

Objective Estimates Improve Risk Stratification for Primary Graft Dysfunction After Lung Transplantation

R. J. Shah^{1,2,*}, J. M. Diamond^{1,2,1}, E. Cantu^{3,1},
J. Flesch^{1,2}, J. C. Lee¹, D. J. Lederer⁴,
V. N. Lama⁵, J. Orens⁶, A. Weinacker⁷,
D. S. Wilkes⁸, D. Roe⁸, S. Bhorade⁹,
K. M. Wille¹⁰, L. B. Ware¹¹, S. M. Palmer¹²,
M. Crespo¹³, E. Demissie², J. Sonnet¹⁴,
A. Shah¹⁵, S. M. Kawut², S. L. Bellamy²,
A. R. Localio² and J. D. Christie^{1,2}

¹Pulmonary, Allergy, and Critical Care Division, University of Pennsylvania School of Medicine, Philadelphia, PA

²Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, PA

³Division of Cardiovascular Surgery, University of Pennsylvania School of Medicine, Philadelphia, PA

⁴Division of Pulmonary, Allergy, and Critical Care Medicine, Columbia University College of Physicians and Surgeons, New York, NY

⁵Division of Pulmonary, Allergy, and Critical Care Medicine, University of Michigan, Ann Arbor, MI

⁶Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, Johns Hopkins University Hospital, Baltimore, MD

⁷Department of Pulmonary and Critical Care, Stanford University, Palo Alto, CA

⁸Division of Pulmonary, and Critical Care Medicine, Indiana University School of Medicine, Indianapolis, IN

⁹Division of Pulmonary and Critical Care Medicine, University of Chicago, Chicago, IL

¹⁰Division of Pulmonary and Critical Care Medicine, University of Alabama at Birmingham, Birmingham, AL

¹¹Division of Allergy, Pulmonary, and Critical Care Medicine, Vanderbilt University Medical Center, Nashville, TN

¹²Division of Pulmonary, Allergy, and Critical Care Medicine, Duke University, Durham, NC

¹³Division of Pulmonary, Allergy, and Critical Care, University of Pittsburgh, Pittsburgh, PA

¹⁴Department of Surgery, Columbia University College of Physicians and Surgeons, New York, NY

¹⁵Department of Surgery, Johns Hopkins University Hospital, Baltimore, MD

*Corresponding author: Rupal J. Shah, rupal.shah@uphs.upenn.edu

¹These authors have contributed equally to this work.

available clinical variables, using a prospective study of 11 centers in the Lung Transplant Outcomes Group (LTOG). Derivation included 1255 subjects from 2002 to 2010; with separate validation in 382 subjects accrued from 2011 to 2012. We used logistic regression to identify predictors of grade 3 PGD at 48/72 h, and decision curve methods to assess impact on clinical decisions. 211/1255 subjects in the derivation and 56/382 subjects in the validation developed PGD. We developed three prediction models, where low-risk recipients had a normal BMI (18.5–25 kg/m²), chronic obstructive pulmonary disease/cystic fibrosis, and absent or mild pulmonary hypertension (mPAP < 40 mmHg). All others were considered higher-risk. Low-risk recipients had a predicted PGD risk of 4–7%, and high-risk a predicted PGD risk of 15–18%. Adding a donor-smoking lung to a higher-risk recipient significantly increased PGD risk, although risk did not change in low-risk recipients. Validation demonstrated that probability estimates were generally accurate and that models worked best at baseline PGD incidences between 5% and 25%. We conclude that valid estimates of PGD risk can be produced using readily available clinical variables.

Abbreviations: CF, cystic fibrosis; COPD, chronic obstructive pulmonary disease; DCD, donation after cardiac death; ECMO, extra-corporeal membrane oxygenation; ILD, interstitial lung disease; LTOG, Lung Transplant Outcomes Group; PAH, pulmonary arterial hypertension; PGD, primary graft dysfunction; PH, pulmonary hypertension; UNOS, United Network of Organ Sharing

Received 16 April 2014, revised 02 February 2015 and accepted for publication 07 February 2015

Introduction

Primary graft dysfunction (PGD) is a form of acute lung injury affecting the lung allograft within 72 h of transplantation (1). PGD is common, occurring in 10–30% of transplant recipients, and it is the leading cause of mortality in the early posttransplant period, accounting for nearly 50% of deaths within the first 30 days (2,3). Additionally, PGD is a risk factor for other long-term complications of lung transplantation, including bronchiolitis obliterans syndrome (4,5).

Given the clear association between PGD and poor outcomes after lung transplantation, the ability to risk stratify recipients prior to transplantation is important for

Primary graft dysfunction (PGD) is a major cause of early mortality after lung transplant. We aimed to define objective estimates of PGD risk based on readily

several reasons. First, better identification of higher- and low-risk recipient groups may allow the caring team to be better prepared for the likelihood of development of PGD. Second, improved preoperative prediction may facilitate attempts to safely expand the donor pool by characterizing lower risk recipient groups. Third, identification of higher-risk groups of recipients will facilitate therapeutic interventions or clinical trials aimed at reducing PGD both before and/or immediately after transplantation.

In this study, we aimed to define and validate objective estimates of the risk of development of PGD based on readily available clinical variables. Furthermore, we used decision curve analysis to demonstrate the net benefit of different models across a range of baseline PGD incidence.

Methods

Study participants

Subjects aged 18–80 years were enrolled from 11 US transplant centers within the Lung Transplant Outcomes Group (LTOG), a multicenter prospective cohort study of lung transplant recipients (bilateral and single) designed to evaluate risk factors and build a prediction model for PGD (3,6–11). The derivation population included subjects enrolled between May 2002 and December 2010; we have previously published on individual risk factors for PGD in the derivation cohort (clinicaltrials.gov identifier: NCT00552357) (3,10,12). A new and separate validation population included subjects enrolled between January 2011 and August 2012. Clinical data were collected prospectively for all subjects (13). The Institutional Review Boards at each site approved our study. Informed consent was obtained from each subject.

Outcome definition

PGD in both the derivation and the validation cohorts was graded according to ISHLT criteria, defined using PaO₂/FiO₂ ratio and the presence of diffuse infiltrates in the lung allograft(s) (14). Two blinded physicians interpreted chest radiographs independently, with adjudication of conflicts by a third reviewer (kappa for agreement on subject-level grade 3 PGD classification = 0.95) (14). The primary outcome was grade 3 PGD at 48 or 72 h after transplantation, which has been validated and has been used as an outcome in previous studies, including clinical trials (3,15,16).

Prediction model development

Variable definitions: We transformed continuous variables into categorical variables for simplification. As much as possible, categories were based on clinically useful algorithms (e.g. WHO classification criteria) and not derived from the data. Body mass index and pulmonary artery pressure were categorized using WHO classification (17). Two variables representing donor smoking were evaluated for use in the prediction model. The first variable was prospectively collected by LTOG personnel, based on Donor Net data and reported history by surrogates, and was categorized as yes, any history of donor smoking, or no, a never smoker (3). The second variable was obtained from UNOS data, which stratified patients as a “yes/no/unknown” history of smoking based on a >20 pack-year smoking history. The UNOS dataset contains additional cigarette smoking variables that had too much missing data to be included as predictors, including a field for “continued cigarette use,” which was missing in 85% of subjects. Pulmonary artery pressures at the time of listing were used preferentially to allow the model to be valid when used prior to transplant; however, if unavailable, pressures at the time of transplant

were used. Pretransplant recipient diagnosis was grouped into three categories based on predicted PGD risk: chronic obstructive pulmonary disease (COPD) or cystic fibrosis (CF), interstitial lung disease (ILD), and all other diagnoses (including sarcoidosis and PAH) (3).

Missing data in the derivation cohort: To avoid incorrect choices in candidate predictors and model mis-specification, we engaged in a formal three-step missing data imputation process in the derivation dataset before modeling. First, we examined each candidate predictor for out of range values or errors and set those values to missing. Second, using the method of chained equations, we created 20 imputed datasets with values for 63 variables using multinomial, ordinal, and linear regression models containing demographic and clinical factors from both donor and recipients (18–20). Third, after the imputation process, we identified out-of-range values and truncated them to fall within clinical bounds.

Statistical analysis in the derivation cohort: For this study, we considered all possible recipient and donor risk factors as potential predictors, not just those studied in prior publications (21). We used several methods to narrow the number of candidate predictors for a final model based on the development dataset. First, we performed all regressions with the set of 20 imputed datasets; then estimated statistical significance based on the sum of between- and within-imputation components of variance to avoid overstating statistical significance. Second, while we initially considered the statistical significance of a factor’s association with PGD, we guarded against overfitting in multivariable models by means of an automated test of stepwise model-building with bootstrap resampling. Within each of 1000 bootstrap samples, we performed logistic regression and noted the number of times that a candidate predictor remained in the resulting model. To reduce the number of candidate recipient predictors, we included in the final model only those predictors that were associated with PGD in the stepwise regressions with p-values <0.05 in >67% of bootstrap resamples. Although different thresholds may be used to identify candidate predictors, we felt 67% was most inclusive (22) and conducted sensitivity analyses of this threshold to include ranges from 25% to 75% of bootstrap resamples.

After identifying candidate predictors, we focused on the ability of competing models using only these predictor variables to generate high- and low-risk patient subgroups with objective estimates of PGD risk. We chose this method to maximize the clinical utility of the prediction estimates for groups of patients, as opposed to methods that focus solely on maximizing discrimination as measured by the area under a ROC curve (or c-statistic), which may be more suitable to ranking individual patients (such as on a waitlist) but often do not generate values that have clinical meaning. To that end, and in addition to calculating c-statistics as measure of model discrimination, we identified candidate risk thresholds for use in clinical practice (23) based on a range of baseline PGD incidences, and generated three different models with varying definitions of low- and high-risk recipients. Standardized risks for PGD were generated in the derivation cohort for the high- and low-risk groups. These risk estimates were generated using postestimation marginalized standardized predictive probabilities from the regression equations, and represent predicted PGD risk assuming all subjects have equal distribution of other clinical covariates, with the only difference being their membership in the low- or high-risk group (24). We identified three potential models: first, a “restrictive” model in which the presence of any predictor categorized a subject as higher-risk; second, an “additive” model, in which the presence of two or more predictors categorized a subject as higher-risk; and third, a “simple” model, which required having only one predictor necessary to be higher-risk, with different variable cutoffs. Both the selection of candidate predictors and the definition of these alternative risk classifications reflected our findings based on the analysis of candidate predictors and their statistical significance.

As donor smoking was the only donor variable identified for inclusion as a potential predictor, we quantified the effect of adding a donor with smoke exposure on the predicted PGD risk over-and-above the estimated risk based on recipient variables alone. We accounted for center as a main effect in our analysis using logistic regression conditioned on center based on previous data that there was a variable incidence of PGD between centers (3). We did not perform a stratified analysis within-center because of limited numbers. In sensitivity analyses, the derivation analysis was also performed within transplant type, acknowledging inherent differences in single and bilateral transplant recipients.

In secondary analyses, we used the same set of candidate models to evaluate prediction of 30-day and 1-year mortality.

Model validation

Study population: The validation set purposely contained data from the same centers, but from a separate time period immediately following the development dataset. This approach permitted use of all available data at the time of model development and then all patients whose data accrued subsequently for the validation. Temporal validation was also chosen because other multicenter validation datasets with well-phenotyped PGD data were not available at the time of study conduct.

We did not employ multiple imputation in the validation dataset to be consistent with clinical practice where complete data on all subjects is unlikely.

Validation of PGD probability estimates: Using the predictors identified in the model development stage, we tested the predicted and observed probabilities of PGD in the separate validation dataset. We primarily focused on whether the observed PGD probabilities in the validation set fell within the confidence intervals of the predicted probabilities that were generated in the derivation set (25). We also employed decision curve analyses to compare alternative models against treating all patients as either "low risk" or "high risk" (26). Decision curves combine traditional measures of sensitivity and specificity with more formal assessment of risk thresholds at which clinicians and their patients might make decisions about risk and treatment choices. They describe (i) whether the candidate prognostic models will perform better than the alternative of classifying everyone (or no one) as high PGD risk and (ii) the net benefit of the model across a range of decision thresholds (where the decision threshold is the predicted risk of PGD from a particular model). The net benefit is a calculated number derived by subtracting the proportion of all patients who are false-positive from the proportion who are true-positive within a range of threshold probabilities (formula: $(\text{true-positive count}/n) - (\text{false-positive count}/n) * (p_t/1 - p_t)$, where p_t represents the threshold probability that would change clinical decision making) (26). Using net benefit at various probabilities allowed us to generate the decision curve in Figure 1. We chose this method as a more clinically useful way of assessing the utility of the prediction model, because it takes into account consequences of actions across risk thresholds (25). Decision curves plotting the net benefit on the y-axis versus the threshold PGD probabilities on the x axis, using the "dca" command in STATA v13.

Sample size: We estimated the number of subjects needed to validate the model, while maximizing the number of subjects in the derivation cohort to optimize power for model generation. Since the primary focus was to identify a group of low-risk patients, we focused on identifying patients whose predicted probability of PGD was 50% of our estimated baseline incidence of PGD. We considered this relative reduction in risk of PGD as clinically significant, which would be commonly accepted as a low-risk phenotype. As our observed incidence was 17% in prior studies (3), we

estimated that a validation sample of 350 with an overall risk of PGD of 8% (a 50% reduction from 17%) would have a risk upper bound of 11% (therefore, be distinct from the overall incidence of 17%).

STATA v11.2-13.0 (STATA Corp. College Station, TX) was used for all analyses.

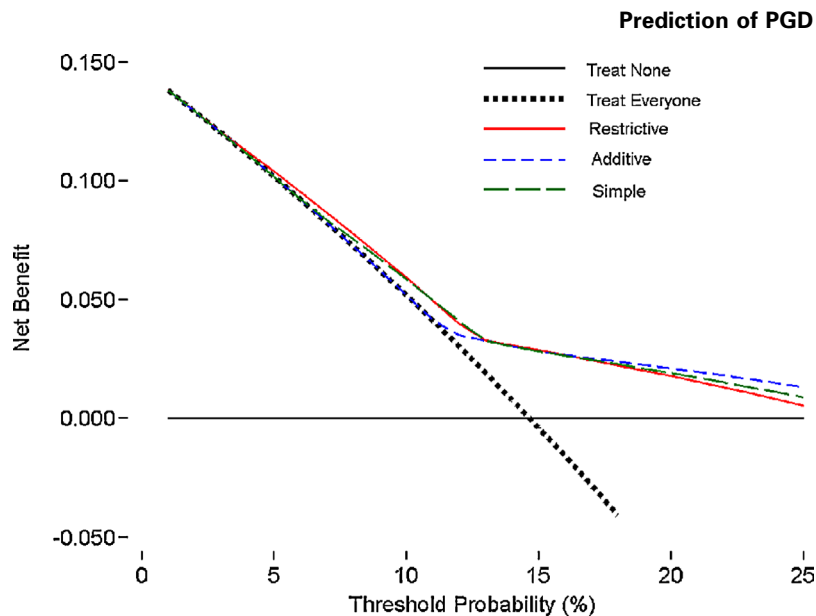
Results

Of the 1255 subjects in the derivation cohort, 211 (17%, 95% CI: 15%, 19%) developed PGD. Of the 382 subjects in the validation cohort, 56 (15%, 95% CI: 11%, 19%) developed PGD. In both cohorts, recipients with PGD were more frequently overweight or obese, had ILD as a pretransplant diagnosis, and more frequently had moderate-severe pulmonary hypertension (PH) (Table 1). Percentages of data imputed for variables in the derivation cohort are listed in Table S1.

Derivation

Table 2 presents the results of bootstrap modeling on variable selection of candidate predictors. Consideration of alternate thresholds ranging from 25% to 75% did not alter the predictor variables selected. Recipient factors with significant predictive utility included a pretransplant diagnosis that was not COPD or CF, abnormal body mass index (BMI), and elevated mean pulmonary artery pressure (mPAP, which was defined as >25 mmHg). Using these three recipient predictors, we were able to categorize recipients into "higher risk" and "low risk." We tested the three recipient prediction models based on different cutoffs using diagnosis, BMI, and PA pressure, summarized in Table 3. Table 4 displays PGD probability estimates for the three models generated. The first model, or the "restrictive model," identified 14% of recipients as low risk, which was defined as no PGD predictors, i.e. normal BMI, no pulmonary hypertension, and a diagnosis of COPD or CF. Low-risk recipients in this model had a predicted risk of PGD of 5% (95% CI: 0, 10). In the high-risk category (86% of recipients), there was a predicted PGD risk of 15% (95% CI: 13, 18%). The second, or "additive model," categorized significantly more recipients as low risk (37%). In this model, a low-risk recipient had at least one of the predictors, and had a predicted PGD risk of 7%, (95% CI: 4, 11%) and a high-risk recipient (63%) had a predicted PGD risk of 18% (14, 21%). Finally, the third model, the "simple model," identified 24% as low risk. This model also categorized a low-risk recipient as having no more than one predictor, but considered high-risk recipients to have a mPAP >40 mmHg (as opposed to 25 mmHg in the restrictive model). In this model, a low-risk recipient had a predicted incidence of PGD of 6% (95% CI: 2, 10%), and the low-risk phenotype was defined by a pretransplant diagnosis of COPD or CF, normal BMI (18.5–25), and either absent or mild pulmonary hypertension (mPAP <40 mmHg). The higher-risk recipient had a predicted PGD risk of 16% (95% CI: 13, 19%). In all models, the goodness-of-fit test indicated good calibration

Figure 1: Decision curve analysis of alternative strategies for prognostic models for primary graft dysfunction after lung transplantation. The plot compares three models against the alternatives of (a) considering everyone to be higher-risk (downward sloping dotted gray line) or (b) foregoing any prognostic modeling and treating no one as being at high risk (solid horizontal line). The y axis is net benefit (the tradeoff between true-positive and false-positive classifications for high-risk status), and the x axis represents the threshold probability for classifying a patient as being at high risk for PGD. Net benefit will vary depending on the threshold risk because that threshold probability reflects the relative loss of missing high-risk patients (false negatives) and of considering too many low-risk patients (false positives). Because the threshold might vary with individual patients and their centers, the decision curves plot net benefit against a range of thresholds. In this setting, typical threshold probabilities might range from 5% to 20%. In that range of thresholds, the proposed prognostic models are not only simple to implement but also superior in net benefit in the classification of higher-risk PGD patients than treating all patients as being at high risk.



($p > 0.05$ for all three). C-statistics were similar in the models (Table 4).

History of any smoking was the only donor risk factor that consistently predicted PGD. We therefore evaluated the impact of adding a lung from a smoking donor to recipients in the higher- and low-risk categories (Table 4). The addition of a donor with smoke exposure to a higher-risk recipient significantly increased the risk of PGD in all three models. However, the addition of a donor with smoke exposure to a low-risk recipient did not significantly increase the risk of PGD in the derivation set. Similar results were found when we evaluated the model stratified by single or bilateral transplant procedure type (Table S3).

Validation cohort

We next evaluated whether observed PGD incidences in the validation cohort were similar to those predicted from the derivation cohort (Table 4). The observed incidences of PGD in the low-risk groups ranged from 5% to 11% and fell within confidence intervals of predicted probabilities in all strata of all models. In the higher-risk groups, again the observed incidences of PGD were 13–16% and fell within predicted risk values from the derivation cohort for these strata. With the addition of donor with smoke exposure to the low-risk groups, the observed risk increased to 13–15%; however, the small number of subjects in these strata limited our ability to determine the significance of the

increase in risk. In particular, there were low numbers of subjects with moderate or severe pulmonary hypertension in the validation cohort. In the higher-risk groups, the addition of donor lungs from smokers significantly increased the PGD risk in all three models, and observed incidences in the validation cohort all fell within the predicted risks from the derivation cohort for these strata.

Next, we classified patients according to the three risk models and assessed whether use of these risk models would increase overall net benefit across a range of thresholds of PGD incidences, compared to the alternatives of defining all patients (or no patients) as high risk. As displayed in Figure 1, for expected ranges of incidences of 5–25%, use of the three prediction models demonstrated benefit over a strategy of treating everyone as either higher or low risk. As shown in Figure 1, the “restrictive model” seemed to perform better at lower PGD thresholds (as low as 5%), while the “additive model” performed best at higher PGD risk thresholds.

Mortality prediction

Given previously demonstrated association of PGD with early mortality, we tested the utility of our model for mortality prediction. Mortality information was available on 1229 of the 1255 subjects (98%). In the derivation cohort, 98 subjects died within 90 days of transplant (8%), and 167 died within 1 year of transplant (13%). Numbers were too small to evaluate the association with mortality

Table 1: Univariate analysis of donor, recipient in peri-operative variables stratified by primary graft dysfunction (PGD) status

Covariates	Derivation		Validation	
	PGD (n = 211)	Non-PGD (n = 1044)	PGD (n = 56)	Non-PGD (n = 326)
Donor variables				
Male gender, n (%)	115 (55)	646 (62)	33 (59)	217 (67)
Age, mean	35.2	34.4	37.1	34.9
Race, n (%)				
Caucasian	135 (65)	660 (64)	35 (65)	180 (55)
African American	43 (21)	214 (21)	12 (22)	83 (26)
Other	29 (14)	160 (15)	7 (13)	62 (19)
Any smoking, yes	92 (48)	335 (35)	22 (39)	64 (20)
Recipient variables				
Male gender	117 (55)	592 (57)	30 (54)	192 (59)
Age, mean	53.3	53.6	54.2	56.7
BMI category, n (%)				
<18.5	11 (9)	77 (12)	3 (5)	29 (9)
18.5–25	30 (25)	281 (43)	23 (41)	126 (39)
25–30	51 (41)	224 (34)	18 (32)	126 (39)
>30	31 (25)	79 (12)	12 (21)	45 (14)
Pulmonary diagnosis, n (%)				
COPD	56 (27)	418 (40)	14 (25)	124 (38)
Idiopathic pulmonary fibrosis	91 (43)	364 (35)	30 (54)	133 (41)
Cystic fibrosis	16 (8)	162 (16)	5 (9)	41 (13)
Sarcoidosis	17 (8)	26 (2)	1 (2)	7 (2)
PAH	12 (6)	28 (3)	5 (9)	9 (3)
Other	19 (9)	45 (4)	1 (2)	12 (4)
mPAP severity category, n (%)				
<25 mmHg (normal)	71 (41)	463 (51)	23 (41)	176 (54)
25–40 mmHg (mild)	70 (40)	380 (42)	27 (48)	122 (37)
41–55 mmHg (moderate)	22 (13)	41 (5)	5 (9)	22 (7)
>55 mmHg (severe)	11 (6)	20 (2)	1 (2)	6 (2)
Race, n (%)				
Caucasian	167 (79)	902 (86)	45 (80)	292 (90)
African American	32 (15)	84 (8)	6 (11)	22 (7)
Other	12 (6)	57 (5)	5 (9)	12 (4)

BMI, body mass index; mPAP, mean pulmonary artery pressure; COPD, chronic obstructive pulmonary disease; ILD, interstitial lung disease; PGD, primary graft dysfunction. Percentages may not exactly equal 100% because of rounding.

PGD is defined as grade 3 PGD on day 2 or 3 after lung transplantation. The distribution of variables presented is from the pre-imputation data in the derivation cohort. The validation data reflects the data used in the analysis

in the validation cohort (only 21 of 376 with mortality information died within 90 days). When using any of the three models in the derivation cohort, low-risk recipients had a lower risk of 90-day mortality than higher-risk

recipients. When we evaluated 1-year mortality, low-risk recipients again had a lower mortality than higher-risk recipients. The addition of a smoking donor did not significantly affect the predicted risk of 90-day or 1-year mortality (Table 5).

Table 2: Results of logistic regression model to evaluate candidate predictors in the recipient in each bootstrap sample of the derivation set (n = 1000)

Candidate predictor variable	% of 1000 datasets predictor p < 0.05
Transplant type	4.70
BMI category	99.40
Recipient diagnosis	99.60
Recipient gender and parity	5.70
Mean PA pressure	99.80
Donor smoking (any)	77.4
>20-pk-year history of donor smoking	13.0

Discussion

In this study, we generated three prediction models for PGD that distinguish low-risk from high-risk recipients for PGD groups based on variables that are readily available prior to transplantation. In a subsequent validation sample, we validated these PGD estimates in recipient groups and then used decision curve analysis to compare alternative models across realistic risk thresholds. We found that abnormal body weight, moderate-severe PH, or a pretransplant diagnosis other than COPD or CF determined a higher-

Table 3: Definition of each model

Model	Low risk	High risk
Restrictive	Normal BMI (BMI<25) AND Mean PAP<25 AND COPD or CF	One of the following: Abnormal BMI (BMI>25) Mean PAP>25 Any diagnosis other than COPD or CF
Additive	Either no risk factors or one of the following: Abnormal BMI OR Mean PAP >25 OR Any diagnosis other than COPD or CF	Two or more: Abnormal BMI Mean PAP >25 Any diagnosis other than COPD or CF
Simple	Normal BMI (BMI<25) AND Mean PAP <40 AND COPD or CF	One of the following: Overweight or obese OR Non-COPD or CF OR mPAP >40

COPD, chronic obstructive pulmonary disease; CF, cystic fibrosis; BMI, body mass index.

risk recipient. Furthermore, we identified history of smoking in a donor as significantly increasing the risk of PGD in higher-risk recipients. Finally, we demonstrated that our prediction models are also useful in predicting mortality based on recipient factors. Predicted probabilities from the model were validated in a separate, temporally distinct, population.

We believe our models are useful in several ways. First, these simple models can be used to identify those recipients who are at higher risk, so that clinicians can prepare patients and families for the possibility of PGD and practitioners can anticipate the need for salvage therapy, such as ECMO. Second, although the low-risk groups are relatively small, we believe the information is useful at the time of transplant, to facilitate the use of emerging

therapies aimed at expanding donors in these lower-risk recipients. Finally, we presented several models, which perform similarly, to allow each center to evaluate the model based on the risk profile of the potential recipients and their threshold of an acceptable PGD risk.

While there are many ways to approach predictive modeling, we chose to focus on identifying subgroups of recipients and generating objective estimates of PGD risk for these groups. We then validated these risk estimates in a separate population, as evidenced by the observed probabilities falling within the confidence intervals of the predicted probabilities. While the models demonstrated good fit and generated reproducible PGD probabilities, the c-statistics were not very high, likely because we chose to define recipient groups yielding two discrete categories of

Table 4: Predicted risk of PGD in derivation and observed risk in the validation cohort for 3 variation of the predictive model

Model	Derivation			Validation	
	% of 1255	Predicted risk of PGD (95% CI)	AUC (95% CI)	% of 382	Observed incidence of PGD (%)
Restrictive			0.67 (0.64, 0.71)		
Low risk	14	5 (0,10)		11	5 (0,16)
Low risk + smoking donor	5	4 (0,10)		9	13 (3,27)
High risk	86	15 (13,18)		89	16 (12,20)
High risk + smoking donor	33	24 (20,28)		20	27 (18,39)
Additive			0.64 (0.60, 0.68)		
Low risk	37	7 (4,11)		35	11 (6,18)
Low risk + smoking donor	14	11 (6,16)		9	15 (5,31)
High risk	63	18 (14,21)		65	16 (12,21)
High risk + smoking donor	24	28 (22,33)		14	32 (20,46)
Simple			0.63 (0.59, 0.67)		
Low risk	25	6 (2,10)		22	6 (2,13)
Low risk + smoking donor	10	9 (4,15)		5	15 (3,40)
High risk	75	16 (13,19)		78	13 (9,17)
High risk + smoking donor	28	26 (21,31)		17	29 (19,42)

The numbers in the derivation cohort represent the predicted risk of PGD based on each model. The numbers in the validation cohort represent the actual risk once the models were used to evaluate how frequently PGD occurred in each risk group. The observed incidences in the validation cohort are similar to the predicted incidences from the derivation cohort.

Table 5: Predicted risk of a) 90-day and b) 365-day mortality in derivation cohort

Model		Predicted risk of 90-day mortality (95% CI)
a)	Restrictive	
	Low risk	6 (1,10)
	Low risk + smoking donor	2 (0,5)
	High risk	8 (6,10)
	High risk + smoking donor	10 (7,13)
	Additive	
	Low risk	6 (3,9)
	Low risk + smoking donor	8 (4,12)
	High risk	8 (6,11)
	High risk + smoking donor	10 (6,13)
	Simple	
	Low risk	5 (2,8)
Low risk + smoking donor	6 (1,10)	
High risk	9 (7,11)	
High risk + smoking donor	10 (6,13)	
Model		Predicted risk of 1-year mortality (95% CI)
b)	Restrictive	
	Low risk	8 (3,13)
	Low risk + smoking donor	10 (2,17)
	High risk	15 (12,17)
	High risk + smoking donor	14 (11,17)
	Additive	
	Low risk	11 (8,15)
	Low risk + smoking donor	13 (8,18)
	High risk	15 (12,18)
	High risk + smoking donor	14 (10,18)
	Simple	
	Low risk	9 (6,13)
Low risk + smoking donor	9 (4,15)	
High risk	15 (12,18)	
High risk + smoking donor	16 (12,20)	

risk. Other methods, such as those based on maximizing c-statistics curves may yield outputs that, while potentially useful for ranking individual patients (such as on a wait list), do not have direct clinical meaning in estimating PGD risk for patient groups (26). In contrast, we chose to create PGD probability estimates for simple subgroups of patients, and used decision curve analysis to demonstrate the utility of the predictive models. Our models worked best at PGD incidences between 5% and 25%. At incidences below or above this range, our models do not add significant prognostic utility over treating everyone as high risk. Based on our prior data, the incidence of PGD is around 15–17%, therefore, we believe that our model will be applicable at most transplant centers.

We chose to include all three models so that the practitioners may choose which model best applies to their practice, based on various factors such as the acceptable threshold in their practice that would characterize a recipient as low or high risk and the incidence of PGD at

their center. For example, if the goal at a particular center is a PGD risk of 5%, the more restrictive model may be best suited for identifying low-risk recipients, as shown in Figure 1. At the time of allocation, the risk of PGD can be readily estimated using our model and allow for the ability to make appropriate management decisions. For example, being prepared for the use of ECMO or other salvage therapies prior to the transplant may facilitate preparation for early interventions and avoid emergent rescue of a potentially rapidly deteriorating patient. Likewise, objective PGD probability estimates may have utility in preparing recipients for the possibility of posttransplant complications. In addition, our model risk stratifies recipients to provide a framework for refinement of use of extended criteria, DCD, or *ex vivo* conditioned donor lungs.

Donor smoke exposure conferred a significant predictive impact for PGD risk in the higher-risk recipient. Prior work has established that use of lungs from donors with smoke exposure led to worse outcomes after lung transplant; however, overall waiting list mortality was increased if lungs from donors with smoke exposure were not used (27). However, in U.S. administrative datasets this effect is not consistent (28), perhaps due to the way smoke exposure is defined. All three of the models demonstrated that while addition of a donor with smoke exposure to the low-risk recipient group increased the risk of PGD; the risk increase did not reach a level that would likely favor the alternative of not being transplanted. In contrast, the use of a donor with smoke exposure in the higher-risk recipient group conferred an increased PGD risk. However, use of lungs from donors with a smoke exposure should be considered acceptable pending more accurate quantification of smoke exposure. Yet, care providers may choose to prepare for a higher risk of PGD when using donors with smoke exposure in high-risk recipients as defined by our models. Importantly, the current methods for defining and quantifying donor smoke exposure seem inadequate, as they rely on proxy reports of timing, quantity, and duration of smoke exposure. Recently, the use of biochemical measures to quantify tobacco exposure more accurately in critically ill patients has yielded promising results and may represent a method that may improve prediction in the lung donor population (29). Likewise, BMI may be an inadequate measure of adiposity, and may be better quantified by biochemical or imaging methods (10,30). In the future, these models may aid in allocation decisions, such as matching a low-risk recipient with a donor who has smoke exposure; however, improved measurement of predictor variables, as well as clarification of the mechanisms by which they increase PGD risk are necessary steps prior to changing policy (29).

The current work builds on our prior studies of PGD by employing advanced statistical methods aimed at selecting those variables with the greatest predictive utility to define subgroups of patients with lower and higher risk, and by using a separate validation population of nearly 400 new

subjects to validate the risk estimates in these groups (3,10). Many prior studies sought to identify individual independent risk factors whereas this study employed different modeling techniques to identify subgroups of patients with distinct predicted probabilities of PGD based on combined variables that are readily available and present prior to the transplant episode. In the validation population, we found that our model was better than treating everyone as either high risk or no risk. Furthermore, we evaluated donor and recipient prediction separately, and then assessed the impact of donor factors on the recipient-based model. Prior to our study, validation of PGD prediction in a multicenter prospectively-assessed population had not been attempted.

There are limitations to this study. We used a temporal validation population, meaning subjects for the validation were taken from the same centers in an ongoing prospective cohort study but transplanted at a different timepoint. However, subjects from this study were recruited from 11 centers throughout the United States; therefore, we believe there is adequate diversity in the validation cohort. Additionally, although temporal validation could introduce bias, there have been minimal changes in donor or recipient management for PGD during this time period. Furthermore, as our study is a large multicenter study, our temporal validation population represents an external population with representation of many centers; alternate multicenter external datasets with well-phenotyped PGD are not currently available. Our validation study was designed and powered to provide probability estimates in the overall population; therefore, we were unable to perform stratified analyses in the validation set (e.g. by transplant type). There was significant variation in the recruitment of patients by center, with both number and percent of subjects enrolled between centers varying which may have contributed to the differences in risk by center. However, we derived our model conditioned on center, with the goal of identifying universal factors not affected by center variations. Our multivariable bootstrap methods for variable selection included a threshold of 67%, which may have been considered restrictive, perhaps excluding individual variables that appear to have univariate associations with PGD (such as race). However, our multivariable models methods were focused on identifying stable and generalizable groups of predictor variables, and did not yield different results when this threshold was varied between 25% and 75%. We were limited in our ability to validate the effects of donor smoke exposure on low-risk recipients due to low numbers in this stratum and ensuing wide confidence intervals. In particular, there were low numbers of subjects with moderate or severe pulmonary hypertension in the validation population. We included both single and bilateral lung transplant recipients in our analysis. PGD grading in single lung transplants can potentially be biased given intrinsic v/q mismatch and shunt after transplant; however, we have published multiple prior studies using both single and bilateral transplants, and have

also presented a sensitivity analysis stratified by transplant type, which did not significantly change results.

In conclusion, we have developed and validated simple prediction models that easily categorize recipients into higher- and low-risk groups with reproducible estimated probabilities of PGD. Logical future directions include impact studies of these models in clinical practice, studies focused on improving measurement of specific predictors such as smoke exposure and obesity, and targeted intervention studies based on risk estimates generated for lower- and higher-risk subgroups.

Acknowledgments

This work was supported by the National Institutes of Health (grants HL087115 [JDC], HL081619 [JDC], HL096845 [JDC/DSW], HL115354 [JDC], HL114626 [DL/JDC], HL088263 [LBW], HL103836 [LBW]).

Disclosure

The authors of this manuscript have no conflicts of interest to disclose as described by the *American Journal of Transplantation*.

References

1. Christie JD, Bavaria JE, Palevsky HI, et al. Primary graft failure following lung transplantation. *Chest* 1998; 114: 51–60.
2. Christie JD, Edwards LB, Kucheryavaya AY, et al. The Registry of the International Society for Heart and Lung Transplantation: 29th adult lung and heart-lung transplant report-2012. *J Heart Lung Transplant* 2012; 31: 1073–1086.
3. Diamond JM, Lee JC, Kawut SM, et al. Clinical risk factors for primary graft dysfunction after lung transplantation. *Am J Respir Crit Care Med* 2013;187: 527–534.
4. Daud SA, Yusef RD, Meyers BF, et al. Impact of immediate primary lung allograft dysfunction on bronchiolitis obliterans syndrome. *Am J Respir Crit Care Med* 2006; 175: 507–513.
5. Christie JD, Kotloff RM, Ahya VN, et al. The effect of primary graft dysfunction on survival after lung transplantation. *Am J Respir Crit Care Med* 2005; 171: 1312–1316.
6. Covarrubias M, Ware LB, Kawut SM, et al. Plasma intercellular adhesion molecule-1 and von Willebrand factor in primary graft dysfunction after lung transplantation. *Am J Transplant* 2007; 7: 2573–2578.
7. Christie JD, Shah CV, Kawut SM, et al. Plasma levels of receptor for advanced glycation end products, blood transfusion, and risk of primary graft dysfunction. *Am J Respir Crit Care Med* 2009; 180: 1010–1015.
8. Christie JD, Robinson N, Ware LB, et al. Association of protein C and type 1 plasminogen activator inhibitor with primary graft dysfunction. *Am J Respir Crit Care Med* 1899; 175: 69–74.
9. Diamond JM, Kawut SM, Lederer DJ, et al. Elevated plasma Clara cell secretory protein concentration is associated with high-grade primary graft dysfunction. *Am J Transplant* 2011; 11: 561–567.
10. Lederer DJ, Kawut SM, Wickersham N, et al. Obesity and primary graft dysfunction after lung transplantation: The Lung Transplant

- Outcomes Group Obesity Study. *Am J Respir Crit Care Med* 2011; 184: 1055–1061.
11. Pires-Neto RC, Morales MMB, Lancas T, et al. Expression of acute-phase cytokines, surfactant proteins, and epithelial apoptosis in small airways of human acute respiratory distress syndrome. *J Crit Care* 2013; 28.
 12. Fang A, Studer S, Kawut SM, et al. Elevated pulmonary artery pressure is a risk factor for primary graft dysfunction following lung transplantation for idiopathic pulmonary fibrosis. *Chest* 2011; 139: 782–787.
 13. Christie JD, Kotloff RM, Pochettino A, et al. Clinical risk factors for primary graft failure following lung transplantation. *Chest* 2003; 124: 1232–1241.
 14. Christie JD, Carby M, Bag R, et al. Report of the ISHLT Working Group on Primary Lung Graft Dysfunction part II: Definition. A consensus statement of the International Society for Heart and Lung Transplantation. *J Heart Lung Transplant* 2005; 24: 1454–1459.
 15. Christie JD, Bellamy S, Ware LB, et al. Construct validity of the definition of primary graft dysfunction after lung transplantation. *J Heart Lung Transplant* 2010; 29: 1231–1239.
 16. Shah RJ, Diamond JM, Cantu E, et al. Latent class analysis identifies distinct phenotypes of primary graft dysfunction after lung transplantation. *Chest* 2013; 144: 616–622.
 17. Obesity: Preventing and managing the global epidemic. Report of a WHO consultation World Health Organization technical report series 2000; 894: 1–253.
 18. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007. 16: 219–242.
 19. Moons KG, Donders RA, Stijnen T, Harrell FE, Jr. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 2006; 59: 1092–1101.
 20. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* 2011; 30: 377–399.
 21. Diamond JM, Lee JC, Kawut SM, et al. Clinical risk factors for primary graft dysfunction after lung transplantation. *Am J Respir Crit Care Med* 2013; 1: 527–534.
 22. Efron B, Tibshirani R. Improvements on Cross-validation: The 632+ bootstrap method. *J Am Stat Assoc* 1997; 92: 548–560.
 23. Greenland S. The need for reorientation toward cost-effective prediction: Comments on “Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond” by M. J. Pencina et al, *Statistics in Medicine* (DOI: 10.1002/sim.2929). *Stat Med* 2008; 27: 199–206.
 24. Graubard BI, Korn EL. Predictive margins with survey data. *Biometrics* 1999; 55: 652–659.
 25. Hosmer DW, Lemeshow S. *Applied logistic regression*. 3rd ed. Malden, MA: Wiley-Interscience; 2000.
 26. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making* 2006; 26: 565–574.
 27. Bonser RS, Taylor R, Collett D, et al. Effect of donor smoking on survival after lung transplantation: A cohort study of a prospective registry. *Lancet* 2012; 380: 747–755.
 28. Shigemura N, Toyoda Y, Bhama JK, et al. Donor smoking history and age in lung transplantation: A revisit. *Transplantation* 2013; 95: 513–518.
 29. Calfee CS, Matthay MA, Eisner MD, et al. Active and passive cigarette smoking and acute lung injury after severe blunt trauma. *Am J Respir Crit Care Med* 2011; 183: 1660–1665.
 30. Lederer DJ, Wilt JS, D’Ovidio F, et al. Obesity and underweight are associated with an increased risk of death after lung transplantation. *Am J Respir Crit Care Med* 2009; 180: 887–895.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Table S1: Percentage of missing data for donor and recipient variables in the derivation cohort.

Table S2a: Results of the prediction model in bilateral lung transplants in the derivation cohort.

Table S2b: Results of the prediction model in single lung transplants in the derivation cohort.