**Exploring A New Framework For US Heart Allocation: Development Of A Continuous Medical Urgency Score For Adult Heart Transplant Candidates**

Grace R. Lyden, PhD1,2 ([Grace.Lyden@cdrg.org](mailto:Grace.Lyden@cdrg.org); ORCID 0000-0003-1408-851X)

Molly White, MS3

Monica M. Colvin, MD1,4

TBA: Heart committee chair and members

Kiran Khush, MD

Sommer E. Gentry, PhD1,5 ([Sommer.Gentry@nyulangone.org](mailto:Sommer.Gentry@nyulangone.org); ORCID 0000-0003-4530-8917)

William F. Parker, MD, MS, PhD3,6

1Scientific Registry of Transplant Recipients, Hennepin Healthcare Research Institute, Minneapolis, Minnesota, USA

2Department of Medicine, University of Minnesota, Minneapolis, Minnesota, USA

3Department of Medicine, University of Chicago, Chicago, Illinois, USA

4Department of Cardiology, University of Michigan, Ann Arbor, MI, USA

5Department of Surgery, New York University Grossman School of Medicine, New York, New York, USA

6Department of Public Health Sciences, University of Chicago, Chicago, Illinois, USA

***Correspondence:*** Grace R. Lyden, PhD

Email: [grace.lyden@cdrg.org](mailto:grace.lyden@cdrg.org)

**Abbreviations**

AIC – Akaike information criterion

AUC – area under the (ROC) curve

API – aortic pulsatility index

BiVAD – biventricular assist device

BNP – brain natriuretic peptide

CI – confidence interval

CPO – cardiac power output

ECMO – extracorporeal membrane oxygenation

HRSA – Health Resources and Services Administration

IQR – interquartile range

LVAD – left ventricular assist device

MCS – mechanical circulatory support

OPTN – Organ Procurement and Transplantation Network

PAPi – pulmonary artery pulsatility index

SRTR – Scientific Registry of Transplant Recipients

ROC – receiver operating characteristic

Word count (abstract xx plus main text xx): xx

Tables: x, Figures: x

**Abstract**

Donor hearts in the United States are allocated to adult transplant candidates according to a categorical 6-status system based primarily on treatment. The US heart transplant community has long advocated for a lab-based risk score to replace or supplement the current system, due to its limited risk stratification and susceptibility to gaming. In March 2024, the OPTN Heart Committee heard one proposal for a continuous US candidate risk score (US-CRS) based on a 6-week predicted probability of mortality without transplant. While the committee supported the concept of a continuous risk score, several members raised concerns about the underlying model, including questions about real-world performance, the impact of LVADs on lab values, and potential bias from informative censoring. This study represents a collaboration with the Heart Committee to respond to those concerns. We propose a new model, US-CRS 2.0, in an updated cohort of adult heart candidates listed 2019-2022 from the Scientific Registry of Transplant Recipients. This model expands on the US-CRS 1.0 with flexible effects for time on LVAD, hemodynamic variables, and interactions of lab values with LVAD use; inverse probability weights adjust for informative censoring at transplant. In cross-sections of candidates competing with each other for transplant in 2023, the US-CRS 2.0 had good calibration and superior discrimination to the 6-status system. The US-CRS 2.0 is a new medical urgency model for adult heart-failure patients that could be used to determine candidate risk level in the allocation of donor hearts.

**1 Introduction**

Donor hearts are exceedingly scarce relative to the need for heart transplants. One in five adult heart candidates listed in 2020 died or was removed without transplant within three years of listing.1 To minimize waitlist deaths, the Final Rule in the United States mandates that organ allocation policies rank-order candidates by medical urgency based on “objective and measurable medical criteria.”2 Current US heart allocation policy uses a categorical system, implemented in October 2018, which rank-orders adult candidates from Status 1 (most urgent) to Status 6 (least urgent).3 Because these statuses are based primarily on treatment and device use, one criticism of the 6-status system is its susceptibility to gaming and potential impacts on physician behavior, as exemplified by the increased use of balloon pumps (Status 2) in heart candidates after the 2018 policy.4,5 Another criticism is the limited risk stratification achieved by 6 discrete categories, which may contribute to the rising use of “exception” requests for candidates who do meet their status by standard criteria.6 As the US transitions to a continuous distribution system of organ allocation, without “hard boundaries” between discrete candidate groups and geographies, a continuous measure of medical urgency for adult heart candidates could be more consistent with that framework and its goals.7,8

The US candidate risk score (US-CRS) has been proposed as a continuous alternative to the 6-status system for measuring medical urgency in adult heart candidates. The US-CRS is a 50-point score based on a 6-week predicted probability of mortality without transplant, from a discrete-time survival model. The score includes 7 patient variables: 5 lab values (albumin, bilirubin, eGFR, B-type natriuretic peptide, sodium) and 2 cardiac device variables (short-term MCS, durable LVAD). In a held-out test set, the US-CRS had superior discrimination compared to the 6-status system and was not outperformed by the French CRS or machine learning approaches.9

In March 2024, the US-CRS was presented to the Heart Committee of the Organ Procurement and Transplantation Network (OPTN) as a possible method for assigning medical urgency points under continuous distribution. While many committee members were receptive to the idea of an objective risk score and acknowledged that the heart transplant community had been desiring one, the committee had some concerns about the model itself and potential unintended consequences of its use in policy. These included questions about real-world model performance, the impact of devices on lab values, and whether the variables in the US-CRS were sufficient to capture mortality risk without transplant. Ultimately, the Heart Committee opted to continue with their proposed medical urgency score based on the 6-status system for the first iteration of continuous distribution, while pursuing further modeling of a continuous risk score to replace or supplement the status system in the future.10

This study represents a collaboration between the US-CRS authors, the Scientific Registry of Transplant Recipients (SRTR), and the OPTN Heart Committee to resolve the committee’s concerns and propose a real-world, policy-ready “US-CRS 2.0” for consideration in allocation policy for adult heart candidates.

Our aim is to introduce and validate a new mortality model that expands on and differs from the US-CRS 1.0 in several key aspects: First, the new model is based on an updated SRTR cohort of all adult heart candidates ever waiting from 2019 through 2022, with a held-out test set of candidates on the list in 2023. Second, we consider the 7 US-CRS variables, but also time on LVAD, hemodynamics, and interactions of the lab values with durable LVAD use, to address the committee’s concern that LVADs would alter the association of clinical labs with mortality. In the new model, a prevalent cohort, longer cohort length and use of a composite outcome increase our statistical power to detect these effects, as well as nonlinear effects. Third, we evaluate real-world performance by comparing cross-sections of candidates who were simultaneously waiting and therefore competing with each other for donor hearts.11 We compare the performance of our new model with the 6-status system including waiting time tiebreakers, which have a significant impact on the order of the match run under current policy. Finally, we use time-varying inverse probability weights (ie, a “marginal structural model”) to adjust for informative censoring at transplant, which may have attenuated the effect sizes in US-CRS 1.0 for predictors highly associated with both transplant and mortality, such as short-term MCS.

**2 Materials and Methods**

**2.1 Study population**

This study used data from SRTR. The SRTR system includes data on all donors, waitlisted candidates, and transplant recipients in the United States, submitted by the members of the OPTN, and has been described elsewhere.12 The Health Resources and Services Administration (HRSA), US Department of Health and Human Services, provides oversight to the activities of the OPTN and SRTR contractors. Work performed by SRTR is exempt from institutional review board review.

This retrospective observational study included a prevalent cohort of all adult (listed at age 18+) heart transplant candidates who were waiting on any day from January 1, 2019, through December 31, 2023. Heart-lung candidates were excluded. Data were split into model training and test sets by calendar time, with 2019-2022 data comprising the training set and 2023 data comprising the test set, to estimate model performance in future heart candidates. Candidates in the training set were censored on December 31, 2022. With this method of data splitting, some patients are included in both the training and test set, but the actual data are nonoverlapping between the training and test sets. We used the Quarter 1 2024 Standard Analysis Files, which include all national data updates through March 2, 2024.

**2.2 Primary endpoint and clinical variables**

The primary endpoint was waitlist mortality, defined as a composite outcome of death or waitlist removal for deteriorated condition within 6 weeks. Deaths after waitlist removal were identified using the Social Security Limited-Access Death Master File and OPTN-verified deaths.13 The 6-week timeframe was chosen to identify patients at short-term risk of death without transplant, similar to the Model for End-Stage Liver Disease.14 The majority (about 60%) of adult heart candidates receive a transplant within three months of listing.1

Available clinical variables included: time-varying medical urgency status under the current system (from Status 6, least urgent, to Status 1, most urgent), serum albumin (g/dL), serum bilirubin (mg/dL), serum or plasma creatinine (mg/dL), serum or plasma sodium (mEq/L), and brain natriuretic peptide (BNP, pg/mL) as measured by a BNP or N-terminal B-type natriuretic peptide pro (NT BNP Pro) test. Estimated glomerular filtration rate (eGFR) was calculated by the 2021 CKD-EPI creatinine equation using creatinine, birth sex, and age at listing.15 Cardiac device variables included durable LVAD use (reported on justification forms) and short-term mechanical circulatory support (MCS), defined as ECMO (at listing or to qualify for Status 1 or 3), temporary surgical LVAD (to qualify for Status 2), or temporary surgical BiVAD (to qualify for Status 1). Hemodynamic variables included cardiac output (L/min), systolic and diastolic blood pressure, central venous pressure, pulmonary artery systolic and diastolic pressure, and pulmonary capillary wedge pressure, all in mmHg. These were used to compute the cardiac power output16 (CPO), aortic pulsatility index17 (APi), and pulmonary artery pulsatility index18 (PAPi). Transplant programs also have to specify whether hemodynamic data were obtained while the patient was on inotrope and/or device support.

**2.3 Marginal structural model of mortality without transplant**

Our goal was to predict 6-week survival without a transplant, for adult heart candidates at any point after listing, based on their disease state at that point in time.

Therefore, similar to US-CRS 1.0, we constructed a landmark survival dataset with a multiple follow-up periods per candidate. For each candidate, a new landmark was generated every 14 days after listing, if the candidate was still waiting, and candidate characteristics were recorded at that time. Most of the urgency statuses under the current system are valid for 14 days, so this represents a standard schedule for data updates.19 After each landmark, candidates were followed until 1) death or removal for deteriorated condition, 2) transplant, or 3) 6 weeks later. Candidates were not censored at removal for other reasons or inactivation, so as to capture deaths shortly after removal and while inactive. All landmarks that started prior to January 1, 2019, were excluded. An example data setup is shown in the Supplement. Unlike previous models that have restricted to one follow-up period per candidate starting at listing, this analytic approach takes advantage of the rich time-varying data collected on all adult heart candidates starting in October 2018, and the resulting predictions represent a weighted average across all landmark timepoints. This aligns with how a heart medical urgency score would be used in practice, with updates over time as a candidate’s condition changes.

We censor at transplant to estimate a “cause-specific hazard” of waitlist mortality, which is appropriate for a medical urgency score.20 That is, we want to quantify a patient’s probability of death if transplant were unavailable, not the probability of death given their access to transplant under current allocation. One challenge for this type of waitlist mortality modeling is that patients who are censored due to transplant tend to be sicker than those who are still waiting, if allocation was based on medical urgency. This is a form of informative censoring and can lead to model coefficients that are biased toward the null.21 To mitigate this informative censoring, we fit a marginal structural model to the landmark dataset described above. Marginal structural models use inverse probability weighting to reweight the distribution of time-varying confounders at each timepoint when a candidate might receive transplant, so that the censored data represent the full population of heart candidates in a hypothetical world where transplant is unavailable.22,23 The weights are based on the probability of receiving a transplant each week, given that a candidate is still alive and waiting (ie, the cause-specific hazard of transplant). Details on the calculation of the weights are in the next subsection.

Marginal structural models can be estimated by any statistical method that allows for time-varying weights, including both regression and machine learning models. We chose a discrete-time survival regression model, ie, a pooled logistic regression, for ease of implementation in OPTN policy and because machine learning methods have not been shown to improve predictions of waitlist mortality in adult heart candidates.9,24,25

As predictors in the mortality model, we included all 7 US-CRS variables: albumin, natural-log-transformed bilirubin, eGFR, natural-log-transformed BNP, sodium, any use of durable LVAD since listing, and any use of short-term MCS since listing. An interaction allowed for different effects of BNP depending on the test performed (BNP or BNP NT Pro). Because our modeling dataset had a larger sample size than the original US-CRS modeling dataset, we also considered: 1) hemodynamic variables, specifically CPO, API, and PAPi; 2) interactions of each US-CRS lab value with LVAD use; and 3) years on LVAD, from earliest reported LVAD implant date to landmark start date, which was truncated at a maximum value of 10 years. We considered a categorical effect for week from landmark, from week 1 to week 6, similar to the baseline hazard in a Cox proportional hazards model. For all other predictors, we used the candidate’s most recent nonmissing value on the landmark start date. Finally, we considered nonlinear effects, specified as natural cubic splines with two interior knots. The final model was chosen by forward selection to minimize the Akaike information criterion (AIC), using the original US-CRS model with linear effects as a starting point.

Implausible values (defined in Supplemental Table 1) were set to missing, and all missing values were replaced by the sample median (see Supplemental Table 2), similar to how missing values are replaced by single imputation in allocation policy. Continuous lab and hemodynamic variables were trimmed to their 1st and 99th percentiles for model fitting and prediction to minimize the impact of extreme values.

Model coefficients were summarized by adjusted cause-specific hazard ratios (aHRs) with robust 95% confidence intervals (CIs) that accounted for candidates contributing multiple landmarks to the dataset. Coefficients were used to calculate the 1-week cause-specific hazard of waitlist mortality, which was converted to a 6-week probability of survival without transplant by the formula: where is the hazard of death in week for a patient with characteristics .

**2.4 Inverse probability weights**

The inverse probability weights of a marginal structural model are based on the probability of transplant, which has to be estimated by a separate model where transplant is the outcome. To fully remove bias caused by informative censoring, this transplant model should include all time-varying confounders, ie, patient variables that predict both heart transplant and waitlist mortality in the observed data. For confounders, we included all variables in US-CRS 1.0 and also considered models with: medical urgency status, hemodynamic variables, and waiting time. We fit a pooled logistic regression from listing to removal in the entire cohort, without landmarks, updating confounders at the start of each week. After removal, the probability of transplant was taken to be 0. We then compared covariate balance between transplant recipients and candidates who were still waiting, weighting by the four possible transplant models (US-CRS, US-CRS + status, US-CRS + status + hemodynamics, and all variables), and chose the model with the best covariate balance.26,27 We applied the chosen model to calculate inverse probability weights in the landmark dataset, as described elsewhere.22 Although landmarks were every two weeks, weights were updated each week to account for transplants each week and the subsequent shifts in the distribution of candidates still waiting.

The “unstabilized” weights with a numerator of 1 can have high variance. We also considered “stabilized” weights where the numerator is based on the probability of transplant given candidate characteristics on the landmark start date. Finally, we considered weights that were both stabilized and truncated at the 1st and 99th percentile each week. We chose the final mortality model by comparing coefficient estimates with these three approaches and summarizing distributions of the weights each week, which should have a mean of 1.28

**2.5 Predictive performance**

Medical urgency allocation points are meant to stratify candidates by need for heart transplant and should be able to discriminate risk between candidates waiting at the same time, as these candidates are in direct “competition” with one another for donor hearts. Therefore, to assess performance of our medical urgency model, we divided the test data into “cross-sections” based on eight calendar dates in 2023 that were each approximately 6 weeks apart. A candidate was selected into a cross-section if they were active and waiting on the cross-section start date, and clinical values were determined at that time. We then compared the model-predicted probability of mortality within 6 weeks of the cross-section start date with the actual pretransplant mortality within 6 weeks, including deaths after removal and while inactive. To assess discrimination, we produced a time-dependent cumulative/dynamic receiver operating characteristic (ROC) curve for 6-week mortality in each cross-section of test data and calculated the area under the curve (AUC).29 For comparison, we also produced ROC curves in each cross-section by medical urgency status and status plus waiting time tiebreakers, which is how candidates are ordered on the match run under current allocation policy.19

To assess the calibration of the new mortality model, we pooled the eight cross-sections of test data. We then rank-ordered candidates by predicted 6-week mortality and divided them into 20 equal-sized groups. In each group, we calculated the average predicted 6-week mortality and the observed 6-week mortality. Observed mortality was estimated by the Kaplan-Meier method, with time-varying inverse probability weights to adjust for informative censoring and robust standard errors. We also compared average predicted mortality with observed mortality in each medical urgency status.

**3 Results**

**3.1 Cohort summary**

23,889 adult heart candidates were eligible for analysis: 18,877 in the 2019-2022 training set and 7,218 in the 2023 test set, with 2,206 candidates included in both datasets (Table 1). The majority of candidates were male (74%) and 37% started at Status 4, with 24% on a durable LVAD and 5% on short-term MCS at baseline. Baseline distributions were similar in the training and test data. In the time-varying dataset by week, there were also no major differences in clinical values or device use between the training and test sets (Supplemental Table 3), and very few candidates had implausible values (<1%). Missing data were minimal (<5%) for most lab values in the model training data but higher for BNP (15% missing) and systolic and diastolic blood pressure (both about 20% missing).

Of the 18,877 candidates in the model training data, there were 1,688 waitlist mortalities: 923 removals due to deteriorated condition and 765 deaths.

**3.2 Marginal structural model of mortality without transplant**

Most landmarks in the modeling dataset started at or shortly after listing, but there were also landmark start dates as late as 5+ years after listing (Supplemental Figure 1). In the unweighted data, there was substantial covariate imbalance between transplant candidates and recipients, with recipients having a poorer prognosis by lab values and hemodynamics, less use of durable LVAD, more use of short-term MCS, and more-urgent status (Supplemental Table 4). The optimal covariate balance was attained with weights adjusting for the US-CRS 1.0 variables, urgency status, and hemodynamic variables (Supplemental Table 7). We compared unstabilized, stabilized, and truncated weights in a simple mortality model with linear effects (Supplemental Figure 2). In general, effects were larger in the weighted models, relative to an unweighted model. The stabilized weights had the best properties in terms of bias-variance tradeoff and distribution (Supplemental Table 9). Therefore, stabilized weights are used for the remainder of this analysis.

The final mortality model chosen by AIC included the following: week from landmark (baseline hazard), albumin, log bilirubin, eGFR, log BNP (with differing effects for BNP versus BNP NT Pro), sodium, any use of durable LVAD since listing, any use of short-term MCS since listing, years on LVAD, API, PAPi, and two-way interactions of durable LVAD use with log bilirubin, eGFR, and log BNP. A nonlinear effect was selected for all continuous variables. Durable LVAD use was associated with lower risk (aHR [95% CI]: .27 (.17, .41) for a candidate with average lab values), while short-term MCS was associated with higher risk (aHR [95% CI]: 4.59 [3.16, 6.66]). The effects of the lab values by whether the candidate had a durable LVAD are shown in Figure 1. Among patients without a durable LVAD (purple solid lines), higher BNP values and higher bilirubin values were associated with a higher risk of mortality; these associations were attenuated or near null for LVAD patients (yellow dashed lines). By contrast, lower eGFR was predictive of a poor outcome in LVAD patients, but this association was attenuated in non-LVAD patients. Lower albumin and lower sodium were associated with higher risk regardless of LVAD status. Hemodynamic effects are shown in Supplemental Figure 3, and all coefficients are available in Supplemental Table 11.

To illustrate the impact of longer time on LVAD, we calculated the predicted 6-week survival probability for a typical LVAD candidate by time since LVAD implant and compared this probability with 6-week survival for Status 3 and 4 candidates (Figure 2). Including all landmarks, LVAD patients in the training data had a median (IQR) of 2.1 (1.2, 3.4) years on LVAD. Predicted survival began to decrease after about a year on LVAD, approaching survival for a Status 3 candidate; after six years, predicted survival increases, in possibly a survivor effect. On an absolute scale, this was a probability range of about 1 percentage point.

Figure 3 shows the distribution of predicted 6-week survival probabilities in the training data by medical urgency status. The majority of Status 1 candidates had predicted 6-week survival of 95% of lower, while survival was mostly above 95% for candidates at Status 2-6. For Status 1-4, the predicted probabilities were generally concordant with the status system, with lower survival for more-urgent status, but there was substantial overlap in predicted probabilities especially for Status 2, 3, and 6. Some Status 1 candidates had high predicted survival (almost 100%), and some Status 6 candidates had very low predicted survival (down to 50%).

**3.3 Predictive performance: Calibration and discrimination**

Time-dependent AUC for the US-CRS 2.0 model ranged from 0.78 (95% CI: 0.66, 0.91) to 0.87 (95% CI: 0.81, 0.93) across the eight cross-sections of test data, compared to a range of 0.69 (95% CI: 0.56, 0.83) to 0.80 (95% CI 0.71, 0.89) for the current 6-status system of risk stratification (Figure 4, Supplemental Figure 4). The AUC was statistically significantly higher for US-CRS 2.0 compared to the current system in three of the eight cross-sections (February, April, August), indicating better risk stratification of 6-week mortality. There were essentially no differences in AUC between the 6-status system versus 6-status plus waiting time tiebreakers.

Pooling all cross-sections together, our model predictions were generally well calibrated, with very small differences between predicted and observed mortality (Figure 5). In the top 5% of candidates by predicted risk, the model underestimated 6-week mortality with an average prediction of 11% (IQR of predictions: 6%, 13%) compared to an observed mortality of 14% (95% CI: 10%, 18%). By medical urgency status, model predictions were generally well calibrated, though mortality was somewhat underestimated for Status 2 candidates and somewhat overestimated for Status 6 candidates (Figure 6).

**4 Discussion**

Points to hit:

* Marginal structural models have been used for proposed medical urgency scores in liver allocation11,30,31 but not previously for heart.
* How this is different from US-CRS 1.0:
  + Bigger sample size, due to longer calendar period since 2018 policy and prevalent (not incident) cohort. Also more events due to composite outcome.
  + Expands US-CRS 1.0 with nonlinear effects, interactions of lab values with LVAD use, hemodynamics, and time on LVAD.
  + We fit a marginal structural model, a causal inference method that is designed to reduce the impact on coefficient estimates of informative censoring due to sicker candidates receiving transplant.

Other thoughts:

* Although LVAD time had a small effect on an absolute scale, its selection into the model suggests that time on LVAD does improve mortality prediction and that not all LVAD patients have the same urgency. These results may support including LVAD time in the urgency attribute of the CAS on top of providing points for time on LVAD as an equity attribute.
* If the natural cubic splines are too complex for coding up in policy, could recreate with linear splines.

**Acknowledgments / Funding**

This work was conducted under the auspices of the Hennepin Healthcare Research Institute (HHRI), contractor for the Scientific Registry of Transplant Recipients (SRTR), under contract no. 75R60220C00011 (US Department of Health and Human Services, Health Resources and Services Administration, Health Systems Bureau, Division of Transplantation). The US Government (and others acting on its behalf) retains a paid-up, nonexclusive, irrevocable, worldwide license for all works produced under the SRTR contract, and to reproduce them, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The data reported here have been supplied by HHRI as the contractor for SRTR. The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by SRTR or the US Government. The authors thank SRTR colleague Anna Gillette for manuscript editing.

**Disclosure**

Dr. Gentry is supported by NIH R01DK132395. The other authors have no conflicts of interest to disclose as described by the *American Journal of Transplantation*.

**Data Availability Statement**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**References**

1. Colvin MM, Smith JM, Ahn YS, et al. OPTN/SRTR 2023 Annual Data Report: Heart. *American Journal of Transplantation*. 2025;25(2):S329-S421. doi:10.1016/j.ajt.2025.01.024

2. US Department of Health and Human Services. *Organ Procurement and Transplantation Network: Final Rule*. https://www.ecfr.gov/current/title-42/chapter-I/subchapter-K/part-121; 1998.

3. Organ Procurement and Transplantation Network. Modify adult heart allocation 2016 2nd round. https://optn.transplant.hrsa.gov/policies-bylaws/public-comment/modify-adult-heart-allocation-2016-2nd-round/. Published 2018. Accessed June 3, 2025.

4. Parker WF, Chung K, Anderson AS, Siegler M, Huang ES, Churpek MM. Practice Changes at U.S. Transplant Centers After the New Adult Heart Allocation Policy. *J Am Coll Cardiol*. 2020;75(23):2906-2916. doi:10.1016/j.jacc.2020.01.066

5. Golbus JR, Gupta K, Colvin M, et al. Changes in Type of Temporary Mechanical Support Device Use Under the New Heart Allocation Policy. *Circulation*. 2020;142(16):1602-1604. doi:10.1161/CIRCULATIONAHA.120.048844

6. Golbus JR, Ahn YS, Lyden GR, et al. Use of exception status listing and related outcomes during two heart allocation policy periods. *Journal of Heart and Lung Transplantation*. May 2023. doi:10.1016/j.healun.2023.05.004

7. Stewart D. Moving Toward Continuous Organ Distribution. *Curr Transplant Rep*. 2021;8(4):301-313. doi:10.1007/s40472-021-00352-z

8. Kasiske BL, Pyke J, Snyder JJ. Continuous distribution as an organ allocation framework. *Curr Opin Organ Transplant*. 2020;25(2):115-121. doi:10.1097/MOT.0000000000000733

9. Zhang KC, Narang N, Jasseron C, et al. Development and Validation of a Risk Score Predicting Death Without Transplant in Adult Heart Transplant Candidates. *JAMA*. 2024;331(6):500. doi:10.1001/jama.2023.27029

10. Organ Procurement and Transplantation Network. *OPTN Heart Transplantation Committee Meeting Summary 3/29/2024*. Houston, Texas; 2024. https://optn.transplant.hrsa.gov/media/pfpcrzc4/20240329\_heart\_committee-meeting-summary.pdf.

11. Schaubel DE, Guidinger MK, Biggins SW, et al. Survival benefit-based deceased-donor liver allocation. *American Journal of Transplantation*. 2009;9(4 PART 2):970-981. doi:10.1111/j.1600-6143.2009.02571.x

12. Leppke S, Leighton T, Zaun D, et al. Scientific Registry of Transplant Recipients: Collecting, analyzing, and reporting data on transplantation in the United States. *Transplant Rev*. 2013;27(2):50-56. doi:10.1016/j.trre.2013.01.002

13. Noreen SM, Patzer RE, Mohan S, et al. Augmenting the United States transplant registry with external mortality data: A moving target ripe for further improvement. *American Journal of Transplantation*. 2024;24(2):190-212. doi:10.1016/j.ajt.2023.09.002

14. Kim WR, Mannalithara A, Heimbach JK, et al. MELD 3.0: The Model for End-Stage Liver Disease Updated for the Modern Era. *Gastroenterology*. 2021;161(6):1887-1895.e4. doi:10.1053/j.gastro.2021.08.050

15. Inker LA, Eneanya ND, Coresh J, et al. New Creatinine- and Cystatin C–Based Equations to Estimate GFR without Race. *New England Journal of Medicine*. 2021;385(19):1737-1749. doi:10.1056/NEJMoa2102953

16. Fincke R, Hochman JS, Lowe AM, et al. Cardiac power is the strongest hemodynamic correlate of mortality in cardiogenic shock: A report from the SHOCK trial registry. *J Am Coll Cardiol*. 2004;44(2):340-348. doi:10.1016/j.jacc.2004.03.060

17. Belkin MN, Alenghat FJ, Besser SA, et al. Aortic pulsatility index predicts clinical outcomes in heart failure: a sub‐analysis of the ESCAPE trial. *ESC Heart Fail*. 2021;8(2):1522-1530. doi:10.1002/ehf2.13246

18. Lim HS, Gustafsson F. Pulmonary artery pulsatility index: physiological basis and clinical application. *Eur J Heart Fail*. 2020;22(1):32-38. doi:10.1002/ejhf.1679

19. Organ Procurement and Transplantation Network. OPTN policies. https://optn.transplant.hrsa.gov/media/eavh5bf3/optn\_policies.pdf. Published March 27, 2025. Accessed June 3, 2025.

20. Lyden GR, Johnson DY, Snyder JJ, Golbus JR, Parker WF. Best practices for statistical analysis of pretransplant medical urgency. *The Journal of Heart and Lung Transplantation*. 2024;43(3):523-526. doi:10.1016/j.healun.2023.11.012

21. Hernán MA, Hernández-Díaz S, Robins JM. A Structural Approach to Selection Bias. *Epidemiology*. 2004;15(5):615-625. doi:10.1097/01.ede.0000135174.63482.43

22. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11(5):561-570.

23. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol*. 2006;98(3):237-242. doi:10.1111/j.1742-7843.2006.pto\_329.x

24. Suresh K, Severn C, Ghosh D. Survival prediction models: an introduction to discrete-time modeling. *BMC Med Res Methodol*. 2022;22(1):1-18. doi:10.1186/s12874-022-01679-6

25. Craig E, Zhong C, Tibshirani R. A review of survival stacking: a method to cast survival regression analysis as a classification problem. *Int J Biostat*. March 2025. doi:10.1515/ijb-2022-0055

26. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679. doi:10.1002/sim.6607

27. Lyden GR, Vock DM, Helgeson ES, Finger EB, Matas AJ, Snyder JJ. Transportability of Causal Inference under Random Dynamic Treatment Regimes for Kidney–Pancreas Transplantation. *Biometrics*. 2023;79(4):3165-3178. doi:10.1111/biom.13899

28. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656-664. doi:10.1093/aje/kwn164

29. Blanche P, Dartigues JF, Jacqmin-Gadda H. Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical Journal*. 2013;55(5):687-704. doi:10.1002/bimj.201200045

30. Keogh RH, Van Geloven N. Prediction under Interventions: Evaluation of Counterfactual Performance Using Longitudinal Observational Data. *Epidemiology*. 2024;35(3):329-339. doi:10.1097/EDE.0000000000001713

31. Gong Q, Schaubel DE. Partly conditional estimation of the effect of a time-dependent factor in the presence of dependent censoring. *Biometrics*. 2013;69(2):338-347. doi:10.1111/biom.12023

**Figure Legends**

Figure 1: Adjusted cause-specific hazard ratios for lab values (albumin, eGFR, bilirubin, BNP NT Pro, BNP, sodium) by durable LVAD use. The outcome is death or removal for deteriorated condition within a week.

Figure 2: Model-predicted probability of 6-week survival without transplant by time on LVAD, for a durable LVAD patient with average lab values, compared to an average Status 3 and average Status 4 candidate.

Figure 3: Distribution of model-predicted probability of 6-week survival without transplant by medical urgency status (1-6) for 2019-2022 candidates. The left panel zooms in on the range from 95% to 100% survival.

Figure 4: Time-dependent area under the ROC curve for 6-week mortality in test data, for the US-CRS 2.0 model compared to the 6-status system and 6-status with waiting time tiebreakers. Nonoverlapping cross-sections were generated every 6 weeks; each cross-section included all candidates active on that start date. ROC, receiver operating characteristic.

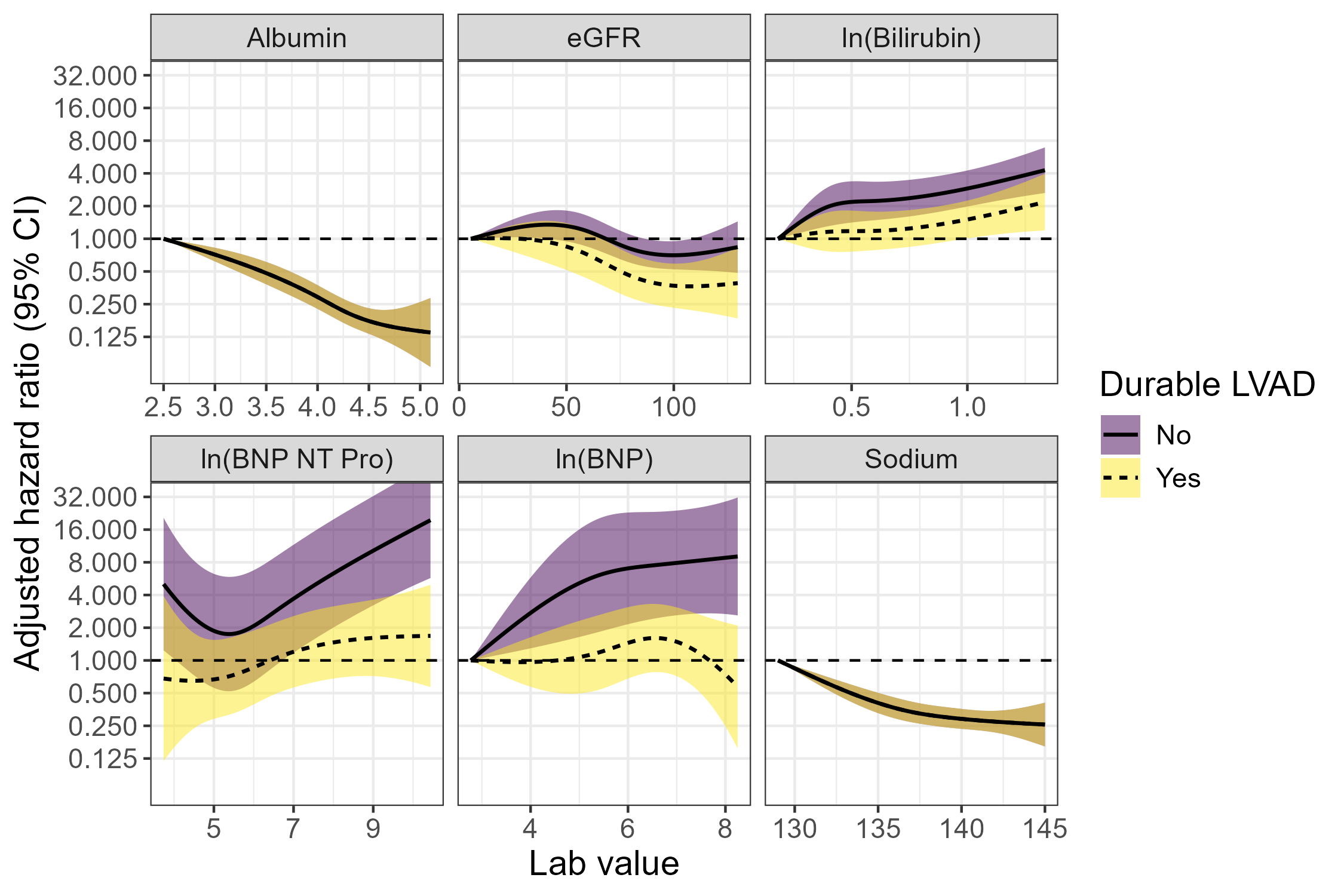
Figure 5: Calibration of US-CRS 2.0 predicted 6-week mortality across all cross-sections of test data. Data were rank-ordered by predicted risk and divided into 20 groups of equal size. The left panel zooms in on the range from 0% to 5% mortality.

Figure 6: Calibration of US-CRS 2.0 predicted 6-week mortality across all cross-sections of test data, by medical urgency status (1-6). Error bars show the 95% confidence interval for observed mortality and the interquartile range for predicted mortality.

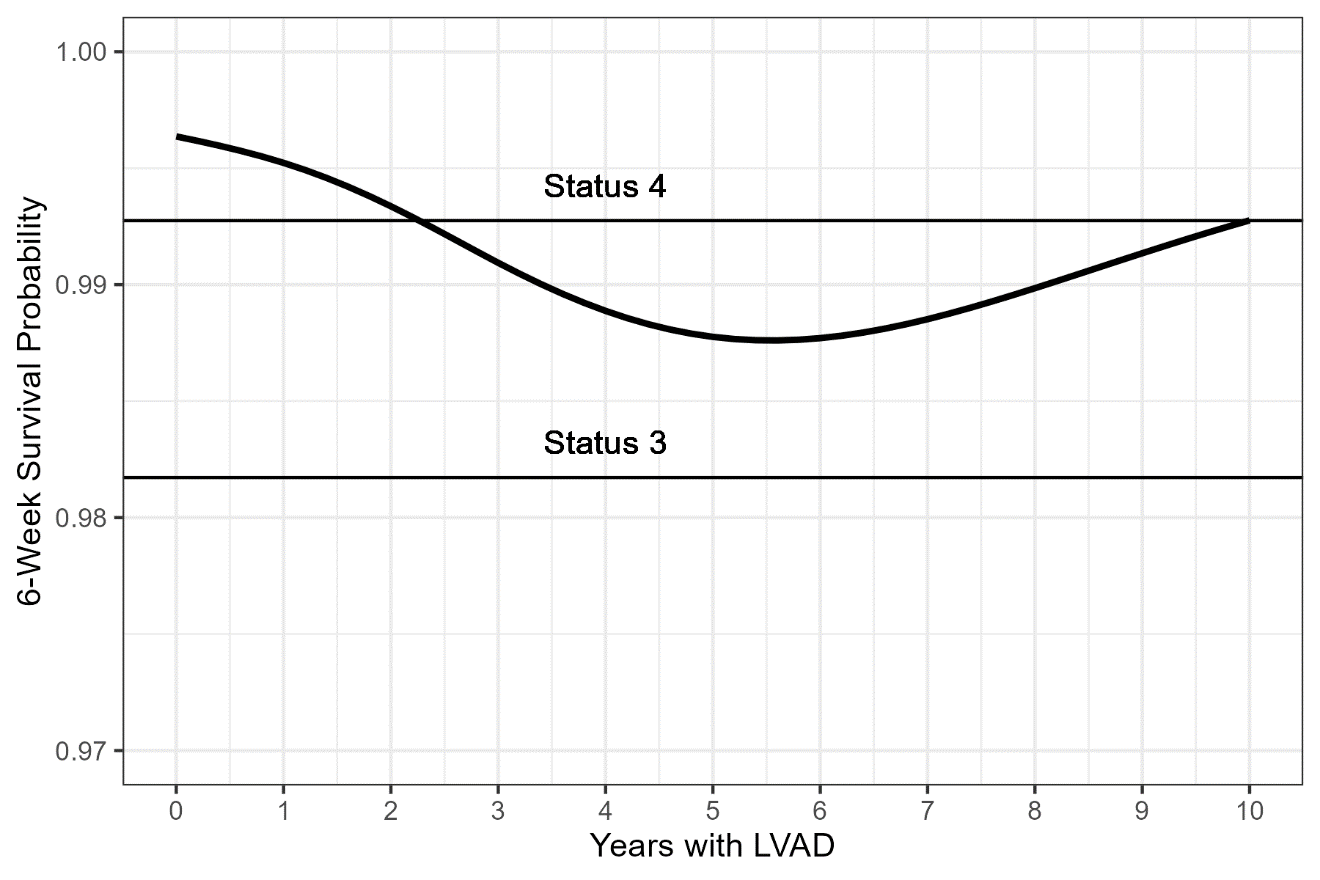
**Table 1:** Baseline characteristics of adult heart candidates, by model training (2019-2022) and test (2023) data split. The medical urgency status and cardiac device use are the candidate’s initial values in each calendar period.

|  | **Training (N=18877)** | **Test (N=7218)** | **Overall (N=23889)** |
| --- | --- | --- | --- |
| **Age at listing (years)** |  |  |  |
| Mean (SD) | 53.1 (12.9) | 52.8 (12.8) | 53.1 (12.9) |
| Median [Min, Max] | 56.1 [18.0, 79.3] | 55.8 [18.0, 76.9] | 56.2 [18.0, 79.3] |
| **Birth sex** |  |  |  |
| F | 4898 (25.9%) | 1784 (24.7%) | 6215 (26.0%) |
| M | 13979 (74.1%) | 5434 (75.3%) | 17674 (74.0%) |
| **Medical urgency status** |  |  |  |
| Status 1 | 807 (4.3%) | 315 (4.4%) | 1120 (4.7%) |
| Status 2 | 3853 (20.4%) | 1619 (22.4%) | 5447 (22.8%) |
| Status 3 | 1892 (10.0%) | 587 (8.1%) | 2374 (9.9%) |
| Status 4 | 7197 (38.1%) | 2679 (37.1%) | 8795 (36.8%) |
| Status 5 | 599 (3.2%) | 266 (3.7%) | 783 (3.3%) |
| Status 6 | 3469 (18.4%) | 1177 (16.3%) | 4207 (17.6%) |
| Inactive | 1057 (5.6%) | 573 (7.9%) | 1160 (4.9%) |
| Missing | 3 (0.0%) | 2 (0.0%) | 3 (0.0%) |
| **Ever on durable LVAD** |  |  |  |
| No | 13955 (73.9%) | 5302 (73.5%) | 18125 (75.9%) |
| Yes | 4922 (26.1%) | 1916 (26.5%) | 5764 (24.1%) |
| **Ever on short-term MCS** |  |  |  |
| No | 18041 (95.6%) | 6904 (95.6%) | 22759 (95.3%) |
| Yes | 836 (4.4%) | 314 (4.4%) | 1130 (4.7%) |

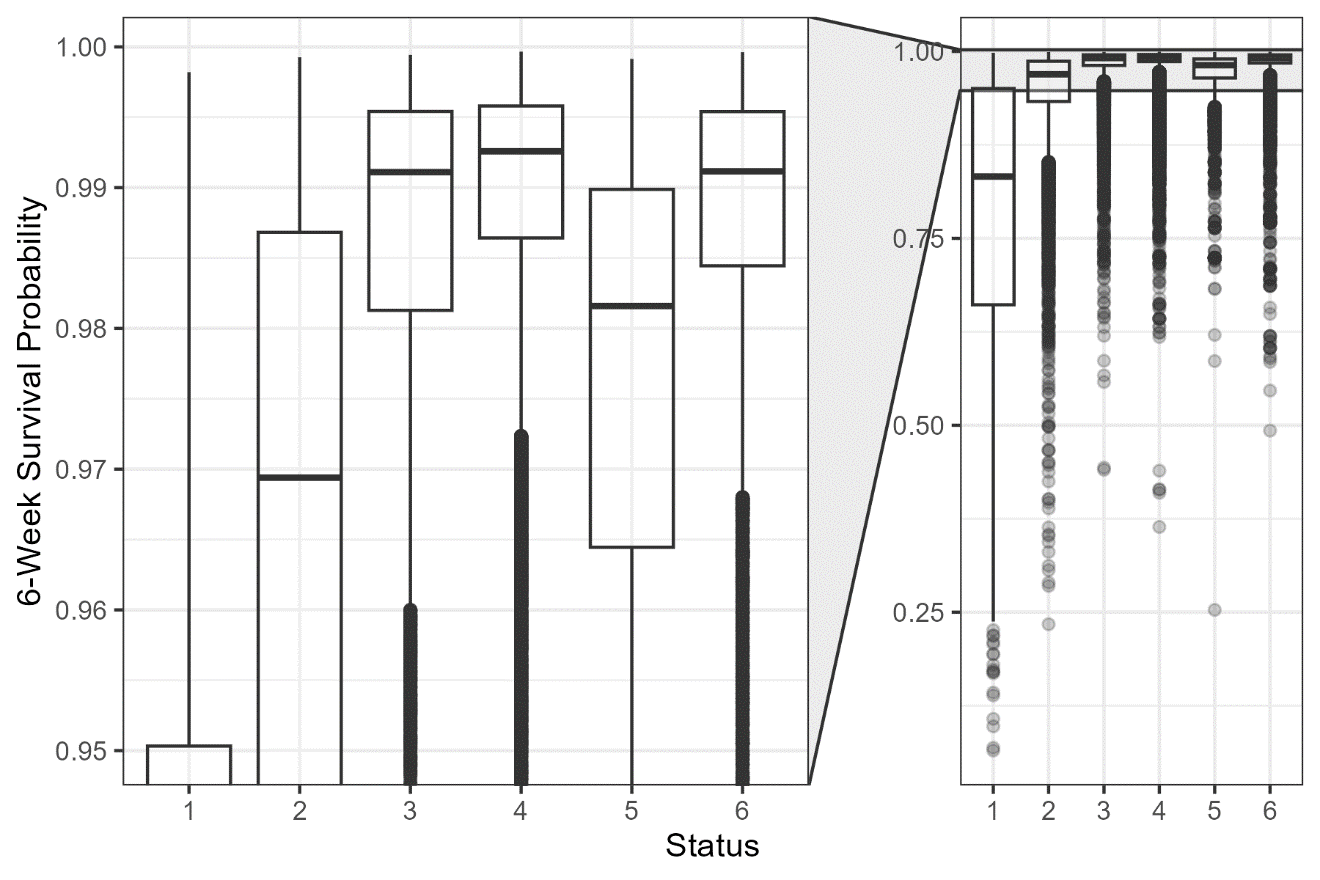
Abbreviations: LVAD, left ventricular assist device; MCS, mechanical circulatory support.

**Figure 1:** Adjusted cause-specific hazard ratios for lab values (albumin, eGFR, bilirubin, BNP NT Pro, BNP, sodium) by durable LVAD use. The outcome is death or removal for deteriorated condition within a week. 

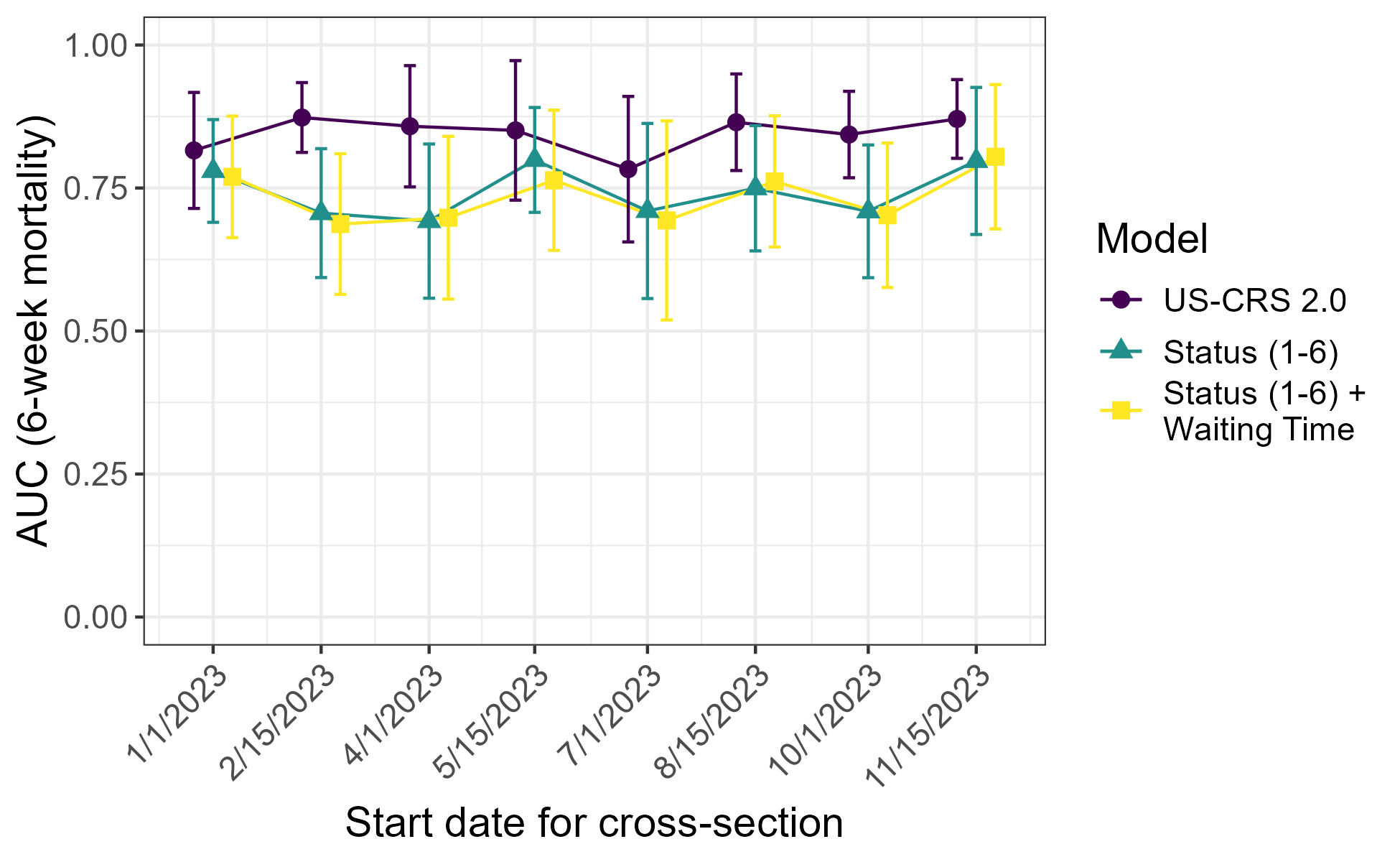
**Figure 2:** Model-predicted probability of 6-week survival without transplant by time on LVAD, for a durable LVAD patient with average lab values, compared to an average Status 3 and average Status 4 candidate.



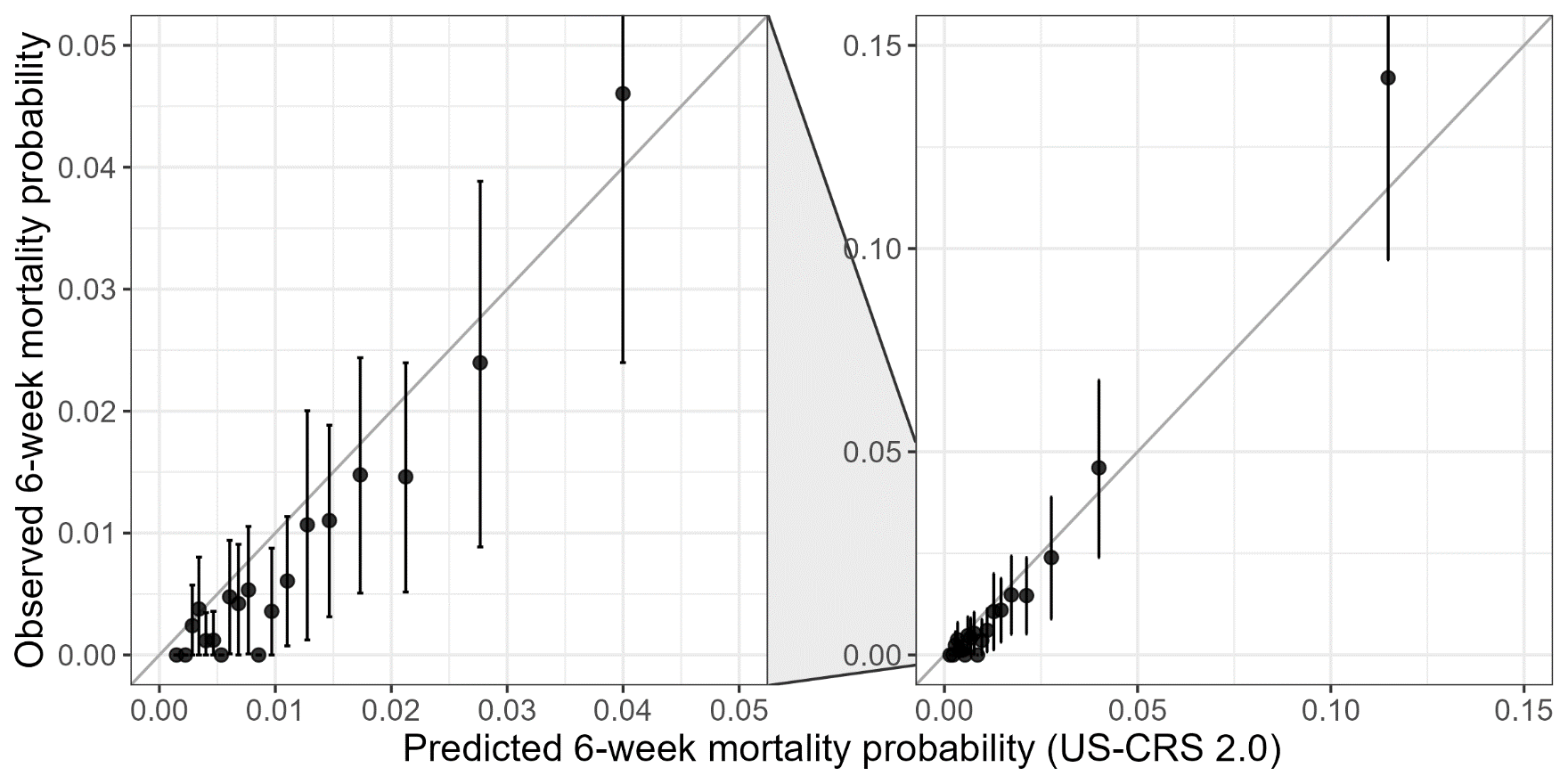
**Figure 3:** Distribution of model-predicted probability of 6-week survival without transplant by medical urgency status (1-6) for 2019-2022 candidates. The left panel zooms in on the range from 95% to 100% survival.



**Figure 4:** Time-dependent area under the ROC curve for 6-week mortality in test data, for the US-CRS 2.0 model compared to the 6-status system and 6-status with waiting time tiebreakers. Nonoverlapping cross-sections were generated every 6 weeks; each cross-section included all candidates active on that start date. ROC, receiver operating characteristic.



**Figure 5:** Calibration of US-CRS 2.0 predicted 6-week mortality across all cross-sections of test data. Data were rank-ordered by predicted risk and divided into 20 groups of equal size. The left panel zooms in on the range from 0% to 5% mortality.



**Figure 6:** Calibration of US-CRS 2.0 predicted 6-week mortality across all cross-sections of test data, by medical urgency status (1-6). Error bars show the 95% confidence interval for observed mortality and the interquartile range for predicted mortality.