UNIVERSITY OF CALIFORNIA

Los Angeles

Fairness in Kidney Exchange

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Applied Statistics and Data Science

by

Sofia Viol Alcazar

2023

ABSTRACT OF THE THESIS

Fairness in
Kidney Exchange

by

Sofia Viol Alcazar
Master of Applied Statistics and Data Science
University of California, Los Angeles, 2023
Professor Xiaowu Dai, Chair

Kidney transplantation is a possibly life-saving and preferred method to treat chronic kidney disease, which affects more than 1 in 7 United States adults. Kidney exchange programs are one popular method to address difficulties of compatibility where recipients with incompatible, but willing donors swap to each receive a compatible kidney donation. This grants each patient the opportunity for the life-saving procedure. This thesis explores whether patients of varying socio-demographics experience different rates of kidney exchange. In this paper, data is used from the United Network of Organ Sharing to analyze the odds of receiving a kidney transplantation and to understand how waiting times vary across diverse groups. Using logistic regression and Poisson regression, we found significant differences in the odds of kidney exchange and waiting times across races, blood groups, and education levels.

The thesis of Sofia Viol Alcazar is approved.

Maryam Mahtash Esfandiari

Frederic R Paik Schoenberg

Xiaowu Dai, Committee Chair

University of California, Los Angeles

2023

To Adrian, Martha, and Lilia -

Thank you for your unwavering support and endless love.

And to my Aunt's kidney donor, thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

In the United States, more than 1 in 7 adults are affected by chronic kidney disease and 2 in 1,000 Americans are struggling with end-stage kidney disease [Nih]. Chronic kidney disease inhibits the kidneys from properly cleaning the blood, which can result in toxic waste accumulating in the body, often leading to high blood pressure, heart failure, stroke, and even early death [Cdc]. Once end-stage kidney disease has developed and the kidneys no longer work at the level necessary to survive, such failure is treated by dialysis or transplantation.

The favored approach to treating kidney failure is to receive a transplant. Because of this, the waiting list for a transplant in the United States is extensive and patients often spend years or even decades on the list. As of early 2023, the Organ Procurement and Transplantation Network (OPTN) records that list being longer than 88,000. To exhibit just how vastly demand outweighs supply, OPTN reported slightly more than 12,500 additions to the waiting list in first 14 weeks of 2023, while only 7,201 Americans were lucky enough to receive a transplant and depart the list [Net].

These transplants are made possible through both living and deceased donors. Over the years, on average, around 30% of these organs are sourced from living donors, while the other 70% come from deceased donors. Family members related by genetics are the most common type of living donor compatible pairs, though not everyone is so lucky to match with someone in their support network. The process to find a match is extensive and complex. Both blood and tissue types need to be compatible along with a plethora of other medical factors.

To address this shortage of available and compatible organs, kidney exchange programs have become increasingly popular. In these exchange programs, incompatible pairs are matched with other incompatible pairs, resulting in the exchange of kidneys; giving both recipients the opportunity for transplant.

The motivation for this thesis stems from the pressing need to ensure that medical treatments are provided equitably to all individuals, regardless of their demographic characteristics. Kidney transplantation is a life-saving procedure for individuals suffering from kidney failure, but it may be possible that certain populations face various barriers to accessing this critical medical care. By examining whether patients with differing socio-demographics have different chances of receiving a kidney exchange, we can identify possible disparities and work towards addressing them. Additionally, kidney exchange is a complex and resource-intensive medical process, so ensuring equity in the allocation of these vital organs is essential in the process' efficiency. By exploring potential differences in access, we have the opportunity to identify areas where the process can be improved to ensure that all individuals receive the life-saving treatments they need in a timely and equitable manner.

Ultimately, our goal is to contribute to the development of a more just and equitable healthcare system for all individuals. In this thesis we explore different possible factors that may alter the likelihood of a patient receiving a kidney transplant using real data collected from the United Network of Organ Sharing, or UNOS for short.

CHAPTER 2

The Data

In order to have confidence in the statistical models produced, we must train them using clean and organized data. Throughout this chapter I will present the data obtained by UNOS (United Network of Organ Sharing) and feature engineering conducted before modeling.

2.1 UNOS Data

The data used in this analysis is collected and reported by UNOS on every organ donor, transplant candidate, recipient, and outcome in the United States. The collected data on kidneys specifically includes one record per kidney waiting list registration and/or transplant and has been recorded through patient forms. Upon entry to the UNOS waiting list an immense amount of data is collected from the patient, such as demographic information, previous health records, and blood type properties. Once a waiting list patient receives a transplant a new record is generated to capture more data relating to the transplantation itself and the donor. The original data consists of 1,097,058 records across 491 different fields. Due to the sheer number of data fields and points, a plethora of research topics could be investigated using this data set, especially in the healthcare and medical field. However, in my analysis only a handful are relevant out of the 491 fields.

2.1.1 Challenges

While having an extensive dataset made available by UNOS offers boundless analyses, it can also pose several challenges. Firstly, the data contained a large amount of missing values and encompassed information on patients seeking not only kidney transplants, but also pancreas transplants. These missing and unnecessary values needed to be addressed before any analysis. Cleaning, pre-processing, and extracting relevant fields in the UNOS data was time-consuming and meticulous.

Without specific medical training, it was also difficult to interpret the vast number of abbreviated medical terms. It was challenging to confidently understand many of the complex and high technically terminology and language that was used in the documentation. To overcome these challenges, it was essential to invest time and research into understanding many of these data elements to prevent the possibility of misinterpretation or lost insights. In the future, we recommend consulting an expert in the kidney transplant field, perhaps a surgeon, to aid in the understanding and selection of variables. This would be an incredibly valuable resource for medical terminology and most importantly, for enhancing the validity of the model.

Lastly, the demographic fields were all numerically encoded, thus for easy interpretability, the levels were re-coded for each. Many of the levels for these features were also repetitive, so summarization was required as well. Overall, the data required extensive pre-processing.

2.2 Feature Engineering

As previously mentioned, the data consisted of 491 distinct fields, but only around a dozen were extracted for this analysis. In Table 2.1 you'll find a summary of those extracted features.

Feature Name	Description
ABO	Recipient blood group at registration
AGE	Recipient age (years)
EDUCATION	Recipient highest education level at registration
ETHCAT	Recipient ethnicity category
ETHNICITY	Recipient ethnicity
GENDER	Recipient gender
GSTATUS_KI	Graft failed
INIT_DATE	Date placed on waiting list
ORGAN	Organ type transplanted
TRR_ID_CODE	Encrypted transplant identifier
TX_DATE	Transplant date
WL_ID_CODE	Encrypted registration identifier
WL_ORG	Organ listed for

Table 2.1: Features

The blood group field was originally grouped into 8 categories to account for the sub types of the A blood group, which include the A1 antigen group and non-A1 antigen group. While testing for the A1 antigen is important for those with A or AB blood, 80% of blood type A and AB are of the sub types A1 and A1B [Net23]. Thus, we felt comfortable using the major blood type categories and not breaking it down to sub type. By only grouping this way it also created a more balanced feature than previously, as you can see in Figure 2.1.

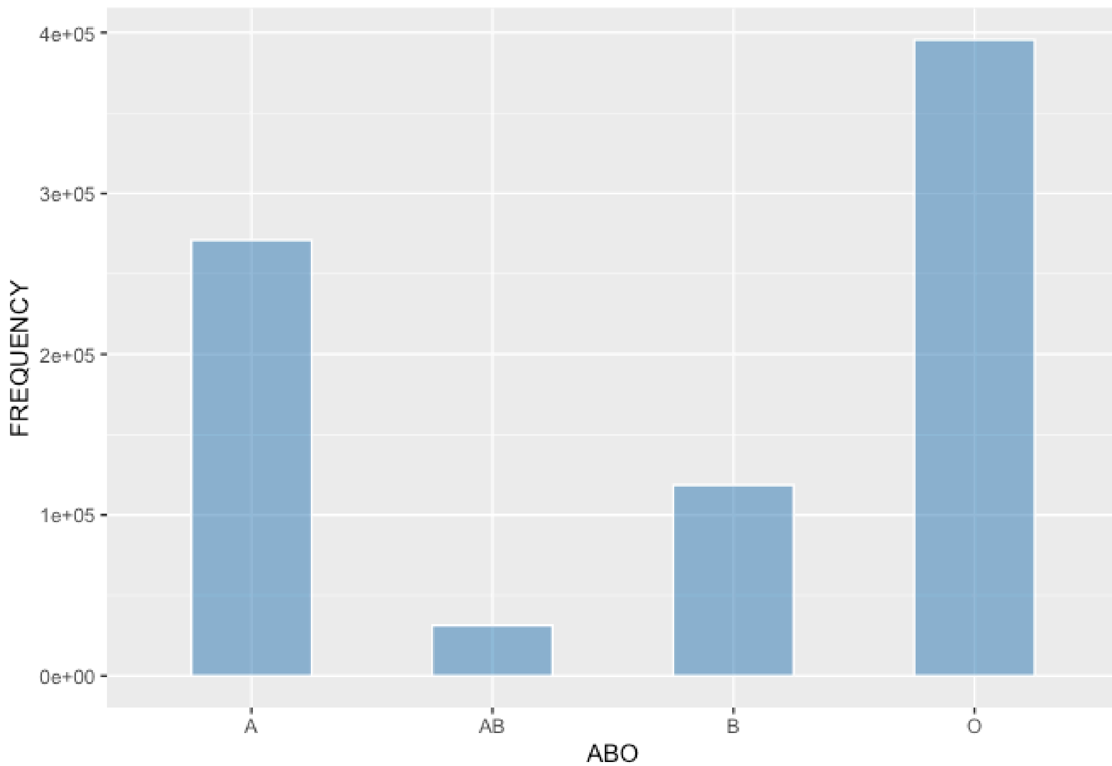

Figure 2.1: Bar chart of Blood Type

The education level of the recipient fell into one of the following categories in the original data: NONE, GRADE SCHOOL (0-8), HIGH SCHOOL (9-12) or GED, ATTENDED COLLEGE/TECHNICAL SCHOOL, ASSOCIATE/BACHELOR DEGREE, POST-COLLEGE GRADUATE DEGREE, N/A (< 5 YRS OLD), UNKNOWN, and Not Reported. After first removing the unknown, not reported, and non applicable values, we further narrowed down these categories since the feature was quite unbalanced as well. We narrowed the factor down to the categories found in Table 2.2 and Figure 2.2.

Old Group	New Group
NONE	NO DEGREE
GRADE SCHOOL (0-8)	NO DEGREE
HIGH SCHOOL (9-12) OR GED	HIGH SCHOOL
ATTENDED COLLEGE/TECHNICAL SCHOOL	SOME COLLEGE
ASSOCIATE/BACHELOR DEGREE	SOME COLLEGE
POST-COLLEGE DEGREE	ADVANCED DEGREE

Table 2.2: Education Level Grouping

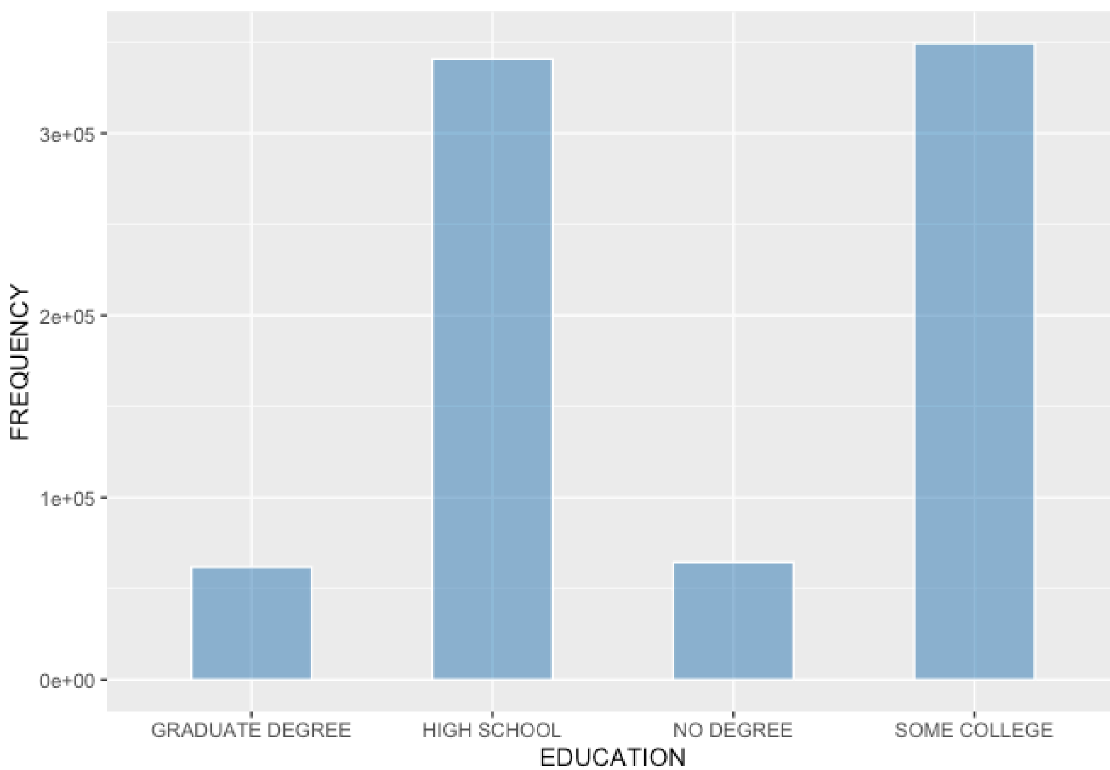

Figure 2.2: Bar chart of Education Level

The UNOS data contained two fields to capture a recipient's ethnicity, ETHCAT and ETHNICITY. ETHCAT, or the ethnic category contained the following fields: WHITE, BLACK, HISPANIC, ASIAN, AMER IND/ALASKA NATIVE, NATIVE HAWAIIAN/ OTHER

PACIFIC ISLANDER, MULTIRACIAL, UNKNOWN, and NOT REPORTED. The ETHNICITY field only recorded whether or not a patient identifies as Hispanic, so we chose to use the more descriptive and in depth ethnicity field, ETHCAT. Similar to the education field, we removed any unknown and not reported values before grouping some of these categories together due to a few very small group sizes. You can find these new groupings in Table 2.3 and Figure 2.3.

Old Group	New Group
WHITE	WHITE
BLACK	BLACK
HISPANIC	HISPANIC
ASIAN	ASIAN
AMER IND/ALASKA NATIVE	OTHER
NATIVE HAWAIIAN/OTHER PACIFIC ISLANDER	OTHER
MULTIRACIAL	OTHER

Table 2.3: Ethnic Category Grouping

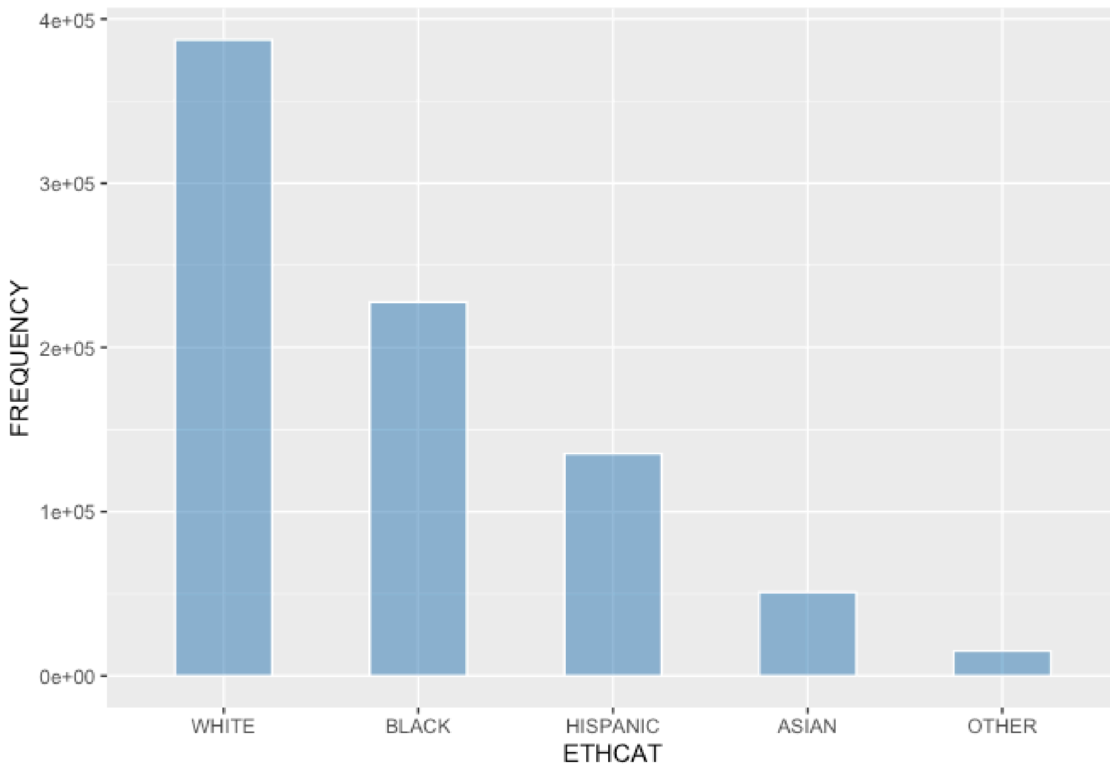

Figure 2.3: Bar chart of Ethnic Category

The data also contained a small number of patients under the age of 18, but for the purpose of this analysis, we are going to focus on only adults. Below is a histogram of the age of waiting list patients. The mean age is around 50 years old.

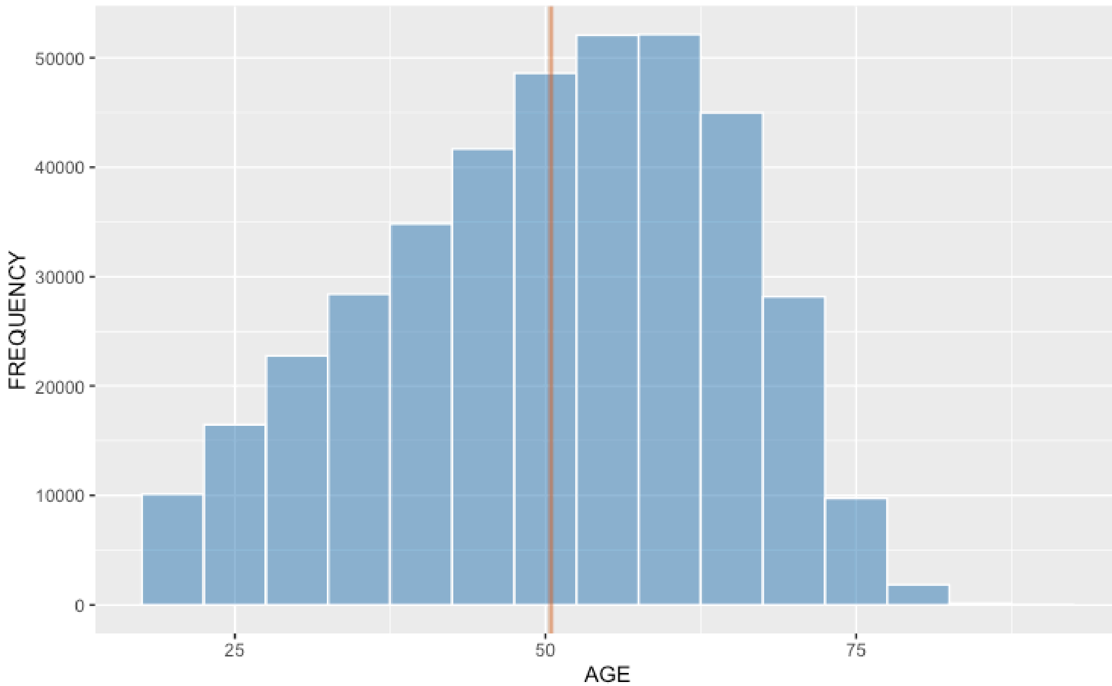

Figure 2.4: Histogram of Age

As previously mentioned, each row in the data set represented a single waiting list registration, identified by an encrypted registration identifier. Should this patient receive a kidney transplant, additional data is added to the record, such as an encrypted transplant identifier and date of transplantation. To easily classify those who have received a transplant from those who have not, we created a binary factor, TRANSPLANT, to use as our dependent variable. Below you can find a bar chart of this response variable and notice that it is nearly perfectly balanced.

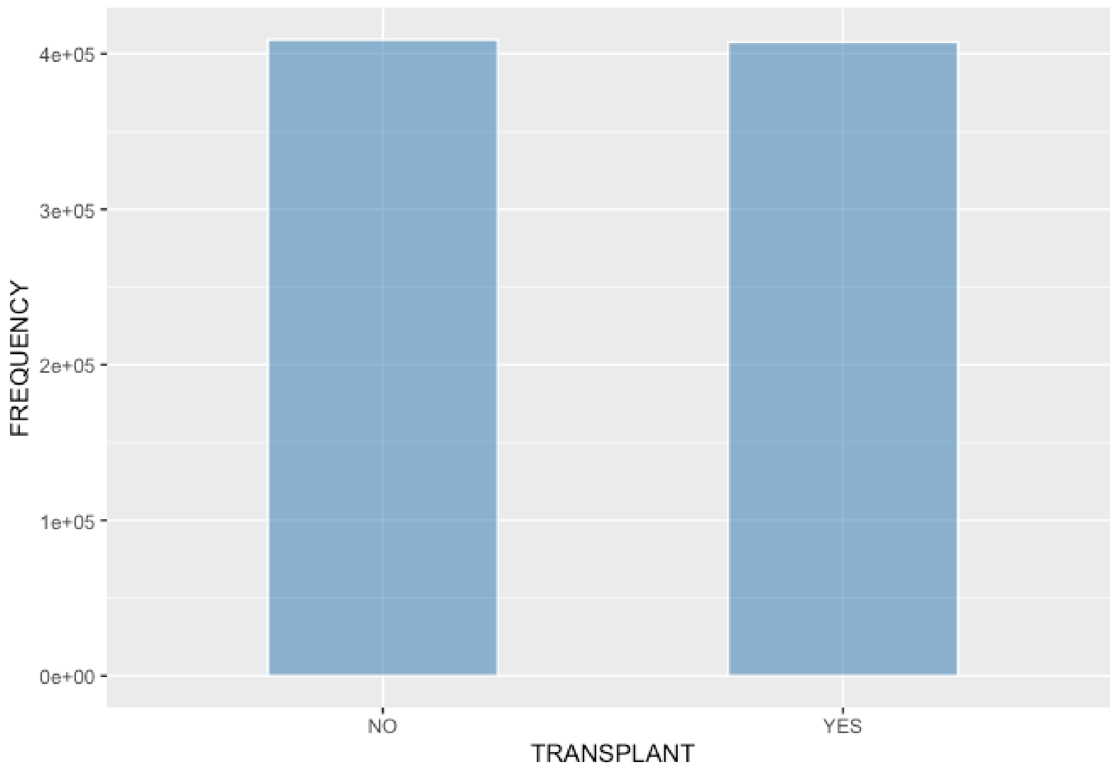

Figure 2.5: Bar chart of Transplant

2.3 Final Dataset

After selecting applicable fields and thoroughly cleaning the data, our final data set consists of 816,566 records across 18 variables. While the original data consisted of information on kidney and pancreas transplants, we narrowed that down to only kidney transplants, and removed unknown values, ultimately reducing the record count by 280,492. Below you can find a table with all of the variables in our final dataset along with their datatype and levels after re-coding.

Feature Name	Datatype	Levels
ABO	Factor	A, AB, B, O
AGE	Numeric	0 - 90 (years)
EDUCATION	Factor	ND, HS, SC, ADV
ETHCAT	Factor	1, 2, 4, 5, OTHER
ETHNICITY	Factor	1, 0
GENDER	Factor	M, F
GSTATUS_KI	Factor	1, 0
INIT_DATE	Date	YYYY-MM-DD
ORGAN	Factor	KI
TRANSPLANT	Factor	1, 0
TRR_ID_CODE	Character	7 character code
TX_DATE	Date	YYYY-MM-DD
WAIT_TIME	Numeric	0 - 1340 (weeks)
WL_ID_CODE	Numeric	7 digit code
WL_ORG	Factor	KI

Table 2.4: Final Dataset Codebook

CHAPTER 3

Exploratory Data Analysis

3.1 Waiting List and Transplant Dates

The data collected contains two key dates, waiting list date and transplantation date, if applicable. The earliest waiting list entry in our data was September of 1979, while the first transplantation was in October of 1987. The most recent transplantation was in June of 2022, so our data spans 7 decades. We first wanted to better understand how long a patient may remain on the waiting list before receiving a transplant. We calculated the number of weeks between the waiting list entry date and transplant date for those lucky enough to receive a kidney. In Figure 3.1 below you'll find the number of weeks patients spent on the waiting list in our data. The mean was around 93 weeks, or 21 months on the waiting list. The longest stay on the waiting list before a kidney transplant was a colossal 1340 weeks, or 25 years. Lastly, the earliest entry with no record of a transplant was in March of 1984.

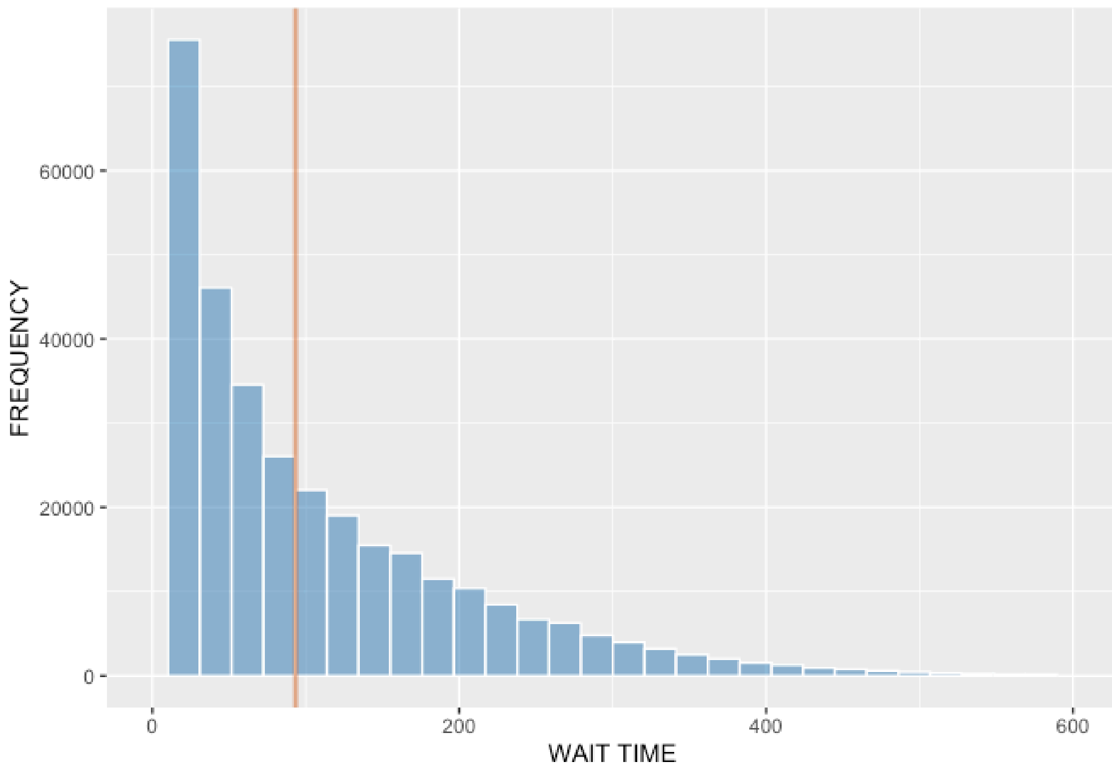

Figure 3.1: Weeks on Waiting List

Next we investigated how waiting times possibly differed by race. In Figure 3.2 you'll find waiting time buckets broken out by each category of the ETHCAT variable. In this figure we see that White patients make up more than half of the transplants that happen within 100 weeks, or 23 months, of joining the waiting list. On the other hand, Black, Hispanic, and Asian patients make up more of the transplants performed after 200 weeks than they do less than 200 weeks. More specifically, Black patients account for only around 25% of the waiting times less than 100 weeks, but nearly 40% of the waiting times over 500 weeks. This can also be reinforced by Figure 3.3, the box plot of waiting time by ethnic category. It appears that the interquartile range of waiting time in weeks for White patients is smaller than the other ethnic categories.

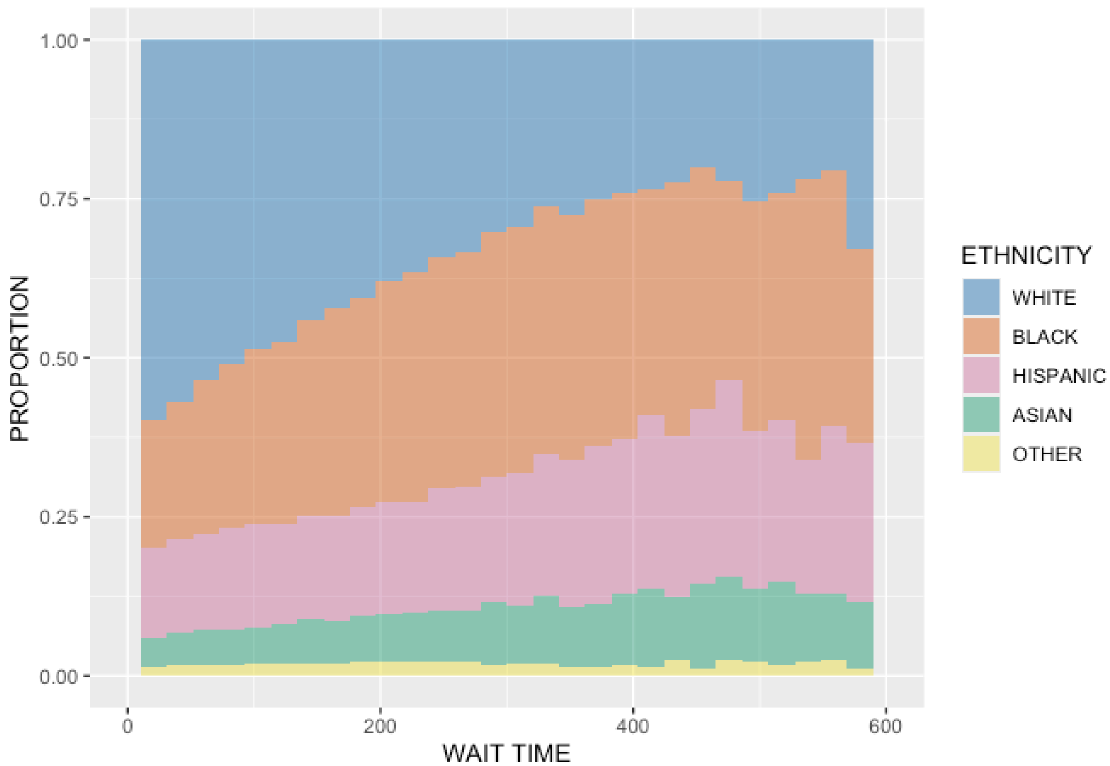

Figure 3.2: Weeks on Waiting List By Ethnicity

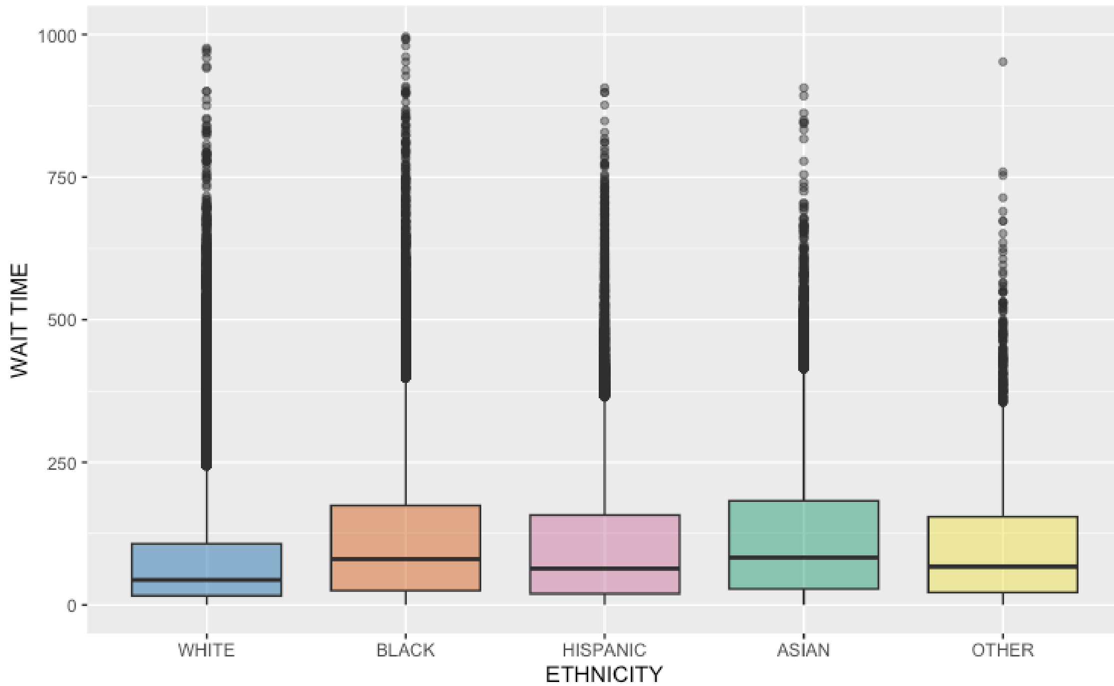

Figure 3.3: Weeks on Waiting List By Ethnicity Box Plot

3.2 Transplant Variable

3.2.1 Transplant and Education Level

By plotting the transplant factor against education level, we see that the proportion of kidney transplants may be higher for patients with no degree. For patients with a high school degree, some college, or an advanced degree the rate of kidney transplantation may not be statistically different.

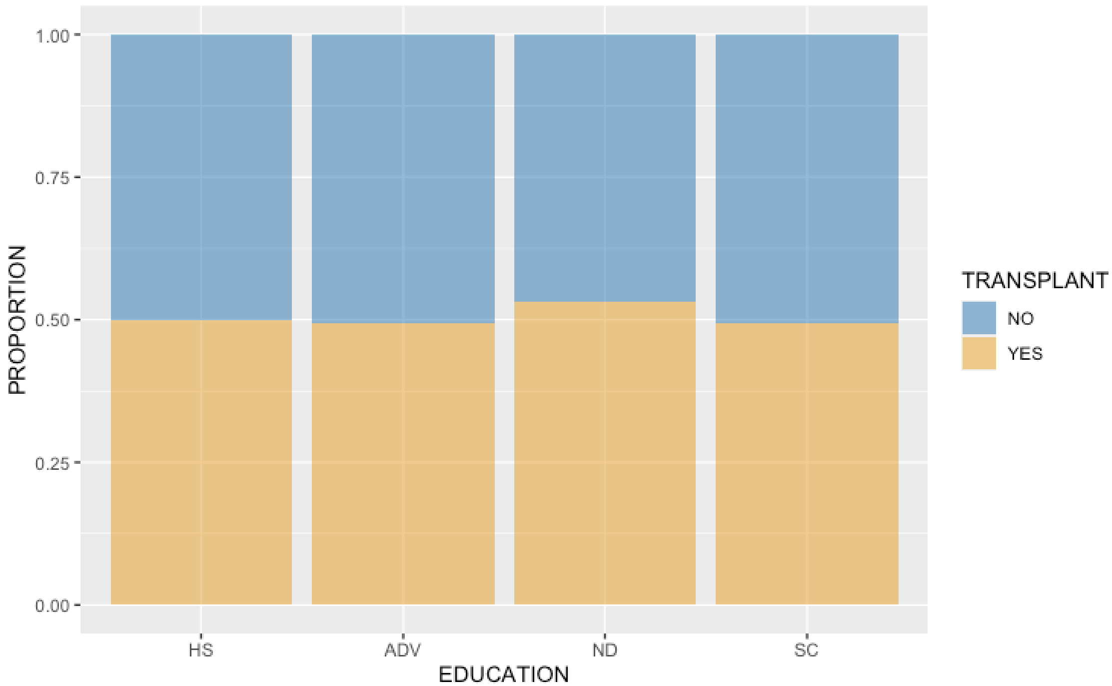

Figure 3.4: Bar chart of Education Level & Transplant

3.2.2 Transplant and Blood Type

Furthermore, using the transplant factor, we can see in Figure 3.4 that it appears different blood types display different proportions of transplants than others. For blood types B and O, the proportion of patients that have received a transplantation is less than those who have not in our data. On the contrary, a majority of patients with blood types A and AB have received a transplantation based on this chart.

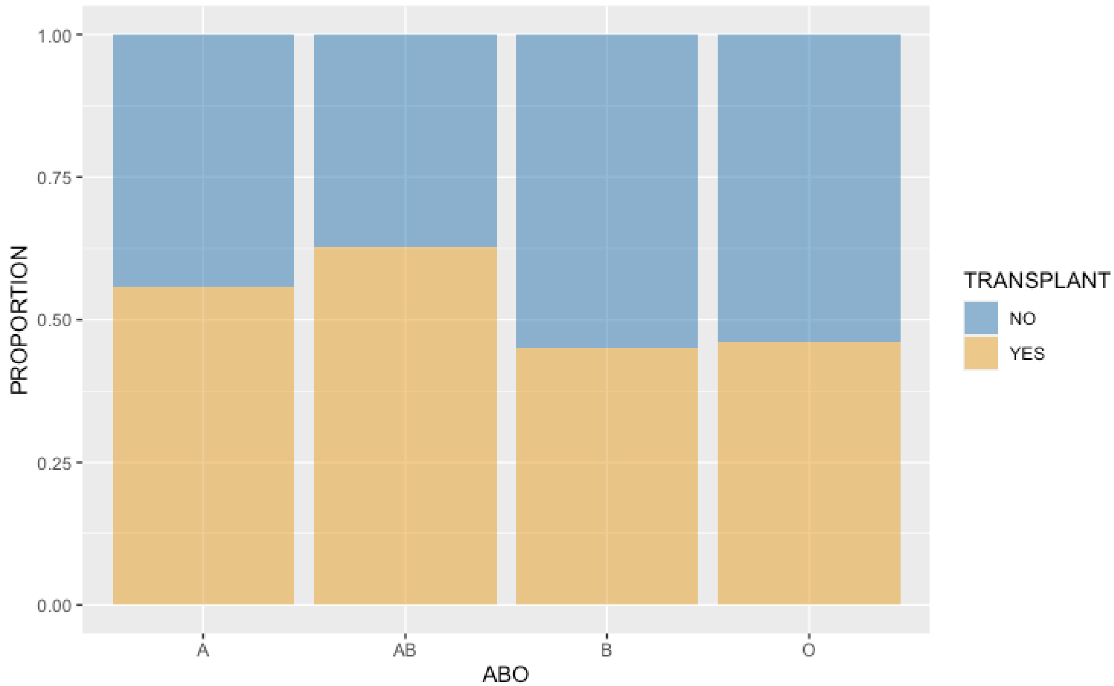

Figure 3.5: Bar chart of Blood Type & Transplant

3.3 EDA Conclusions

Through our exploratory data analysis, we were able to extract important key insights prior to modeling. In summary, we gained insights into time spent on the waiting list and possible disparities in transplant rates. The average time spent on the waiting list was approximately 100 weeks, or 23 months across all patients in the UNOS (United Networking of Organ Sharing) data. We found that White patients account for most of the transplants occurring after only a short stay on the waiting list, while other races shared a large proportion of the longer waiting times. Specifically, Black patients accounted for about 25% of the waiting times less than 2 years, but nearly 40% of the waiting times over 9 years. By plotting transplant rates across education levels we discovered patients with no degree may have higher rates. Similarly, transplants across blood types may not be equal. We identified that blood types A and AB appear to have higher transplant rates than B and O blood types. The potential disparities across socio-demographics found in this chapter will be further explored in the next.

CHAPTER 4

Statistical Modeling

In this chapter we investigate if different combinations of demographic factors have an effect on the likelihood of receiving a kidney transplant using a Chi-squared Test of Independence and Logistic Regression. We also analyzed how the number of weeks on the waiting list differs across varying demographics using a Poisson regression.

4.1 Chi-squared Test of Independence

The Chi-squared test is a non-parametric tool intended to investigate disparities between groups when the dependent variable is nominal [McH13]. Additionally, it does not depend on a specific distribution. Thus, we chose it to initially analyze possible disparities we previewed in our exploratory data analysis before fitting a logistic regression.

In a Chi-squared test of independence the null and alternative hypothesis are as follows:

H_O (null hypothesis): The variables are independent.

H_A (alternative hypothesis): The variables are not independent.

4.1.1 Chi-squared: Transplant and Blood Type

First we tested the transplant and blood type variables. The education level and ethnicity category variables will follow similarly. Below is a table displaying the frequency of each group.

	A	AB	B	O
YES	151,458	19,715	53,480	182,850
NO	119,488	11,652	65,146	212,777

Table 4.1: Transplant & Blood Type Counts

The Pearson's Chi-squared test resulted in the following statistics,

$$\chi^2 = 9253, \text{ df} = 3, \text{ p-value} < 2.2e^{-16}$$

Based on this statistically significant p-value, we were able to conclude there is evidence of a difference in the occurrence of transplantation amongst different blood types.

4.1.2 Chi-squared: Transplant and Education Level

Next we tested whether transplant and education level are independent. The contingency table for these two factors is below.

	ADV	HS	ND	SC
YES	30,503	170,075	34,148	172,777
NO	31,373	170,711	30,286	176,693

Table 4.2: Transplant & Education Level Counts

Our test yields the following results,

$$\chi^2 = 285.8, \text{ df} = 3, \text{ p-value} < 2.2e^{-16}$$

The results rejected the null hypothesis again and we concluded there is evidence of a difference in the occurrence of transplantation amongst patients with various education levels. Although this test yielded the same p-value as the previous, it is important to note the χ^2 value is significantly smaller. Therefore, the difference amongst education level may not be as great as amongst blood type, though we suspected this after our exploratory data analysis.

4.1.3 Chi-squared: Transplant and Ethnic Category

Lastly, our Chi-square test of independence between the ethnic category and transplant follows similarly. You may find the contingency table and Pearson's Chi-squared results below.

	WHITE	BLACK	HISPANIC	ASIAN	OTHER
YES	211,252	102,304	64,406	22,512	7,029
NO	176,173	125,407	70,903	28,312	8,268

Table 4.3: Transplant & Ethnic Category Counts

$$\chi^2 = 6591.4,\ \text{df} = 4,\ \text{p-value} < 2.2e^{-16}$$

As previously, we are able to conclude there is evidence of a difference in the occurrence of transplantation amongst patients of differing ethnicities. In addition to blood type, the χ^2 value for this test leads us to believe the differences between ethnicities are significantly large.

Once we found evidence of a difference in occurrence of transplantation for blood types, education, and ethnic categories, we fit a logistic regression using all three of these demographic variables to decipher which groups differ and by how much.

4.2 Logistic Regression

Our response variable, transplant, was balanced with 407,503 patients receiving a kidney transplant and 409,063 not yet receiving a transplant, thus no re-sampling was required. We chose a train/test split of 70/30 which created a training set of 571,597 observations and a testing set of 244,969 observations.

Now that we've cleaned, explored, and split the data, we can feed it into our statistical analysis. Seeing that the response variable is binary, our baseline model was chosen to be a logistic regression model to predict kidney transplantation. Logistic regression is a powerful, easily interpretable and popular tool for modeling a binary outcome variable and making predictions based on the independent variables. The purpose of this type of model is to estimate the probability that a given observation belongs to a particular class of the response variable, based on the values of the independent variables. Because our Chi-squared tests between the independent variable and each of the predictors were significant, we included each variable in our logistic regression model.

4.2.1 Results

The results of our logistic regression model ran on the training data can be found in Table 4.4. The intercept estimate represents the log odds of receiving a kidney transplant for a White patient with blood type A and an high school degree or GED. The estimates for all of the other predictors represent the log odds of receiving a kidney transplant for the level specified, holding all else constant. In order to interpret these estimates we exponentiated the values to obtain the odds ratio for each level. The odds ratios and their confidence intervals can be found in Table 4.5. The p value column represents the probability of obtaining a test statistic as extreme as observed in our training data assuming there is no association between the independent and dependent variable. All values are followed by three asterisks, indicating there is evidence of a significant association between the independent and dependent variables and the estimate is statistically significant. Additionally, when using our model to predict the outcome of transplantation on our test data, we observed around a 56% accuracy. The p-value for the test comparing the accuracy to the no information rate was $< 2.2e\text{-}16$. This indicates that the model predicts kidney transplantation significantly better than random guessing.

	Estimate	Std. Error	z value	p value
(Intercept)	0.385602	0.006166	62.540	$<$ 2e-16 ***
ABO[AB]	0.333271	0.014784	22.542	$<$ 2e-16 ***
ABO[B]	-0.356137	0.008525	41.777	$<$ 2e-16 ***
ABO[O]	-0.347623	0.006054	-57.424	$<$ 2e-16 ***
ETHCAT[BLACK]	-0.344398	0.006461	-53.308	$<$ 2e-16 ***
ETHCAT[HISPANIC]	-0.296516	0.008031	-36.921	$<$ 2e-16 ***
ETHCAT[ASIAN]	-0.387043	0.011565	-33.468	$<$ 2e-16 ***
ETHCAT[OTHER]	-0.314984	0.019890	-15.836	$<$ 2e-16 ***
EDUCATION[ADV]	-0.084738	0.010641	7.963	1.68e-15 ***
EDUCATION[ND]	0.200248	0.010732	8.659	$<$ 2e-16 ***
EDUCATION[SC]	-0.041984	0.005849	-7.178	7.05e-13 ***

Table 4.4: Logistic Regression Results

	Odds Ratio	2.5%	97.5%
(Intercept)	1.4704997	1.4528400	1.4883816
ABO[AB]	1.3955260	1.3557069	1.4365969
ABO[B]	0.7003765	0.6887703	0.7121753
ABO[O]	0.7063652	0.6980332	0.7147953
ETHCAT[BLACK]	0.7086467	0.6997293	0.7176762
ETHCAT[HISPANIC]	0.7434035	0.7317922	0.7551963
ETHCAT[ASIAN]	0.6790616	0.6638381	0.6946241
ETHCAT[OTHER]	0.7298005	0.7018840	0.7587999
EDUCATION[ADV]	0.9187528	0.8997879	0.9381152
EDUCATION[ND]	1.2217062	1.1962815	1.2476800
EDUCATION[SC]	0.9588851	0.9479558	0.9699399

Table 4.5: Odds Ratios & Confidence Intervals

The estimated odds ratio measures the strength of association between each factor and the outcome, kidney transplantation. Estimates for the odds ratios are deemed statistically significant if the ratio is significantly different than 1. A negative association is indicated if the odds ratio is less than 1, meaning the factor is associated with a decrease in odds of kidney transplantation. Conversely, if the estimated odds ratio is greater than 1, we conclude a positive association and the factor is associated with an increase in odds of kidney transplantation. Lastly, an estimated odds ratio close to 1 indicates there is no association between the factor and kidney transplantation. In Table 4.5 we can view the estimated odds ratio as well as the 95% confidence interval for each. We already concluded each dependent variable has a significant association with the independent variable, but we can further confirm this through the confidence intervals. None of the estimated odds ratio confidence intervals for each variable contain 1, suggesting the true value of the odds ratio is unlikely to be 1 and association between each variable and kidney transplantation is statistically significant. The intercept represents the reference category chosen to be White patients with blood type A and a high school degree or GED. It is easier to interpret the intercept in the context of probability rather than odds, so we calculated this below.

$$\frac{e^{0.385602}}{1+e^{0.385602}} = 0.5952236$$

Our results indicate that for White patients with blood type A and a high school degree or GED, the probability of receiving a kidney transplantation is approximately 0.595. As you can see, each ETHCAT level has an odds ratio less than 1, meaning each of these ethnic categories has a decreased odds of kidney transplantation compared to White patients. The largest difference was observed for Asian patients, where they experienced a 32.1% decrease in the odds of receiving a kidney transplant, holding all else constant, compared to White patients in our model. The largest positive difference, was for patients with blood type AB. They experienced a 39.5% increase in the odds of receiving a kidney transplant, holding all else constant, compared to blood type A in our model. This disparities are more easily visualized in Figure 4.1 below. It is clear from this plot that patients with blood type AB have a much higher odds of kidney transplant than patients with blood type B and O. You can also see that each ethnic category has odds much lower than 1.

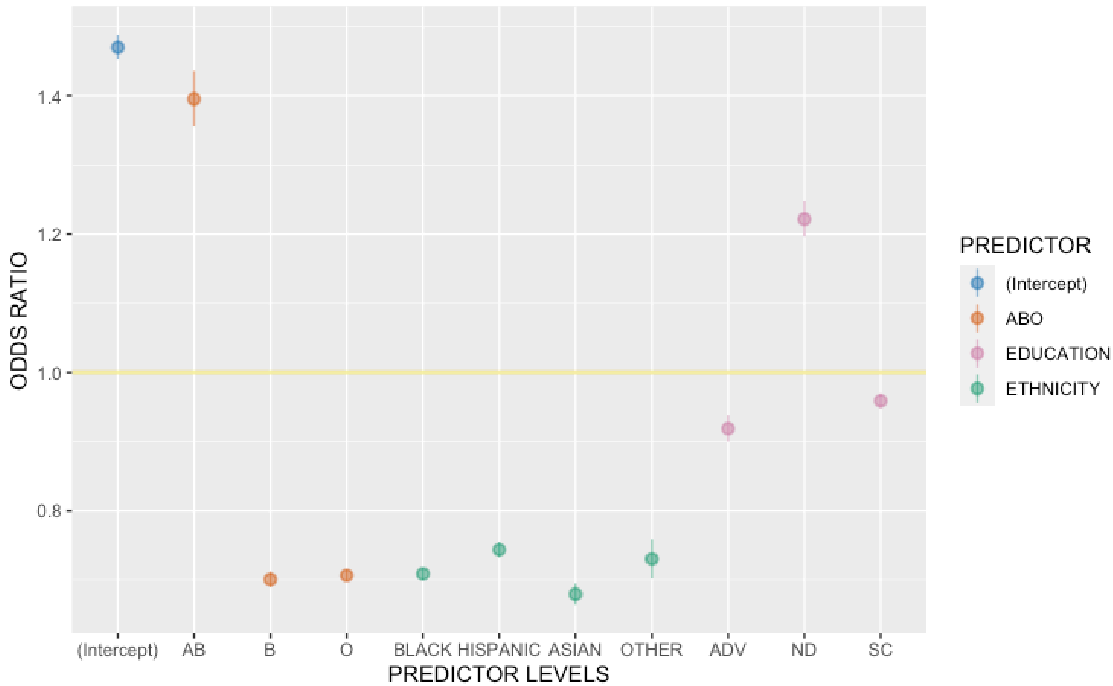

Figure 4.1: Odds Ratio by Predictor Levels

4.2.2 Estimated Marginal Means

To further explore and demonstrate the differences across groups seeking kidney exchange in our model, we've included the estimated marginal means for the blood type and ethnic category variables. These pairwise contrasts for each variable can be found in Table 4.6 and Table 4.7. For each result, the values are averaged over the levels of the other two factors and p-values are calculated on the log odds ratio scale.

Contrast	Odds Ratio	p value
A / AB	0.717	<.0001
A / B	1.428	<.0001
A / O	1.416	<.0001
AB / B	1.993	<.0001
AB / O	1.976	<.0001
B / O	0.992	0.7164

Table 4.6: Blood Type Pairwise Estimated Marginal Means

As you can see, each blood type experiences significantly different odds of kidney exchange when compared in our model, besides the contrast between blood types B and O. Highlighting the largest difference, the results indicate that the odds of receiving a kidney transplantation for patients with AB blood are approximately 99.3% higher than the odds for patients with blood type B.

Contrast	Odds Ratio	p value
WHITE / BLACK	1.411	<.0001
WHITE / HISPANIC	1.345	<.0001
WHITE / ASIAN	1.473	<.0001
WHITE / OTHER	1.370	<.0001
BLACK / HISPANIC	0.953	<.0001
BLACK / ASIAN	1.044	0.0034
BLACK / OTHER	0.971	0.5888
HISPANIC / ASIAN	1.095	<.0001
HISPANIC / OTHER	1.019	0.8994
ASIAN / OTHER	0.930	0.0110

Table 4.7: Ethnic Categories Pairwise Estimated Marginal Means

We previously mentioned that each ethnic category has a decreased odds of kidney transplantation when compared to White patients; that along with all other ethnic category comparisons can be found in Table 4.7. Our model expresses the odds of receiving a kidney transplant for White patients are roughly 47% higher than the odds for Asian patients and 41% higher than the odds for Black patients. Four contrasts were not found to be statistically significant in our analysis and those are the comparisons between Black and Asian patients, Black and other race patients, Hispanic and other race patients, and lastly Asian and other race patients.

4.3 Poisson Regression

For those patients that were lucky enough to receive a kidney transplant, we also analyzed the number of weeks they spent on the waiting list using a Poisson Regression. Compared to a linear regression on normally distributed data, Poisson regression is used for data that is right skewed, or always greater than 0. A simple linear model can't be used since the model parameter, μ, can only take on positive values; thus a log transformation is required. [YB15] In our case, since waiting times can't be less than 0, we chose this style of regression.

4.3.1 Results

Our model estimated the weeks spent on the waiting list using patients' blood type, ethnic category, education level, gender, and age as predictors. The results of our Poisson regression model using our training data can be found in Table 4.8. Similar to the logistic regression model, the intercept in these results represents a White, female patient in their late 40s with blood type A and a high school degree (or GED). The estimates for each of the predictors represents the difference in log of expected weeks on the waiting list represented by a one unit change in the predictor variable. For the predictors that are factor variables, a one unit change is equivalently moving from the reference level to the level specified. As a reminder, all values are followed by three asterisks, indicating there is evidence of a statistically significant association between that the independent and dependent variables and the estimate is statistically significant. Also, to evaluate the performance of the model, we analyzed the null deviance and residual deviance, which is used as a measure for goodness-of-fit. We computed the Chi-squared statistic to be χ^2 = Null Deviance - Residual Deviance = 1928544 across 12 degrees of freedom, since there was 12 predictors in our model. This resulted in a statistically significant p-value, meaning the model fits the data significantly better than the model with only an intercept term. This model also had a lower AIC (Akaike Information Criterion) than the model with only blood type, ethnic category, and education level as predictors. This is important because AIC penalizes more complex models to account for overfitting. Thus, the model we chose described more of the variance in wait time, without harsh overfitting. To put these results in more interpretable terms, we exponentiate the estimates to receive incidence rate ratios for each predictor. These estimates and confidence intervals can be found below as well, in Table 4.9.

	Estimate	Std. Error	z value	p value
(Intercept)	3.777e+00	9.036e-04	4180.69	$<$ 2e-16 ***
ABO[AB]	-3.013e-01	1.181e-03	-255.09	$<$ 2e-16 ***
ABO[B]	2.227e-01	6.404e-04	347.69	$<$ 2e-16 ***
ABO[O]	2.494e-01	4.620e-04	539.74	$<$ 2e-16 ***
ETHCAT[BLACK]	4.002e-01	4.832e-04	828.14	$<$ 2e-16 ***
ETHCAT[HISPANIC]	3.197e-01	6.026e-04	530.55	$<$ 2e-16 ***
ETHCAT[ASIAN]	4.699e-01	8.132e-04	577.83	$<$ 2e-16 ***
ETHCAT[OTHER]	3.014e-01	1.478e-03	203.95	$<$ 2e-16 ***
EDUCATION[ADV]	-1.858e-01	8.582e-04	-216.48	$<$ 2e-16 ***
EDUCATION[ND]	-1.010e-02	7.741e-04	-13.05	$<$ 2e-16 ***
EDUCATION[SC]	-5.327e-02	4.408e-04	-120.84	$<$ 2e-16 ***
GENDER[M]	-1.470e-02	4.120e-04	-35.68	$<$ 2e-16 ***
AGE	8.965e-03	1.405e-05	638.19	$<$ 2e-16 ***

Table 4.8: Poisson Regression Results

	IRR	2.5%	97.5%
(Intercept)	43.7055459	43.6282092	43.7830095
ABO[AB]	0.7398899	0.7381788	0.7416040
ABO[B]	1.2493972	1.2478299	1.2509663
ABO[O]	1.2832366	1.2820750	1.2843992
ETHCAT[BLACK]	1.4920932	1.4906806	1.4935070
ETHCAT[HISPANIC]	1.3767311	1.3751059	1.3783580
ETHCAT[ASIAN]	1.5998312	1.5972829	1.6023827
ETHCAT[OTHER]	1.3517748	1.3478631	1.3556944
EDUCATION[ADV]	0.8304544	0.8290584	0.8318521
EDUCATION[ND]	0.9899487	0.9884477	0.9914515
EDUCATION[SC]	0.9481252	0.9473064	0.9489448
GENDER[M]	0.9854088	0.9846135	0.9862048
AGE	1.0090056	1.0089778	1.0090334

Table 4.9: Incidence Rate Ratios & Confidence Intervals

Our results indicate that Black patients have 1.49 times more weeks spent on the waiting list compared to White patients, holding all else constant. Similarly, Hispanic patients and Asian patients have 1.38 and 1.60 times more weeks spent on the waiting list compared to White patients, respectively. In terms of education level, patients with an advanced degree have 0.83 times less weeks spent on the waiting list than those with only a high school degree. Lastly, patients with AB blood type experienced 0.74 times less weeks on the waiting list compared to blood type A patients, in our model. As you can see from Table 4.9, the previously stated incidence rate ratios as well as all of the others are statistically significant and the confidence intervals do not contain 0.

4.4 Conclusions

To conclude, this analysis and paper intended to examine whether patients of different demographics have unequal access to kidney exchange. We addressed this question through a logistic regression to predict the odds of receiving a kidney transplantation and a Poisson regression to understand waiting times for those fortunate enough to receive the life-saving treatment. Overall, these disparities across demographics need to be addressed by researchers, policymakers, or healthcare providers.

In our logistic regression model, we found many key insights. First, we saw that each ethnic category had a decreased odds of kidney transplantation compared to White patients, with Asian patients experiencing the largest negative difference. Also, from our estimated marginal means, we saw that most of the comparisons between ethnic categories not including White patients were not statistically significant, meaning we did not find a significant difference in odds in our model between the other ethnic categories. Additionally, patients with blood type AB experienced a large increase in the odds of receiving a kidney transplant compared to blood type A. Though, this is not surprising, since blood type AB is considered the universal recipient and they can receive organs from any blood type.

Finally, our Poisson regression also shed some light on the difference in waiting times across demographics. Most significantly, Black and Asian patients experienced far longer waiting times than White patients in our model. Again, blood type AB patients experienced much shorter waiting times in our model compared to blood type A. Since these patients are considered universal recipients, there are not as many challenges posed in finding a donor, which is reflected in the shorter waiting times. The greatest disparity across education level came from patients with advanced degrees, who experienced shorter wait times in our model compared to those with only a high school degree or GED. Overall, each one of these disparities across demographics found in our models need to be addressed by researchers, policymakers, or healthcare providers.

CHAPTER 5

Future Work & Impact

The impact of addressing the questions in this thesis has the potential to yield valuable insights that can inform and hopefully improve our current kidney exchange process. By pointing out possible disparities in access to kidney exchange based on demographic factors, policymakers, healthcare providers, and researchers can work towards providing equitable access for all individuals. This could involve implementing policies to address racial disparities, or optimizing the currently used matching algorithm to account for discrepancies related to blood type.

There are many different aspects that future work could focus on, and here are a couple. First off, researchers could conduct a comprehensive analysis of any underlying factors that may contribute to the disparities uncovered in this thesis to help gain a deeper understanding of the root causes. For example, examining things like cultural beliefs or socioeconomic factors could shed light on the reasons we see differing rates of kidney exchange. Secondly, research could explore the long-term outcomes and impact of kidney exchange on different groups. This could involve evaluating post-transplant survival rates or quality of life measures. Transplant survival rates could even be studied with the UNOS data analysed in this paper. Such investigations would help to address these disparities and enhance the overall fairness and effectiveness of our kidney exchange programs.

REFERENCES

[Cdc] "Chronic Kidney Disease in the United States, 2021." https://www.cdc.gov/kidneydisease/publications-resources/ckd-national-facts.html.

[McH13] Mary L. McHugh. "The Chi-square test of independence." *PubMed Central*, 2013.

[Net] Organ Procurement & Transplantation Network. "OPTN Metrics." https://insights.unos.org/OPTN-metrics/.

[Net23] Organ Procurement & Transplantation Network. "Guidance for ABO Subtyping Organ Donors for Blood Types A and AB.", 2023.

[Nih] "Kidney Disease Statistics for the United States - NIDDK." https://www.niddk.nih.gov/health-information/health-statistics/kidney-disease. [Accessed 12-May-2023].

[YB15] Shengping Yang and Gilbert Berdine. "Poisson Regression." *The Southwest Respiratory and Critical Care Chronicles*, **3**:61, 2015.