PrediCare

A tool for predicting & optimizing patient attendance

Team 15
Sabrina Aberman (SSE), Claire Frankel (SSE),
Bari Gordon (SSE), Sonia Kumar (SSE)

Advisor: Megan Ryerson

II. Executive Summary

Healthcare facilities servicing low-income patients face abnormally high no-show rates. Missed appointments cost the U.S. healthcare system $1.5 billion each year. This is an incredibly complex problem with many contributing factors. Transportation to primary care is a well-documented barrier for patients with Medicaid coverage. According to the National Academy of Sciences, about 3.6 million patients miss medical appointments each year because of transportation barriers. The ability to show up to appointments remains one of the biggest challenges facing the healthcare industry.[1][2]

Our team worked with the primary care clinic on 3701 Market in order to optimize patient attendance. Since the clinic's patient demographic consists primarily of low-income, older patients on Medicaid, it faces high no-show rates and high associated costs. Our team built a predictive algorithm – using patient inputs such as past attendance history, available transportation options, and demographic data – to identify which patients would benefit from transportation scheduling by clinic personnel.

We designed a dashboard, incorporating the results from our algorithm, to facilitate clinic social workers with scheduling transportation intervention. By offering social workers an easy way to identify savable patients proactively, we aim to reduce the number of patients missing appointments due to transportation barriers. We believe that this will allow for improved patient attendance, improved overall health, reduced visits to the ER, and decreased operating costs of Penn's healthcare system.

Our algorithm is able to predict high priority patients who could benefit from transportation with a 9% false positive rate and low priority patients with a 30% false negative rate. At this clinic, we identified 162 high priority patients in a sample of 549. Of those identified, 43% were not able to complete their appointment on time. Overall, this is 13% of the overall population that could possibly have made it to their appointment with transportation.

III. Overview of Project

Our team aims to reduce the number of patients missing appointments due to transportation barriers. Our design objectives are to get patients to their appointments on time, identify patients who would benefit from transportation intervention, and assist social workers in scheduling patient transportation.

The goal of this algorithm is not to provide a set-in stone answer to who will benefit from transportation. Rather, it is to help a social worker understand his/her patients and make a more informed decision on whether to spend time and resources in getting a patient to their

[1] http://www.epic.com/epic/post/3156
[2] https://hitchhealth.co/

appointment. Missed appointments cost both the patient and healthcare facility. Improving this problem in any way is a huge benefit for doctors, nurses, social workers, and patients.

IV. Method of Solution

a. Specification and Requirements

1. Population Selection: We began by selecting a subset of the patient populations: those with a appointment completed rate of less than 0.95. This number covers almost all patients who had ever missed appointments. We would like to get more specific and define a lower completed rate cut-off for those designated as "high" rate of missed appointments. However, due to our dataset and quality, we were afraid that by setting the boundary too low that we would miss out on too much data and our algorithm would lose strength.

Of those with missed appointments, we separated patients who were offered a Lyft ride to get them to their appointment into those who accepted the ride and those who were offered a ride yet did not accept. Of those who accepted a ride, 97.6% made it to their appointment on time and of those who did not, only 50% made it to their appointments on time. These two subpopulations were key to our analysis as we focused on finding differences between them.

2. T-Tests: To determine which variables had significant differences between the two populations, we conducted t-tests for each variable between the two populations. Our target significance level was 0.4. In the real world, we do not expect very significant statistics. What we were looking for was a better understanding of human behavior, and each insight gained was valuable. (See Appendix Figure 8 for details of t-tests)

3. Advanced Data Preparation: After selecting our significant variables via t-tests and our own holistic knowledge, we performed a few transformations on our variables in order to accentuate their extreme values. Numeric variables were squared to create new variables for which low values were made lower and high values were made higher (note that no value for any variable was negative so we did not need to worry about variable sign). We then scaled our variables. We normalized them to put them on the scale scale as the range of the different numeric variables differed greatly. This was done by subtracting the mean from each value and then dividing by the standard deviation.

4. Logistic Regressions: We then conducted a logistic regression to answer the question: What is the likelihood that a given individual will accept a ride? Our initial model input was all the significant variables from the t-tests as well as a few more insignificant variables (See Appendix Figure 9 for full list of initial model variables). We chose to include these additional variables in the initial model because they still had some small effect when interacting with other variables even though they did not show statistical significance. This smaller subset of data might not reflect the full effect of these variables but a larger population might. There are not hard and fast rules to play by from a qualitative perspective.

The final model for the logistic regression is as follows:

Coefficients:	Estimate
(Intercept)	0.3395
Walk1	-1.4227
drive_self1	17.6936
transit_time	0.7404
ns_rate_yr	-0.5244
MONTHS_IN_PRACTICE	0.2824
emp_disabled1	1.0402
risk_score	-0.3043

Response: first_ride_taken	LR Chisq	Df	Pr(>Chisq)
Walk	2.8305	1	0.092487
drive_self	7.5756	1	0.005916
transit_time	3.0743	1	0.079539
ns_rate_yr	2.5698	1	0.108921
MONTHS_IN_PRACTICE	0.8175	1	0.365912
emp_disabled	2.4513	1	0.117428
risk_score	0.9060	1	0.341177

Those predicted to have at least a 0.70 chance of needing transportation intervention were assigned as needing a ride. This boundary was chosen to minimize time/cost lost to social worker and clinic. We wanted to minimize the false positive rate such that clinic resources are not wasted. We recognize that in turn we sacrifice a low false negative rate. More generally, this tradeoff is one between economic and social welfare. By minimizing false positive rates, we maximize economic welfare to the clinic by improving their attendance rate but not coordinating and paying for rides that are not needed. By disregarding the false negative rate, we are not able to maximize social welfare by getting as many people to their appointments as possible.

We tested our algorithm on a small testing subset of the population.

Transportation Intervention		**Benefits**	**Does not Benefit**
Priority Level	High	19	2
	Medium	20	15
	Low	3	7

False Positive Rate for High Priority: 0.095
False Negative Rate for Low Priority: 0.3

5. Clustering: The logistic regression predictive algorithm is good at capturing the extreme cases which would or would not benefit from transportation intervention. However, because this is a highly complex problem, there are more nuanced cases that are predicted by the algorithm to be medium priority. In order to identify the patient profiles that associate with these cases, we cluster the population classified as Medium priority by the algorithm so far.

The cluster analysis produced 4 different populations. The acceptance rates in each population were 0.33, 0.56, 0.63, and 0.75 with the overall rate of acceptance as 0.57. There is a significant difference between the means of the low and high acceptance rate populations. BSS/TSS = 94.7%. We build profiles based on the cluster centers of each population and whether or not the cluster center means are significantly different across the low and high acceptance populations.

Those who identify with the high acceptance rate cluster are assigned High Priority while those identify with the low rate cluster are assigned Low Priority. The remaining patients are assigned Medium Priority. The end user, the social worker, can then make a final decision on whether or not to offer intervention based on qualitative information they know about their client.

6. Testing Results: Due to our limited data, we needed as much data as possible in order to draw conclusions about human behavior. Therefore, we were not able to put aside a validation set. The goal of this algorithm is not to provide a set-in stone answer to who will benefit from transportation; it is to help a social worker in understanding his/her patients and make a more informed decision on whether to spend time and resources in getting a patient to their appointment.

We had 549 patients for whom we had information not related to an offer or acceptance of a ride. Therefore, we predicted priority levels for these patients without validation. We predicted the following distribution: High – 162, Medium – 280, Low – 106. Of those with a predicted High Priority, 43.2% (70 individuals) were not able to complete their given appointment on time. Because we know that 97% of individuals who accepted a ride made it to their appointment on time, it is very likely that, with transportation intervention, these individuals would make it to their given appointment. This is about 13% of the overall population that could possibly make their appointment with transportation; this is a huge increase overall.

7. Qualitative Findings and Identifiable Patient Profiles

See Appendix Figure 6 for detailed outline of patient profiles and their ranking of priority.

b. What specific classes and knowledge does the project depend on?

In completing this design project, we largely drew on knowledge from three main Electrical and Systems Engineering classes: ESE 302 - Engineering Applications of Statistics, ESE 305 - Introduction to Data Science, ESE 204 - Decision Models, and ESE 680 - Human Systems Engineering.

In building our predictive model, we needed a thorough understanding of the statical methods used in predictive modeling. The background knowledge for building a logistic regression was largely obtained from ESE 302, while the methods of parameter optimization, variable selection, and model building were learned from ESE 305. We also used knowledge gained from ESE 204 to perform a cost-benefit analysis. We needed to understand decision models, and the expected costs and benefits to be derived from different scenarios. For example, we needed to understand the expected cost a hospital will incur by a patient missing their scheduled appointment and the costs associated with the hospital paying for transportation intervention. In building the user interface of the prototype, it was crucial to have an understanding of how to best create a system that is easily usable and understandable for users. We implemented specific interviewing methods learned from ESE 680 and subsequently built user workflow diagrams to help us identify where our project could have the largest impact.

V. Self-learning

We had to expand our knowledge in the areas of predictive modeling software development. In the area of predictive modeling, we had a lot of preexisting knowledge related to supervised learning. However, we needed to learn more about the field of unsupervised learning for our clustering analysis. We also had to think of new ways to approach our data given the poor data quality so we could still extract meaning. We purchased books on both Predictive Modeling and Data Science in R in order to improve both our modeling and coding abilities as well as learn about unsupervised learning.

We also had to develop our prototype using the tool Justinmind. We were able to learn the ins and outs of Justinmind through watching extensive prototyping tutorials. Additionally, using Penn's subscription with Lynda.com, we went through online courses to cover basics of web and product design as well as product management and UX/UI foundations to help with the build out and design of the prototype.

VI. Design and Iteration

The Design of our Project

The design of our project was focused on the data gathering, modelling, and prototyping. We gathered data from two key sources: a Lyft study run by Dr. Krisda Chaiyachati and patient records from the clinic on 3701 Market. In the Lyft study, 816 low-income Medicaid patients in the West Philadelphia area were offered one-time free Lyft rides to and from their appointment at the clinic on 37th and Market. The study collected data on whether or not the patient accepted this offer as well as demographic data about the patient, such as age, address, employment status, and modes of transportation typically used to get to appointments. This data was then merged with patient records to include patient history.

Before we could start modelling, we took the time to fully understand the different categories of data and the actual pain points of the patients. To do this we reviewed our four categories of data: (1) transportation, (2) employment, (3) demographics, and (4) attendance history. We also met with a variety of subject matter experts (more details of these meetings can be found in *VIII. Summary of Meetings*). Additionally, we mapped our data in Tableau to uncover themes that would potentially further strengthen our algorithm. The data visualization not only helped us get a feel for where our patients are located but also allowed us to start seeing trends and differentiate groups of patients. Mapping in Tableau assisted us in determining patient clusters necessary for refining our predictive algorithm.

Using this knowledge we began to build our predictive algorithm to determine which patients were missing appointments due to transportation barriers and would benefit from transportation intervention (as outlined in *IV. Method of Solution*). This algorithm classified patients as high, medium, or low priority based on patient medical data and personal characteristics. We then created a prototype of an app for a social worker to be able to use the results from this algorithm.

We met with a social worker, Ricardo Santos Martinez, and built a workflow diagram of his “typical” day. This was crucial in determining the requirements for our prototype. Customer and end user input are indispensable in building any product or prototype. The goal of this product is not only to bring economic and social impact to the healthcare facilities and to the general medicaid population but also to simplify the workflow of social workers. While at the end of the day Ricardo himself may not benefit from this system as a patient, the system could make his job magnitudes simpler which in turn can help his patients directly and free up his time to help his patients more indirectly. Additionally, Justinmind has incredible sharing capabilities, so we were able to send Ricardo the prototype throughout the process to get his input and feedback in UX/UI testing.

Our prototype lists all of the clinic’s patients (see *Figure 5*) that the social worker would be working with and orders them by next appointment date and includes the output of the algorithm (priority: high, medium, low) as well as ride status (unscheduled, scheduled, patient confirmed, N/A). There are filters available on patient name or priority level so the social worker can find the patient specific needed. The social worker can then open the patient profile (see *Figure 6*) which would link to Epic (more information on this can be found in *XIII. Standards and Compliance*) to pull medical data from the clinic, as well as editable fields that the social worker/clinic would get from the patient (i.e. do they walk to their appointments?). The medical data in the prototype is not editable as it contains patient data (such as risk score, appointment history, etc.) pulled from the clinic’s database. However, there are a number of editable fields such as the priority level, ride status, notes field, and specific characteristic data relating to the patient (i.e. what methods of transportation they use to get to appointments). Through our user testing and customer feedback gathering with Ricardo, we found that social workers essentially live in the ‘notes’ field of their database. Therefore, it was essential to include space on this platform for the social worker to add their own information manually. Again, because missed appointments due to transportation is a social problem as well as economic one, we needed to allow for sometimes unexplainable human behavior to be noted.

One of the most important aspects of the prototype is the fact that the priority level is an editable field. Our predictive algorithm initially populates this attribute based on the results from the algorithm, however, we realize that this is a human system and that the social worker knows these patients better than our algorithm can. Therefore, the social worker has the autonomy to edit the priority level. In the future, we will track how often the social worker has to change the priority level and improve our algorithm using this information. Additionally, our prototype has a very important page that lists out common clusters we found in the data so the user can reference this and get a better understanding of how our algorithm is classifying the priority level of our patients (see *Figure 7*).

A full report of all technical specifications, the HTML code, and the prototype .vp file are all included with our final submission. The .vp file include CSS code for all attributes on each page.

Iterations and Pivots

Our vision for this project took a few turns along the way, but ultimately our goal to improve patient attendance remained the same. Initially, we wanted to determine if rideshare options such as Lyft or Uber could improve the attendance rate of patients. Other thoughts included providing patients with an optimized mode of transportation or an algorithm to optimize patient scheduling and overbooking (similar to how airlines work). However, we ran into a variety of issues including a lack of available data. Our feedback sessions with Professor Won proved invaluable as he reminded us of the importance of providing a useful deliverable and tackling a single problem. He steered us in the direction of instead building a predictive algorithm to determine which patients would benefit from transportation intervention. Professor Won noted that all clinics, doctors, nurses, and patients would benefit from increased patient attendance. We then consulted with the social worker in the Penn Clinic to confirm that this would be valuable for our end user.

Additionally, once we had our project idea more refined due to the help of Professor Won, we wanted to not only build a predictive algorithm to determine the likelihood of a patient missing their appointment, but also an algorithm to recommend the best possible transportation mode for a patient (given geographic location, employment status, past attendance history, and costs of transportation for both the patient and the hospital). However, given the amount of work left on the no-show algorithm as well as the secondary task of building out a dashboard for clinic use, we decided to cut the idea of the personalized transportation recommendation. Instead, we focused our efforts on perfecting the predictive algorithm and ensuring that our portal for social worker use met the users' needs.

VII. Societal, Global, and/or Economic Impact

Patient no-shows significantly affect the delivery and cost of care. Thus, an attempt at improving attendance at clinic appointments has potential global societal and economic impact. We improve quality of life in our communities when patients receive better healthcare treatment and improve our communities' economies when our hospitals can conduct more business, which adds not only revenues but also jobs to these hospitals and healthcare clinics as well as reduce wasted taxpayer money.

Societal and Global Impact

According to the World Health Organization, tackling health inequalities is a global health priority.[3] It is important to better understand the risks and needs of patients who do not engage effectively with the country's healthcare system in order for providers to ameliorate its services and meet those needs. Our project specifically targets the low income population of West Philadelphia (i.e., patients who benefit from Medicaid). In completing this project, we worked closely with social workers to efficiently indicate which patients may need additional support for attending appointments. Our aim in working directly with the social worker in the Penn Clinic is that we design our product for what our end user wants. We went through many iterations of our design to gather user feedback. Ultimately, our product can be sold to a electronic medical records software such as Epic. While this is our end customer, our end user differs. We hope that by designing the product for social workers, more patient rides will be able to successfully be scheduled leading to higher appointment attendance. Clinics and hospitals will then be more inclined to buy and use Epic with this additional feature.

Our product is designed for patients living in major urban cities like Philadelphia. Although we have not been able to to test the product in other cities, the information used to build our models is not specific to the Philadelphia area. If the same patient characteristics were gathered from cities similar to Philly in clinics similar to the Penn clinic at 37th and market, we are confident our product would be just as useful. There are 190M patients with records in Epic.[4] About 80% of the US population resides in urban areas[5] and about 74 million people qualify for medicare[6] or about 23% of the population. Therefore, we can estimate that our product has the potential to help about 35 million patients. Based on the results we found from implementing on system at the Penn clinic, we found that our product had the potential to increase patient attendance by 13%. Therefore, if implemented in all Epic systems and assuming that the majority of urban cities are similar to Philadelphia, our platform can help an additional 4.55 million patients nationally get to their appointments. This number may be an overestimate as Philly tends to have an unusually big transportation desert which could be an extra hindrance in getting patients to their appointments. However, at the same time, there are many individuals who do not even recognize that their reason they miss appointments classifies as due to transportation barriers so the reported number of missed appointments due to transportation barriers may in fact be an underestimate.

Economic Impact and Partial Business Analysis

As previously mentioned, there are expected decreases in lost clinic revenue. On the larger system-wide scale, this project has the potential to improve attendance rates at primary care appointments thus improving patient health, and reducing overall hospital admissions or

---

[3] http://apps.who.int/iris/bitstream/10665/43943/1/9789241563703_eng.pdf
[4] https://www.epic.com/about
[5] https://www.census.gov/geo/reference/ua/uafacts.html
[6] https://www.medicaid.gov/medicaid/program-information/medicaid-and-chip-enrollment-data/report-highlights/index.html

emergency department visits. The unnecessary use of emergency departments is an expensive burden on hospitals and patients. Further this cost is not only a direct burden to hospitals and patients but also a burden on ordinary citizens as taxpayers. Ultimately, any unpaid costs by patients are covered by the government and taxes. Especially for Medicaid patients, taxes are what pay for this coverage. Overall in 2014, Medicaid coverage accounted for 9% of the US budget.[7] This is a huge overall cost. If we could decrease by even a tiny fraction, it is definitely worth pursuing. If more expenses are incurred by those who cannot afford it, ordinary people will bear the cost. Therefore, the economic impact has direct and indirect impact on the community.

There are unquantifiable impacts to this project which we cannot put a price on. First, the impact on the working routine of the social worker whose job it is to interact with patients, understand their transportation needs, and schedule rides. Ultimately, while scheduling rides is not specifically in the job description of these workers, it falls on them to look out for their clients/patients well being and make sure they make it to their appointments. The time and effort on the part of the social worker cannot be measured. It is a waste of time and effort to contact every single patient with an appointment in a given period of time and offer them a ride. Not only will some not accept it, others may accept it and still cancel their appointment later regardless. Over time, the social worker can learn about his patients to understand their needs and who exactly to offer transportation to. Ricardo, the social worker we worked with at 37th and market said in his own words: “I know who i know; i don’t know who i don’t know.” While Ricardo might know the needs of some of his patients, he cannot possibly have memorized the needs of every patient who walks through his doors, especially those who are new to the practice or do not visit frequently. Second, the other unquantifiable aspect of this project is the cost of human life. We cannot say how much one should spend in order for a patient to reach their appointment. We never know which appointment missed could be tremendously harmful to a human life. While ideally we could spend up to the profit margin of an appointment to get patient to their appointment, this simply isn’t feasible. Therefore, there is a tradeoff between human life and social worker resources on which should be be a priority in deciding who could possibly benefit from transportation intervention.

In order to address this tradeoff, we had to consider whether we’d rather have an overage of patients offered transportation who may not need it or an overage of patients not offered transportation who do need it. Ultimately, we decided to minimize our false positive rate when predicting those who do need intervention. This means that we minimize the number offered transportation who do not need it and suffer from an overage of those not offered who could benefit. In iterating through our designs, we saw so many inefficiencies in the way that social workers gathered information and scheduled rides along with his many other tasks. Therefore, it would just be ineffective for their role and for the resources of the hospital to over offer rides. Further, our algorithm as stated previous is not a final decision. It is merely a suggestion which can be overridden manually by a social worker.

[7] https://www.investopedia.com/articles/personal-finance/082015/how-much-medicaid-and-medicare-cost-americans.asp

There is quantifiable impact of our work as well. We can estimate the profit from the total appointments scheduled as well as the total profit earned through improved attendance rates.

Profit Estimate per Appointment

The revenue of the appointment itself ranges from $155-268 according to Krisda H. Chaiyachati, MD, MPH, a primary health provider working with Penn's clinic. Each patient is assigned a risk score by the clinic to indicate their health and concern levels as determined by the clinic. This number is rather arbitrary and will differ clinic to clinic as we attempt to scale this product to other clinics. However, because this number is relative, it is likely simple to find a substitute statistic at other clinics. The patient revenues are scaled within this range according to the patient's risk score. About 50% of revenue from doctor appointments is retained as profit after overhead, most variable costs, and salaries.[8] The 50% profit that comes from a completed appointment can go toward paying for the cost of a ride to that appointment.

Cost Estimate of Transportation to Appointment

For estimation purposes, we will estimate cost of transportation using the estimate cost of a Lyft ride, given that at its most expensive, the method of transportation will be rideshare (Lyft, Uber, etc.) and at its least expensive, it would be a HUP shuttle. The estimated cost of a Lyft ride in Philadelphia is equal to $1.38 base fare + $0.20/minute * # minutes + $1.27/mile * #miles.[9] We have addresses of all patients and the address of the clinic which makes this cost easily estimated.

Overall, we found that of those who total profit obtained through offering Lyft rides to patients classified as High Priority for transportation intervention but did not make it to their appointment could have brought in a profit of about $5465. Of those who were classified as High Priority but were shown to have made it to their appointment anyway, the accumulated transportation cost was about $671. While these patients did in fact make it to the given appointment, they have the tendencies of those who miss appointments due to transportation barriers and therefore could easily miss appointments in the future. Therefore, with our product, the clinic could likely have brought in an additional $4794 in profit. While the clinic would need to forgo about $671 of unnecessary costs, this is a very small portion of the overall profit to be gained by offering transportation intervention to this small subset of the population.

[8] https://www.aafp.org/fpm/2007/0600/p19.html#fpm20070600p19-b1
[9] https://www.lyft.com/cities/philadelphia-pa

VIII. Summary of Meetings

Team and Advisor Meetings

Our first semester, our team organized standing weekly meetings with our advisor Megan Ryerson on Monday mornings at 9am. At least three members of our team would attend the meeting each week and one person would be in charge of taking detailed notes and writing a follow up email to Megan. In these meetings, our goals were to summarize the past week's progress and outline the next week's goals. These meetings were critical when we felt like we needed to realign on the overarching goal. In addition to the standing meeting with our advisor, our team met at least twice a week to work through weekly updates, deliverables, and other tasks. We set aside Wednesday from 3-4:30pm and Sun from 8-10pm to hold our team meetings.

Megan also put us in touch with a few key people who were working with her on the Lyft study done by 3701 Market (i.e., offering free rideshare transportation to Medicaid patients).

Moving into the second semester, our team organized standing weekly meetings with Megan Ryerson on Monday afternoons at 3pm. These meetings were attended by Sabrina, Claire, and Bari. Because of other class schedules, Sonia planned separate meetings with Megan that acted as R working sessions. In the weekly meetings attended by Sabrina, Claire, and Bari, our goals were to summarize the past week's progress and outline the next week's goals. These meetings were critical as we were in the requirements gathering stage and needed advice on design. In addition to the standing meeting with our advisor, our team met at least 2-3 times a week to work through weekly updates, deliverables, and other tasks. We set aside Monday from 11am-1pm, Thursday from 5:30-6:30pm, and Sun from 8-10pm to hold our team meetings.

Krisda Chaiyachati

Our main point of contact for data collection purposes has been Dr. Krisda Chaiyachati. Krisda H. Chaiyachati, MD, MPH, MSHP is the Medical Director of Penn Medicine's FirstCall Virtual Care, a VA Advanced Fellow, and Associate Fellow at the Leonard Davis Institute of Health Economics at the University of Pennsylvania. Dr. Chaiyachati studies and designs innovative strategies for improving healthcare accessibility and patient engagement. In his work, Krisda looks for opportunities to reduce social influences that can pose as a barrier to accessible health care and experiments with initiatives to remove such barriers. Krisda has been working with our advisor, Megan Ryerson, to determine if rideshare-based (e.g., Lyft) medical transportation improves show rates for appointments among low-income patients; they are also exploring the impact of rideshare-based medical transportation on healthcare utilization.

Our first meeting with Krisda took place on 10/16/17 and all team members were present for this initial meeting. In this meeting, we learned that 40-50% of Medicaid patients miss their primary care appointments. Furthermore, 20-30% say that transportation is the main reason for missing. Krisda's goal is to see if eliminating the transportation barrier would lead to fewer missed appointments. His research shows that there are many different factors that influence a missed appointment and that he suggests that the solution is more complex than simply offering a Lyft ride.

Krisda pointed out that these patients create travel habits and may not be willing to change the way in which they get to an appointment. He also mentioned that many of these patients have few transit options that are (subjectively) close enough to their home or clinic.

Krisda connected us with a few other SMEs: Mike Serpa, Ricardo Santos Martinez, and Emily Brown. We dedicated the week following our first meeting with Krisda to meeting these three SMEs and Penn Professor James Won.

Throughout the second semester, we had additional meetings via phone and email with Krisda in order to clarify the additional data received on patient attendance history and to discuss the options for attaining more data.

<u>Ricardo Santos Martinez</u>

Ricardo is the social worker at Penn's clinic at 3701 Market. Bari and Sonia met with Ricardo on 10/25/17 at his office in the clinic. We learned from Ricardo that our patient data set primarily involves older adults in West Philadelphia on Medicaid. We learned about different transportation options available to these patients through Medicaid and SEPTA, as well as some of the limitations of these transport options. We also learned that patients can be late to an appointment at least twice a week and this is largely in part due to the logistics of the Medicaid-offered NEMT services.

It also became clear that Ricardo is involved in a lot of manual work to coordinate rides for certain patient. Social workers are good at "knowing who they know, but not at knowing who they don't know". We therefore hope that our predictive algorithm will shed light on those patients that go "unknown."

Moving into second semester, Ricardo became a key stakeholder for our project a our team decided to focus our efforts on how best to help social workers more effectively schedule transportation for patients in need. Sabrina and Bari met with Ricardo on 2/6/18. In this meeting Sabrina and Bari conducted a pointed interview with the purpose of understanding Ricardo's daily activities. They then used this information to build a workflow diagram in order to identify the points throughout Ricardo's day that could be improved or made more effective. From this diagram we were able to see how complex this issue really is and how many points of conflict and stress Ricardo must face on a day to day basis. We were able to identify that our algorithm could have the largest impact in helping Ricardo identify who needs transportation intervention. Ricardo's workflow diagram can be found in our Appendix *Figure 1*.

Previously, Ricardo was only made aware of patients in need of transportation scheduled via referrals from patients, physicians, or receptionists. We decided to build an algorithm to let Ricardo be more proactive with transportation scheduling by identifying right away who would benefit from this intervention.

Professor James Won

Bari and Claire met with Professor Won on 10/25/17. Professor Won has experience working as a systems engineer at CHOP and teaches a course on human system engineering at Penn. He has a good understanding of the medical system and studies optimal healthcare designs.

Professor Won helped us comb through and evaluate the abundant information gathered in our previous SME meetings. In particular, he helped us refocus our objectives and consider a design output for the clinic or the patient.

He emphasized that the healthcare system is constantly planning around unpredictability. We concluded that there is significant value for doctors and clinic operations if patients show up to their appointments on time. More generally, this discussion begged the question: can a component of unpredictability be eliminated?

This meeting helped our team shift our focus to design. In particular, the most significant outcome was that we would now design a predictive algorithm for a patient's likelihood to miss an appointment.

Moving into second semester, we kept close contact with Professor Won to further hone down on our algorithm's goal, as well as understand how best to interview our stakeholders and design a product that was user friendly and compliant with human systems interactions. On 2/5/18, Bari and Claire met with Professor Won. This meeting was held right before our workflow session with Ricardo. Bari and Claire used this meeting to develop a game plan in terms of how to go about interviewing Ricardo effectively. Professor Won was crucial in helping guide us toward an effective algorithm and design that was focused on human engineering.

Mike Serpa

Mike is an innovation associate at the Penn Medicine Center for Health Care Innovation. Our whole team met with Mike on 11/1/17. Mike introduced us to the concept of transportation deserts. Mike also recommended we contact Imran Cronk, Penn alumni and founder of RideHealth, to better understand how rideshare services are currently helping healthcare providers.

Emily Brown

Emily Brown is a Penn Med student who conducted the qualitative interviews for Krisda's study. Claire and Sabrina met with Emily on 10/23/17. Emily was able to identify financial and organizational factors that contributed to patients' barriers to healthcare. While the conversation was interesting, it didn't feel like it necessarily added significant value to our upcoming quantitative analysis.

IX. Final Schedule with Milestones

Please see table below for our main milestones and responsibilities and reference detailed descriptions in Section X.

	Sonia	**Claire**	**Sabrina**	**Bari**
Algorithm build-out	x			
Prototype build-out		x		
Meeting w/ subject matter experts	x	x	x	x
Dashboard design requirements		x	x	x
Workflow diagrams, graphics, telling our "story"			x	x
Standards			x	x

X. Discussion of Teamwork

Our team has developed a very solid and communicative work style. Over the past two semesters, we have been able to determine each member's strengths and weaknesses and feel that everyone picks up her share of the work and complements each other in a team setting. Sonia has taken the lead in our algorithmic build up in R. She was often in close contact with our advisor in order to work through the more technical components of our project. Claire has led the prototype design effort in Justinmind. Through self learning and close collaboration with Sonia, she has been critical in our design phase. Sabrina has been expert in design and communications with our advisor. Her aesthetic and organizational skills have been very helpful throughout the semester. Bari has been key in communicating with our stakeholders. She is great at planning meetings and interviewing subject matter experts in order to obtain necessary data and insights to help drive forward the design of our project. Despite these clear contributions, all members were involved in all activities throughout the course of this project. Overall we feel that we are honest and comfortable with one another as well as enjoy each other's company.

XI. Budget and Justification

The Justinmind license provided us with premium features and was crucial in creating a realistic and data-driven prototype that had all the features and functionality necessary. Additionally, we were able to share the prototype publically with a link so we could perform UX/UI tests and gather user feedback easily. The books were immensely helpful in strengthening our understanding of various topics. Most important, the book on R and building out predictive models assisted us in answering some nuanced questions related to cleaning up the data and creating our algorithm. The book on Tableau helped us visualize our data and create the interactive maps which allowed both our team and our peers to better understand the geographic spread and various characteristics of our data.

Justinmind Software Monthly License	$58.00 (2 mo. at $29/mo.)
R Student License	$0.00
Tableau Student License	$0.00
R for Data Science (Book)	$30.29
Communicating Data with Tableau (Book)	$33.69
Applied Predictive Modeling (Book)	$55.00
Total	**$176.98**

XII. Work for Second Semester

During the spring semester, our team focused in on our main deliverables. We clarified our objective to create a tool to be used by social workers that will identify patients that would benefit from transportation. This was done through numerous meetings with our advisor, Megan Ryerson, and the social worker at 3701 Market, Ricardo Santos Martinez. Milestones included creating a workflow diagram of Ricardo's daily activities, identifying the problem that we want our algorithm to address, data mining, building up our algorithm in R, and finalizing our prototype in Justinmind. We decided to switch from displaying interactive maps in Tableau, to designing a prototype in Justinmind to be used by social workers. This prototype includes the results of our algorithm and is an easy way for social workers to identify their "high priority" patients.

XIII. Standards and Compliance

Our senior design project had to comply with three main standards: Algorithmic Bias Considerations (IEEE), The Security Rule (HIPAA), and The Privacy Rule (HIPAA).

Algorithmic Bias Considerations (IEEE 7003) is designed to ensure that organizations designing algorithms can clearly articulate how their algorithm targets and influences the users and stakeholders of said algorithm. Moreover, this standard certifies that creators communicate the use of best practices and testing methods to users and regulatory authorities so as to avoid "unjustified differential impact" on stakeholders[10]. This standard played an important role in framing our prototype design and requirements. For instance, the dashboard allows for social workers to override any priority status assigned to a patient by the algorithm. This is crucial in avoiding unjustified targeting of patients, especially in this particular setting; that is, social workers may learn new information about patients over time that can influence one's need, desire, or disinterest in transportation scheduling, thus making revision of algorithm results an important feature of our prototype. The Predicare tool is an effective vehicle in aiding decision-making. As such, it is an adjunct to a human-made decision and not a stand-alone product. In keeping IEEE 7003 in mind, we acknowledge the need for human input in this particular setting.

The HIPAA Security Rule (45 CFR Part 160 and Subparts A and C of Part 164) is a national standard that ensures protection of electronic health records that are "created, received, used, or maintained by a covered entity"[11]. As such, HIPAA requires administrative, technical, and physical security measures to protect electronic health information. Similarly, the HIPAA Privacy Rule (45 CFR Part 160 and Subparts A and E of Part 164) requires safeguards to protect the transaction of health information between parties[12]. This entails setting limits on the use and disclosure of patient information without patient authorization and giving patients rights over their personal and sensitive information. These standards limited our access to data during the design process (i.e., all information was unidentifiable and we couldn't pull live data from Epic, the clinic's electronic medical records software). Going forward, these standards will affect the use and implementation of Predicare by a clinic. If the dashboard is integrated as an Epic functionality, Epic and the respective hospital or clinic will manage the location of data storage. Access management also becomes critical; it is important to manage who has access to the data. In other words, only a social worker's login should allow for access to our feature page. Login tracking and monitoring will be managed by hospital IT, and so will be the ability to terminate someone's login information.

[10] https://standards.ieee.org/develop/project/7003.html
[11] https://www.hhs.gov/hipaa/for-professionals/security/index.html
[12] https://www.hhs.gov/hipaa/for-professionals/privacy/index.html

XIV. Conclusion

In conclusion, we have developed a prototype to be used by social workers that uses our predictive algorithm to identify patients that would benefit from scheduled transportation to primary care appointments. This prototype is an output of our algorithm, interviews with stakeholders, and knowledge accumulated over this past year.

Some challenges that we foresee facing include selling this software to Epic and having it be properly integrated and incorporated within the patient management system. We will use our technical specifications and detailed HTML and CSS code to help this process run as smoothly as possible. Additionally, because of the complexity of this problem, many factors may come into play that our algorithm does not account for. In such cases we rely on the social worker's experience and judgement to make final decisions regarding these patients.

This project was an incredible learning opportunity. We learned about the complexities of the US healthcare system, the obstacles involved when working with low-income populations, and the rigorous tasks held by a social worker. We learned of the importance of interviewing stakeholders and creating a user friendly experience when dealing with human and system design. Lessons regarding standards allowed us to better understand our limitations and constraints within our design process. These lessons also taught us that each industry is unique in its requirements and standards. We learned of algorithmic bias, privacy and security standards, and data norms in the healthcare and transportation industries.

We hope to continue to foster this knowledge of the healthcare and transportation industries to further our careers and allow us to ultimately provide beneficial impact down the road.

XV. Appendices

Note: the technical specifications report, HTML code, prototype .vp file, and R code for algorithm are all included in .zip file in report submission on Canvas.

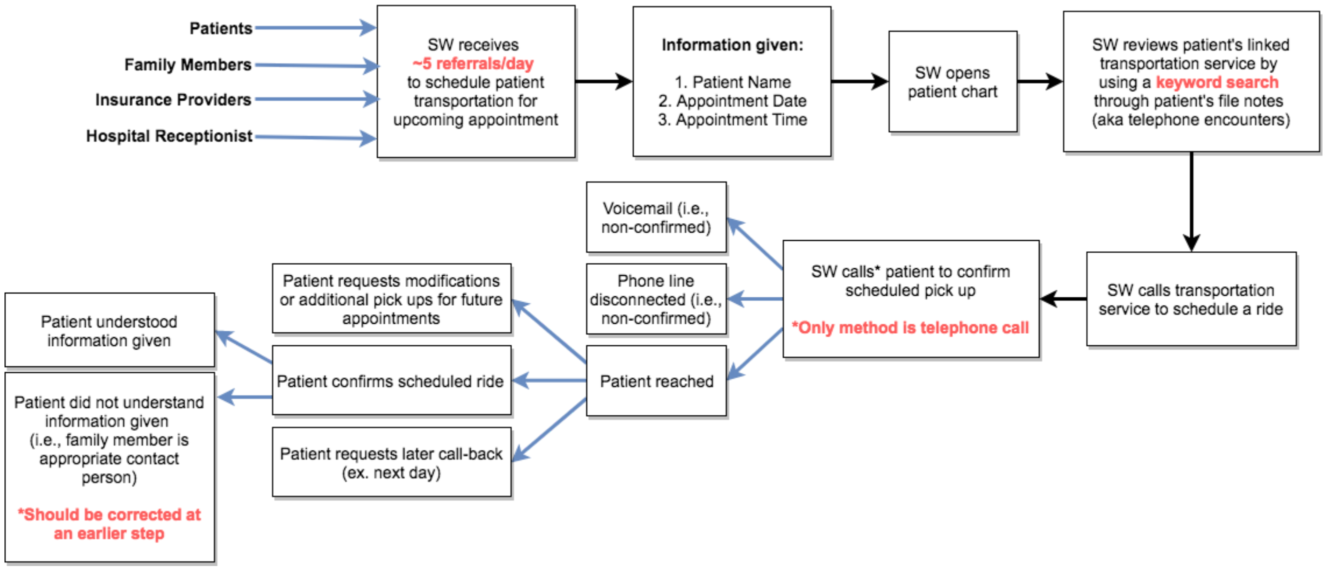

Figure 1: Ricardo's Workflow Diagram

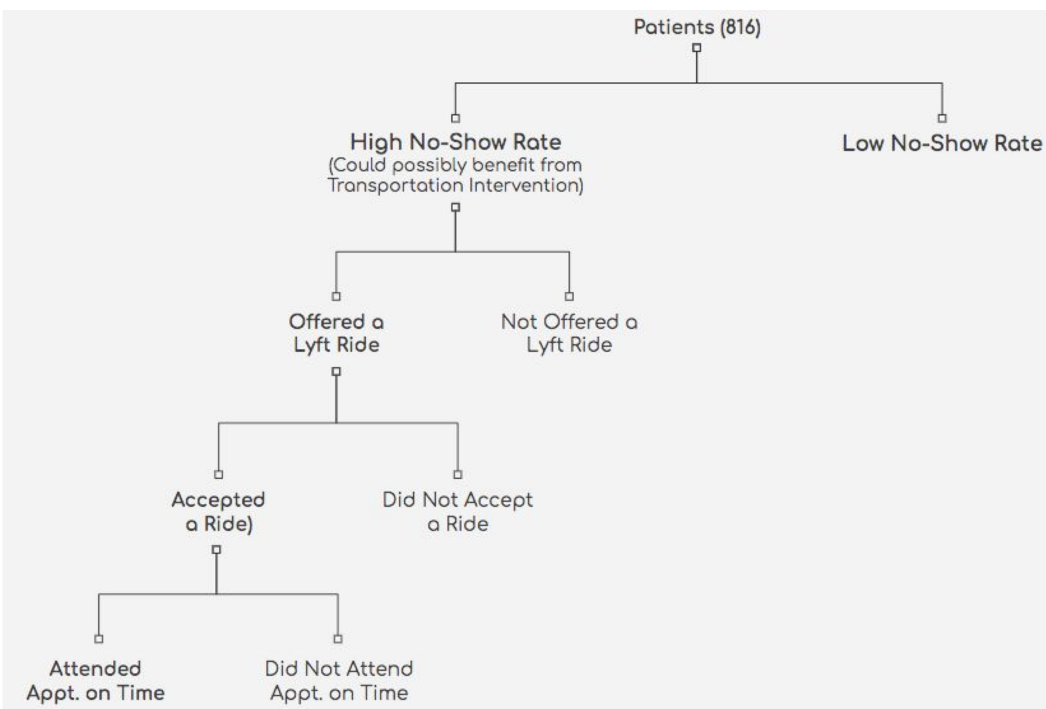

Figure 2: Data Tree used in Algorithm

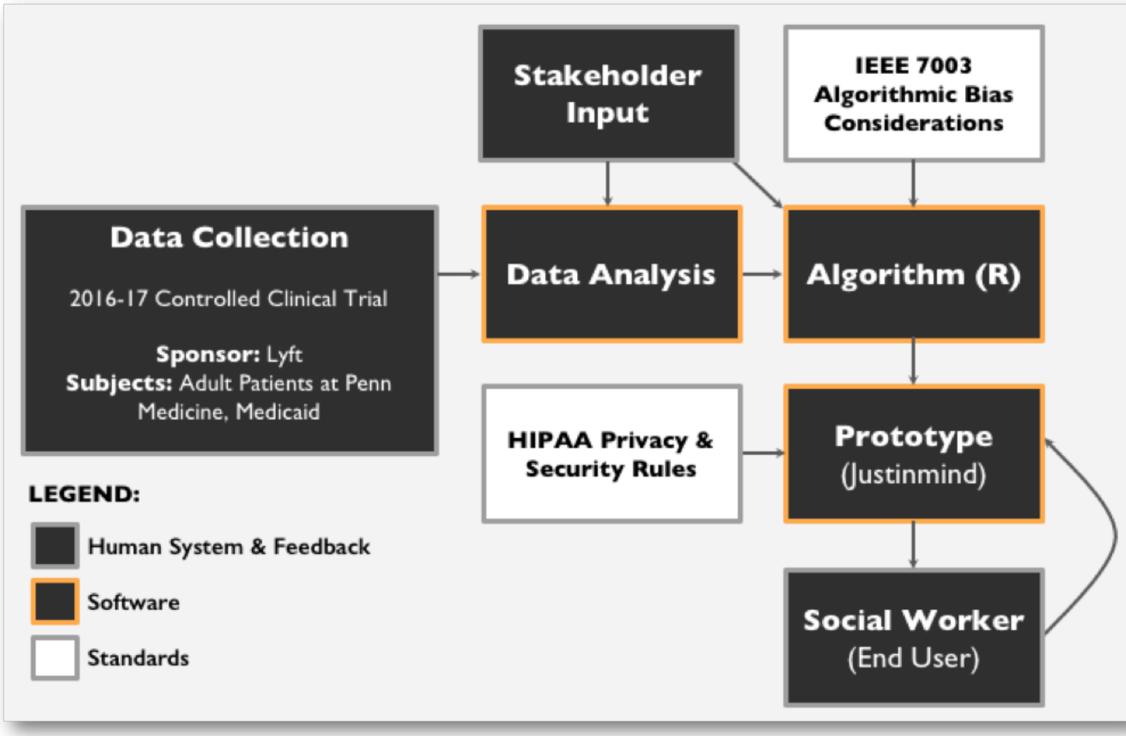

Figure 3: High Level Design Flow

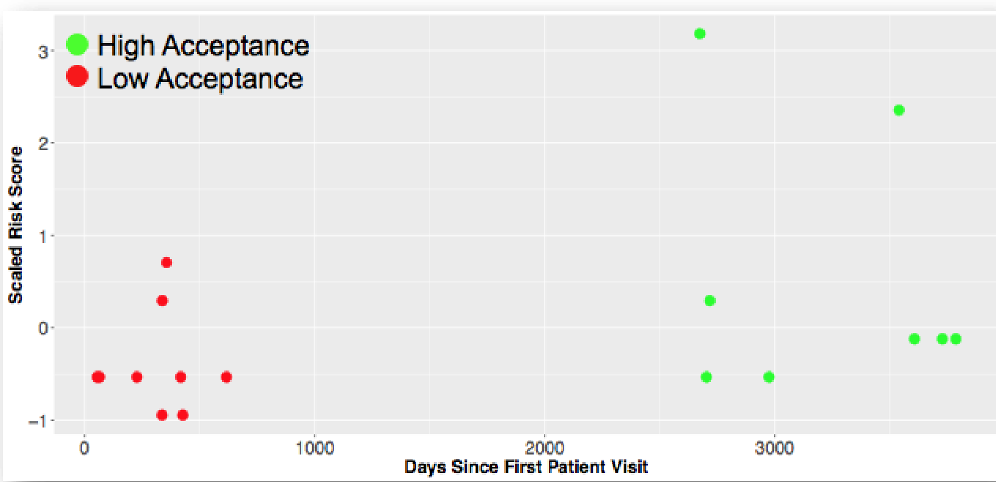

Figure 4: Results of clustering analysis for high and low acceptance patients along the variables of Scaled Risk Score and Days Since First Patient Visit

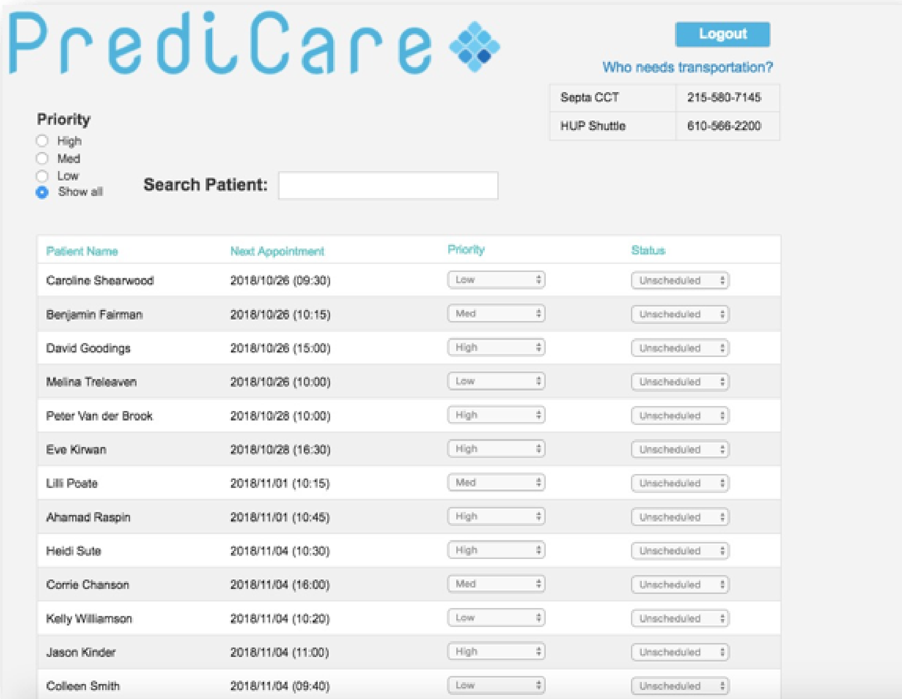

Figure 5: Prototype Home Screen

Figure 6: Prototype Patient Information

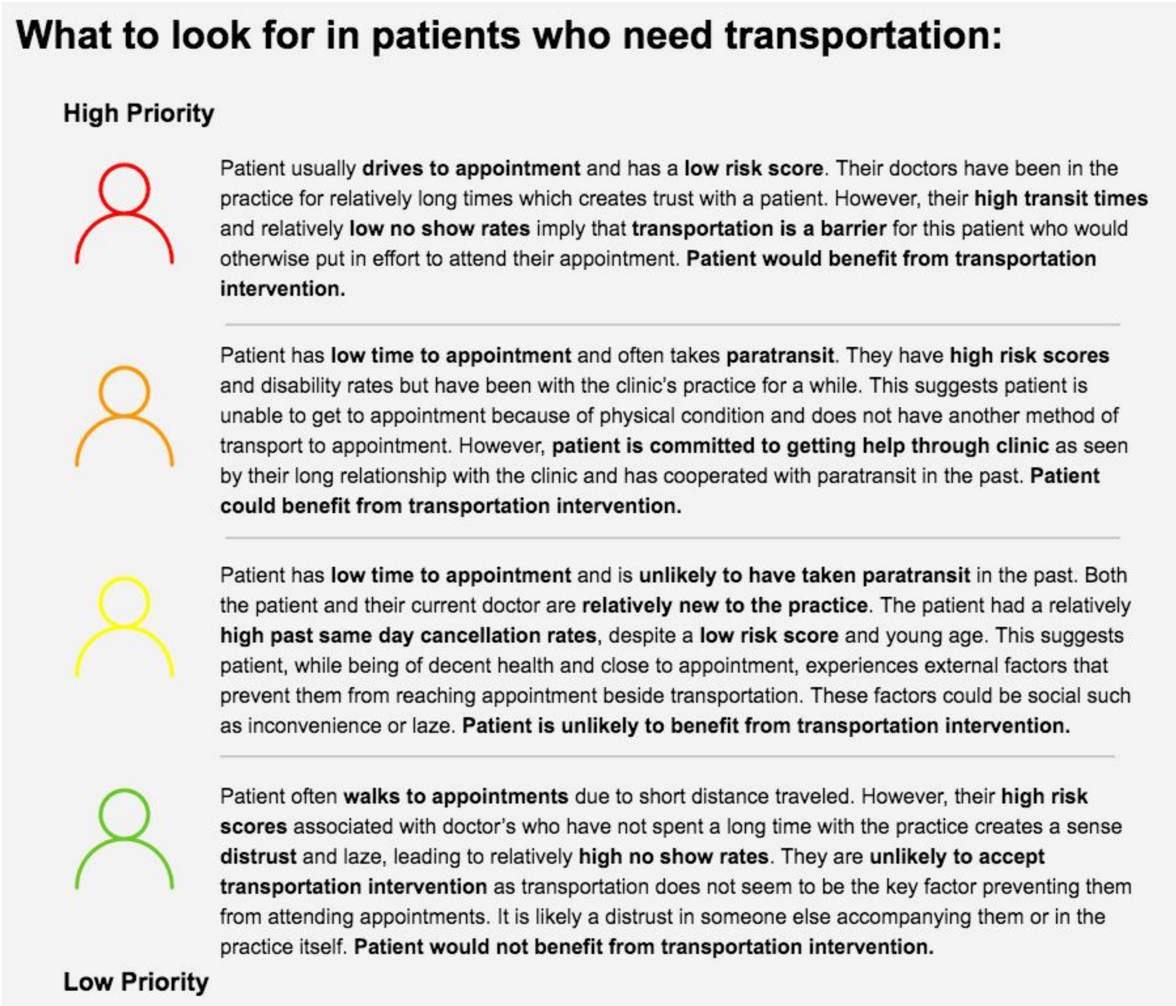

Figure 7: Patient Profiles

Significant Variables:

i. Transit time: 0.06
ii. Driving Time: 0.24
iii. Walking Time: 0.22
iv. Transit Distance: 0.16
v. Driving Distance: 0.23
vi. Walking Distance: 0.21
vii. Number of Bus Stops Available: 0.27
viii. Late Rate (yr): 0.43
ix. Late Rate (6m): 0.41
x. Age: 0.40
xi. Months Doctor has been in practice: 0.17
xii. Drives self as method of transportation: 0.13
xiii. Walks as method of transportation: 0.18
xiv. Employment Status (unemployed): 0.41
xv. Employment Status (unemployed because of disability): 0.31

Insignificant Variables:

i. SDC Rate (6m): 0.53
ii. SDC Rate (yr): 0.70
iii. NS Rate (yr): 0.67
iv. NS Rate (6m): 0.85
v. Days since first visit: 0.83
vi. Risk Score: 0.87
vii. Bus Frequency: 0.90
viii. Emergency Hospital Admissions: 0.73
ix. Takes paratransit as method of transportation: 0.68
x. Takes Taxi as method of transportation: 1
xi. Gender: 1

Figure 8: Details of Initial T-Tests

Transit Time, Driving Time, Walking Time, Transit Distance, Driving Distance, Walking Distance, Number of Bus Stops Available, Late Rate in Past Year, Late Rate in Past 6 Months, Age, Months Patient's Doctor has been with the Practice, Whether or Not Patient Drives Self, Whether or not Patient Walks to Appointments, Patient's Employment Status, Patient's No Show Rate in the Past Year, Risk Score

Figure 9: Variables Included in Initial Model