ISYE 6501 Course Project

Nick DiNapoli, ndinapoli6@gatech.edu

Due: July 22, 2021

1 Alliance for Paired Kidney Donation

1.1 Background

The Alliance for Paired Kidney Donation (APKD) is a non-profit organization that organizes multi-hospital kidney donations. They aim to solve or mitigate the issue of long (and growing) waiting lists, and hence patient fatalities. The idea behind a kidney exchange is having living donors whose kidney is incompatible with their intended recipient to donate a kidney such that their intended recipient receives a compatible one from another donor [1]. See Figure 1 for a schematic of how this is completed.

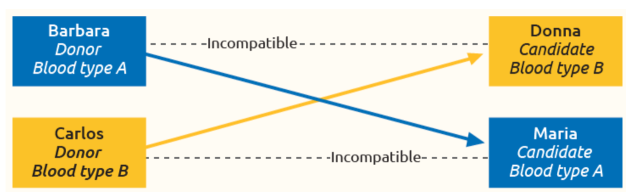

Figure 1: Here, Barbara wants to donate to Donna but is incompatible. Carlos wants to donate to Maria but they are also not compatible. By swapping donors, Carlos matches Donna and Barbara matches Maria, and the two transplants are made possible [2].

There are also other types of donations: 1) Directed donations (donating to a family member or friend) and 2) Non-directed donations (donating to an unknown person). The APKD was the first to organized non-simultaneous chains. Chains can be completed such that every patient receives a kidney no later than their intended donor donates one. As it can be seen in Figure 2, this process can lead to potentially very long chains.

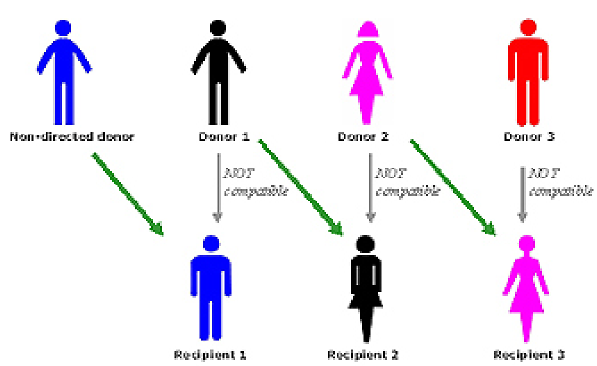

Figure 2: The potential of donation chains [3].

The APKD adopted integer programming to identify optimal sets of disjoint exchanges. Through the use of Alvin Roth's Nobel-prize winning algorithm, operations research and market design, 6% of all living-donor transplants stem from non-simultaneous chains. According to [4], transplants from living donors increase survival rates by nearly 12 years compared to those who receive kidneys from deceased donors. Each transplant saves Medicare more than $270,000 over five years and the unquantifiable cost of a life [1].

1.2 Necessary Data and Collection

Not to take away from other applications of analytics but the importance of a good analytics model when dealing with human lives is immense. Because of this, the reliance on good data is also huge. The beauty of this problem is that the APKD had access to past data coming from directed and non-directed donations. Also, because their proposed paired donation model is also a way of optimizing patient outcomes, not much data on top of that collected by previous models needs to be considered.

1.2.1 Data

There are several attributes/predictors as well as other relevant data that need(s) to be collected for a successful model in this project by the APKD. Data that should be considered and collected for donors and recipients are summarized in Tables 1 & 2.

Predictor	Example Data	Variable Type
Age	30 years	discrete
Sex	Female	categorical
Weight	150.5 lbs	continuous
Blood type	A+	categorical
Intended recipient	Joel Sokol	unbounded, unique
Location	Atlanta, GA	categorical
Physical health	7	discrete
Mental health	8	discrete
Previous kidney disease	0	binary
Previous heart disease	0	binary
Previous liver disease	0	binary
Previous lung disease	0	binary
Previous high blood pressure	1	binary
Previous diabetes	0	binary
Previous cancer	0	binary
Previous hepatitis	0	binary
Previous HIV	0	binary
Good standing blood test	1	binary
Good standing urine test	1	binary

Table 1: Necessary data to be collected for donors and recipients.

Predictor	Example Data	Variable Type
Type of kidney disease	Chronic kidney disease	categorical
Length of time on waiting list	2.25 years	continuous
Anticipated time before not viable for transplant	7.75 years	continuous

Table 2: Additional data to be collected for recipients.

1.2.2 Data Collection

All of the data previously mentioned is certainly not easy to collect which makes this analytics problem particularly challenging. Collecting attributes like age, sex, weight, etc. may quite simply be submitted by each donor/recipient but others like physical/mental health, previous diseases, and blood/urine tests require the intervention and expertise of medical professionals. The simple data can be collected via some online or paper form but the latter must be imputed after some medical professional intervention which could be quite costly. This also barks the question of how to handle missing data and depending on the importance of each predictor in the proposed model, this question can be tackled at some later time. The key to the data collection in this project would be having one central database where individuals and medical professionals can impute data in the same format as they obtain the correct information.

1.2.3 Data Maintenance

Assuming there is in fact one central U.S. database that contains all data points, the data and therefore model should be re-run or updated at least daily. I say this because on average there are about 47 new transplants every day in the U.S. [5] and these surgeries will provide 94 new labeled data points (one for donor + one for recipients) for a supervised learning model. Additionally, because of the importance of the data involved there should be daily tests to find both missing data and outliers.

1.3 Analytics Models

Now that the predictors used by analytics models have been discussed, it is key to understand what appropriate labels may be. This problem is particularly challenging because the labels (outcomes) depend on so many factors. The obvious choice is success (lived) vs. no success (died), but what is truly successful? There seems to be a spectrum of successes i.e. some patients (both donor and recipient) may survive the transplant surgery but have minor or even severe complications later in life. It is important to consider a spectrum of labels for different parts of this analytics project. For example, a recipient with a small period remaining before it is too late to receive a transplant would be thrilled to obtain a kidney even if there is a high probability that there are complications later in life. I will explain in the succeeding sections.

1.3.1 Donor Selection

Before taking physical location, or donor-specific data into consideration, there needs to be some model to rank the donors by most attractive donor to least. Because the result (or label) of a transplant surgery depends so heavily on the health and attributes of the recipient, it wouldn't be necessary, or even helpful, to try and fit the donor data to results using a SVM or logistic regression model. Also, if data for both donors and recipients were used in training these types of models, the significance of the recipient predictors would dominate those of the donor. Yes, one could certainly use all of the data in Tables 1 & 2 and complete a logistic regression model to receive probabilities of surviving using the donor-recipient pair BUT for making predictions this would require data from two people and not just one. Again, there are several different approaches to this problem. A linear regression model using backward elimination could be used to start with all factors of both the donor and recipient but even if one predictor of the donor proves to be significant then it is necessary to always use two people when making future predictions.

Because of these facts, for donor selection I suggest using an integer program optimization model to rank each donor by best to worst. Some of the variables in Table 1 simply act as filters for even being considered a donor and other variables have specific weights that the program takes into consideration when optimizing. Table 3 provides an example of what this may look like.

Predictor	Example Weight (if not a filter)
Age	1.0
Sex	0.1
Weight	3.0
Blood type	filter
Intended recipient	n/a
Location	n/a
Physical health	5.0
Mental health	0.5
Previous kidney disease	filter
Previous heart disease	filter
Previous liver disease	filter
Previous lung disease	0.05
Previous high blood pressure	0.05
Previous diabetes	filter
Previous cancer	0.05
Previous hepatitis	filter
Previous HIV	filter
Good standing blood test	7.5
Good standing urine test	7.5

Table 3: Predictors with example weight to be maximized in an integer program optimization model

If the goal is to maximize some linear objective function, Table 3 shows examples of how the predictors may be weighted after scaling the data. For example, the importance of having a good standing blood and urine test may be 1.5 times more important than having good physical health or 2.5 times more important than one's weight etc. Even though integer programs are a bit more difficult to solve even with software packages, after this is completed there is a ranking of optimal donors.

1.3.2 Recipient Selection

Just as donor selection is a finicky problem with several options for which analytics model to use, recipient selection is the same way. As mentioned in the previous section, the outcome of the transplant surgery is so heavily dependent on the predictors of the recipient so now it is appropriate to train some sort of supervised learning model to predict who might be a good candidate for receiving a kidney. Because classes of this model are so drastic (survive vs. die) and hence the cost of a false positive or false negative prediction are extremely high, it would be most appropriate to obtain probabilities of belonging to each class. For this reason it would be an intelligent choice to use the medical background and predictors from Tables 1 & 2 and previous data of recipients to run a logistic regression model to obtain the probabilities of a successful surgery for those on the wait list. Recall the statement in the introduction of this section, "It is important to consider a spectrum of labels for different parts of this analytics project". I say this because it might be a smart choice in this part of the project to consider the further breakdown of labels and that all successful surgeries do not necessarily mean long-term success. To assess this fact, assuming there is access to this data, one could build a logistic regression tree following the schematic shown in Figure 3. Now, for each person on the wait list, there is an assigned probability of a successful surgery, a probability of complication less than 5 years post-op and a probability of never having a complication.

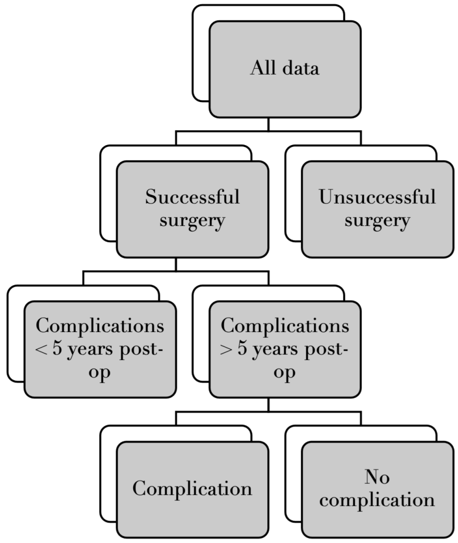

Figure 3: Logistic regression tree.

1.3.3 Pair and Chain Selection

Lastly, there is the magnificent feat of using the outcomes of the previous two sections, the physical location of the donors and recipients, and the length of time a recipient has been on the wait list and has left on the wait list to determine an optimal pair or chain of donations. Just like the previous sections, there are numerous ways to tackle this problem. Naturally, this project leans towards using an optimization model with the goal of maximizing survival probability, maximizing the probability of having no complications, maximizing the length of a donation chain, and minimizing the distance that patients have to travel. One idea might be to create a network but the emphasis on location in this problem is small because people are willing to travel in order to receive a crucial transplant. Because the donors are already ranked, the recipients are all assigned probabilities and locations are flexible, the most difficult part of this problem is determining chain length. To determine this, one might follow the logic in the pseudo code below. Note that this does not consider the other variables or objective function that the optimization model will.

```
Repeat
    Choose next person with longest wait list time
    Are chances of survival above defined threshold?
        Yes -> continue
        No -> loop and try next recipient
    Repeat
        Does next top ranked donor have compatible blood type?
            Yes -> Do they have non-compatible intended recipient?
                Yes -> loop and use their non-compatible recipient
                No -> Go to next ranked donor with recipient
            No -> Go to next ranked donor with correct blood type
```

The optimization model will contain the following:

1. Variables
   (a) donors chosen
   (b) recipients chosen
2. Constraints
   (a) donors and recipients must have compatible blood type
   (b) assuming data is pre-filtered i.e. age ≥ 18 and age ≤ 65, location in U.S., blood and urine tests meet minimum requirement etc.
3. Objective function
   (a) maximizing survival probability
   (b) maximizing the probability of having no complications
   (c) maximizing the length of a donation chain
   (d) minimizing the distance that patients have to travel

A pseudo mathematical expression representing this objective function is shown in Equation 1 where i is each recipient selected, Distance is the distance to paired donor, L is the length of the chain and α, β, γ, and ζ are tunable parameters that the user can specify to put additional weight and importance on one aspect. Again, for simplicity, we only take into consideration the recipient's health although one could also add that of the donor's. Additionally, we can use all of the probability information from the regression tree to further enhance the complexity of the model but it might not always be necessary because taking complication probabilities into consideration is a whole different beast but is worth mentioning. Again, this is because some might not care if the long-term outlook is grim if they are just trying to survive. Notice the negative sign in front of the distance aspect because the goal is to minimize it.

$$Maximize \quad \alpha \sum_i P(survival)_i + \beta \sum_i P(complication)_i - \gamma \sum_i Distance_i + \zeta L \tag{1}$$

1.4 Summary

These type of analytics problems are anything but easy. There are so many options for how to handle the data, select the model(s) for each step in the process, and for piecing it all together. There is such a great deal of programming and mathematical rigor but even more so human expertise and interaction. How does one even begin to tune the objective function from Equation 1? How can one ethically take away from the importance of human survival? The beauty of optimization and proper model building is knowing when it is done right, the math and art of human intervention will lead to solutions that will be better for all in the long run.

References

[1] https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Improving-Organ-Donations

[2] https://unos.org/transplant/living-donation/

[3] https://www.kidney.org/transplantation/livingdonors/incompatiblebloodtype

[4] https://paireddonation.org/explore-kidney-donation/

[5] https://paireddonation.org/about-us/algorithm/