + \gamma_4 Z_2\}$, 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 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.

We consider two scenarios: $\beta = (\beta_0 = 0.8, \beta_1 = \beta_2 = 0)$ (i.e., true baseline covariate effects are zero) and $\beta = (\beta_0 = 1, \beta_1 = -0.8, \beta_2 = -0.5)$ (i.e., non-zero baseline covariate effects). Figures 1(a) and 1(b) show the examples of the assumed piecewise exponential survival curves for these two scenarios, assuming $Z_2 = 0.5$. A vertical line indicates that the time Z_1 is measured. Piecewise hazards driving mortality are displayed over the piecewise exponential curves used in simulation, and overall survival curves for $Z_0 = 0$ and $Z_0 = 1$ are superimposed in bold for this case with $Z_2 = 0.5$. Because Z_1 is simulated from a Bernoulli(0.5), the overall curves shown for $Z_0 = 0$ and $Z_0 = 1$ are arithmetic averages of the two piecewise exponential curves sharing the same value of Z_0 . In Figure 1(a), because there is no effect of Z_0 on the restricted mean, the restricted means corresponding to $Z_0 = g$, $g = 0, 1$ are algebraically

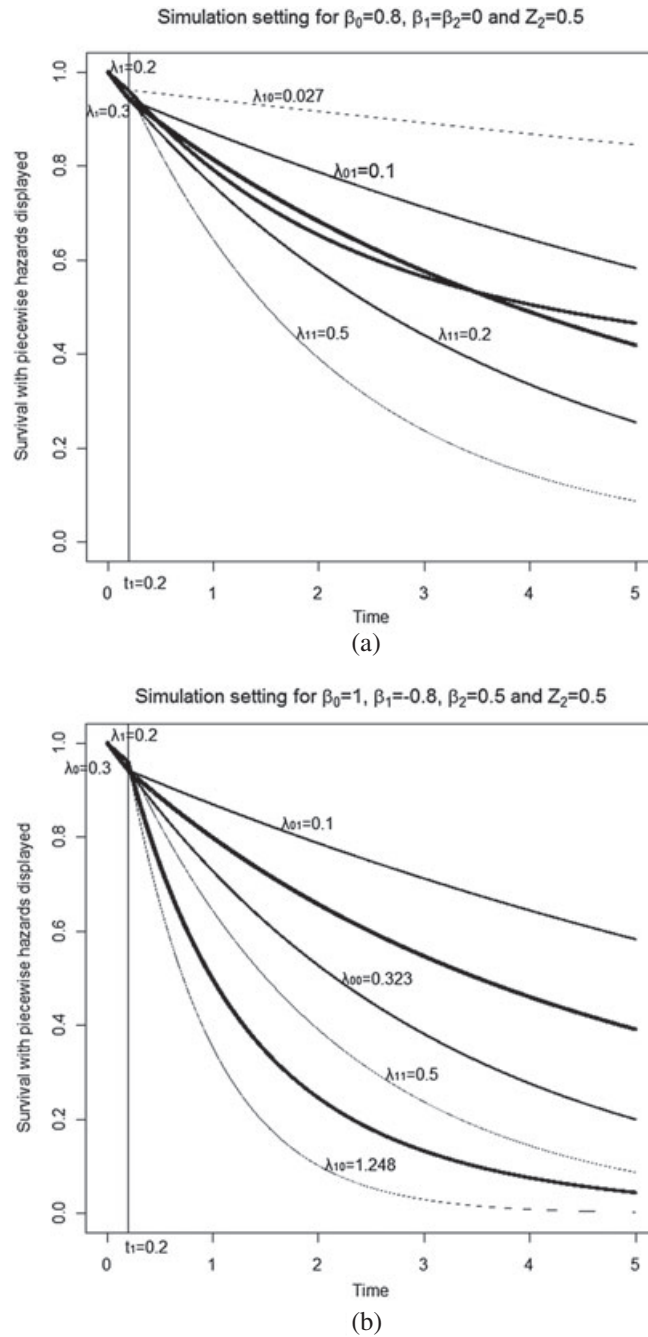


Figure 1. Survival curves under two simulation settings.

identical. Areas under the bold curves in Figure 1(b) capture the predictive value of Z_0 used in simulation for scenario 2.

In interpreting results, it is useful to consider the degree of bias. 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, KM) will suffer in either setting 1 or 2. For each scenario, 1000 simulations are run with $n = 150$ subjects and $\tau = 5$ years.

We begin by summarizing the results under scenario 1. The results for fitting model (2) under scenario 1 ($\beta_0 = 0.8, \beta_1 = \beta_2 = 0$) are located in part (A) of Table I. The PO method that does not adjust for dependent censoring gives biased estimates for all parameters, with bias $(\hat{\beta}_0^{PO}) > \text{bias}(\hat{\beta}_1^{PO}) > \text{bias}(\hat{\beta}_2^{PO})$.

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

Parameters	Method	Bias	SE ^a	ESD ^b	CP ^c (%)
(A) Covariate effects are zero.					
$\beta_0 = 0.8$	Uncensored	0.016	0.210	0.216	94.1
	PO	-0.651	0.421	0.569	69.9
	MI	0.088	0.213	0.229	88.5
	MI_b	0.082	0.214	0.223	91.0
	PO^W	-0.091	0.245	0.229	94.8
	MI^W	0.039	0.215	0.234	91.0
	MI_b^W	0.043	0.223	0.226	94.1
$\beta_1 = 0$	Uncensored	-0.001	0.188	0.183	95.4
	PO	-0.291	0.376	0.374	88.8
	MI	-0.234	0.193	0.203	74.5
	MI_b	-0.229	0.193	0.209	75.3
	PO^W	0.043	0.219	0.208	96.3
	MI^W	-0.017	0.193	0.204	93.5
	MI_b^W	-0.022	0.202	0.210	93.7
$\beta_2 = 0$	Uncensored	-0.015	0.326	0.325	94.1
	PO	0.135	0.652	0.681	69.9
	MI	-0.087	0.333	0.356	88.5
	MI_b	-0.091	0.336	0.350	91.0
	PO^W	-0.019	0.380	0.364	94.8
	MI^W	-0.044	0.335	0.370	91.0
	MI_b^W	-0.078	0.359	0.369	94.1
(B) Covariate effects are non-zero.					
$\beta_0 = 1$	Uncensored	0.007	0.202	0.205	95.0
	PO	-0.230	0.308	0.320	90.9
	MI	0.002	0.211	0.220	93.6
	MI_b	-0.011	0.212	0.213	94.0
	PO^W	-0.100	0.239	0.223	94.7
	MI^W	0.004	0.206	0.223	92.5
	MI_b^W	-0.011	0.207	0.214	93.7
$\beta_1 = -0.8$	Uncensored	-0.008	0.181	0.185	94.8
	PO	-0.230	0.276	0.349	85.7
	MI	0.004	0.191	0.205	93.1
	MI_b	0.011	0.191	0.200	93.3
	PO^W	-0.006	0.214	0.199	97.2
	MI^W	-0.013	0.186	0.201	92.2
	MI_b^W	-0.001	0.187	0.196	93.4
$\beta_2 = -0.5$	Uncensored	-0.007	0.313	0.320	93.3
	PO	0.038	0.478	0.496	93.9
	MI	0.087	0.332	0.353	93.1
	MI_b	0.107	0.333	0.341	93.8
	PO^W	0.058	0.372	0.351	95.3
	MI^W	0.007	0.321	0.367	90.5
	MI_b^W	0.026	0.326	0.350	93.7

^aAverage standard error based on individual analysis.

^bEmpirical standard deviation of parameter estimates across 1000 simulations.

^cEmpirical coverage probability.

Parameter estimates, $\hat{\beta}^{PO^W}$, using the IPCW-adjusted PO method, labeled PO^W , have comparably small bias, with a slight tendency to underestimate β_0 . When conducting MI using only baseline covariates, bias in estimating β_1 is on the order of a standard error, with more modest bias seen in estimating β_0 and β_2 . The IPCW-adjusted MI methods, labeled MI^W and MI_b^W , overall have the best track records of removing bias.

The empirical standard deviations were comparable with the standard errors averaged across simulations with the exception of the intercept term for the PO method. The results are comparable with and without a bootstrap step incorporated into the MI algorithms to the degree that we are comfortable recommending inference either with or without this step. In particular, only very minor differences are seen in the average standard errors based on individual analyses (SE column) and the coverage rates (CP column) of Table I. If we consider that the number of simulation iterations is 1000, then an approximate 95% confidence interval for a 0.95 coverage rate is (0.936, 0.964). MI^W and MI_b^W coverage estimates from Table I run slightly low, observed between 0.905 and 0.941. The added bootstrap step pushed coverage rates slightly above the 0.936 threshold in all but one case, whereas without the bootstrap step, the estimated coverage rates fell below the 0.936 threshold. This gives some support to the idea that adding the bootstrap step improves undercoverage slightly and may be preferred, at least in principle, as a ‘proper’ imputation strategy.

Recall that in setting 1, censored individuals tend to have longer times to event so that the methods struggling with bias will tend to underestimate restricted lifetimes. In the case with no censoring, that is, no bias, the proportion of times that model (2) estimates a lower restricted lifetime than the true time to event is 31%. This percentage increases as bias from estimation of model (2) increases. The proportion of times that estimated restricted means undershoot the observed times to event is highest for PO (55%), followed by MI , PO^W , and MI^W (34%, 34%, and 31%, respectively).

For an individual with $Z_0 = 1$ and $Z_2 = 0.5$, the PO , MI , and PO^W methods underestimate the time lived during the 5-year period by 10 months, 2 months, and 21 days on average, respectively, while the MI^W method is off by less than 1 day on average over the 5-year period.

In addition to fitting model (2), the MI^W procedure can perform other traditional analyses of interest. For instance, MI^W and its counterpart that incorporates a bootstrap step, MI_b^W , give nearly unbiased estimates for survival for all t shown in part (A) of Table II for scenario 1. This table gives marginal survival probability estimates at years 1, 2, 3, 4, and 5 for all imputation methods as well as the censored version of the data using KM survival. At 5 years, the KM underestimates the 5-year survival by 7%, and the imputation methods using only baseline covariates (MI and MI_b) underestimate survival by approximately 10%.

Recall that in scenario 2, ($\beta_0 = 1$, $\beta_1 = -0.8$, $\beta_2 = -0.5$), censored patients tend to have shorter times to event. The PO method that does not properly adjust for dependent censoring once again gives strong bias for β_0 and β_1 (Table I(B)). MI methods incorporating information from only baseline covariates (MI and MI_b) show lingering bias primarily for the β_2 term. For the remaining procedures (PO^W , MI^W and MI_b^W), the magnitude of bias is minimal.

The average bias for estimating the time lived during the 5-year period was approximately 5 months for the PO method, 1 month for the PO^W method, 18 days for the MI method, and 2 days for the MI^W method over the 5-year period for a patient with $Z_0 = 1$, $Z_2 = 0.5$. Table II(B) shows that the 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.

4. Example

We now return to the Organ Procurement and Transplantation Network (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.

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 and other restrictive lung disease, has the highest censoring percentage (77%),

Table II. Comparison of survival estimates using uncensored observations (Uncensored), Kaplan–Meier (KM), unadjusted MI (MI), unadjusted MI with bootstrap (MI_b), IPCW-adjusted MI (MI^W), and IPCW-adjusted MI with bootstrap (MI_b^W) under two scenarios.

Time	Truth	Uncensored	KM	MI	MI _b	MI ^W	MI _b ^W
(A) Covariate effects are zero.							
1	0.806	0.806 (0.000) ^a [0.032, 0.033, 100] ^b	0.798 (−0.008) [0.035, 0.037, 100]	0.796 (−0.010) [0.035, 0.037, 99.9]	0.794 (−0.012) [0.035, 0.038, 100]	0.806 (0.000) [0.033, 0.036, 99.8]	0.801 (−0.005) [0.036, 0.038, 100]
2	0.664	0.671 (0.007) [0.038, 0.039, 100]	0.645 (−0.019) [0.046, 0.047, 99.9]	0.635 (−0.029) [0.044, 0.050, 95.7]	0.632 (−0.032) [0.044, 0.051, 96.4]	0.672 (0.008) [0.041, 0.046, 99.0]	0.664 (0.000) [0.045, 0.048, 99.6]
3	0.569	0.574 (0.005) [0.040, 0.040, 100]	0.531 (−0.038) [0.052, 0.053, 97.1]	0.511 (−0.058) [0.046, 0.058, 74.8]	0.507 (−0.062) [0.047, 0.057, 78.2]	0.576 (0.007) [0.043, 0.055, 96.5]	0.566 (−0.003) [0.048, 0.058, 97.8]
4	0.501	0.501 (0.000) [0.041, 0.040, 100]	0.442 (−0.059) [0.056, 0.055, 92.4]	0.412 (−0.089) [0.046, 0.062, 52.1]	0.414 (−0.087) [0.047, 0.061, 55.6]	0.499 (−0.002) [0.044, 0.068, 90.7]	0.494 (−0.007) [0.049, 0.067, 93.3]
5	0.450	0.445 (−0.005) [0.040, 0.040, 100]	0.377 (−0.073) [0.058, 0.058, 86.3]	0.344 (−0.106) [0.044, 0.063, 40.1]	0.347 (−0.103) [0.046, 0.063, 43.7]	0.446 (−0.004) [0.043, 0.075, 86.7]	0.443 (−0.007) [0.049, 0.073, 89.8]
(B) Covariate effects are non-zero.							
1	0.648	0.646 (−0.002) [0.039, 0.039, 100]	0.667 (0.019) [0.041, 0.041, 99.7]	0.661 (0.013) [0.041, 0.042, 100]	0.659 (0.011) [0.041, 0.042, 100]	0.647 (−0.001) [0.041, 0.043, 100]	0.646 (−0.002) [0.041, 0.042, 100]
2	0.458	0.457 (−0.001) [0.041, 0.039, 100]	0.495 (0.037) [0.047, 0.045, 98.6]	0.483 (0.025) [0.044, 0.046, 98.7]	0.482 (0.024) [0.044, 0.046, 99.2]	0.456 (−0.002) [0.043, 0.048, 99.7]	0.456 (−0.002) [0.044, 0.048, 99.6]
3	0.350	0.348 (−0.002) [0.039, 0.037, 100]	0.393 (0.043) [0.048, 0.047, 97.1]	0.378 (0.028) [0.043, 0.047, 97.5]	0.375 (0.025) [0.043, 0.047, 97.7]	0.347 (−0.003) [0.041, 0.048, 99.1]	0.345 (−0.005) [0.042, 0.047, 99.2]
4	0.279	0.278 (−0.001) [0.036, 0.035, 100]	0.325 (0.046) [0.048, 0.047, 95.4]	0.310 (0.031) [0.041, 0.047, 95.2]	0.307 (0.028) [0.042, 0.047, 95.9]	0.279 (0.000) [0.039, 0.047, 97.5]	0.278 (−0.001) [0.040, 0.048, 98.5]
5	0.230	0.229 (−0.001) [0.034, 0.033, 100]	0.276 (0.046) [0.048, 0.047, 94.6]	0.263 (0.033) [0.039, 0.047, 92.1]	0.260 (0.030) [0.040, 0.048, 93.3]	0.236 (0.006) [0.037, 0.047, 94.9]	0.234 (0.004) [0.038, 0.048, 96.7]

^aEstimate (bias).

^b[Average standard error based on individual analysis, empirical standard deviation of parameter estimates across 1000 simulations, coverage]

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

Parameter	Hazard ratio	95% CI	<i>p</i> -value
Characteristic at listing			
Female (vs. male)	0.72	(0.63, 0.82)	< 0.0001
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)	< 0.0001
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)	< 0.0001
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)	< 0.0001
Inactive status (vs. active)	0.00	(0, > 1000)	0.8780
Off the waitlist (vs. active)	0.00	(0, > 1000)	0.9410

LAS, lung allocation score; CI, confidence interval.

^aHazard ratio (HR) corresponding to one unit increase for LAS 30+ relative to those with 0 < LAS < 30.

^bHR corresponding to spline term for LAS 35+, giving HR due to one unit increase in LAS in the range 35 ≤ LAS < 40 of 1.12 * 0.98 = 1.10 relative to 0 < LAS < 30.

^cHR corresponding to spline term for LAS 40+, giving HR due to one unit increase in LAS in the range 40 ≤ LAS < 60 of 1.12 * 0.98 * 0.95 = 1.04 relative to 0 < LAS < 30.

^dHR corresponding to spline term for LAS 60+, giving HR due to one unit increase in LAS in the range LAS ≥ 60 of 1.12 * 0.98 * 0.95 * 0.97 = 1.01 relative to 0 < LAS < 30.

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 (1). Parameter estimates are displayed in Table III. Although one might suppose that the probability of transplant should increase monotonically with higher LAS value, the parameter estimates from Table III 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. The risk set is further restricted to having a similar urgency estimate based on $\hat{\beta}^{PO^W}$ as shown in the second column of Table IV. The predictors used in this model precisely match those proposed by the OPTN Thoracic Committee [22]; some predictors are included based on significant association with survival seen in previously studied cohorts.

Ten imputed data sets are built from the MI procedure outlined in Section 2.4. Estimates, $\hat{\beta}^{MI^W}$, based on the imputed data sets are shown in the rightmost column of Table IV. 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 body mass index of 25, has no

Table IV. Lung waitlist models using two different methods for 3701 lung candidates.

	PO^W e^{β^a} (95% CI), p -value	MI^W e^{β} (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
Lymphangiomyomatosis	1.25 (0.10, 16.41), 0.8639	1.03 (0.69, 1.55), 0.8729
Obliterative bronchiolitis	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 > 30 mmHg	0.46 (0.16, 1.37), 0.1641	1.18 (0.95, 1.46), 0.1280
Sarcoidosis and PA mean ≤ 30 mmHg	0.84 (0.22, 3.13), 0.7893	0.85 (0.69, 1.06), 0.1454
Physiologic reserve		
Age (years)	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 (ref = some/total assistance with ADL)	1.08 (0.65, 1.80), 0.7609	1.01 (0.93, 1.10), 0.8762
Six-minute walk (per 100 ft)	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 groups A, C, and D (L/min)	0.87 (0.83, 0.91), < 0.0001	0.92 (0.91, 0.93), < 0.0001
PA systolic (per 10 mmHg) for group A	0.91 (0.71, 1.16), 0.4331	0.99 (0.95, 1.04), 0.6922
PCO ₂ increase of ≥ 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

CI, confidence interval; COPD, chronic obstructive pulmonary disease; iPAH, idiopathic pulmonary arterial hypertension; CF, cystic fibrosis; IPF, interstitial pulmonary fibrosis; PA, pulmonary artery; BMI, body mass index; FVC, force vital capacity.

^aFor risk factors, e^{β} acts multiplicatively on the number of days lived in a year.

^bThese diagnoses were grouped into larger diagnosis groups (A, B, C, and D) by the OPTN Thoracic Committee for the purpose of modeling risk factors that may vary by diagnosis group. Bronchiectasis, lymphangiomyomatosis, and sarcoidosis and PA mean ≤ 30 mmHg share risk factor parameters with diagnosis group A; Eisenmenger with group B; and obliterative bronchiolitis, pulmonary fibrosis other, and sarcoidosis and PA mean > 30 mmHg with group D.

^cADL, activities of daily living.

diabetes, requires no assistance with ADL, walks 300 ft in 6 minutes, requires 2 L/min of O₂ at rest, has PCO₂ of 50 mmHg, is not on a ventilator, has a stable creatinine of 0.8 mg/dL, and has a cardiac index > 2 L/min/min².

The *PO^W* method seems to estimate less waitlist days lived in a year compared with the *MI^W* method. For comparison, we also include urgency estimates built from an integrated IPCW Cox model [7, 8], where the partial likelihood score function incorporating censoring weights becomes

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

Parameter estimates for the IPCW Cox model are located in Table V. Boxplots of patient urgency by diagnosis group are located in Figure 2(a). Although all three methods account for dependent censoring,

Table V. Lung waitlist models using Cox proportional hazards model with inverse weight for 3701 lung candidates.			
Parameter	Hazard ratio	95% CI	p-value
Diagnosis group (ref = group A, primarily COPD)			
Group B (primarily iPAH)	2.47	(1.78, 3.44)	< 0.0001
Group C (primarily CF)	1.46	(1.09, 1.95)	0.0105
Group D (primarily IPF)	2.39	(1.75, 3.25)	< 0.0001
Diagnosis ^a			
Bronchiectasis	1.19	(0.86, 1.65)	0.3027
Lymphangioleiomyomatosis	0.93	(0.44, 1.93)	0.8358
Obliterative bronchiolitis	1.39	(0.89, 2.16)	0.1429
Pulmonary fibrosis other	0.80	(0.67, 0.96)	0.0180
Sarcoidosis and PA mean > 30 mmHg	0.89	(0.73, 1.10)	0.2803
Sarcoidosis and PA mean ≤ 30 mmHg	1.03	(0.76, 1.39)	0.8513
Physiologic reserve			
Age (years)	1.00	(1.00, 1.01)	0.4194
BMI (kg/m ²)	0.97	(0.96, 0.98)	< 0.0001
Diabetes	1.38	(1.25, 1.52)	< 0.0001
No assistance with ADL ^b (ref = some/total assistance with ADL)	1.00	(0.88, 1.14)	0.9882
Six-minute walk (per 100 ft)	0.97	(0.96, 0.98)	< 0.0001
Severity			
FVC for group D (per 10% predicted)	0.94	(0.90, 0.97)	0.0003
O ₂ requirement for groups A, C, and D (L/min)	1.10	(1.09, 1.11)	< 0.0001
PA systolic (per 10 mmHg) for group A	1.05	(0.99, 1.12)	0.0924
PCO ₂ increase of ≥ 15%	1.38	(1.08, 1.75)	0.0095
PCO ₂ (mmHg)	1.01	(1.01, 1.01)	< 0.0001
Ventilator	4.29	(3.49, 5.28)	< 0.0001
Creatinine (mg/dL)	1.68	(1.43, 1.98)	< 0.0001
Cardiac index < 2.0 (L/min/min ²)	1.30	(1.11, 1.52)	0.0010

CI, confidence interval; COPD, chronic obstructive pulmonary disease; iPAH, idiopathic pulmonary arterial hypertension; CF, cystic fibrosis; IPF, interstitial pulmonary fibrosis; PA, pulmonary artery; BMI, body mass index; FVC, force vital capacity.

^aThese diagnoses were grouped into larger diagnosis groups (A, B, C, and D) by the OPTN Thoracic Committee for the purpose of modeling risk factors that may vary by diagnosis group. Bronchiectasis, lymphangioleiomyomatosis, and sarcoidosis and PA mean ≤ 30 mmHg share risk factor parameters with diagnosis group A; Eisenmenger with group B; and obliterative bronchiolitis, pulmonary fibrosis other, and sarcoidosis and PA mean > 30 mmHg with group D.

^bADL, activities of daily living.

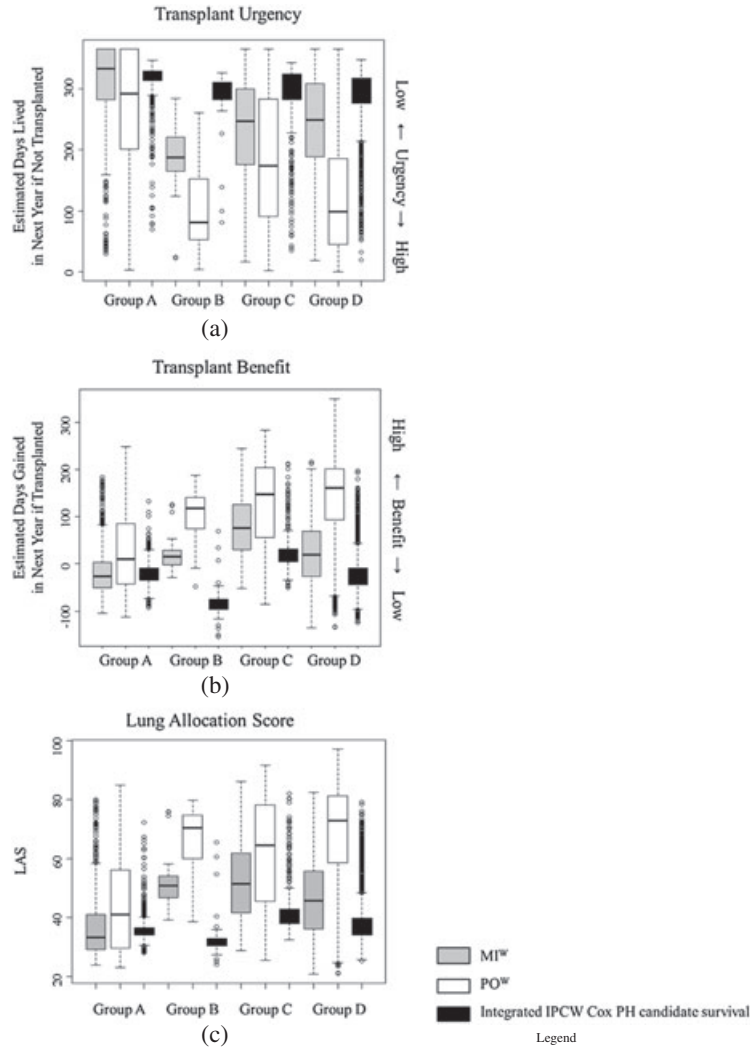


Figure 2. Estimated transplant urgency, benefit, and lung allocation score (LAS) at time of listing by diagnosis group and estimation method for 3701 lung transplant candidates.

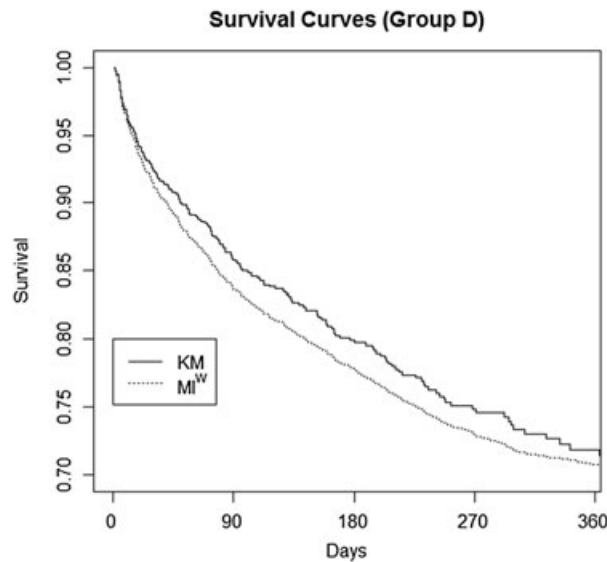


Figure 3. Survival curves for group D patients using KM and MI^W .

differences in urgency are seen based on the method used. The 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 the first year of listing.

A few model diagnostics are 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 the observed C_i 35% of the time, while the proportions are 20% and 16% for the Cox and MI^W methods, respectively. Among observed failures, the sum of squared residuals between the observed (restricted) failure times 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 Cox urgency estimates (15,373,491) and the PO^W estimates (17,921,077). We also look at the concordance index [23, 24], which is the percent of times that a model predicts two pairs of data (X_i, Δ_i) , (X_j, Δ_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 Cox methods, respectively.

Table VI. Lung post-transplant model for 4784 transplant recipients (no censored data).

	$e^{\beta a}$	95% CI	<i>p</i> -value
(Intercept)	344.42	(303.96, 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
Lymphangioliomyomatosis	1.24	(0.89, 1.74)	0.2060
Obliterative bronchiolitis	1.25	(0.93, 1.69)	0.1437
Pulmonary fibrosis other	1.01	(0.89, 1.15)	0.8734
Sarcoidosis and PA mean > 30 mmHg	0.90	(0.74, 1.08)	0.2561
Sarcoidosis and PA mean ≤ 30 mmHg	1.00	(0.79, 1.26)	0.9927
Physiologic reserve			
Age > 45 spline ^c (years)	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 100 ft)	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 and 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 groups B, C, and D (L/min)	0.99	(0.98, 1.00)	0.2129
Change in creatinine ≥ 150%	0.78	(0.65, 0.95)	0.0132

CI, confidence interval; COPD, chronic obstructive pulmonary disease; iPAH, idiopathic pulmonary arterial hypertension; CF, cystic fibrosis; IPF, interstitial pulmonary fibrosis; PA, pulmonary artery; BMI, body mass index; FVC, force vital capacity.

^aFor risk factors, e^{β} acts multiplicatively on the number of days lived in a year.

^bThese diagnoses were grouped into larger diagnosis groups (A, B, C, and D) by the OPTN Thoracic Committee for the purpose of modeling risk factors that may vary by diagnosis group. Bronchiectasis, lymphangioliomyomatosis, and sarcoidosis and PA mean ≤ 30 mmHg share risk factor parameters with diagnosis group A; Eisenmenger with group B; obliterative bronchiolitis, pulmonary fibrosis other, and sarcoidosis and PA mean > 30 mmHg with group D.

^cAge > 45 spline: the maximum of 0 and age-45.

^dADL, activities of daily living.

Marginal survival analysis shows lower survival curve estimates in group D using MI^W versus the KM (Figure 3). This group also experiences the shortest average 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 [25]).

4.2. Lung recipient analysis, transplant benefit, and lung allocation score

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 1 year from the time of transplant, that is, no censoring of the 1-year restricted mean, with 816 (17%) deaths within that first year. The results from fitting model (2) in this uncensored case are shown in Table VI. 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 VI. 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 $266 - 227 = 39$ days during that year.

Figure 2(b) shows the 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 2(c).

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 data sets that adjust for dependent censoring. Availability of these completed data sets, along with the standard ways of combining the 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 restricted 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 IPCW PO 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 opportunity in time. The 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 with the others. Whenever any one model influences public policy to this degree, model comparisons similar to those performed in Section 4, where the PO^W method gave consistently improved results, are important. A personal favorite for comparing model fit across vastly different modeling paradigms in the organ allocation setting is the index of concordance approach [23, 24]. This metric captures the spirit of correctly ordering urgency for any two patients being considered for an organ.

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.

Acknowledgements

This research was funded in part by the Health Resources and Services Administration, US Department of Health and Human Services, Scientific Registry of Transplant Recipients contract number 234-2005-37009C. The views expressed herein are those of the authors and not necessarily those of the US Government. The authors express appreciation to Tempie Shearon, Kathryn Meyer, and Ying Qian, formerly of the Scientific Registry of Transplant Recipients (SRTR), for their support.

References

1. Egan TM, Murray S, Bustami RT, Shearon TH, McCullough KP, Edwards LB, Coke MA, Garrity ER, Sweet SC, Heiney DA, Grover FL. Development of the new lung allocation system in the United States. *American Journal of Transplantation* 2006; **6**(Part 2):1212–1227.
2. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. *Journal of American Statistics Association* 1958; **53**:457–481, 562–563.
3. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**:187–200.
4. Savage IR. Contributions to the theory of rank-order statistics – the two sample case. *Annals of Mathematical Statistics* 1956; **27**:590–615.
5. Mantel N. Evaluation of survival data and two new rank-order statistics arising in its consideration. *Cancer Chemotherapy Reports* 1966; **50**:163–170.
6. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**:779–788.
7. Robins JM. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, American Statistical Association: Alexandria, Virginia, 1993, 24–33.
8. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology-Methodological Issues*, Jewell N, Dietz K, Farewell B (eds). Birkhäuser: Boston, 1992; 297–331.
9. Satten GA, Datta S, Robins JM. Estimating the marginal survival function in the presence of time dependent covariates. *Statistics and Probability Letters* 2001; **54**:397–403.
10. Scharfstein D, Robins JM, Eddings W, Rotnitzky A. Inference in randomized studies with informative censoring and discrete time-to-event endpoints. *Biometrics* 2001; **57**:404–413.
11. Xiang F, Murray S. Restricted mean models for transplant benefit and urgency. *Statistics in Medicine* 2012; **31**:561–576.
12. Taylor JMG, Murray S, Hsu CH. Survival estimation and testing via multiple imputation. *Statistics and Probability Letters* 2002; **58**:221–232.
13. Hsu CH, Taylor JM, Murray S, Commenges D. Survival analysis using auxiliary variables via non-parametric multiple imputation. *Statistics in Medicine* 2006; **25**:3503–3517.
14. Faucett CL, Schenker N, Taylor JMG. Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics* 2002; **58**:37–47.
15. Wei GCG, Tanner MA. Application of multiple imputation to the analysis of censored regression data. *Biometrics* 1991; **47**:1297–1309.
16. Liu LX, Murray S, Tsodikov A. Multiple imputation based on restricted mean model for censored data. *Statistics in Medicine* 2011; **30**:1339–1350.
17. Schenker N, Taylor JMG. Partially parametric techniques for multiple imputation. *Computational Statistics and Data Analysis* 1996; **22**:425–446.
18. Little RJA, Rubin D. *Statistical Analysis with Missing Data*, 1st edn. Wiley: New York, 1986.
19. Andersen PK, Hansen M, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. *Life Data Analysis* 2004; **10**:335–350.
20. Rubin DB, Schenker N. Multiple imputations in health-care database: an overview and some applications. *Statistics in Medicine* 1991; **10**:585–598.
21. Heitjan DF, Little RJA. Multiple imputation for the fatal accident reporting system. *Applied Statistics* 1991; **40**:13–29.
22. OPTN Thoracic Organ Transplantation meeting minutes March 23, 2010.
23. Harrel FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *Journal of the American Medical Association* 1982; **247**:2543–2546.
24. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 1996; **15**:361–387.
25. Yusen RD, Shearon TH, Qian Y, Kotloff R, Barr ML, Sweet S, Dyke DB, Murray S. Lung transplantation in the United States, 1999–2008. *American Journal of Transplantation* 2010; **10**(Part 2):1047–1068.