Identifying clinical phenotypes of heart transplant candidates using machine learning

Final Report for Unsupervised Machine Learning MACS 40800

Authors: Lilly Reich, William Parker

Contribution: Both authors contributed equally to the project conception, data analysis, and writing of the manuscript.

Introduction:

Heart transplantation is a life-saving treatment for end-stage heart failure, a devastating disease which kills over 250,000 Americans each year.[1] Unfortunately, the supply of deceased donor hearts cannot meet demand, and over a third of candidates will die or be delisted without transplant.[2] In the context of such scarcity, the Department of Health and Human Services charges the Organ Procurement and Transplant Network (OPTN) to make the "best use" of scarce deceased donor hearts by ranking candidates from "most to least medically urgent."[3] In contrast to other organ transplant systems, there is currently no objective score used to rank candidates on the heart waiting list. Instead, each candidate's priority for transplantation is based on "Status," a varied designation determined by the supportive therapy prescribed by their transplant center. It has been previously shown that some heart transplant centers appear to overtreat candidates with intensive therapies at far higher rates than other centers,[4–6] presumably in the effort to manipulate the waitlist priority of their candidates. Therefore, there is clear need for an objective system that precisely identifies the candidates who benefit the most from heart transplant.

One of the major barriers to an objective score-based allocation system in heart transplantation are the complicated interactions between the support therapy that keeps the candidate alive while waiting for a transplant and the important clinical variables (e.g. laboratory measurements, test results, and exam findings like patient functional status). There are

physiologic reasons to suspect that each treatment will have a varying non-linear effect on the patient's clinical state and that treatment effects are likely to be modified by other clinical variables. For example, cardiac output (the quantity of blood pumped by the heart in a minute) should be increased by differing degrees based on the intensity of the support treatments (which range from medications to mechanical pumps). But based on other physiologic measurements (filling pressures in the heart), these treatments may be more or less effective at increasing cardiac output. By using all standard survival analysis regression based approaches, and considering pre-transplant support treatments as fixed characteristics of a candidate, previous attempts[7–9] to develop objective prediction models for heart transplant candidates have therefore excluded active treatment that modifies other candidate variables. Inconveniently, none of these models has been adopted for policy purposes due to these shortcomings.

In this project, we applied unsupervised machine learning clustering techniques to identify natural clinical phenotypes of heart transplant candidates. Our hypothesis was that clustering approaches could better define clinically important groups than the therapy-based Status designation. We evaluated the performance of our clusters relative to the Status categories using the association of each group with survival on the waitlist.

Methods

Data source and study population

We used data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR is a complete national registry that includes all donors, wait-listed candidates, and transplant recipients in the United States. Death records are supplemented from a link to the linked Social

Security Death Master File to capture the outcomes of candidates who were delisted alive. The accuracy of these data are confirmed with continuous audits by Medicaid/Medicare Services (CMS) and only low-levels of missingness are tolerated.[10,11]

The study population is all adult heart transplant candidates listed for heart transplantation in the United States since 2000. The number of variables measured by SRTR dataset expanded considerably in 2000[3] including much greater detail about the type of treatments. Pediatric candidates will be excluded initially as they are subject to separate policy rules and are a minute portion of the transplant population. Due to memory constraints when calculating the distance matrix, we took a random sample of 1,000 candidates listed during the study time period. We also excluded candidates who were listed in "inactive" status who were unable to immediately receive a transplant.

Feature selection and weighting- Designing a "Survival-weighted" feature space

The SRTR dataset contains over 50 clinically important candidate variables such as hemodynamic measurements, laboratory data, demographics, and type of therapeutic support (including details on exact type of mechanical circulatory support devices) (**Table 1)**. These variables are recorded at initial listing and then dynamically updated as the candidate's medical condition evolves on the waitlist. If the candidate receives a transplant, all variables are re-measured prior to transplantation.

While many of these variables have already been shown to be strong independent predictors of mortality in both heart transplant candidates and recipients,[12–14] many likely represent clinically unimportant noise. Therefore, we constructed a novel "survival-weighted" feature space to select clinically important variables and weigh them by their relevance. We first standardized continuous variables by dividing them by their range (after dropping top and

bottom 5% outlier values that are medically unfeasible and likely data entry errors). We then ran a cox proportional hazards model on the standardized dataset with all candidate variables as predictors

$$h(t) = h_o(t)\exp(\boldsymbol{X\beta})$$

We then extracted the beta coefficients and their significance levels. Variables with a coefficient that did not meet statistical significance by the Wald test were dropped from the list of features. For variables with statistically significant coefficients, we constructed a weighted Gower's distance with the absolute value of the coefficient as the weight.

$$d_{gower} = \frac{\sum_k s_{pqk} * |\beta_k|}{\sum_k |\beta_k|}$$

Where if the variable is continuous

$$s_{pqk} = \frac{p_k - q_k}{\max(k) - \min(k)}$$

And if the variable is categorical

$$s_{pqk} \begin{cases} 1 \; if \; p_k = q_k \\ 0 \; if \; p_k \neq q_k \end{cases}$$

Clustering Methodology:

We performed visual assessment of tendencies via an Ordered Dissimilarity Image (ODI) of the distance matrix with both the survival weighted feature space and an unweighted Gower's feature space. We then performed hard partitioning clustering using k-mediods (partitioning around medoids) in order to ensure our centroids represented real patients and observable values of categorical variables. We *a priori* chose the number of clusters to be 3 to match the number of priority statuses in the heart allocation system. However, we examined the choice of cluster number by calculating the average silhouette width by cluster number.

Cluster characteristics and association with survival

For each cluster, we calculated descriptive statistics for each of the feature variables. We also determined the relationship between the fitted clusters and the listing Status of each candidate. Finally, we performed a survival analysis to visualize the success of our clustering by estimating each clusters survival with the Kaplan-Meir Survival estimator.

Results:

Patient Sample and survival analysis to generate weights

There were 52,960 adult heart-only candidates listed during the study time period. The cox proportional hazards model was fit to this entire sample with all available candidate variables (**Table 2**). The significant variables were age, body mass index, treatment (ECMO, IABP, high-dose inotropes, low-dose inotropes, other MCS), renal function, functional status, diabetes, cardiac diagnosis, private insurance, and race. The largest weights were for renal function, ECMO support, and functional status.

Clustering diagnostics

A random sample of 10,000 patients was selected for clustering due to memory constraints. The ODI of the Gower distance matrix for both the "survival weighted" feature space and the unweighted gower's distance space are shown in **Figure 1**. On visual inspection, neither plot has clear visual clusters, however the unweighted gower's distance matrix has more homogeneity.

Optimal cluster number K

The plot of average silhouette width by cluster number is shown in **Figure 2.** The minimum silhouette width was at k =5, significantly lower than 3 groups (number used in the heart allocation system).

Cluster features

The distribution of various features by cluster are displayed in **Figure 3-5** and **Table 3**. In the survival weighted feature space, the cluster 1 was strictly composed of patients with "severe impairment" in functional status (**Figure 3a**). In comparison, without survival weights, cluster 1 was distributed across the functional status levels (**Figure 3b**). Correlating with the worse functional status, Cluster 1 candidates had higher pulmonary capillary wedge pressure, suggestive of worse clinical status. However, cluster 1 candidates were younger (cluster 3 had the oldest candidates) and had better renal function. The survival weighted clusters were distributed across all 3 statuses (**Figure 4**). The more intense treatments (ECMO, high-dose inotropes, IABP) tended to be found mostly in cluster 1 **(Figure 5)**.

Association of clusters with survival

There was a strong association between cluster membership and survival **(Figure 6A**). Cluster 1 (young but with severe impairment in functional status) candidates had the lowest survival, followed by cluster 3 (old with better functional status). Cluster 2 had the best survival. In comparison with Status 1A, 1B, and 2 (**Figure 6B**), the clusters produced better separation between the KM estimates and identified a larger higher risk group (more cluster 1 than Status

1A). **Figure 6C** shows the survival of the unweighted clusters. While the groups are still significantly different, the differences between the groups are much smaller. **Figure 6D** shows the survival weighted clusters with k =5 (suggested optimum by the silhouette width). The added clusters do not identify groups with meaningfully different survival.

Discussion:

Our clustering approaches identified clinically important phenotypes of heart transplant candidates and cluster membership was associated with the clinically important outcome of like survival on the waitlist. Arguably our clusters were superior to the three Status levels used in allocation with greater separation between the groups in estimated survival and larger group size for the highest risk group (cluster 1). Interestingly, the cluster 1 candidates are not simply the sickest, i.e. with the worst possible combination of all the clinically relevant variables. While they have some serious risk factors for death on waitlist, these factors are balanced by lower risk profiles in other variables. This fits well with clinical practice where a patient with too many "strikes" against them will not get listed for transplant in the first place. For example, a severely ill patient who truly requires invasive support like ECMO to survive needs to be young and have normal renal function to be considered for listing by the center.

Limitations:

Our choice of weights for feature space was somewhat arbitrary. If we re-fit the cox model after dropping the non-significant features, the weights would likely change meaningfully. A more elegant approach would be to use a penalized regression model to more efficiently shape the feature space in one step. Furthermore, weighing by the association of variables with a

clinically important outcome begs the questions of why not simply fit a supervised learner to directly predict the outcome that you care about. It is hard to imagine that our approach is superior to a supervised ML method for creating distinct clinical groups, this would have to be proved in a rigorous comparison study.

The fact that k =5 yielded a lower average silhouette width but the new clusters did not have significantly different survival is interesting. Why the improved clusterability in this group did not translate to new groups with significantly different survival is not clear

Finally, as shown by the ODI plot, our weighting approach did not create an obviously "clusterable" space. instead of weighing by an outcome, the feature space could be weighted to maximize clusterability. If a measure of clusterability can be written analytically as function of the data and the weight vector, then a derivative (analytic or numeric) could be calculated and stochastic gradient descent used to find optimal weights with respect to the measure of clusterability (and a given k). Some preliminary work in this area shows this might be a promising technique.[15-16]

.

Table 1: Candidate variables available in the SRTR dataset

Demographics	Medical History and comorbidities	Clinical and Laboratory	Therapeutic support history
Age	Primary Cardiac Diagnosis	Invasive cardiac hemodynamics (cardiac index, pulmonary capillary wedge pressure, central venous saturation, etc.)	Durable mechanical circulatory support
Sex	Cardiac surgical history		Inotropes (including dosage)
Race	Hypertension		Intra-aortic balloon pump
Education	End-stage renal disease		Temporary Mechanical circulatory support
Work history	Diabetes	Glomerular filtration rate	Extra-corporeal membrane oxygenation
Insurance	Vascular disease	Bilirubin	Total Artificial Heart
Height, Weight, BMI	Medication history	Albumin	Mechanical Ventilation
ABO blood type	Substance abuse history	Functional status	Implantable defibrillator

Table 2: Cox proportional hazards model output. Model was estimated with the outcome of death on the waitlist and used to generate survival weights for the gower's distance matrix.

	Dependent variable:
	Time
Age	0.163***
	(0.064, 0.261)
Bmi	-0.247***
	(-0.326, -0.169)
Ecmo	1.350***
	(1.147, 1.552)
Female	0.003
	(-0.049, 0.056)
high_dose_inotropes	0.427***
	(0.346, 0.509)
Iabp	0.362***
	(0.266, 0.459)
low_dose_inotropes	0.161***
	(0.099, 0.222)
Lvad	0.042
	(-0.028, 0.112)
other_mcs	0.354***
	(0.255, 0.454)
cardiac_index	0.039
	(-0.044, 0.122)
eGFR	-1.115***
	(-1.208, -1.023)
Pcwp	0.519***
	(0.437, 0.600)
blood_typeAB	0.132*
	(-0.004, 0.269)
blood_typeB	-0.055
	(-0.131, 0.021)
blood_typeO	0.022
	(-0.026, 0.070)
diabetesNon-diabetic	-0.144***
	(-0.192, -0.096)
diabetesUnknown	-0.047
	(-0.442, 0.347)
functional_statusModerate Impairment, 50-60%	0.081***
	(0.024, 0.138)
functional_statusSevere Impairment ≥ 40%%	0.461***
	(0.398, 0.525)
functional_statusUnknown	0.425***
	(0.349, 0.501)
simple_diagnosisIschemic cardiomyopathy	0.056**

	(0.003, 0.109)
simple_diagnosisOther	0.289***
	(0.218, 0.361)
simple_diagnosisRestrictive cardiomyopathy	0.132***
	(0.049, 0.215)
payorMedicare	0.035
	(-0.043, 0.113)
payorOther	-0.049
	(-0.170, 0.071)
payorPrivate	-0.188***
	(-0.261, -0.115)
raceHispanic	-0.080
	(-0.177, 0.017)
raceOther	-0.071
	(-0.208, 0.066)
raceWhite	-0.089***
	(-0.146, -0.032)
Observations	41,911
R_2	0.058
Max. Possible R_2	0.974
Log Likelihood	-75,091.590
Wald Test	2,783.900*** (df = 29)
LR Test	2,507.248*** (df = 29)
Score (Logrank) Test	2,999.934*** (df = 29)
Note:	*p<0.1;* ***p<0.05;*** p<0.01

Table 3: Means of continuous variables by cluster (survival weighted)

cluster	*Age (years)*	*Estimated Glomerular Filtration Rate*	*Pulmonary Capillary Wedge Pressure*	*Body Mass Index*
1	51.8	67.2	21.4	27.2
2	51.6	65.6	19.2	27.4
3	54.0	64.9	18.9	27.7

Figure 1a: Ordered Dissimilarity Index (ODI) plots of the weighted feature space

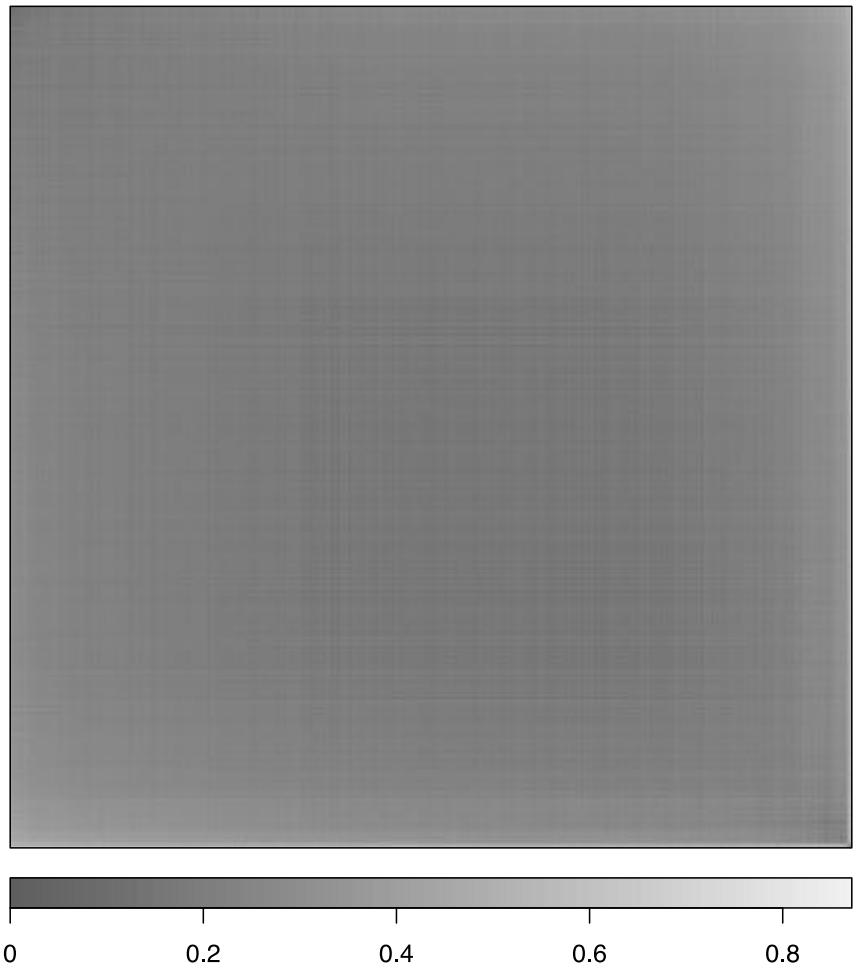


Figure 1b: Ordered Dissimilarity Index (ODI) plots of the unweighted feature space

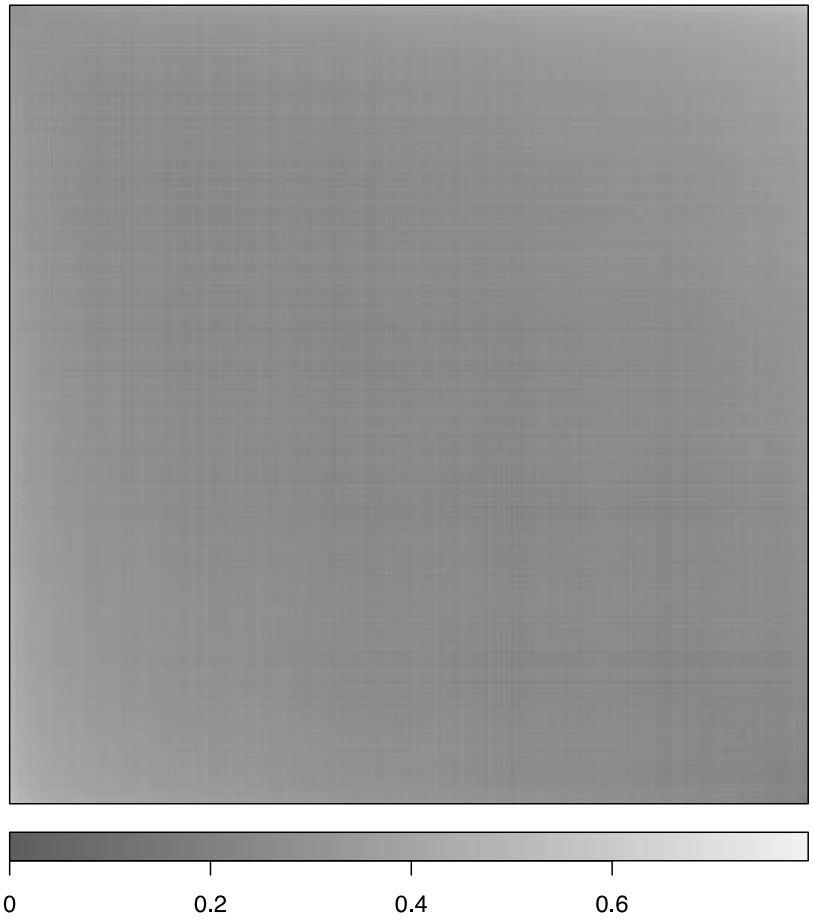


Figure 2: Average silhouette width by number of clusters

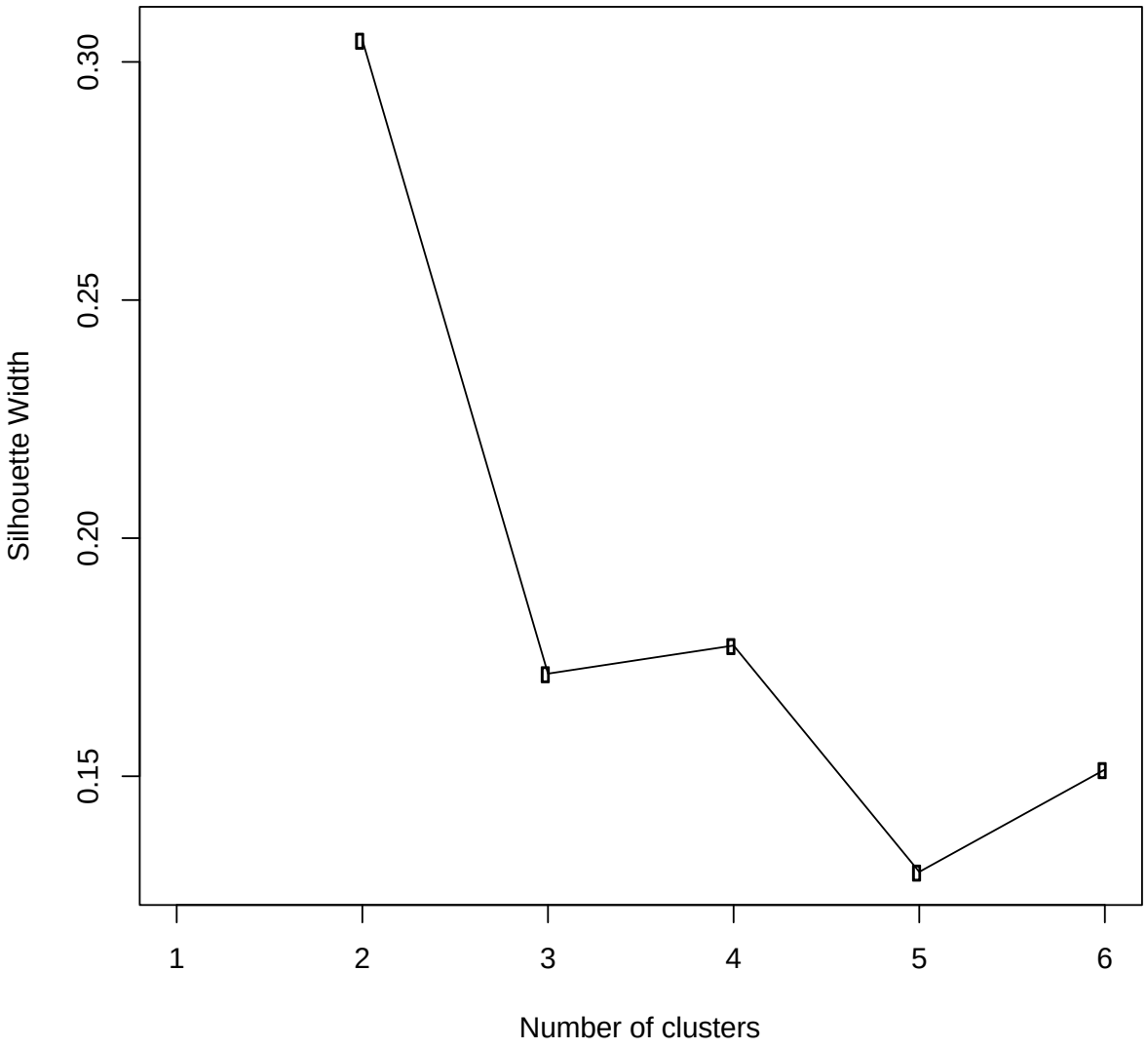


Figure 3a: Functional Status by Cluster with weights

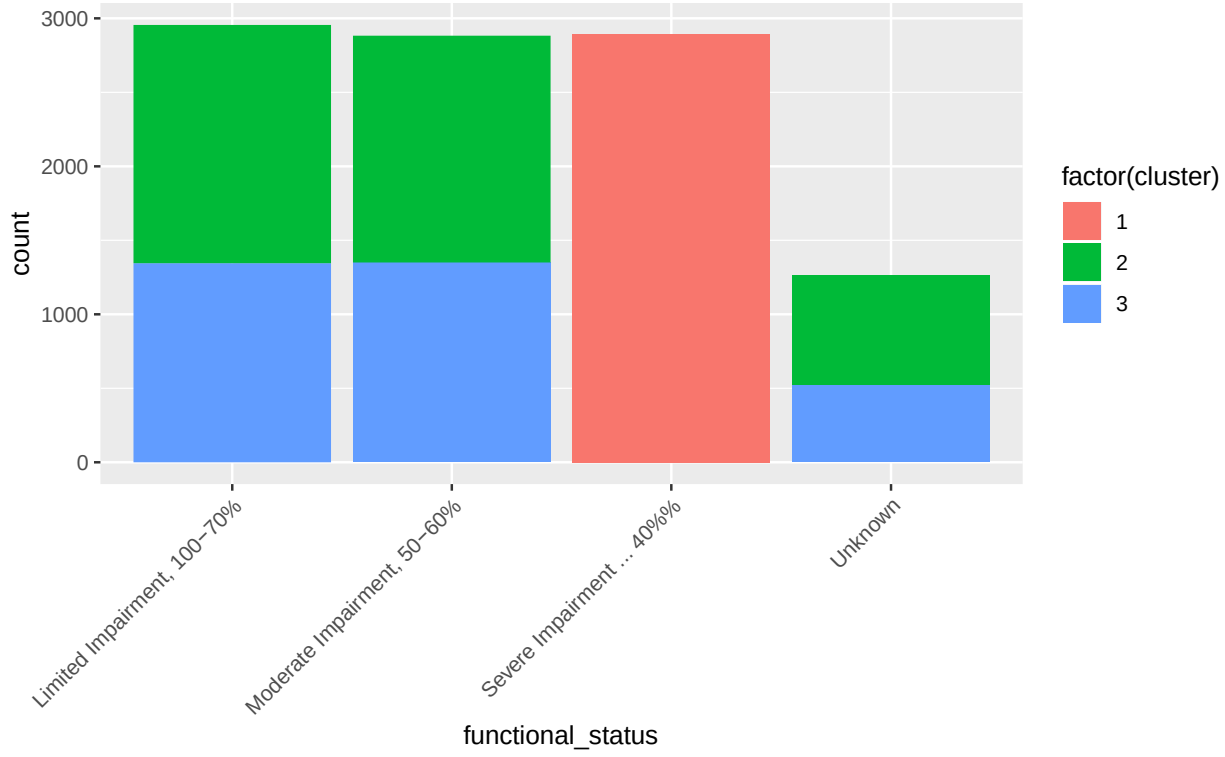


Figure 3b: Functional Status by Cluster without weights

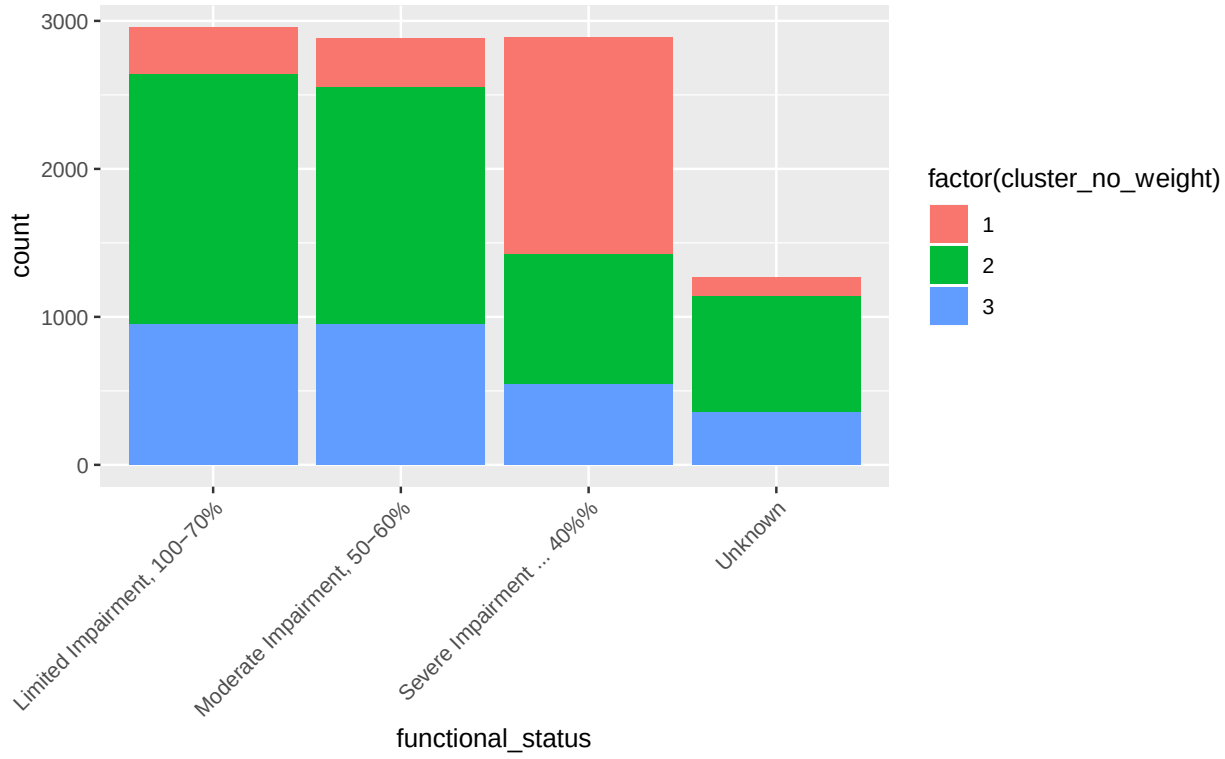


Figure 4: Status by cluster

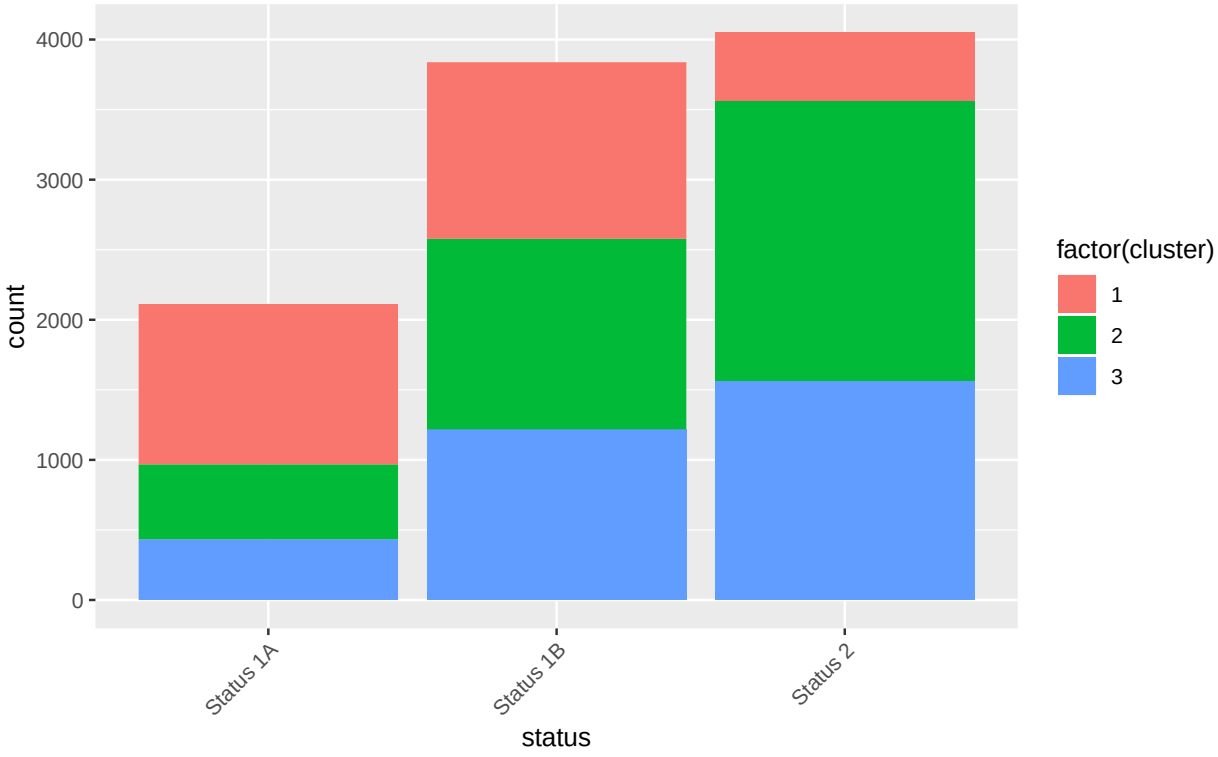


Figure 5: Treatment distribution by Cluster

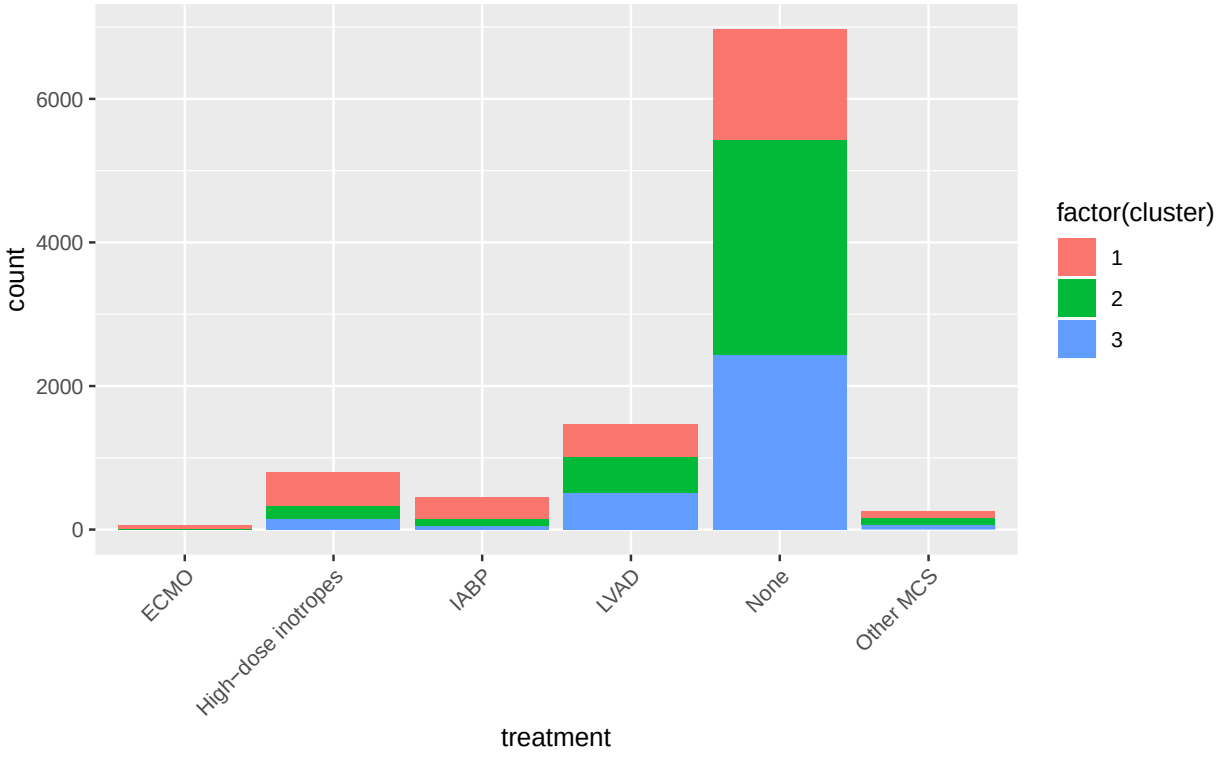


Figure 6A: KM survival estimates by cluster (weighted)

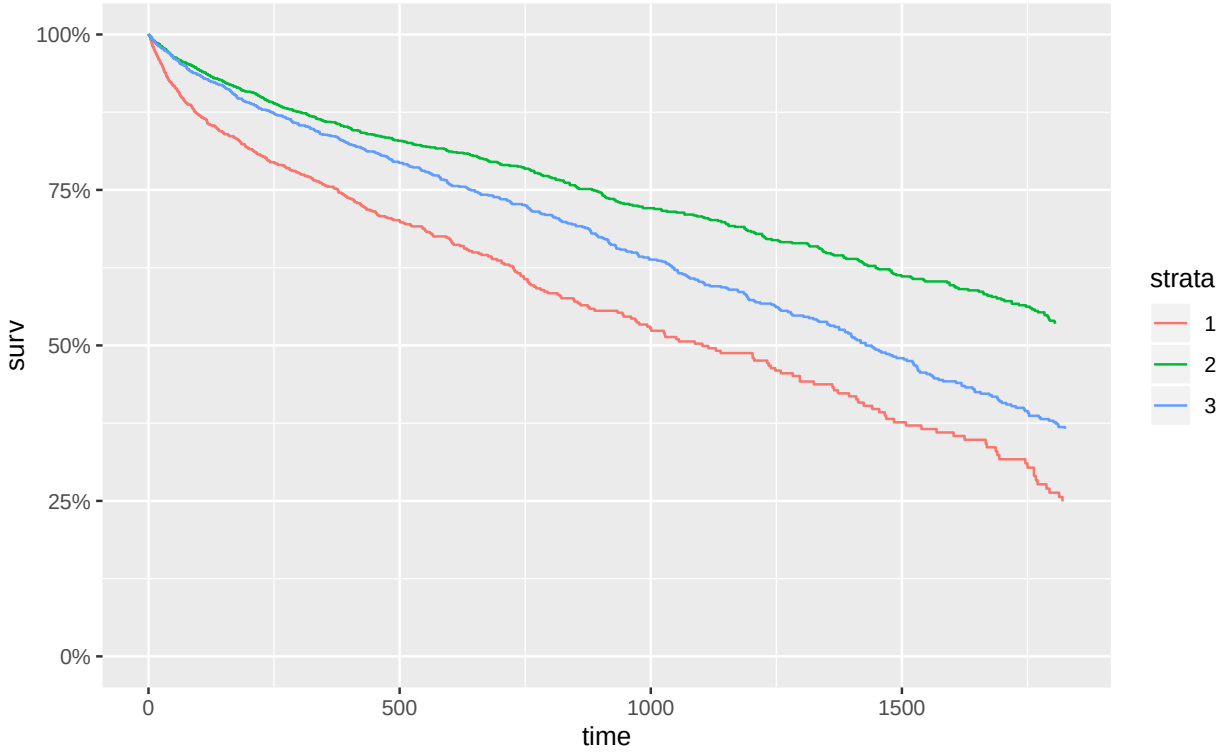


Figure 6B: KM survival estimates by Status

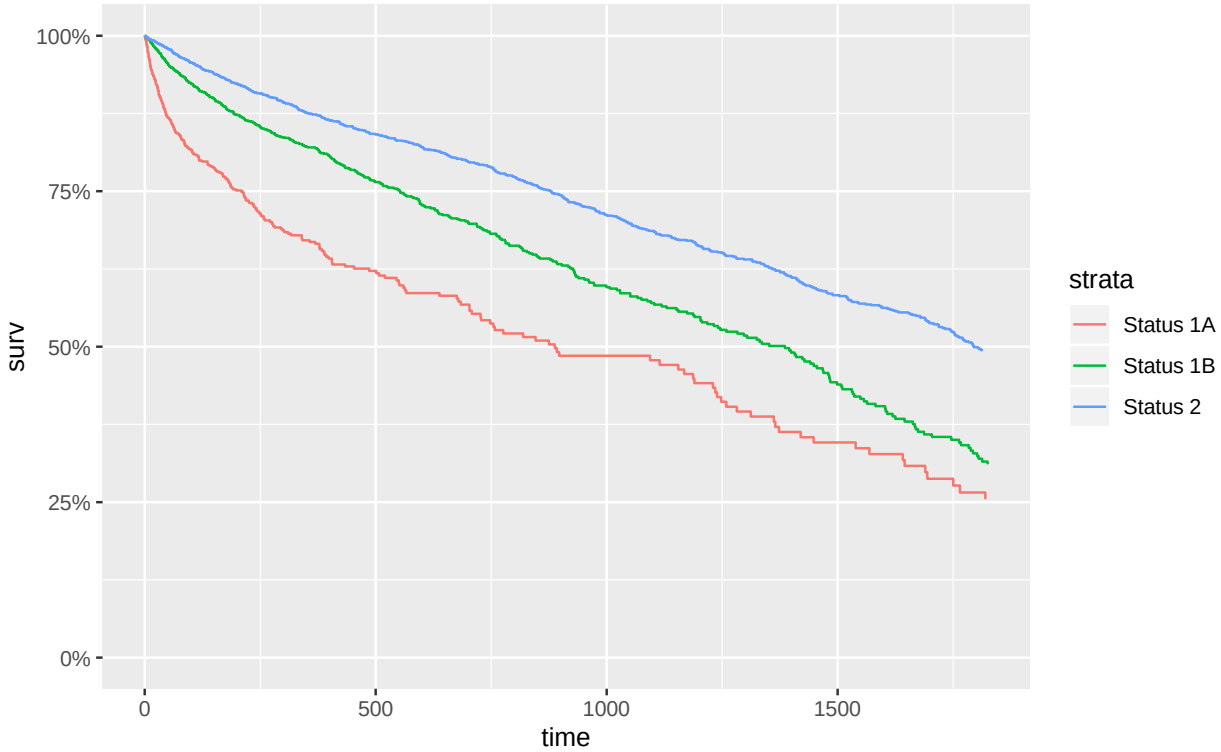


Figure 6C: KM survival estimates by cluster (unweighted)

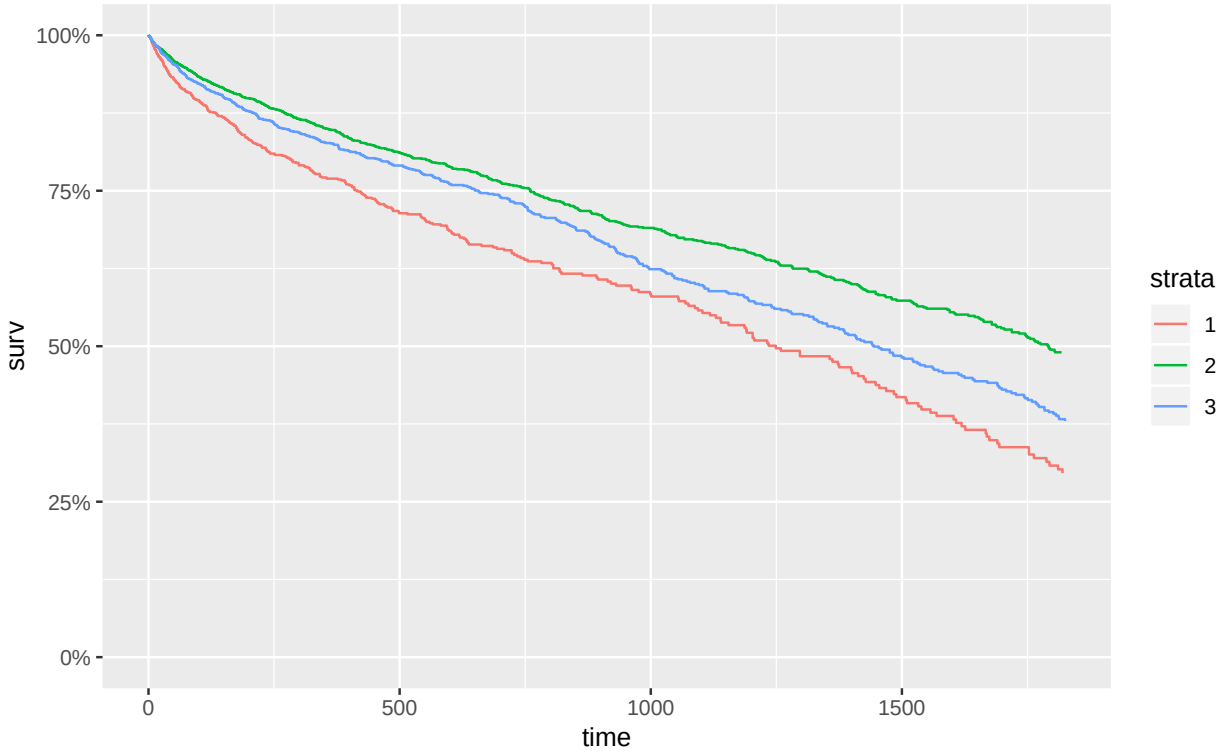
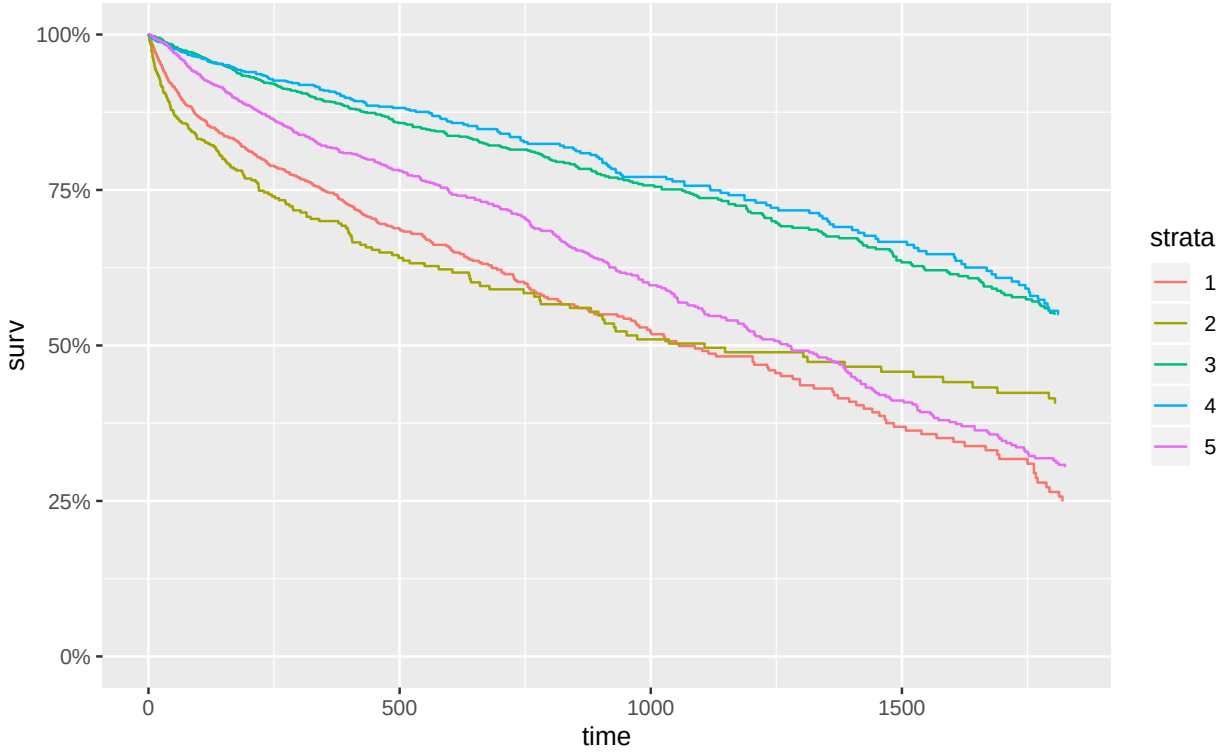


Figure 6D: KM survival estimates by cluster (weighted, k = 5)



References:

1. Benjamin EJ, Blaha MJ, Chiuve SE, et al. Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. *Circulation*. January 2017:CIR.0000000000000485. doi:10.1161/CIR.0000000000000485

2. Colvin M, Smith JM, Hadley N, et al. OPTN/SRTR 2016 Annual Data Report: Heart. *American Journal of Transplantation*. 18(S1):291-362. doi:10.1111/ajt.14561

3. Health Resources and Services Administration (HRSA), Department of Health and Human Services (HHS). Organ procurement and transplantation network. Final rule. *Fed Regist*. 2013;78(128):40033-40042.

4. Parker WF, Garrity ER, Fedson S, Churpek MM. Trends in the Use of Inotropes to List Adult Heart Transplant Candidates at Status 1A. *Circulation: Heart Failure*. 2017;10(12):e004483. doi:10.1161/CIRCHEARTFAILURE.117.004483

5. Parker WF, Garrity ER, Fedson S, Churpek MM. Potential impact of a shock requirement on adult heart allocation. *The Journal of Heart and Lung Transplantation*. 2017;36(9):1013-1016. doi:10.1016/j.healun.2017.05.015

6. Parker WF, Anderson AS, Hedeker D, et al. Geographic Variation in the Treatment of U.S. Adult Heart Transplant Candidates. *Journal of the American College of Cardiology*. 2018;71(16):1715-1725. doi:10.1016/j.jacc.2018.02.030

7. Krakauer H, Lin MJ-Y, Bailey RC. Projected Survival Benefit as Criterion for Listing and Organ Allocation in Heart Transplantation. *The Journal of Heart and Lung Transplantation*. 2005;24(6):680-689. doi:10.1016/j.healun.2004.04.015

8. Singh TP, Milliren CE, Almond CS, Graham D. Survival Benefit From Transplantation in Patients Listed for Heart Transplantation in the United States. *J Am Coll Cardiol*. 2014;63(12):1169-1178. doi:10.1016/j.jacc.2013.11.045

9. Schnitzler MA, Whiting JF, Brennan DC, et al. The life-years saved by a deceased organ donor. *Am J Transplant*. 2005;5(9):2289-2296. doi:10.1111/j.1600-6143.2005.01021.x

10. Medicare C for, Baltimore MS 7500 SB, Usa M. Transplant-Laws-and-Regulations. https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/GuidanceforLawsAndRegulations/Transplant-Laws-and-Regulations.html. Published July 19, 2017. Accessed October 25, 2018.

11. Centers for Medicare & Medicaid Services (CMS), HHS. Medicare program; hospital conditions of participation: requirements for approval and re-approval of transplant centers to perform organ transplants. Final rule. *Fed Regist*. 2007;72(61):15197-15280.

12. Jasseron C, Legeai C, Jacquelinet C, et al. Prediction of Waitlist Mortality in Adult Heart Transplant Candidates: The Candidate Risk Score. *Transplantation*. 2017;101(9):21752182. doi:10.1097/TP.0000000000001724

13. Hsich EM, Thuita L, McNamara DM, et al. Variables of importance in the Scientific Registry of Transplant Recipients database predictive of heart transplant waitlist mortality. *American Journal of Transplantation*. 0(ja). doi:10.1111/ajt.15265

14. Weiss ES, Allen JG, Arnaoutakis GJ, et al. Creation of a quantitative recipient risk index for mortality prediction after cardiac transplantation (IMPACT). *Ann Thorac Surg*. 2011;92(3):914-921; discussion 921-922. doi:10.1016/j.athoracsur.2011.04.030

15. Bektas A, Schumann R. How to Optimize Gower Distance Weights for the k-Medoids Clustering Algorithm to Obtain Mobility Profiles of the Swiss Population. In: *2019 6th Swiss Conference on Data Science (SDS)*. ; 2019:51-56. doi:10.1109/SDS.2019.000-8

16. Sinclair C. Improving Clustering Performance Using Feature Weight Learning. Medium. https://towardsdatascience.com/improving-clustering-performance-using-feature-weight-learning-d65d4fec77cb. Published December 20, 2018. Accessed December 10, 2019.