

# Restricted mean models for transplant benefit and urgency

Fang Xiang and Susan Murray\*<sup>†</sup>

The US lung allocation policy estimates each individual's urgency and transplant benefit in defining a lung allocation score (LAS). Transplant benefit, as defined by the Organ Procurement and Transplantation Network Thoracic Committee, is the days of life gained over the following year if transplanted versus not transplanted. Urgency is measured by days of life during the next year without transplant. In both definitions, accurate estimation of wait list days lived, or a wait list restricted mean lifetime, is required. Risk factors are available to estimate patient urgency when listed, with more urgent patients removed from the wait list upon death or transplant. As a patient progresses, priority for transplant (censoring) changes accordingly. Therefore, it is crucial to adjust for dependent censoring in modeling days of life. We develop a model for the restricted mean as a function of covariates, by using pseudo-observations that account for dependent censoring linked to a series of longitudinal measures (LAS). Simulation results show that our method performs well in situations comparable with the LAS setting. Applying wait list and post-transplant model results that account for dependent censoring to wait list patients, we obtain estimates of transplant benefit that are larger for many of the more urgent patients in need of transplant. The difference in LAS for an individual, when properly accounting for dependent censoring, has high impact on the priority and timing of an organ offer for these patients. Copyright © 2012 John Wiley & Sons, Ltd.

**Keywords:** dependent censoring; pseudo-observation; restricted mean life; survival; transplant benefit

## 1. Introduction

For the statistical aspects of lung transplant candidate data to be appreciated, some background is required. To get a lung transplant in the USA, candidates register with the Organ Procurement and Transplantation Network (OPTN) to obtain placement on a lung waiting list. When these transplants were infrequent, a first come, first served policy seemed equitable to those waiting for transplant. But as the demand increased, so did the average waiting time to transplant, and an increasing number of end-stage lung disease patients died while waiting for an organ offer. Published in 1998 and enacted in 2000, a Final Rule, crafted by the Health Resources and Services Administration of the US Department of Health and Human Services, dictated, among other things, that a more equitable organ allocation algorithm needed to be created and maintained based on objective medical data [1].

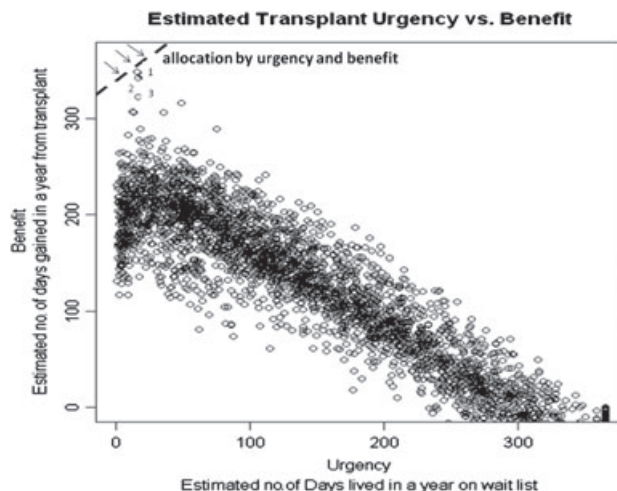
In the case of patients waiting for a lung transplant, a statistical algorithm (lung allocation score or LAS) for ranking patients was implemented on May 4, 2005 [2]. The LAS includes measures of the net benefit of the transplant to the candidate as well as the candidate's clinical urgency over the upcoming year. The measure for net transplant benefit is calculated by subtracting the patient's estimated number of days lived on the waiting list without a transplant over the next year (i.e., transplant urgency) from the estimated number of days lived during the first year following transplantation (i.e., post-transplant survival measure). This is an individual measure of transplant benefit rather than a collective measure of transplant benefit that is sometimes obtained through the use of a time-dependent covariate for transplant, as in analyses carried out for the original Stanford Heart Transplant study [3].

Figure 1 shows the estimated patient-specific urgency by anticipated transplant benefit for a group of lung candidates actively listed between September 1, 2006 and September 30, 2008. It was recognized

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

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

<sup>†</sup>E-mail: skmurray@umich.edu



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

that ordering patients based on urgency alone, that is, from left to right in Figure 1, might prioritize patients with little or no transplant benefit, whereas ordering patients solely based on higher benefit, that is, from top to bottom of Figure 1, would likely result in many deaths of urgent patients who would not live until an organ offer. In the end, a compromise was reached so that both benefit and urgency were taken into consideration, that is, allocation according to the diagonal line moving from top left to bottom right of Figure 1. The LAS takes the difference between the net transplant benefit and the transplant urgency, with the final score normalized to produce a range from 0 to 100.

Estimates for both urgency and benefit depend on accurate estimation of wait list days lived during the year following listing. Patients' risk factors measured at listing include diagnosis, age, body mass index (BMI), diabetes, assistance with activities of daily living (ADL), 6-min walk distance (6MWD), forced vital capacity (FVC), oxygen ( $O_2$ ) requirement at rest, pulmonary artery (PA) systolic pressure, partial pressure of carbon dioxide in the blood ( $PCO_2$ ), continuous mechanical ventilation, creatinine, and cardiac index. For estimating days lived in the year following transplant, risk factors used in the LAS include diagnosis, age, assistance with ADL, 6MWD, FVC, continuous mechanical ventilation, cardiac index,  $O_2$  requirement at rest, and creatinine. More details on statistical methodology will be given shortly.

The LAS has been largely successful since its implementation. The number of deaths on the waiting list and the waiting time for transplant have decreased. As opposed to 512 wait list deaths in 2004, there were only 266 deaths in 2008, in spite of more urgent patients being listed in 2008 [4]. Listing behavior of end-stage lung patients has changed dramatically since the implementation of LAS. With no advantage to accruing waiting time in the new allocation score, the number of patients actively listed for transplant decreased from 2163 candidates at the end of 2004 to 1089 patients at the end of 2008 [4]. That is, patients not yet ready to accept an organ offer began to remove themselves from the active candidate pool and delay entering the pool until further progression of disease. As a consequence, the median waiting time has dropped from 792 days in 2004 to 200 days or less after the LAS was used [4]. As successful as the LAS has been, national policy dictates that the algorithm must be continually updated to reflect more recent cohorts of patients, and this is occurring right now in a post-LAS implementation cohort.

The estimated number of days lived during a year in the calculation of LAS is sometimes called the restricted mean life. When estimated nonparametrically, under independent censoring, it is typically defined using the area under a Kaplan–Meier (KM) survival curve [5] for the period of interest (0 to 1 year). In the original development of the LAS, a Cox proportional hazards (PH) model [6] was used to estimate each individual's survival curve, and the area under the first year of the survival curve was used to estimate the restricted mean life. We feel that a more appropriate model would target the restricted

mean more directly, rather than modeling the hazard ratio. Not only would regression parameters be more directly linked to the restricted mean of interest, but also increased transparency of how scores for the LAS are produced would be welcomed by patients and physicians following allocation scores.

Andersen *et al.* [7] introduced one modeling strategy for the restricted mean. First, they generated pseudo-observations (POs) that have the same conditional mean of interest for regression modeling as the original individual level data. The advantage of creating POs in the first modeling step is that they can be modeled using traditional uncensored linear models. Pseudo-observations for mean restricted life, as defined by Andersen, Hansen, and Klein, are created using marginal estimates of restricted mean life; that is,  $\hat{\alpha}_0 = \hat{E}[\min(\tau, T)] = \int_0^\tau \hat{P}(T > t)dt$ , where  $T$  denotes the failure time,  $\tau$  is the upper limit of a time window of interest, and  $\hat{P}(T > t)$  is the KM estimate. Then, the pseudo-observation (PO) for each individual, also known from jackknife methodology, is calculated as

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

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

The intuition behind POs given in (1) is that any nonparametric estimator of  $\alpha_0 = E[\min(\tau, T)]$  is also implicitly an estimator of

$$E_Z[E[\min(\tau, T)|Z]], \tag{2}$$

where the inner expectation is of interest in regression modeling. In the case where the outermost expectation is viewed with respect to the empirical distribution of  $Z$ , with  $\tilde{\alpha}_0 = \frac{1}{n} \sum_{i=1}^n E[\min(\tau, T)|Z_i]$ , POs take the form  $n\tilde{\alpha}_0 - (n - 1)\tilde{\alpha}_0^{-i} = n[\frac{1}{n} \sum_{i=1}^n E[\min(\tau, T)|Z_i]] - (n - 1)[\frac{1}{n-1} \sum_{j=1, j \neq i}^n E[\min(\tau, T)|Z_j]] = E[\min(\tau, T)|Z_i]$ , the quantity of interest in regression modeling. Andersen, Hansen, and Klein make the case that  $\hat{\alpha}_0$  and  $\tilde{\alpha}_0$  are both consistent for  $\alpha_0$ ; hence, POs based on (1), which are estimable from censored data, can be used to estimate regression parameters predicting  $E[\min(\tau, T)|Z_i]$  by using readily available linear models. That is, models based on individual specific POs in (1),  $i = 1, \dots, n$  will have regression parameters similar to a model fit to  $\min(\tau, T)$  values,  $i = 1, \dots, n$ , if these values were available (uncensored). Graw *et al.* [8] formalize this argument and verify appropriate asymptotics of parametric estimates.

Any modeling strategy for estimating restricted means requires taking into account an especially interesting dependent censoring issue when updating the LAS to a more current cohort of patients. By removing more urgent patients from the waiting list to get transplanted (via the LAS), our resulting analysis data set is dependently censored in direct relationship to daily changing LAS of individual patients. In considering a restricted mean model approach, KM estimates used in creating POs are especially subject to dependent censoring bias. Inverse probability of censoring weighted (IPCW) methods, such as those discussed by Robins and Finkelstein [9], Robins [10], Robins and Rotnitzky [11], Satten *et al.* [12], and Scharfstein *et al.* [13] among others, have been successful in counteracting this type of bias and can be used to consistently estimate  $S_T(t) = P(T > t)$ , cumulative hazard functions, and other quantities of interest.

In this paper, we propose to estimate transplant urgency and benefit by using a PO approach to estimate 1-year restricted mean life separately in wait list and post-transplant cohorts. Our approach will modify each wait list restricted mean PO by including IPCW-based survival estimates in place of KM estimates to account for dependent censoring linked to time-dependent LAS of patients. We will estimate a 1-year transplant benefit for each patient by using a restricted mean model estimate of days lived in a year following transplant minus a separate restricted mean model estimate of days lived in a year following listing without transplant.

We structure the rest of the manuscript as follows. In Section 2, we formally describe the mean structure for restricted life given covariates, an appropriate PO approach to fit this mean structure, and the IPCW implementation of POs required to account for censoring via time-dependent LAS. Section 3 briefly shows simulation studies that ensure that our overall analysis approach in the presence of dependent censoring is sound in finite sample populations. In Section 4, we present a restricted mean model for lung wait list candidates and separately for post-transplant recipients to be used in constructing an LAS for each patient. We provide results on estimated days of life without transplant in a 1-year period (urgency) as well as estimated days gained from transplant over the following 1-year period (benefit). We also display estimated LAS for lung transplant candidates by using the new methodology and also

provide results by using estimates of restricted means for lung candidates based on integrating both traditional and IPCW-adjusted Cox PH model survival curves. Discussion follows in Section 5.

## 2. Estimating restricted mean life by using inverse probability of censoring weighted pseudo-observation

### 2.1. Mean structure for restricted mean life

The mean structure for the restricted mean life is

$$E[\log\{\min(\tau, T)\}] = \beta^T Z, \tag{3}$$

where  $\tau$  is fixed and is within the range of the observed data. When there is no censoring, uncensored data applied to model (3) becomes a standard linear model on  $\log\{\min(\tau, T)\}$ . However when censoring is present and informative, the use of observed data will lead to biased results.

Andersen *et al.* [7] formulated POs by using (1) and then fitted model (3) to the resulting POs by using standard linear models. An equally appropriate approach would be to fit the model of the restricted mean by using a log link. We have found that the intercept estimator is somewhat improved upon creating POs based on the transformed random variable  $\log\{\min(\tau, T)\}$ . That is, instead of creating POs based on  $\hat{\alpha}_0 = \hat{E}[\min(\tau, T)] = \int_0^\tau \hat{S}(t)dt$  and log transforming the pseudo-values, we will create POs based on marginal estimates of  $\hat{\delta}_0 = \hat{E}[\log\{\min(\tau, T)\}]$ . Let  $Y = \log\{\min(\tau, T)\}$ , ranging from  $-\infty$  to  $\log \tau$ . Assume for the moment the simplest form of model (3), where  $E[Y] = \delta_0$ , that is, the marginal mean of  $\log\{\min(\tau, T)\}$  that does not depend on any covariates. We may derive the mean of  $Y$  as follows:

$$\begin{aligned} E(Y) &= \delta_0 = E[\log\{\min(\tau, T)\}] = \int_0^\infty \log[\min(\tau, T)]dF_T(t) \\ &= \int_0^\tau \log t dF_T(t) + \int_\tau^\infty \log \tau dF_T(t) \\ &= \int_0^\tau \log t d(1 - S_T(t)) + \log \tau(1 - F_T(\tau)) \\ &= - \int_0^\tau \log t dS_T(t) + \log \tau \cdot S_T(\tau). \end{aligned} \tag{4}$$

### 2.2. Pseudo-observation approach

From Equation (4) in Section 2.1, the marginal mean of  $\log\{\min(\tau, T)\}$  can be estimated via

$$\hat{\delta}_0 = \hat{E}(Y) = - \int_0^\tau \log t d\hat{P}(T > t) + \log \tau \cdot \hat{P}(T > \tau) \tag{5}$$

where  $\hat{P}(T > t)$  is some marginal survival estimate on the original time scale. In the context of dependent censoring, we will describe a consistent estimate for  $S_T(t)$  in Section 2.3,  $\hat{S}_T^W(t)$ , that uses an inverse weight approach and show that its use in (5) gives a consistent estimate of  $\delta_0$ .

Arguments justifying the use of POs in fitting (3) proceed similarly to the original justification made for the PO approach. That is, any estimator of  $\delta_0$  is implicitly an estimator of  $E_Z[E[\log\{\min(\tau, T)\}|Z]]$ . When the outermost expectation is viewed with respect to the empirical distribution of  $Z$  with  $\hat{\delta}_0 = \frac{1}{n} \sum_{j=1}^n E[\log\{\min(\tau, T)\}|Z_j]$  and  $\hat{\delta}_0^{-i} = \frac{1}{n-1} \sum_{j=1, j \neq i}^n E[\log\{\min(\tau, T)\}|Z_j]$ , POs,  $n\hat{\delta}_0 - (n-1)\hat{\delta}_0^{-i}$ , reduce to  $E[\log\{\min(\tau, T)\}|Z_i]$ , which matches in expectation the quantity that we wish to model. Although both  $\hat{\delta}_0$  and  $\hat{\delta}_0^{-i}$  are consistent for  $\delta_0$ , the latter gives the most useful form for estimating  $\delta_0$  based on censored survival data. So, similar to the strategy employed on the scale of  $\alpha_0$ , we base our inference on POs,  $n\hat{\delta}_0 - (n-1)\hat{\delta}_0^{-i}$ , where  $\hat{\delta}_0^{-i}$  is estimated from (5) leaving out individual  $i$ .

Once POs,  $\mathcal{PO} = (\mathcal{PO}_1, \mathcal{PO}_2, \dots, \mathcal{PO}_n)$  are obtained, the regression model (3) can be estimated using  $\mathcal{PO}$  as the response. Our parameter estimates become  $\hat{\beta} = (Z^T Z)^{-1} Z^T \mathcal{PO}$  with estimated covariance matrix  $\hat{V}(\hat{\beta}) = \hat{\sigma}^2 (Z^T Z)^{-1}$ , where  $\hat{\sigma}^2$  is computed in the usual way as  $(\mathcal{PO} - Z\hat{\beta})^T (\mathcal{PO} - Z\hat{\beta}) / (n - p)$  for  $p$  parameters in the model. These results can be estimated from nearly any statistical software package once  $\mathcal{PO}_i, i = 1, \dots, n$  are obtained.

2.3. Inverse probability of censoring weighted estimates of survival probability

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

First, one estimates the censoring survival function at any fixed time  $t$ , denoted by  $\hat{K}_i^V(t) = P(C_i > t | \bar{V}_i(t))$ , where  $C_i$  is the censoring time for patient  $i$  and  $\bar{V}_i(t) = \{\mathbf{V}_i(u); 0 \leq u \leq t\}$  is the patient's recorded history up to time  $t$  of a vector of possibly time-dependent covariates,  $\mathbf{V}_i$ , that predict the censoring time  $C_i$ . In the case of the lung allocation data,  $C_i$  is the time a patient is removed from the lung wait list for transplant, and  $\bar{V}_i(t)$  consists of patient LAS used to rank patients for transplant from time 0 (listing time) to time  $t$  as well as a few additional predictors including race, gender, blood type, height, and active waiting status. In calculating the contribution of a subject at risk at time  $t$ , the subject is given a weight inversely proportional to his or her estimated probability of remaining uncensored until time  $t$  with a history of  $\bar{V}_i(t)$ , that is,

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

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

$$\lambda_Q\{t|\bar{V}(t)\} = \lambda_{Q0}(t) \exp\{\gamma'\mathbf{V}(t)\}. \tag{6}$$

In the case of the lung wait list data,  $\gamma'\mathbf{V}(t)$  becomes  $\gamma_1 \text{LAS}(t) + \gamma_2 \text{race} + \gamma_3 \text{gender} + \gamma_4 \text{blood type} + \gamma_5 \text{height} + \gamma_6 I(\text{active waiting status})$ . Then, a consistent estimate of the probability that subject  $i$  gets censored after time  $t$ ,  $K_i^V(t)$ , becomes

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

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

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

An IPCW version of Nelson–Aalen estimator for  $\Lambda(t)$  is calculated using

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

where  $N_{T_i}(u) = I(X_i \leq u, \delta_i = 1)$  is the observable counting process for death. Then, the survival probability is estimated with  $\hat{S}^W(t) = \exp\{-\hat{\Lambda}^W(t)\}$ . The adjusted POs described in Section 2.2 use  $\hat{P}(T > t) = \hat{S}^W(t)$  in Equation (5). Product-integral versions of inverse weighted survival functions such as those described by Satten and Datta [14] would also be appropriate for use as an alternative to  $\hat{S}^W(t)$ .

Proof of consistency of  $\hat{S}_T^W(t) = \exp(-\hat{\Lambda}^W(t))$  for  $S_T(t)$  proceeds from consistency of  $\hat{\Lambda}^W(t)$  for  $\Lambda(t)$ , a property that was studied extensively by Robins [10] and Robins and Finkelstein [9]. Conditions required for this consistency to hold are that (i)  $\lambda_Q(t|\bar{V}(t))$  follows the form given in Equation (6) and (ii)  $\lambda_Q(t|\bar{V}(t), T, T > t) = \lambda_Q(t|\bar{V}(t), T > t)$ . Consistency of  $\hat{E}(Y)$ , used in creating POs in this manuscript, follows from noting that  $\int_0^\tau \log t \, d\hat{S}_T^W(t)$  in (5) can be written as

$$\lim_{m \rightarrow \infty} \sum_{j=1}^m \log t_j \Delta \hat{S}_T^W(t_j) \xrightarrow{P} \lim_{m \rightarrow \infty} \sum_{j=1}^m \log t_j \Delta S_T(t_j) = \int_0^\tau \log t \, dS_T(t).$$

### 3. Simulation study

To validate the method used in analyzing the lung candidate wait list data subjected to dependent censoring, we conduct a simulation study comparing parameter estimates of model (3) by using linear regression 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 (1), and (iii)  $\log[\min(\tau, T)]$  is subject to censoring and is replaced by IPCW-adjusted POs.

In each simulation, we perform the following procedures.

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

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

Step 3: We also generate dependent censoring times  $C_i$  from piecewise exponential distributions and obtain hazard rates 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  $\gamma_4 = 1$ . So,  $Z_0, Z_2$ , and time-dependent  $Z_1$  influence censoring.

For each scenario of  $\beta$ , we run 1000 simulations with  $\tau = 5$  years and either  $n = 150$  or  $n = 300$  patients. Results for the scenario with  $\beta_0 = 0.8, \beta_1 = \beta_2 = 0$  are located in part (1) of Table I. In this

**Table I.** Comparison of estimates from model (3) by using uncensored observations (uncensored), the traditional pseudo-observation (PO) approach, and the inverse probability of censoring weighted (IPCW) PO approach under two scenarios, 1000 iterations.

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

\*SE is the average of estimated standard errors across 1000 iterations.

†ESD is the empirical standard deviation of 1000 parameter estimates.

case, the true baseline covariate effects on survival are zero, but dependent censoring is being driven by the time-dependent covariate. The unadjusted PO method gives more biased estimates for all the parameters, especially for  $\beta_0$  and  $\beta_1$ . The IPCW-adjusted PO method reduces bias substantially and also has smaller standard error after adjusting for dependent censoring. When both baseline covariates are equal to 0.5, the PO method underestimates the time lived during the 5-year period by 14 months on average whereas the IPCW PO method is off by only 2 months over the 5-year period. Empirical standard deviations were comparable with standard errors averaged across simulations, with the exception of the intercept term for the traditional PO method. Coverage for the traditional PO intercept was 66.7%, in spite of its much wider confidence interval width, because of the increased bias and underestimated variability for that term.

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

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

## 4. Example

We organize this section into three components. Section 4.1 summarizes analyses for the lung wait list candidates, Section 4.2 summarizes analyses for the post-transplant cohort, and Section 4.3 interprets these analyses in terms of urgency, benefit, and LAS calculated for the wait list patient cohort. Results are typically reported by the four defined diagnosis groups A, B, C, and D. Diagnosis group A is obstructive lung disease, primarily chronic obstructive pulmonary disease. Group B consists of pulmonary vascular diseases, primarily idiopathic pulmonary arterial hypertension. Group C consists of cystic fibrosis, as well as immunodeficiency disorders. Group D is restrictive lung disease, primarily interstitial pulmonary fibrosis. The OPTN Thoracic Committee classifies all lung wait list patients into one of these four diagnosis groups for the purpose of estimating diagnosis group influence on urgency and benefit, and modeling interactions across diagnosis groups. A few diagnoses, such as bronchiectasis, are given a parameter to distinguish their estimated days of life from that of their overall diagnosis group. These parameters have historically not been significantly different from those of their overall diagnosis group, and yet patient advocates have actively pursued the ability to estimate urgency and benefit more specifically for their patients to the extent that enough data are available to do so.

### 4.1. Lung candidate analysis

The wait list candidate data contain 3701 lung candidates aged 12 years and above who were newly listed in the lung wait list during September 1, 2006 and September 30, 2008. Censoring within 1 year of listing only occurs when a candidate is transplanted, which was the case with 2698 (73%) of the candidates. By diagnosis group, 923 (70%) of 1317 group A candidates, 67 (58%) of 116 group B candidates, 294 (69%) of 428 group C candidates, and 1414 (77%) of 1840 group D candidates were transplanted. Historically, of the four groups, diagnosis group D has had the poorest wait list survival, and with the LAS based in part on urgency, this group also currently experiences the shortest time to transplant. The median time to transplant for group D is only 71 days, as opposed to 170 days for group A, 221 days for group B, and 126 days for group C. We show baseline characteristics by primary diagnosis group in Table II. At listing, group D patients typically have very high severity and poor physiologic reserve. In contrast, group A patients have historically had much lower urgency for transplant as measured by survival. These patients are often seeking a transplant based on improving quality of life as opposed to lengthening life.

**Table II.** Baseline characteristics by diagnosis group for 3701 lung wait list patients.

Characteristics	Group A (primarily COPD) <i>n</i> = 1317	Group B (primarily iPAH) <i>n</i> = 116	Group C (primarily CF) <i>n</i> = 428	Group D (primarily IPF) <i>n</i> = 1840
Physiologic reserve				
Age (years)	57.6 ± 8.1*	45.9 ± 15.1	27.9 ± 10.3	55.9 ± 11.6
BMI (kg/m <sup>2</sup> )	24.6 ± 4.3	24.8 ± 4.3	19.1 ± 2.9	26.7 ± 4.6
Diabetes	11.0% <sup>†</sup>	10.3%	48.4%	22.9%
No assistance with ADL	10.4%	9.4%	20.9%	9.6%
6-min walk distance (feet)	770.5 ± 356.0	733.3 ± 480.8	902.2 ± 508.0	755.8 ± 482.0
Severity				
FVC (% predicted)	52.4 ± 17.2	69.3 ± 23.1	38.1 ± 11.1	47.0 ± 16.7
O <sub>2</sub> requirement at rest (l/min)	3.1 ± 2.5	4.6 ± 5.1	3.4 ± 4.1	4.6 ± 4.7
PA systolic (mmHg)	38.0 ± 10.8	78.5 ± 24.0	38.8 ± 10.6	42.9 ± 17.0
PCO <sub>2</sub> (mmHg)	50.3 ± 11.7	43.5 ± 7.0	56.2 ± 20.0	45.9 ± 11.6
Continuous mechanical ventilation	1.1%	2.6%	7.5%	6.1%
Serum creatinine (mg/dl)	0.82 ± 0.2	0.94 ± 0.3	0.67 ± 0.2	0.91 ± 0.3
Cardiac index < 2.0 (l/min/min <sup>2</sup> )	5.5%	35.6%	1.6%	7.8%

\*For continuous variables, the numbers shown are mean ± standard deviation.

<sup>†</sup>For binary variables, the numbers shown are proportions.

ADL, activities of daily living; BMI, body mass index; CF, cystic fibrosis; COPD, chronic obstructive pulmonary disease; FVC, forced vital capacity; iPAH, idiopathic pulmonary arterial hypertension; IPF, interstitial pulmonary fibrosis; PA, pulmonary artery; PCO<sub>2</sub>, partial pressure of carbon dioxide.

**Table III.** Proportional hazards censoring model (3701 candidates).

Parameter	HR	95% CI	<i>p</i> -value
Characteristic at listing			
Female (versus male)	0.72	(0.63, 0.82)	<.0001
Race: Black (versus White)	0.81	(0.68, 0.95)	0.0116
Race: Other (versus White)	0.91	(0.77, 1.08)	0.3016
Height: <5'3" (versus >5'9")	0.54	(0.45, 0.65)	<.0001
Height: 5'3" to 5'6" (versus >5'9")	0.73	(0.62, 0.86)	0.0001
Height: 5'6" to 5'9" (versus >5'9")	0.80	(0.71, 0.90)	0.0001
Blood type: B (versus A)	1.06	(0.91, 1.23)	0.4801
Blood type: O (versus A)	0.92	(0.84, 1.02)	0.1118
Blood type: AB (versus A)	1.07	(0.85, 1.33)	0.5669
Time-dependent patient condition and listing status			
LAS = 0 (versus LAS > 0)	0.16	(0.02, 1.18)	0.0728
LAS: linear spline for 30+	1.12*	(1.06, 1.19)	<.0001
LAS: linear spline for 35+	0.98 <sup>†</sup>	(0.91, 1.06)	0.6779
LAS: linear spline for 40+	0.95 <sup>‡</sup>	(0.91, 0.99)	0.0070
LAS: linear spline for 60+	0.97 <sup>§</sup>	(0.95, 0.98)	<.0001
Inactive status (versus active)	0.00	(0, >1000)	0.8780
Off the wait list (versus active)	0.00	(0, >1000)	0.9410

Inverse weights based on this model are capped at 20.

\*Hazard ratio (HR) corresponding to one unit increase for lung allocation score (LAS) 30+ relative to those with 0 < LAS < 30.

<sup>†</sup>HR 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.

<sup>‡</sup>HR 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.

<sup>§</sup>HR 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.

We use the P(censoring occurs after time *t*|candidate's history up to and including *t*) to calculate an inverse weight used in consistent estimation of survival curves and adjusted POs as described in Sections 2.2 and 2.3. In particular, we use a time-dependent Cox model for time to censoring, which



includes patients' daily updated LAS, gender, race, blood type, status (active, inactive, offlist), and height, as given in Equation (6). We summarize parameter estimates from the Cox model on the censoring hazard in Table III. Probability of transplant is strongly influenced by current LAS values. Although one might expect the probability of transplant to increase monotonically as current LAS increases, in fact, the higher transplant priority is tempered by a lower chance of surviving until an organ becomes available for those with the highest LAS values. This feature is reflected in the parameter estimates shown.

To estimate lung candidate urgency, we fit model (3) by using both IPCW-adjusted POs and traditional POs. For comparison, we estimate restricted means for lung candidates by integrating traditional and IPCW-adjusted survival curves based on Cox model [6]. Predictors included in all lung candidate models are the same as those proposed by the OPTN Thoracic Committee in modeling wait list survival for this cohort [15]. The Thoracic Committee has vetted extensively all predictors as being worthy of inclusion in the algorithm based on statistical and/or clinical validity based on either the current or

**Table IV.** Lung wait list results for model (3) by using IPCW PO and traditional PO methods for 3701 lung candidates.

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

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

†The OPTN Thoracic Committee grouped these diagnoses into larger diagnosis groups (A, B, C, and D) 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; and obliterativebronchiolitis, pulmonary fibrosis other, and sarcoidosis and PA mean >30 mmHg with group D.

ADL, activities of daily living; CF, cystic fibrosis; COPD, chronic obstructive pulmonary disease; FVC, forced vital capacity; iPAH, idiopathic pulmonary arterial hypertension; IPCW PO, inverse probability of censoring weighted pseudo-observation; IPF, interstitial pulmonary fibrosis; PA, pulmonary artery; PCO<sub>2</sub>, partial pressure of carbon dioxide.

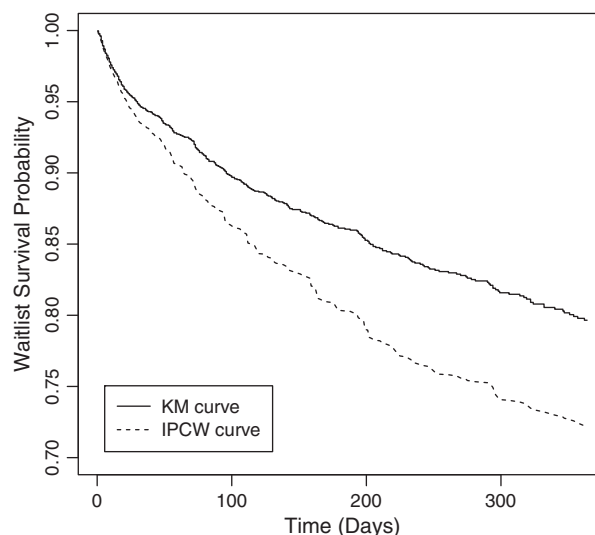
a prior wait list cohort studied. In some cases, statistically insignificant parameters are maintained as placeholders with the expectation that statistical significance will re-assert itself in future cohorts; age, assistance with ADL, PA systolic,  $PCO_2$ , and creatinine fall into this category.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

We show parameter estimates for fitting model (3) to patients awaiting transplant in Table IV. For risk factors,  $e^{\hat{\beta}}$  acts multiplicatively on the number of days lived in a year. For instance, the estimated number of days lived is  $102.15 \times 0.11 \times 1.07^{25} \times 0.49 \times 1.07^3 \times 1.25^{5.5} \times 0.87^2 \times 0.60^{0.8} = 63$  days based on IPCW PO and  $301.71 \times 0.40 \times 1.03^{25} \times 0.92 \times 1.04^3 \times 1.06^{5.5} \times 0.92^2 \times 0.91^{0.8} = 283$  days based on traditional PO for a 55-year-old diagnosis group D patient who has a BMI of 25, has diabetes, requires assistance with ADL, walks 300 ft in 6 min, has 55% predicted FVC, requires 2 l/min of  $O_2$  at rest, is not on a ventilator, has a stable creatinine of 0.8 mg/dl, has no partial pressure of  $CO_2$  in their blood, and has a cardiac index  $>2$  l/min/min<sup>2</sup>. So, the unadjusted PO overestimates the number of days lived in the next year without transplant by 220 days for this very urgent group D patient. We show estimated KM and IPCW survival curves used in PO calculations in Figure 2; overestimation of days lived using the traditional PO model stems from the overestimation of survival by the KM method.

We show traditional and IPCW-adjusted Cox model hazard ratios estimated from the lung candidate data in Table V. The difference in area under the baseline survival curves over a year for the two methods was 11 days, with a more favorable survival profile without adjustment. Integrating a survival curve estimated from a traditional Cox PH model for this patient yields an estimated number of days lived of 330, which exceeds the estimate based on the unadjusted PO model. The IPCW-adjusted Cox model estimates a restricted mean of 301 for the same patient. Because bias due to dependent censoring has been accounted for in both IPCW PO method and IPCW Cox method, different modeling restrictions account for the observed differences in estimation. We provide additional information on urgency estimates based on different modeling paradigms in Section 4.3.

## 4.2. Lung recipient analysis

The post-transplant data contain 4784 patients aged 12 years and above who received a lung transplant between May 4, 2005 and September 3, 2008. One-year event rates from the time of transplant are perfectly known, that is, no censoring, with 816 (17%) deaths within the first year. We show results from fitting model (3) in the uncensored case in Table VI. For the same group D wait list patient described in Section 4.1, the estimated days lived in the first year following transplant is 254 days based on the model in Table VI. Recall that the estimated days gained in the first year following transplant is calculated using the estimated days lived 1 year post-transplant as in Section 4.2 minus the estimated days lived



**Figure 2.** Wait list survival probabilities estimated using KM and IPCW.

**Table V.** Lung wait list survival model by using IPCW Cox and traditional Cox methods for 3701 lung candidates.

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

\*The OPTN Thoracic Committee grouped these diagnoses into larger diagnosis groups (A, B, C, and D) 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; and obliterativebronchiolitis, pulmonary fibrosis other, and sarcoidosis and PA mean >30 mmHg with group D.

ADL, activities of daily living; BMI, body mass index; CF, cystic fibrosis; COPD, chronic obstructive pulmonary disease; FVC, forced vital capacity; iPAH, idiopathic pulmonary arterial hypertension; IPCW PO, inverse probability of censoring weighted pseudo-observation; IPF, interstitial pulmonary fibrosis; PA, pulmonary artery; PCO<sub>2</sub>, partial pressure of carbon dioxide.

1 year without transplant as in Section 4.1. So, the IPCW PO method estimates  $254 - 63 = 191$  days gained during the first year after a transplant for this patient. All other methods indicate days of life lost if transplanted. The IPCW Cox method, the traditional PO method, and the unadjusted Cox method give  $254 - 301 = -47$  days,  $254 - 283 = -29$  days, and  $254 - 330 = -76$  days, respectively.

#### 4.3. Urgency, benefit, and lung allocation scores

Figure 3 shows boxplots of estimated transplant urgency for the 3701 wait listed patients by diagnosis group, using the two PO modeling paradigms laid out in Table IV and the two Cox modeling paradigms in Table V. By adjusting for dependent censoring, POs based on IPCW survival curves estimated a higher urgency in group D patients, which better matches group D survival experience seen before organ allocation took urgency into account (i.e., before LAS-induced dependent censoring was introduced). The integrated IPCW Cox survival curves also show more urgency than the integrated Cox survival curves that do not take into account dependent censoring. The interquartile range for urgency estimates based on integrated Cox PH survival curves is decidedly more narrow than the range of estimated restricted means based on model 3 for either the PO or the IPCW PO methods. In addition, the Cox modeling approaches tend to estimate many more days lived without transplant compared with the PO methods.

	$e^{\hat{\beta}^*}$	95% CI	<i>p</i> -value
(Intercept)	344.42	(303.36, 391.03)	<0.0001
Diagnosis group (ref = group A, primarily COPD)			
Group B (primarily iPAH)	0.65	(0.52, 0.81)	0.0002
Group C (primarily CF)	0.92	(0.82, 1.04)	0.1713
Group D (primarily IPF)	0.84	(0.73, 0.96)	0.0107
Diagnosis <sup>†</sup>			
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
Obliterativebronchiolitis	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 <sup>‡</sup> (years)	0.99	(0.99, 1.00)	0.0139
No assistance with ADL	1.02	(0.94, 1.11)	0.6648
6-min 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 <sup>2</sup> )	0.86	(0.74, 1.00)	0.0496
O <sub>2</sub> at rest for dgn group A (l/min)	0.97	(0.96, 0.99)	0.0063
O <sub>2</sub> 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

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

†The OPTN Thoracic Committee grouped these diagnoses into larger diagnosis groups (A, B, C, and D) 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; and obliterativebronchiolitis, pulmonary fibrosis other, and sarcoidosis and PA mean >30 mmHg with group D.

‡Age >45 spline: the maximum of 0 and age -45.

ADL, activities of daily living; CF, cystic fibrosis; COPD, chronic obstructive pulmonary disease; FVC, forced vital capacity; iPAH, idiopathic pulmonary arterial hypertension; IPF, interstitial pulmonary fibrosis; PA, pulmonary artery; PCO<sub>2</sub>, partial pressure of carbon dioxide.

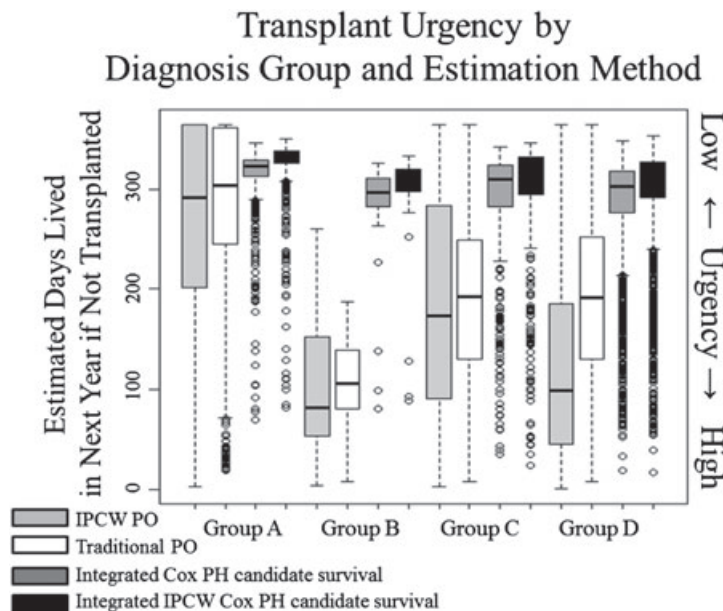


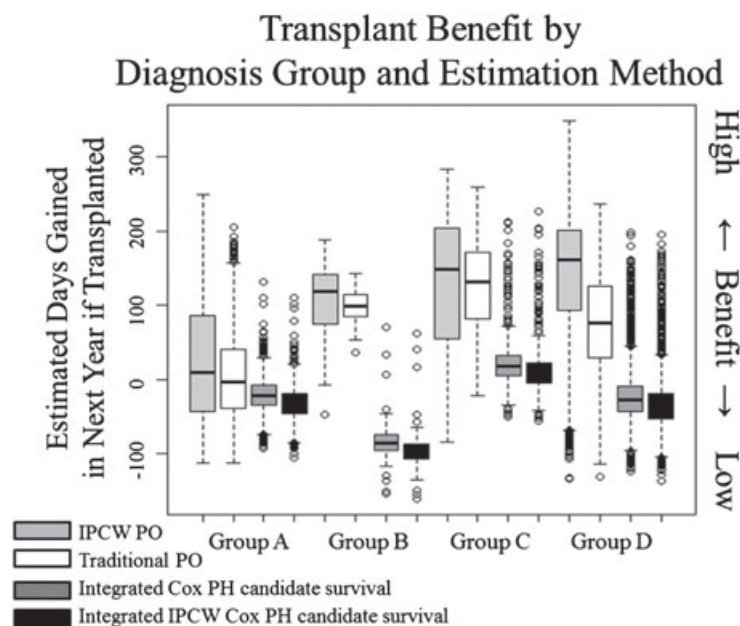
Figure 3. Urgency by diagnosis group.

Transplant benefit calculations similar to those carried out for the hypothetical group D wait list patient described in Sections 4.1 and 4.2 were performed for all 3701 patients in our wait list cohort. Figure 4 shows boxplots of estimated transplant benefit by diagnosis group by using IPCW PO, traditional PO, integrated IPCW Cox PH survival, and integrated Cox PH survival for this cohort. In each case, the lung recipient model used model (3) applied to the original (perfectly observed) data for this setting. Patient transplant benefit calculations incorporating the IPCW PO model for urgency identified more benefit in group D patients than when using any other modeling paradigm. The use of integrated IPCW Cox survival curves also exhibit more estimated benefits than the use of integrated Cox survival curves not adjusting for dependent censoring. However, benefit estimates remain low, with tight interquartile ranges, when compared with either of the PO methods.

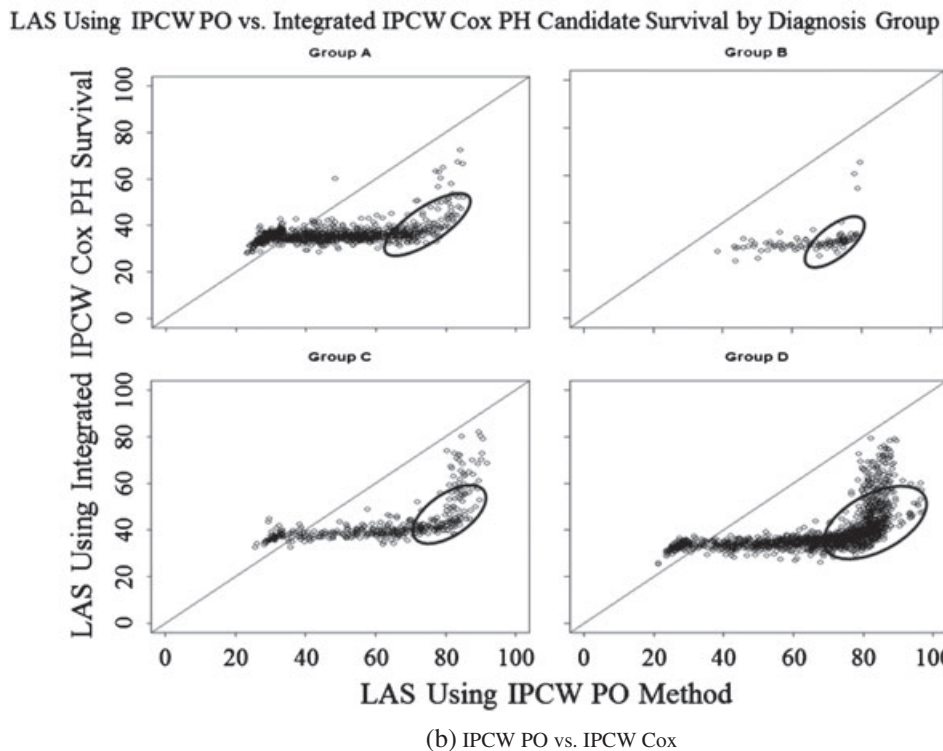
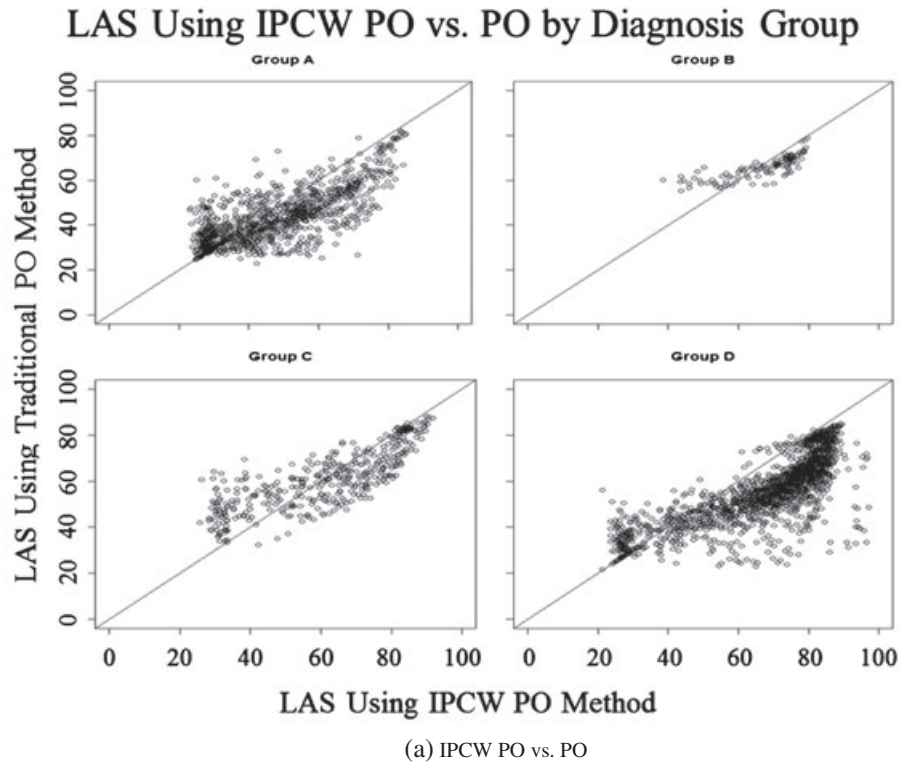
Figure 5(a) shows scatter plots of LAS calculated using IPCW PO versus traditional PO by diagnosis group, with a 45° line superimposed on the plot. For each diagnosis group, the LAS changes substantially when taking into account dependent censoring. When looking at the top 100 ranked patients based on their IPCW PO derived LAS, their scores estimated using traditional PO methods dropped by approximately 16 points on average (0.8 standard deviations of the estimated LAS distribution) when not adjusting for dependent censoring. Similarly, scores dropped by approximately 36 points on average using the traditional Cox model integrated wait list survival curves. Model paradigm selection and adjustment for dependent censoring have a serious impact on time to transplant for those top priority candidates identified using IPCW PO methodology. Figure 5(b) shows a scatter plot of LAS values when calculated using the IPCW PO method versus using integrated IPCW Cox survival curves for wait list urgency. Circled values represent patients who would move from having a low allocation priority by using PH assumptions to a very high allocation priority by using model (3). We rank only 30 patients in the top 100 scores regardless of the IPCW method used, IPCW PO or IPCW Cox integrated survival curves.

## 5. Discussion

We present new methodology for estimating restricted means in the presence of dependent censoring captured by longitudinal covariates. Upon estimation of  $S_T(t)$  by using inverse weight methodology, remaining inference becomes very straight forward using our suggested approach. In particular, it is not necessary to program complicated variances of inverse weighted estimates, because the PO approach



**Figure 4.** Estimated transplant benefit at time of listing by diagnosis group and estimation method for 3701 lung transplant candidates.



**Figure 5.** LAS at listing calculated using IPCW PO, PO, and Cox PH model by diagnosis group for 3701 lung candidates.

merges nicely into use of more standard software package for regression in evaluating parameter estimates. Hence, this approach can realistically be implemented by statistical practitioners.

Statistical input into properly modeling components of the LAS in the USA has high impact on perhaps 1000 patients at any given time, as the rate of new listings and wait list removals seems to balance at that level. The success of devising and maintaining an intelligent and practical allocation system

for urgent patients introduces a uniquely interesting set of statistical issues. Defining 1-year transplant urgency and benefit at the individual level can be achieved successfully using our described methodology, even when subjected to dependent censoring by transplant for more urgent patients. Hence, the LAS can be updated now and in the future using the most recent cohort of patients with minimal bias.

When applied to the lung candidate data, the IPCW PO method gives a broader range of urgency estimates than when estimating urgency based on integrated IPCW Cox model survival curves. This in turn leads to a broader range of LAS values with which to prioritize the candidates. Our feeling is that parameterization on the scale of the restricted mean leads to more appropriate urgency estimates than parameterization based on constant hazard ratios over time, particularly after viewing the range of scores from using different modeling paradigms in Figure 5(b).

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

## APPENDIX A.

In Section 3, we simulate settings when dependent censoring affects estimation of (3) unless adjustments are made via the IPCW PO method. The survival function for piecewise exponential failure times used in Section 3 is

$$S_T(t) = \begin{cases} e^{-\lambda z_0 t} & 0 \leq t \leq t_1 \\ e^{-\lambda z_0 t_1} e^{-\lambda z_0 z_1 (t-t_1)} & t > t_1 \end{cases},$$

with pdf

$$f_T(t) = \begin{cases} \lambda z_0 \cdot e^{-\lambda z_0 t} & 0 \leq t \leq t_1 \\ \lambda z_0 z_1 \cdot e^{-(\lambda z_0 - \lambda z_0 z_1) t_1} \cdot e^{-\lambda z_0 z_1 t} & t > t_1 \end{cases}.$$

In step 2 of Section 3, parameters  $\lambda_0 = 0.3$ ,  $\lambda_1 = 0.2$ ,  $\lambda_{01} = 0.1$ , and  $\lambda_{11} = 0.5$  are fixed. The remaining parameters  $\lambda_{00}$  and  $\lambda_{10}$  are chosen to satisfy (3) as described in the succeeding equation. Recall that  $Z_1$  is measured at time  $t_1 = 0.2$  and that  $\tau = 5$ . Also,  $Z_0$  and  $Z_1$  are generated from independent Bernoulli(0.5).

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

When  $Z_0 = 0$ , we have

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

We solve the aforementioned equation for  $\lambda_{00}$ . Similarly, when  $Z_0 = 1$ , we solve the following equation for  $\lambda_{10}$ :

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

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

## 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 of the SRTR for their support.

## References

1. Department of Health and Human Services. Organ Procurement and Transplantation Network; Final Rule. In: Federal Register 42 CFR, Part 121, 56649–56661, October 20, 1999.
2. 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.
3. Crowley J, Hu M. Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* 1977; **72**:27–36.
4. 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.
5. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. *Journal of American Statistics Association* 1958; **53**:457–481. 562–563.
6. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**:187–200.
7. Andersen PK, Hansen M, Klein JP. Regression analysis of restricted mean survival time based on pseudo observations. *Life Data Analysis* 2004; **10**:335–350.
8. Graw F, Gerds T, Schumacher M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Analysis* 2009; **15**:241–255.
9. 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.
10. 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: American Statistical Association, 1993; 24–33.
11. 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.
12. Satten GA, Datta S, Robins JM. Estimating the marginal survival function in the presence of time dependent covariates. *Statistics & Probability Letters* 2001; **54**:397–403.
13. 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.
14. Satten GA, Datta S. The Kaplan–Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician* 2001; **55**:207–210.
15. OPTN Thoracic Organ Transplantation Meeting Minutes, March 23 2010.