Title: Identifying clinical phenotypes of heart transplant candidates using machine learning

Project Proposal for Unsupervised Machine Learning MACS 40800

Lilly Reich, William Parker

Introduction and Motivation:

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 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. All previous attempts[7–9] to develop objective prediction models for heart transplant candidates have used standard survival analysis regression based approaches and consider pre-transplant support treatments as fixed characteristics of a candidate, rather than active treatment that modifies other candidate variables. None of these models has been adopted for policy purposes because of these shortcomings

Proposed Project Aims:

1) Apply unsupervised machine learning clustering techniques to identify natural clinical phenotypes of heart transplant candidates
2) Determine the association between identified clusters and clinically important outcomes like survival (both before and after transplant)
3) Apply dimension reduction techniques like principle components analysis as a tool to explore the between-center variation in the candidate characteristics and treatment

Approach:

Data source and study population

We will use 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 will be 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 very small portion of the transplant population.

Candidate variables

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)**. Many of these variables have already been shown to be strong independent predictors of mortality in both heart transplant candidates and recipients.[12–14] Importantly, 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 just prior to transplantation.

Demographics	Medical History and comorbidities	Clinical and Laboratory	Therapeutic support history
Age	Primary Cardiac Diagnosis	Invasive cardiac hemodynamics	Durable mechanical circulatory support
Sex	Cardiac surgical history	(cardiac index, pulmonary	Inotropes (including dosage)
Race	Hypertension	capillary wedge pressure,	Intra-aortic balloon pump
Education	End-stage renal disease	central venous saturation, etc.)	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 1: A sample of important

Aim 1: Cluster heart transplant candidates

We will employ two general strategies when clustering the unlabeled ~ 50-dimensional candidate feature space: Hard partitioning (each candidate assigned to a specific cluster) and Soft partitioning (each candidate has a probability of membership in each cluster).[15] The exact methodologies we employ for will be determined as we learn about them in the course and through iterative feedback with the class group. However, we will begin with k-means clustering as a hard-partitioning methodology and Gaussian finite mixture models as a soft partitioning methodology.

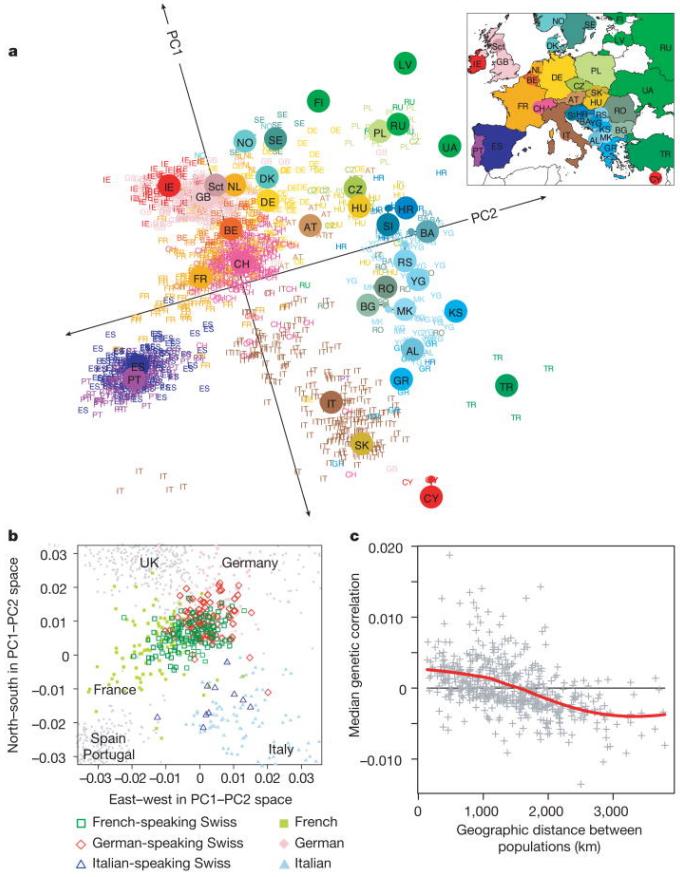
Aim 2: Association of clusters with survival

If our clustering approaches have identified clinically important phenotypes of heart transplant candidates, cluster membership should be associated with clinically important outcomes like survival on the waitlist (before transplantation) and post-transplant survival. After we have determined the optimal hard-partitioned clustering approach, we will estimate survival by cluster by the standard Kaplan Meir method. After we have determined the optimal soft partitioning approach, we will use estimated weights for each candidate as regressors in a cox proportional hazards model.

Aim 3: Principle Component Analysis and Transplant Center Practices

In the current allocation system, heart transplant centers have tremendous leeway to both select candidates for transplantation and choose the treatments designed to support them while they wait on the list for their transplantation. As a result, candidate treatments and clinical variables are highly correlated within transplant centers and candidates are treated substantially differently

based on their geographic location.[6] In this aim we will apply principle component analysis to the high dimensional candidate feature space (see Table 1). We will then plot candidates on the first two principle components and examine the plot for obvious patterns. If the impact of geography on candidate selection and treatments is strong enough, we may see something resembling a geographical map of the US analogous to the famous map of Europe recreated with genetic information in similar fashion (see Figure 1 below).[16]



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):2175-2182. 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. Clustering Introduction & different methods of clustering. *Analytics Vidhya*. November 2016. https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/. Accessed October 16, 2019.

16. Novembre J, Johnson T, Bryc K, et al. Genes mirror geography within Europe. *Nature*. 2008;456(7218):98-101. doi:10.1038/nature07331