Statistical consequences of a successful lung allocation system – recovering information and reducing bias in models for urgency

Nabihah Tayob[a*†] and Susan Murray[b]

The lung allocation system has reduced the number of waitlist deaths by ranking transplant candidates on the basis of a lung allocation score that requires estimation of the current 1-year restricted mean waitlist survival (urgency). Fewer waitlist deaths and the systematic removal of candidates from the waitlist for transplantation present statistical challenges that must be addressed when using recent waitlist data. Multiple overlapping 1-year follow-up windows are used in a restricted mean model that estimates patient urgency on the basis of updated risk factors at the start of the window. In simulation studies, our proposed multiple imputation procedure was able to produce unbiased parameter estimates with similar efficiency to those obtained if censoring had never occurred. The analysis of 10,740 lung transplant candidates revealed that for most risk factors incorporating additional follow-up windows produced more efficient estimates. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: multiple imputation; dependent censoring; correlated follow-up windows; pseudo-observations; lung allocation score

1. Introduction

Since 2005, national lung allocation policy in the United States of America for those aged 12 and over has relied on the statistical estimation of a lung transplant candidate's 1-year restricted mean lifetime without opportunity for transplant (urgency) and the number of days to be gained in the next year if a transplant is offered immediately (benefit) [1]. The United Network for Organ Sharing (UNOS) is charged with collecting and updating patient risk factors so that the lung allocation score (LAS) that determines transplantation priority can change with a candidate's prognosis. In fact, patients are required to update their allocation factors every 6 months or be penalized with a zero LAS value that effectively puts them at the end of the candidate list (OPTN Policy 3.7; https://optn.transplant.hrsa.gov/PoliciesandBylaws2/policies/pdfs/policy_9.pdf).

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

[a]*Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, U.S.A.*
[b]*Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.*
Correspondence to: Nabihah Tayob, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston, TX 77030, U.S.A.
[†]*E-mail: ntayob@mdanderson.org*

because listing recommendations tend to catch patients at a similar state of urgency at the time these listing risk factors are collected, whereas patients progress at quite different rates thereafter. In general, we believe that being able to model the restricted mean survival across a mixture of multiple short-term windows of follow-up provides an alternative and potentially superior understanding of time-dependent risk factors.

Gong and Schaubel [2] model the distribution of survival time from a set of specified calendar times, conditional on the risk factors measured at each specified calendar time, through Cox-regression models. They account for dependent censoring through inverse probability of censoring weights. An estimate of restricted mean survival can be obtained from a Cox-regression model by integrating the estimated survival curve over the follow-up window of interest. Because our interest lies in estimation of a 1-year restricted mean survival for the lung transplant urgency model, we have chosen to model the restricted mean survival directly.

We extend a multiple imputation procedure developed by Xiang, Murray and Liu [3] for dependently censored data to address issues of removal from the lung candidate pool based on LAS values involving urgency. While they were able to use time-dependent LAS scores to obtain consistent estimates of survival via the inverse probability of censoring weighted (IPCW) method of Robins [4], they were not able to incorporate updated covariate information directly into their restricted mean model for transplant urgency.

The designers of the LAS were specifically tasked with producing a score that does not change according to the length of time spent on the waitlist. However, they are equally interested in using information collected beyond the first year of listing. Our approach is to build a restricted mean model that estimates transplant urgency across the mixture of available follow-up windows. The (possibly) censored survival times within each follow-up window are treated as correlated multivariate (longitudinal) outcomes. The updated risk factors available at the start of each follow-up window become the covariates for the corresponding restricted time-to-event recorded for that window. Although there is no technical restriction from including the waiting time accrued by the start of each follow-up window as a covariate, our goal of estimating each patient's urgency score in a manner that does not depend on waitlist time precludes the use of this covariate.

A clear advantage of our proposed model is the broader range of patient experience that contributes to estimation of 1-year restricted mean lifetime. Standard methods do not take additional follow-up after the first year from listing into account. This is unfortunate, because follow-up extending beyond the initial year of listing has a great deal of information to contribute regarding candidate transplant urgency. Our method takes advantage of available, but neglected, follow-up data when fitting the model for the estimating 1-year restricted mean lifetime. In terms of using the final fitted model to produce urgency estimates for currently listed patients, a patient's urgency estimate would be calculated using parameters from the final fitted model with predictors taken from that patient's most recent risk factor update. As this patient's risk profile changes, the calculation is repeated, substituting the most recent predictors into the linear predictor of the model. Arguably, the mixture distribution of 1-year follow-up windows used to produce our urgency model's parameter estimates is a more relevant sample of follow-up times than data from the first year of listing alone because the urgency model is used to score all waitlisted patients for lung allocation, including patients that continue to wait beyond the first year from listing.

Methods are summarized in Section 2. Notation and the data structure induced by using multiple follow up windows are described in Sections 2.1 and 2.2. An outline of our multiple imputation procedure is given in Section 2.3, and the details are provided in Section 2.6. The multiple imputation procedure results in imputed datasets with no censoring, which allows us to take advantage of generalized estimating equation (GEE) methods [5] to account for the particular flavor of correlation induced from incorporating (overlapping) follow-up windows from the same patient. The use of GEE methods applied to imputed datasets in this setting is described in Section 2.4. Methods for combining the analyses from multiply imputed datasets are described in Section 2.5. Section 3 assesses our approach versus alternative approaches for estimating lung candidate urgency via simulation. We then analyze a recent release of lung transplant data collected by UNOS in Section 4, providing updated urgency measures in this cohort as well as evaluating urgency changes over time. A discussion follows in Section 5.

2. Multiple Imputation Methodology

2.1. Notation

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

Our proposed method incorporates information from several follow-up windows of length 1 year, spaced 6 months apart; that is, windows start at $\{0, 6, 12, \ldots\}$ months until removal from the candidate list for transplantation. More generally, windows of length τ start at times $\{t_1, \ldots, t_{n_i}\}$, where the ith patient contributes n_i follow-up windows. For notational convenience, we define b as the maximum number of follow-up windows any patient could potentially contribute over the entire study duration, that is, $b = \max\{n_1, \ldots, n_n\}$ subject to having at least 25 events in the final window.

2.2. Model and data structure

Our data structure is similar in nature to longitudinal data with potentially time-varying covariates. The key differences between our data structure and the standard longitudinal data analysis structure are as follows: (1) our time-dependent outcomes are short-term restricted times-to-event in the different follow-up windows; (2) our outcomes are subject to dependent censoring; and (3) our potentially time-dependent covariates, $\mathbf{Z}$, are captured at the start of the corresponding follow-up window (as opposed to being measured concurrently as in a standard longitudinal data structure). As in longitudinal data analysis, our goal is to model the average outcome based on covariates.

The outcome variable of interest, T^*, denotes a τ-restricted survival time when sampling from follow-up windows of length τ among patients at-risk in the cohort. Each patient contributes n_i longitudinal measures of T^*, regularly spaced by design for all patients in the cohort. Our urgency model is then

$$E\{\log(T^*)|\mathbf{Z}\} = \boldsymbol{\beta}^T \mathbf{Z}. \tag{1}$$

Although time and time by covariate interactions theoretically may be modeled as part of $\mathbf{Z}$, we do not include these terms when estimating lung waitlist urgency in our application. This modeling choice satisfies the mandate that lung allocation formulas not involve waiting time in a formulaic manner, which patients could use to their advantage.

For bookkeeping purposes, we find it convenient to include window-start-time notation to keep track of unique components in the data structure. For individuals with available follow-up during a window starting at t, where $t \in \{t_1, \ldots, t_b\}$, we define $T_i^*(t) = \min(T_i - t, \tau)$ as the τ-restricted time to event from t. The multivariate data for each individual become $\{T_i^*(t_1), \mathbf{Z}_i(t_1)\}$, $\{T_i^*(t_2), \mathbf{Z}_i(t_2)\}, \ldots, \{T_i^*(t_{n_i}), \mathbf{Z}_i(t_{n_i})\}$.

In Figure 1, we illustrate the relationships between T_i, $T_i^*(t)$ and $\mathbf{Z}_i(t)$ for $t = 0, 6, 12,$ and 18 months. In our first example, patient 1 dies at 20 months post listing ($T_1 = 20$). Hence, patient 1 contributes information on 1-year survival via the data pairs $\{T_1^*(0) = 12 \text{ months}, \mathbf{Z}_1(0)\}$, $\{T_1^*(6) = 12 \text{ months}, \mathbf{Z}_1(6)\}$, $\{T_1^*(12) = 8 \text{ months}, \mathbf{Z}_1(12)\}$, and $\{T_1^*(18) = 2 \text{ months}, \mathbf{Z}_1(18)\}$. Patient 2 is has $T_2 = 7$ months and is therefore observed for two follow-up windows with start times $\{0, 6\}$ and corresponding data pairs $\{T_2^*(0) = 7 \text{ months}, \mathbf{Z}_2(0)\}$ and $\{T_2^*(6) = 1 \text{ month}, \mathbf{Z}_2(6)\}$. Patient 3 was transplanted at $C_3 = 7$ months and is therefore also observed for two follow-up windows with start times $\{0, 6\}$. In Section 2.3, we describe a multiple imputation procedure for missing failure times that will be used in our analysis. Hence, if a failure time of 10 months is imputed for patient 3, this patient would contribute data pairs $\{T_3^*(0) = 10 \text{ months}, \mathbf{Z}_3(0)\}$ and $\{T_3^*(6) = 4 \text{ months}, \mathbf{Z}_3(6)\}$ to the analysis for one of the M multiply imputed datasets. The corresponding longitudinal data structure that we would use to fit Model (1) is shown at the bottom of Figure 1, where the columns ID, T^* and $\mathbf{Z}$ are the subject ID, outcome and covariate vector respectively. In our setting, the waitlist time/follow-up window start time (column t in the data structure) cannot be included in Model (1). However, the modeling procedure has no such limitation and time and interactions between time and other covariates can be included in Model (1).
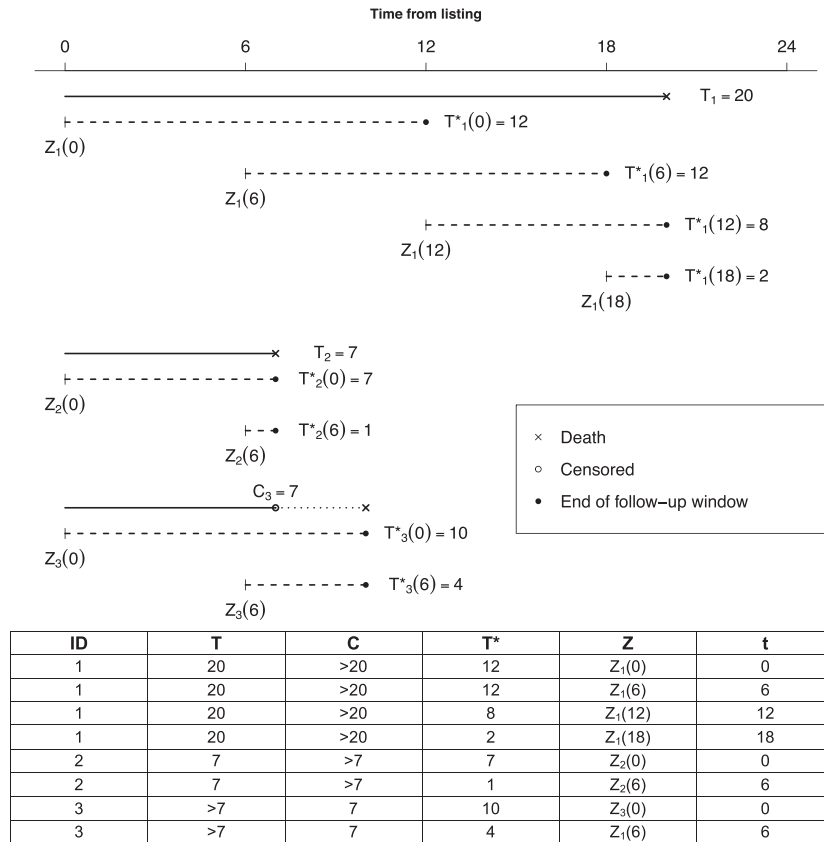
Figure 1. Example of how to construct the random variables $T_i^*(t) = \min\{T_i(t), 12\}$ in each of the follow-up windows for which the patients are under observation in three cases. Patient 1 dies at 20 months post-listing; patient 2 dies at 7 months post-listing; and patient 3 is transplanted at 7 months post-listing. For patient 3, we impute M death times, one of which is equal to 10 months and use the imputed death times in the longitudinal data analysis with M complete datasets. The associated data structure for the example patients is shown at the bottom of the figure.

2.3. Outline of imputation algorithm

Each censored patients' current follow-up window has measured covariates $\mathbf{Z}(t_{n_i})$ and a censored event time $C(t_{n_i}) = \min(C_i - t_{n_i}, \tau)$ that requires multiple imputation for generation of convenient complete datasets for analysis. There are many different approaches for multiple imputation in the presence of dependent censoring. We extend the inverse transform imputation algorithm used by Xiang, Murray and Liu [3] to the setting with multiple follow-up windows.

2.3.1. Inverse transform imputation.
The simplest case of the inverse transform imputation method is based on a Kaplan–Meier estimate, $\hat{S}_T(t)$, that is consistent for the survival function, $S_T(t)$. In this case, Taylor, Murray and Hsu [6] showed that multiple imputation reproduces the the Kaplan–Meier on average. We review this simplest case to avoid additional notation in delivering the concept.

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

2.3.2. Proposed imputation.
Our proposed multiple imputation method follows the same inverse transform imputation procedure but with a different risk set definition for R_i and a consistent survival estimate

that accounts for dependent censoring within the risk set, both of which will be given in later in Section 2.6. An outline of the steps needed to generate our imputed datasets is as follows:

(1) Obtain a consistent estimate of the survival function in the presence of dependent censoring due to transplantation.

    (a) Fit a Cox model for censored outcomes.
    (b) Construct inverse weights based on the Cox model above.
    (c) Compute the IPCW survival estimate.

(2) Estimate model parameters $\boldsymbol{\beta}$ via the pseudo-observation approach to use in risk set formation.
(3) For each patient censored at C_i,

    (a) Form risk set R_i by including at-risk patients similar to the ith patient at censoring based on model (1) and the pseudo-observation estimate of $\boldsymbol{\beta}$.
    (b) Compute the IPCW survival estimate within the risk set.
    (c) Sample a valid impute.

2.4. Generalized estimating equations for complete data

Once we have constructed the longitudinal data structure as described in Section 2.2 via the imputation procedure in Section 2.3, the correlation between observations from different follow-up windows within each patient must be accounted for when fitting model (1). GEE provides a framework that easily allows the correlation between overlapping and non-overlapping follow-up windows to differ via the unstructured working correlation matrix. In addition, the robust sandwich variance provides protection against misspecification of working correlation matrix. Currently, available statistical software for correlated censored survival outcomes assumes that correlated event times follow an exchangeable correlation structure between any two pairs of outcomes, which does not accommodate our data well. For example, the overlapping follow-up windows should have a different correlation structure than the windows that do not overlap. The easy implementation of unstructured correlation matrices makes `GEE` software ideal.

2.5. Analysis of the M multiply imputed datasets

We repeat the imputation procedure until we obtain M completed datasets. In practice $M=10$ is usually sufficient to produce valid results. The analysis of the M multiply imputed datasets is given by Li, Raghunathan and Rubin [7]. For each complete dataset, we fit model (1) using GEE methods described in Section 2.4. The parameter estimates for dataset m are denoted by $\hat{\boldsymbol{\beta}}_m^{MI}$, and the estimate of their variance–covariance matrix is denoted by $\hat{\boldsymbol{\Sigma}}_m^{MI}$, $m = 1, \dots, M$.

The estimates of $\boldsymbol{\beta}$ based on the multiple imputation procedure are $\hat{\boldsymbol{\beta}}^{MI} = \sum_{m=1}^{M} \hat{\boldsymbol{\beta}}_m^{MI}/M$. The associated variance–covariance matrix estimate is $\hat{\boldsymbol{\Sigma}}^{MI} = (1 + r)\boldsymbol{W}$ where $r = (1 + M^{-1})\mathrm{tr}(\boldsymbol{B}\boldsymbol{W}^{-1})/p$ and p is the length of $\boldsymbol{\beta}$. The average within imputation variance–covariance is $\boldsymbol{W} = \sum_{m=1}^{M} \hat{\boldsymbol{\Sigma}}_m^{MI}/M$ and the between imputation variance–covariance is $\boldsymbol{B} = \sum_{m=1}^{M}(\hat{\boldsymbol{\beta}}_m^{MI} - \hat{\boldsymbol{\beta}}^{MI})(\hat{\boldsymbol{\beta}}_m^{MI} - \hat{\boldsymbol{\beta}}^{MI})^T/(M - 1)$.

The hypothesis tests and 95% confidence intervals for $\hat{\boldsymbol{\beta}}^{MI}$ are constructed based on the statistic $(\hat{\boldsymbol{\beta}}^{MI} - \boldsymbol{\beta})^T(\hat{\boldsymbol{\Sigma}}^{MI})^{-1}(\hat{\boldsymbol{\beta}}^{MI} - \boldsymbol{\beta})/p$ that asymptotically follows an F-distribution with p and $\nu = (M - 1)(p + 1)(1 + r^{-1})^2/2$ degrees of freedom.

2.6. Details of multiple imputation algorithm

In Section 2.3.2, we gave an outline of the proposed multiple imputation procedure. We now provide details of each step of the algorithm.

2.6.1. Step 1: IPCW survival estimation.
The method to construct inverse weighted survival estimates, $\hat{S}_T^W(t)$, that are consistent for the survival function, $S_T(t)$, in the presence of dependent censoring proceeds as follows. First, we fit a Cox model for censored outcomes via the model,

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

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

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

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

We can then estimate the IPCW cumulative hazard by

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

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

2.6.2. Step 2: restricted mean model via pseudo-observations adjusted for dependent censoring.
Andersen, Hansen and Klein [8] developed a pseudo observation approach to modeling (1), with a single follow up window at time $t = 0$. Xiang and Murray [9] later extended the methodology by incorporating the inverse weighted survival estimates, described in Step 1, to account for dependent censoring. Although pseudo observation methods sometimes struggle with intercept bias in small samples, the resulting restricted mean estimates are useful in defining risk sets for imputation because these sets are invariant to intercept estimation. These models also have the advantage of being extremely easy to implement. With the ultimate goal of defining risk sets that will be used in our imputation procedure, we now summarize how to estimate pseudo observations from each follow-up window.

For each patient $i = 1, \ldots, n$ at the start of each follow-up window, t_j where $j = 1, \ldots, n_i$, the follow-up window specific pseudo-observation is defined to be

$$\mathscr{PO}_{ij} = \tilde{n}_j \hat{\delta}_j - (\tilde{n}_j - 1)\hat{\delta}_j^{-i} \tag{2}$$

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

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

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

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

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

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

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

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

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

inserting these into equation (2) to obtain $\mathscr{PO}_{ij}$, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. GEE methodology applied to data pairs $\{\mathscr{PO}_{ij}, \mathbf{Z}_i(t_j)\}$, $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$ give us model (1) parameter estimates, $\hat{\boldsymbol{\beta}}^{PO^W}$.

2.6.3. Step 3(a): Risk set formation. As in the simplest case of risk set formation, the minimal constraint for belonging to R_i is $T_k > C_i$, $k = 1 \ldots, n$. The covariates $\mathbf{Z}_i(t)$ are related to survival, and we can improve our imputation by selecting patients with similar urgency at the censoring time C_i based on our linear model (1). The second constraint for belonging to risk set R_i is then $|\hat{\boldsymbol{\beta}}^{PO^W T} \mathbf{Z}_k(C_i) - \hat{\boldsymbol{\beta}}^{PO^W T} \mathbf{Z}_i(C_i)| < \epsilon$ where ϵ is the parameter that controls how closely the linear predictors should match at C_i. The choice of epsilon is based on defining a risk set large enough to produce valid multiple imputes but as homogeneous as possible with respect to urgency. In simulations, epsilon was chosen to ensure at least five patients in the risk set. In the lung transplant setting, we require that patient k is in the same diagnosis group as patient i and that $LAS_k(C_i) = LAS_i(C_i)$ so that patients have similar urgency and transplant probability. The parameter epsilon was chosen to ensure at least 20 patients in the risk set.

2.6.4. Step 3(b): Inverse weighted survival estimation within risk set. Within the risk set, the inverse probability of censoring weight for the kth patient is defined as

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

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

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

2.6.5. Step 3(c): Sampling a valid impute. We require an impute for an individual censored at $C_i(t_{n_i}) < \tau$, where imputes are sampled from the survival distribution of $T_i^*(t_{n_i})$ within the risk set, R_i. The impute, t, is sampled by (a) generating a uniform(0,1) random variable, u, and (b) finding the smallest value t where $\hat{S}^W_{T_i^*(t_{n_i})}(t|R_i) \leqslant u$. Then by (c) identifying the observed event time, $T_k^*(t_{n_i})$, that corresponds to t, we can solve for the associated residual ε using model (1) where $\log\{T_k^*(t_{n_i})\} = \hat{\boldsymbol{\beta}}^{PO^W T} \mathbf{Z}_k(t_{n_i}) + \varepsilon$. Lastly, (d) if $t = \tau$ then impute $\tilde{T}_i^*(t_{n_i}) = \tau$, otherwise impute $\tilde{T}_i^*(t_{n_i}) = \exp\{\hat{\boldsymbol{\beta}}^{PO^W T} \mathbf{Z}_i(t_{n_i}) + \varepsilon\}$. If $\tilde{T}_i^*(t_{n_i}) < C_i(t_{n_i})$ then sample another impute using steps (a)-(d) until the imputed value is greater than $C_i(t_{n_i})$.

3. Simulations

In order to better understand finite sample behavior of our methods in relation to other available approaches, we summarize simulation results from 500 Monte Carlo iterations with $n = 300$ patients. Each iteration gives GEE parameter estimates for model (1), via (a) the IPCW pseudo-observation method applied to the first follow-up window as in Xiang and Murray [9], (b) Xiang, Murray and Liu [3]'s multiple imputation method applied to the first follow-up window and (c) our proposed multiple imputation method that incorporates information from multiple follow up windows. As a benchmark, we also present (d) results in the absence of censoring when multiple follow-up windows are used in estimation. Each method assumes $\tau = 1$ year. Methods using follow-up beyond year 1 in estimation incorporate information from 1-year windows starting at $t_1 = 0$, $t_2 = 6$ months and $t_3 = 12$ months.

In the first simulation scenario, the data are generated as follows.

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

Step 2: Each failure time T_i is simulated from a piecewise exponential distribution with hazard equal to λ_{i1} in the interval $[0, 6]$, λ_{i2} in the interval $(6, 12]$ and λ_{i3} in the interval $(12, \infty)$. The patient-specific hazards are chosen to ensure that the model coefficients are the same, regardless of the number of follow-up windows used. This allows us to compare existing methods that use only the first follow-up window to our proposed method that uses all the follow-up windows. In Appendix A, we give further details on the algebra involved.

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

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

On the basis of the results in Table I, we observe that the parameter estimates from all methods have minimal bias under both the null hypothesis and for non-zero β's. Inclusion of follow-up windows starting at $t_2 = 6$ and $t_3 = 12$ months results in more efficient estimates. The asymptotic relative efficiency (ARE) of our proposed method versus the IPCW pseudo-observation method is between 1.79 and 2.52 for each of the parameters. The ARE of our proposed method versus the multiple imputation method of Xiang and Murray is between 1.72 and 2.49. The ARE comparing our proposed method with versus without censoring is between 0.96 and 0.99, indicating that our method effectively handles dependent censoring and produces parameter estimates with nearly the same efficiency as if censoring never occurred. We have also conducted simulations with larger sample size ($n = 400$), and the results are similar (see Supporting Information).

In the second set of simulation scenarios, we highlight the operating characteristics of the various approaches in the case where there is an important predictor (for example, waiting time) that is not

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

Parameter	IPCW-PO	MI	Uncensored	MI*
$\beta_0 = 2.1$	2.110	2.110	2.078	2.058
	[0.010, 0.125, 0.120, 0.968]	[0.010, 0.124, 0.120, 0.964]	[−0.022, 0.085, 0.086, 0.944]	[−0.042, 0.087, 0.085, 0.940]
$\beta_1 = 0$	−0.010	−0.010	−0.001	0.002
	[−0.010, 0.163, 0.161, 0.946]	[−0.010, 0.163, 0.160, 0.948]	[−0.001, 0.100, 0.100, 0.950]	[0.002, 0.104, 0.101, 0.954]
$\beta_2 = 0$	−0.008	−0.008	−0.003	−0.003
	[−0.008, 0.204, 0.211, 0.958]	[−0.008, 0.204, 0.210, 0.958]	[−0.003, 0.150, 0.153, 0.936]	[−0.003, 0.152, 0.155, 0.936]
$\beta_0 = 2.1$	2.096	2.096	2.067	2.051
	[−0.004, 0.127, 0.126, 0.960]	[−0.004, 0.127, 0.126, 0.962]	[−0.033, 0.088, 0.082, 0.954]	[−0.049, 0.090, 0.085, 0.932]
$\beta_1 = -0.125$	−0.121	−0.120	−0.123	−0.118
	[0.004, 0.167, 0.175, 0.934]	[0.005, 0.167, 0.174, 0.930]	[0.002, 0.103, 0.110, 0.942]	[0.007, 0.107, 0.110, 0.952]
$\beta_2 = 0.1$	0.111	0.111	0.120	0.111
	[0.011, 0.208, 0.207, 0.960]	[0.011, 0.208, 0.207, 0.964]	[0.020, 0.153, 0.152, 0.960]	[0.011, 0.155, 0.155, 0.950]

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

Table II. Comparison of parameter estimates when there is an interaction between the follow-up window start time (t_j) and the time-dependent covariate ($Z_{1i}(t_j)$) with coefficient set to -0.05 but time is excluded from the fitted models.

Parameter	IPCW-PO	MI	Uncensored	MI*
	Excluding time in the fitted model			
$\beta_0 = 2.1$	2.198	2.199	2.188	2.165
	[0.098, 0.118,	[0.099, 0.118,	[0.088, 0.079,	[0.065, 0.081,
	0.115, 0.848]	0.115, 0.842]	0.075, 0.762]	0.078, 0.860]
$\beta_1 = -0.125$	-0.127	-0.128	-0.371	-0.342
	[-0.002, 0.154,	[-0.003, 0.154,	[-0.246, 0.097,	[-0.217, 0.101,
	0.164, 0.938]	0.164, 0.938]	0.100, 0.290]	0.101, 0.430]
$\beta_2 = 0.1$	0.091	0.089	0.080	0.074
	[-0.009, 0.192,	[-0.011, 0.192,	[-0.020, 0.141,	[-0.026, 0.144,
	0.189, 0.960]	0.188, 0.962]	0.146, 0.936]	0.147, 0.938]

We compare the IPCW pseudo-observation method (IPCW-PO) and the multiple imputation method (MI), which both use one follow-up window; and using uncensored observations (Uncensored) and our multiple imputation method (MI*), which both use three follow-up windows, with 500 Monte Carlo simulations. M=10 in both multiple imputation methods.
Empirical Mean
[Bias, Empirical Mean Standard Error,
Empirical Standard Deviation, Coverage of 95% Confidence Interval]

included in model (1). For each of the simulations summarized in Tables II and III, the outcomes are generated assuming an interaction between follow-up window time t_j and the time-dependent predictor $Z_{1i}(t_j)$ with coefficient $\beta_4 = -0.05$. The coefficient of the main effect of t_j is assumed to be 0 (i.e $\beta_3 = 0$). All remaining coefficients are the same as those used to generate Table I results. The coefficients corresponding to $Z_{1i}(0)$, $Z_{1i}(6)$ and $Z_{1i}(12)$ are -0.125, -0.425 and -0.725, respectively, after incorporating the influence from the interaction term in the urgency model. Excluding the interaction term from model (1) in this scenario results in a mixture coefficient of -0.37 for $Z_{1i}(t_j)$ that describes the association between $Z_{1i}(t_j)$ and the restricted mean across the mixture of $Z_{1i}(t_j)$ values seen in the cohort at times 0, 6 and 12 months. Further details regarding data generation for this setting are relegated to Appendix A.

Table II gives results when model (1) is fit without the interaction term between follow-up window time t_j and the time-dependent predictor $Z_{1i}(t_j)$. Results show very little bias in estimating the mixture coefficient using our method. Alternatively, the IPCW pseudo-observation (IPCW-PO) and MI methods that use only information during the first follow-up window give estimates for β_1 that do not reflect later changes in risk associated with $Z_{1i}(t_j)$. Table III shows good model performance when the interaction term between t_j and $Z_{1i}(t_j)$ that generated the data is included in model (1).

4. Example

The lung waitlist consists of 10,740 transplant candidates aged 12 years and older who were newly listed between September 1, 2006 and March 2, 2012; 7,359 of these patients received a transplant, 884 died while on the waitlist, 1,125 dropped off the waitlist without a transplant, and 1,372 were alive on the waitlist on March 2, 2012. In addition to mortality data maintained by UNOS, the Social Security Death Master File was used to supplement vital status information in our cohort. From these records, 503 of 1,125 patients who chose to leave the waitlist died during the approximate 5.5-year follow-up period and 622 remained at risk at the end of this same period. Risk factors used to model LAS urgency are given by the OPTN Thoracic Committee (OPTN Policy 3.7) and have been vetted as worthy of inclusion in the algorithm. Most have proven historical statistical significance in at least one previous analysis of lung candidate data.

The data reported here have been supplied by the Minneapolis Medical Research Foundation as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government. This study used data from the SRTR. The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the USA, submitted by the members of the Organ Procurement and Transplantation Network (OPTN) and has been described

Table III. Comparison of parameter estimates when there is an interaction between the follow-up window start time (t_j) and the time-dependent covariate ($Z_{1i}(t_j)$) with coefficient set to -0.05.

Parameter	Uncensored	MI*
	Including time in the fitted model	
$\beta_0 = 2.1$	2.236	2.234
	[0.136, 0.097,	[0.134, 0.099,
	0.095, 0.672]	0.096, 0.688]
$\beta_1 = -0.125$	-0.121	-0.122
	[0.004, 0.138,	[0.003, 0.140,
	0.145, 0.936]	0.146, 0.934]
$\beta_2 = 0.1$	0.087	0.082
	[-0.013, 0.148,	[-0.018, 0.152,
	0.152, 0.944]	0.154, 0.946]
$\beta_3 = 0$	-0.012	-0.017
	[-0.012, 0.012,	[-0.017, 0.013,
	0.013, 0.820]	0.013, 0.720]
$\beta_4 = -0.05$	-0.043	-0.038
	[0.007, 0.022,	[0.012, 0.023,
	0.022, 0.926]	0.023, 0.914]

We compare using uncensored observations (uncensored) and our multiple imputation method (MI*), which both use three follow-up windows, with 500 Monte Carlo simulations. M=10 in the multiple imputation method.
Empirical Mean
[Bias, Empirical Mean Standard Error, Empirical Standard Deviation, Coverage of 95% Confidence Interval]

elsewhere. The Health Resources and Services Administration, U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors.

Patients are divided into four overarching diagnosis groups, A through D, by the OPTN Thoracic Committee that are considered to be similar with respect to waitlist and post-transplant survival. The details of the diagnoses that comprise each group are given in OPTN Policy 3.7. In our dataset at listing, of the 3618 patients in Group A, 2924 (81%) were diagnosed with chronic obstructive pulmonary disease; in Group B, 262 (56%) out of 468 were diagnosed with primary pulmonary hypertension; in Group C, 1284 (99%) out of 1296 were cystic fibrosis patients; and in Group D, 3633 (68%) out of 5358 patients were diagnosed with idiopathic pulmonary fibrosis. A few group A and D diagnoses are allowed to enter the urgency model as their own risk factors. For group A, these are bronchiectasis, lymphangioleiomyomatosis and sarcoidosis with PA mean$\leqslant$30mm Hg. For group D, these are obliterative bronchiolitis, pulmonary fibrosis other and sarcoidosis with PA mean$>$30 mm Hg. Eisenmenger syndrome, from group B, is also listed as a risk factor in LAS. Most of these smaller diagnosis groups are not statistically different from their larger conglomerate group designation but having a separate parameter has been important in obtaining public approval of the algorithm.

In Table IV, we summarize the risk factors at listing within each of the four diagnoses groups. These factors are age, body mass index (BMI), cardiac index prior to any exercise, central venous pressure (CVP) at rest, whether they were on continuous mechanical ventilation, serum creatinine, whether they were diabetic, percent predicted forced vital capacity (FVC), whether they required assistance with the activities of daily living (ADL), O_2 requirement at rest needed to maintain adequate oxygen saturation, partial pressure of carbon dioxide (PCO_2), pulmonary artery (PA) systolic pressure at rest, and 6-minute walk distance obtained while receiving supplemental oxygen to maintain oxygen saturation of 88% or greater at rest. Those familiar with the LAS may recall that bilirubin has recently been approved as an urgency risk factor. However, this measure has only recently started being collected by the OPTN and was unavailable in the March 2012 release data used for our analyses. In our cohort, those who become inactive or otherwise remove themselves from the waitlist have often deteriorated to the point of having

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

LAS Covariates	Group A n=3618	Group B n=468	Group C n=1296	Group D n=5358
Age (years)	58.1 (8.3)	45.7 (14.7)	28.4 (10.3)	57.1 (11.0)
BMI (kg/m^2)	24.6 (4.3)	25.1 (4.7)	19.3 (2.9)	27.0 (4.4)
Cardiac Index<2.0 (L/min/m^2)	150 (4.1%)	82 (17.5%)	16 (1.2%)	271 (5.1%)
CVP (mm Hg)	7.7 (4.2)	10.8 (6.5)	5.2 (3.8)	5.5 (4.3)
Continuous Mechanical Ventilation	26 (0.7%)	10 (2.1%)	82 (6.3%)	196 (3.7%)
Creatinine (serum mg/dL)	0.8 (0.2)	1.0 (0.6)	0.7 (0.3)	0.9 (0.3)
Diabetes	430 (11.9%)	64 (13.7%)	627 (48.4%)	1249 (23.3%)
FVC (% predicted)	54.0 (17.5)	66.4 (22.7)	40.0 (11.8)	47.6 (17.0)
No assistance with ADL	368 (10.2%)	32 (6.8%)	249 (19.2%)	506 (9.4%)
O$_2$ requirement at rest (L/min)	3.0 (2.5)	4.0 (4.4)	2.9 (3.5)	4.9 (5.2)
PCO$_2$ (mm Hg)	49.9 (10.8)	42.4 (6.4)	53.4 (17.4)	44.7 (8.2)
PA systolic (mm Hg)	38.3 (10.8)	76.7 (25.7)	39.0 (10.6)	43.2 (16.2)
6-min walk distance (feet)	783.4 (347.9)	776.7 (438.4)	970.6 (465.8)	797.2 (464.2)

[†]For continuous variables, numbers shown are mean (standard deviation)
For binary variables, numbers shown are number (proportions)
Body mass index (BMI); cardiac index (CI); central venous pressure (CVP)
Activities of daily living (ADL); pulmonary artery (PA)

poor surgical risk. Because urgency estimates in our application are meant to capture the natural course of mortality without a transplant, all time at risk after listing is included in urgency estimation regardless of active or inactive status for accepting organ offers.

Patients in group C are on average the youngest transplant candidates (mean age is 29.4 years), and this is expected because the group consists almost entirely of patients with cystic fibrosis, a genetic disorder that results in lung disease from a very young age. The main diagnoses in the other groups are lung diseases that develop over time, so most patients are older when they require a lung transplant. Having a cardiac index less than 2 L/min/m^2 is considered to be an indicator that the heart is not functioning well. On the basis of this measure, patients in group B tend to be the most severely ill with 17.5% of the patients having a cardiac index < 2 L/min/m^2. The proportion of patients requiring continuous mechanical ventilation is highest in group C (6.3%) and lowest in group A (0.7%). FVC % predicted is a measure of lung function, and in terms of this measure, the diagnoses groups are ranked C, D, A, and B from most severely ill to least severely ill. However, the patients in group C are also most likely to need no assistance with activities of daily living (19.2%) and are able to walk much long distances in the 6-minute walk test compared with the other groups. Group B consists of patients with various hypertensive disorders and as we would expect, patients in this group have the highest average PA systolic pressure at rest (76.7 mm Hg). Patients from different diagnosis groups are transplanted at different rates according to their LAS scores, which incorporates diagnosis as a predictor. For example, in our cohort described in Table 2, the number (%) of transplants was 2359 (65.2%), 238 (50.9%), 873 (67.4%), and 3889 (72.6%) for groups A–D, respectively.

As urgent patients receive transplants, they are removed from the waitlist. This systematic removal of patients creates a problem of dependent censoring that we adjust for using inverse probability of censoring weights in the survival estimation procedure discussed in Section 2.6.1. For each patient on the waitlist, the probability of being censored is estimated from a time-dependent Cox model, the results of which are presented in Table V. The covariates that influence censoring are gender, race (white, black, and other), height, blood type (A, B, O, and AB) and time-dependent LAS and listing status (active and inactive).

Gender, race, and height are all seen to be highly significant characteristics for differentiating which patients will be censored. Blood types A, B, and AB were observed to be similar in terms of censoring hazard but patients with blood type O had a lower hazard of censoring compared with blood type A (hazard ratio=0.9, 95% confidence interval: 0.92–0.99). Patients with an LAS of 0 have the lowest possible score with a very low chance of transplant (censoring) that obtains ameliorated a bit by geography when higher risk patients are not in competition for an organ. A one-unit increase in LAS when the 0<LAS≤30 results in a decreasing hazard that reflects the low probability of being offered a transplant for low LAS scores. The effect of a one unit increase in LAS decreases for higher ranges of LAS scores. This may

Table V. Proportional hazards censoring model for 10,740 lung transplant candidates

	Hazard ratio	95% Confidence interval	p-value
Time-independent characteristics			
Female (vs Male)	0.76	(0.72,0.81)	< 0.0001
Black (vs White)	0.78	(0.72,0.84)	< 0.0001
Other (vs White)	0.79	(0.73,0.85)	< 0.0001
Height: < 5′3″ (versus > 5′9″)	0.60	(0.56,0.65)	< 0.0001
Height: 5′3″-5′6″ (versus > 5′9″)	0.77	(0.72,0.83)	< 0.0001
Height: 5′6″-5′9″ (versus > 5′9″)	0.86	(0.82,0.91)	< 0.0001
Blood type: B (versus A)	1.01	(0.94,1.08)	0.8118
Blood type: O(versus A)	0.95	(0.91,0.99)	0.0194
Blood type: AB (versus A)	1.08	(0.96,1.20)	0.2048
Time-dependent characteristics			
LAS> 0 (versus LAS= 0)	1190	(7.03, > 3000)	0.0068
Unit increase in LAS: 0<LAS⩽30	0.77	(0.65,0.92)	0.0031
Unit increase in LAS: 30<LAS⩽35	1.14	(1.11,1.17)	< 0.0001
Unit increase in LAS: 35<LAS⩽40	1.10	(1.08,1.11)	< 0.0001
Unit increase in LAS: 40<LAS⩽60	1.04	(1.03,1.04)	< 0.0001
Unit increase in LAS: LAS>60	1.02	(1.02,1.02)	< 0.0001
Active vs inactive status	4.73	(4.34,5.15)	< 0.0001
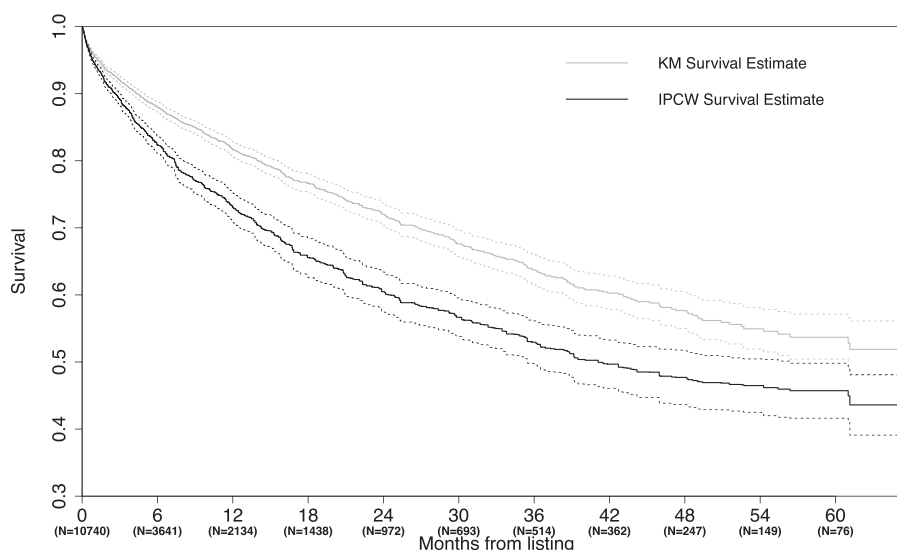
Figure 2. Waitlist probability of survival estimated by the Kaplan-Meier method and the IPCW-survival method. The 95% point-wise confidence intervals for each estimated survival curve are depicted by dashed lines. N = number of patients at risk at the start of each follow-up window.

seem counterintuitive; however, the probability of being censored is counterbalanced by the probability of surviving until a transplant becomes available.

Figure 2 compares the IPCW survival estimate of waitlist survival to the Kaplan-Meier estimate of waitlist survival from listing up to five years post listing. The Kaplan-Meier estimate does not adjust for dependent censoring resulting from transplantation and therefore estimates higher waitlist survival probabilities compared to the IPCW survival. The Wald type 95% point-wise confidence interval for the IPCW survival estimate is calculated using 100 bootstrap replicates to estimate the variance of the IPCW survival estimate.

We define 1-year follow-up windows starting at 0, 6, 12, 18, 24, 30, and 36 months, where the start time of the final follow-up window was chosen to ensure we had at least 25 risk set deaths in each follow-up window. UNOS required patients on the waitlist to update their covariates every 6 months, and therefore our choice of follow-up windows that start every 6 months is appropriate in this setting. The IPCW pseudo-observations are constructed and used to fit model (1), which describes lung candidate urgency based on LAS risk factors, and the parameter estimates are then used to define risk sets.

Statist. Med. **2017,** 36 2435–2451

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

	IPCW-PO			MI			MI*		
	e^β	95% CI	p-value	e^β	95% CI	p-value	e^β	95% CI	p-value
Intercept	663.84	(508.83,866.06)	< 0.0001	593.58	(477.76,737.48)	< 0.0001	573.88	(487.81,675.13)	< 0.0001
Age (per 20 year increase)	0.91	(0.86,0.96)	0.0011	0.93	(0.89,0.97)	0.0025	0.92	(0.89,0.95)	< 0.0001
BMI (per 1 kg/m² decrease if BMI<20kg/m²)	0.98	(0.95,1.01)	0.1148	0.98	(0.95,1.00)	0.0420	0.98	(0.96,1.00)	0.0176
Cardiac Index<2.0 (ref=CI≥ 2)	0.83	(0.73,0.94)	0.0036	0.85	(0.77,0.94)	0.0031	0.88	(0.81,0.95)	0.0035
CVP in Group B (per 10 mm Hg)	0.79	(0.60,1.02)	0.0709	0.78	(0.60,1.01)	0.0620	0.75	(0.63,0.90)	0.0028
Continuous Mechanical Ventilation (ref=other/no ventilation)	0.17	(0.12,0.24)	< 0.0001	0.14	(0.11,0.19)	< 0.0001	0.23	(0.18,0.30)	< 0.0001
Creatinine (Age> 18 years)	0.94	(0.86,1.02)	0.1264	0.94	(0.87,1.01)	0.0956	0.95	(0.90,1.00)	0.0488
Diabetes (ref=no diabetes)	0.93	(0.86,1.00)	0.0514	0.95	(0.90,0.99)	0.0319	0.94	(0.90,0.97)	0.0010
Diagnosis group (ref=group A)									
Group B	1.05	(0.79,1.39)	0.7567	1.07	(0.83,1.38)	0.5817	1.05	(0.87,1.27)	0.5993
Group C	0.82	(0.71,0.96)	0.0126	0.86	(0.77,0.96)	0.0117	0.88	(0.81,0.96)	0.0075
Group D	1.09	(0.94,1.26)	0.2611	1.05	(0.92,1.20)	0.4211	1.03	(0.93,1.13)	0.5620
Bronchiectasis	0.93	(0.82,1.05)	0.2371	0.95	(0.85,1.06)	0.3715	0.97	(0.90,1.04)	0.4219
Eisenmenger Syndrome	1.39	(1.08,1.78)	0.0105	1.20	(0.96,1.51)	0.1061	1.06	(0.81,1.38)	0.6725
Lymphangioleiomyomatosis	1.00	(0.83,1.20)	0.9618	1.02	(0.83,1.25)	0.8348	1.05	(0.95,1.17)	0.3176
Obliterative Bronchiolitis	1.25	(1.00,1.56)	0.0515	1.19	(0.97,1.46)	0.0904	1.08	(0.93,1.26)	0.2572
Pulmonary Fibrosis Other	0.91	(0.78,1.06)	0.2315	1.01	(0.91,1.11)	0.9057	1.02	(0.95,1.11)	0.5585
Sarcoidosis with PA mean>30 mm Hg	1.19	(0.95,1.49)	0.1273	1.29	(1.15,1.44)	0.0001	1.28	(1.18,1.39)	< 0.0001
Sarcoidosis with PA mean≤30 mm Hg	0.93	(0.83,1.05)	0.2511	0.92	(0.83,1.03)	0.1285	0.91	(0.83,0.99)	0.0358
FVC (per 10% decrease if FVC %predicted<80%) in Group D	0.94	(0.91,0.96)	< 0.0001	0.95	(0.93,0.97)	0.0002	0.96	(0.95,0.98)	0.0004
No assistance with ADL (ref=some/total assisstance)	1.02	(0.97,1.07)	0.3791	0.99	(0.95,1.03)	0.5457	0.99	(0.96,1.01)	0.2787
O₂ requirement at rest in Group B	0.94	(0.90,0.98)	0.0022	0.93	(0.89,0.97)	0.0014	0.95	(0.91,0.98)	0.0014
O₂ requirement at rest in Groups A, C, or D	0.92	(0.91,0.93)	< 0.0001	0.92	(0.91,0.93)	< 0.0001	0.93	(0.92,0.94)	< 0.0001
PCO₂ (per 10 mm Hg)	0.96	(0.92,0.99)	0.0252	0.97	(0.94,1.00)	0.0917	0.98	(0.95,1.00)	0.0396
PCO₂ increase of ⩾ 15% (PCO₂ increase of < 15%)	0.97	(0.65,1.44)	0.8857	0.92	(0.65,1.31)	0.6264	1.02	(0.93,1.12)	0.6648
PA systolic in Group A (per 10 mm Hg increase if PA systolic> 40 mm Hg in Group D	0.99	(0.94,1.04)	0.6038	0.98	(0.93,1.04)	0.5132	0.98	(0.95,1.02)	0.3888
PA systolic in Group B, C, or D (per 10 mm Hg increase)	0.97	(0.94,0.99)	0.0129	0.97	(0.95,0.99)	0.0035	0.97	(0.95,0.99)	0.0007
6-min walk distance (per 1000 feet)	1.22	(1.13,1.32)	< 0.0001	1.16	(1.09,1.23)	0.0002	1.14	(1.10,1.20)	< 0.0001

Effect size, e^β, per unit increase in the continuous risk factor unless otherwise stated; Effect size for reference categories of categorical risk factors is 1.
Body mass index (BMI); cardiac index (CI); central venous pressure (CVP); activities of daily living (ADL); pulmonary artery (PA)

In Table VI, we present the results of fitting the restricted mean model to estimate lung candidate urgency using three different methods, (a) the IPCW pseudo-observation method applied to the first follow-up windows [9], (b) the multiple imputation method applied to the first follow-up window [3] and (c) our proposed multiple imputation method that incorporates information from multiple follow up windows. The exponentiated parameters ($e^{\hat{\beta}}$), 95% confidence intervals, and p-values are presented for each risk factor and are therefore interpreted in terms of their multiplicative effect on the number of days lived in the next year. Note that results presented in [9] and [3] are based on an older release of waitlist candidate data.

The intercept estimated from the IPCW-PO method is larger than that of the multiple imputation method applied to the first follow-up window. In addition, most of the effect size confidence intervals from the IPCW-PO method are wider than those obtained using the multiple imputation method applied to the first follow-up window. In general, the conclusions reached regarding statistically significant risk factors are the same even if the effect sizes sometimes differ between the two methods. The exceptions include BMI, Diabetes, Eisenmenger Syndrome, Sarcoidosis with PA mean>30 mm Hg and PCO$_2$ (per 10 mm Hg).

Our proposed multiple imputation procedure estimates the effect of the risk factors across the seven follow-up windows viewed during the 4 years since listing. Hence, our parameter estimates should be comparable with the other methods if time since listing does not play a strong role. Again, the intercept was the most different between our proposed method, and the others, approximately 20 days smaller than the multiple imputation method based on only the first follow-up window. Our proposed method gave shorter confidence intervals compared with the other methods indicating increased efficiency resulting from the incorporation of additional follow-up windows. Comparing the two multiple imputation procedures, we observe that time since listing plays a role for the risk factor indicating whether they were on continuous mechanical ventilation. The reduction in expected number of days lived for patients on continuous mechanical ventilation decreases over time. This observation is confirmed when we include interactions between risk factors and time into the model. The additional information contained in later follow-up windows also allows us to confirm the statistical significance of some of the risk factors including CVP in group B, Creatinine, Sarcoidosis with PA mean ⩽30 mm Hg and PCO$_2$.

5. Discussion

LAS implementation has successfully reduced waitlist deaths, which reflects vitally important improvements for end stage lung disease care. This same reduction in deaths results in less power to estimate waitlist survival in current cohorts. It is therefore critical to develop statistical methodology that is able extract as much information as possible from available data. As we saw in Section 4, incorporating additional follow-up windows to estimate transplant urgency resulted in greater efficiency, as evidenced by narrower confidence intervals for parameter estimates. We were also able to confirm that low central venous pressure remained a statistically significant risk factor for patients in diagnosis group B for the current cohort. Measuring this risk factor is an invasive procedure, so it is reassuring that its collection is useful for ranking patients.

UNOS requires lung waitlisted patients to update their risk factors every 6 months and therefore 1-year follow-up windows timed to begin at 6-month intervals are a natural choice in this setting. In the special case where there are no covariates used in τ-restricted mean estimation via our method, we have previously shown that windows starting every $\tau/2$ units have attractive operating characteristics [10]. In settings where covariate data is updated more frequently, the only operational cost to including more frequently spaced follow-up windows is computational time.

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

A key challenge addressed by our method is how to account for dependent censoring while minimizing model assumptions. We followed the multiple imputation approach of Xiang, Murray and Liu in imputing the dependently censored values. One advantage of this approach is that each censored outcome is

imputed from a risk set of patients, with similar urgency to the censored patient, who also remain at risk beyond the patient's censoring time. Risk set selection that additionally restricts patients to have similar censoring risk has also been proposed in the literature with an argument that the additional restriction confers double robustness to model misspecification of either the mortality or censoring model [11]. Along these lines, we additionally required our risk set to have a similar LAS to the censored patient at the corresponding follow-up time, which was the most influential factor driving a loss-to-follow-up on the waitlist. Because the multiple imputation approach of Xiang, Murray and Lin additionally adjusts for dependent censoring within the risk set via an inverse weighting method, we feel comfortable that our approach has sufficiently addressed dependent censoring in our application. It should be noted, however, that the inverse weighting method used to produce an impute from each censored patient's comparable urgency risk set does not employ doubly-robust inverse weighting methods seen in the literature, for instance, [12–16]. So it is possible that further modest improvements to our method could be found using different inverse weighting approaches.

Notation used in developing our model overlaps to some degree with notation used in restricted mean residual life estimation [17] and in landmark analyses methods [2, 18–20]. Restricted mean residual life estimation focuses on the conditional estimation of survival among progressively more restrictive risk sets. Similarly, the landmark analyses approach initially proposed by Anderson *et al.* [18] also focuses on conditional analyses from various fixed landmark times and neither attempts to combine information across the multiple follow-up windows defined by the landmark times. More recent approaches [2, 20] do consider combining information across the follow-up windows within the context of the proportional hazards modeling framework. Therefore, our approach could be considered to fall within the broad spectrum of research that encompasses landmark analyses methods.

In the example data analysis, we have presented results for the analysis that uses only the first follow-up window and the analysis that uses all the available follow-up windows. We believe that including the later follow-up windows gives a more accurate picture of the distribution of risk factors among patients on the waitlist rather than just capturing the characteristics of the waitlist at listing.

6. Software

Software, in the form of R code, is available in the Supporting Information online or on request from the corresponding author (ntayob@mdanderson.org).

Appendix A: Patient specific hazards

Let T_i for $i = 1, \ldots, n$ be the failure times of the patients in the simulated dataset. We assume that T_i follows a piecewise exponential model with subject-specific hazards equal to λ_{i1} in the interval $[0, 6]$, λ_{i2} in the interval $(6, 12]$ and λ_{i3} in the interval $(12, \infty)$. In our simulation setting, we consider three follow-up windows of length $\tau = 12$ months, starting at times $t_1 = 0, t_2 = 6$, and $t_3 = 12$ months. The subject-specific hazards are chosen to ensure the following is true:

$$E[\log\{\min(T_i - t_3, \tau)\}|T_i > t_3, \bar{Z}_{1i}(T_i), Z_{1i}(t_3) = z_1^*, Z_{2i} = z_2^*]$$
$$= E[\log\{\min(T_i - t_2, \tau)\}|T_i > t_2, \bar{Z}_{1i}(T_i), Z_{1i}(t_2) = z_1^*, Z_{2i} = z_2^*]$$
$$= E[\log\{\min(T_i - t_1, \tau)\}|T_i > t_1, \bar{Z}_{1i}(T_i), Z_{1i}(t_1) = z_1^*, Z_{2i} = z_2^*]$$
$$= \beta_0 + \beta_1 z_1^* + \beta_2 z_2^*,$$

for constants z_1^* and z_2^*. This results in coefficients for model (1) that are the same regardless of the number of follow-up windows used and allows us to compare the results from methods that apply to only the first follow-up window to our proposed method, which applies to all three follow-up windows.

For the piecewise exponential distribution,

$$E[\log\{\min(T_i - t_j, \tau)\}|T_i > t_j, \bar{Z}_{1i}(T_i), Z_{2i}]$$
$$= -\int_0^\tau \log(u)dP\{T_i - t_j > u|T_i > t_j, \bar{Z}_{1i}(T_i), Z_{2i}\}$$
$$+ \log(\tau)P\{T_i - t_j > \tau|T_i > t_j, \bar{Z}_{1i}(T_i), Z_{2i}\},$$

where $P\{T_i - t_j > u | T_i > t_j, \bar{Z}_{1i}(T_i), Z_{2i}\} = \exp\{-\int_{t_j}^{t_j+u}[\lambda_{i1}I(t_1 \leqslant v \leqslant t_2) + \lambda_{i2}I(t_2 < v \leqslant t_3) + \lambda_{i3}I(t_3 < v < \infty)]dv\}$.

Therefore, we have three nonlinear equations with three unknown parameters that we need to solve for: λ_{i1}, λ_{i2} and λ_{i3}. In each case, the solution is obtained using numerical algorithms for nonlinear equations.

In the first simulation scenario (Table I), the data are generated as follows. Starting with the follow-up window $t_3 = 12$, we have

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

where λ_{i3} is the only unknown quantity.

For $t_2 = 6$, we have

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

where λ_{i2} is the only unknown quantity.

For $t_1 = 0$, we have

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

where λ_{i1} is the only unknown quantity.

In the second set of simulations (Tables II and III), the data are generated as follows. Starting with the follow-up window $t_3 = 12$, we have

$$\beta_0 + \beta_1 Z_{1i}(12) + \beta_2 Z_{2i} + \beta_3 * 12 + \beta_4 * 12 * Z_{1i}(12)$$
$$= \int_0^{12} \log(u)\lambda_{i3}\exp\{-u\lambda_{i3}\}du + \log(12)\exp\{-12\lambda_{i3}\},$$

where λ_{i3} is the only unknown quantity.

For $t_2 = 6$, we have

$$\beta_0 + \beta_1 Z_{1i}(6) + \beta_2 Z_{2i} + \beta_3 * 6 + \beta_4 * 6 * Z_{1i}(6)$$
$$= \int_0^6 \log(u)\lambda_{i2}\exp\{-u\lambda_{i2}\}du + \int_6^{12} \log(u)\lambda_{i3}\exp\{-6\lambda_{i2} - u\lambda_{i3} + 6\lambda_{i3}\}du$$
$$+ \log(12)\exp\{-6\lambda_{i2} - 6\lambda_{i3}\},$$

where λ_{i2} is the only unknown quantity.

For $t_1 = 0$, we have

$$\beta_0 + \beta_1 Z_{1i}(0) + \beta_2 Z_{2i} + \beta_3 * 0 + \beta_4 * 0 * Z_{1i}(0)$$
$$= \int_0^6 \log(u)\lambda_{i1}\exp\{-u\lambda_{i1}\}du + \int_6^{12} \log(u)\lambda_{i2}\exp\{-6\lambda_{i1} - u\lambda_{i2} + 6\lambda_{i2}\}du$$
$$+ \log(12)\exp\{-6\lambda_{i1} - 6\lambda_{i2}\},$$

where λ_{i1} is the only unknown quantity.

Conflict of Interest

None declared.

Acknowledgements

References

1. Egan TM, Murray S, Bustami RT, Shearon TH, McCullogh 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**(2):1212–1227.
2. Gong Qi, Schaubel DE. Partly conditional estimation of the effect of a time-dependent factor in the presence of dependent censoring. *Biometrics* 2013; **69**(2):338–347.
3. Xiang Fang, Murray Susan, Liu Xiaohong. Analysis of transplant urgency and benefit via multiple imputation. *Statistics in Medicine* 2014; **33**(26):4655–4670.
4. 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*, Alexandria, Virginia, 1993, 24–33.
5. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
6. Taylor JMG, Murray S, Hsu C-H. Survival estimation and testing via multiple imputation. *Statistics & Probability Letters* 2002; **58**(3):221 –232.
7. Li KH, Raghunathan TE, Rubin DB. Large-sample significance levels from multiply imputed data using moment-based statistics and an f reference distribution. *Journal of the American Statistical Association* 1991; **86**(416):1065–1073.
8. Andersen P, Hansen M, Klein J. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Analysis* 2004; **10**(4):335–350.
9. Xiang F, Murray S. Restricted mean models for transplant benefit and urgency. *Statistics in Medicine* 2012; **31**(6):561–576.
10. Tayob N, Murray S. Nonparametric restricted mean analysis across multiple follow-up intervals. *Statistics & Probability Letters* 2016; **109**:152 –158.
11. Hsu C-H, Taylor JMG, Murray S, Commenges D. Survival analysis using auxiliary variables via non-parametric multiple imputation. *Statistics in Medicine* 2006; **25**(20):3503–3517.
12. Robins JM. Robust estimation in sequentially ignorable missing data and causal inference models. *Proceedings of the American Statistical Association Section on Bayesian Statistical Science 1999*, Baltimore, M.D., U.S.A., 2000, 6–10.
13. Robins JM, Rotnitzky A, Laan MVD. Comment. *Journal of the American Statistical Association* 2000; **95**(450):477–482.
14. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for nonignorable drop-out using semiparametric nonresponse models: Rejoinder. *Journal of the American Statistical Association* 1999; **94**(448):1135–1146.
15. Van Der Laan M, Robins JM. *Unified methods for censored and longitudinal data and causality*. Springer-Verlag: New York, 2003.
16. Robins JM, Rotnitzky A. Discussion of the paper by Bickel and Kwon, inference for semiparametric models: some questions and an answer. *Statistica Sinca* 2001; **11**(4):863–960.
17. Ghorai JK, Rejto L. Estimation of mean residual life with censored data under the proportional hazard model. *Communications in Statistics- Theory and Methods* 1987; **16**:2097–2114.
18. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *Journal of clinical oncology* 1983; **1**(11): 710–719.
19. Van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics* 2007; **34**(1):70–85.
20. Zheng Y, Heagerty PJ. Partly conditional survival models for longitudinal data. *Biometrics* 2005; **61**(2):379–391.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article.