Semester Project: Oregon Health Insurance Experiment Analysis

Alec Wick and Colin Buckley

College of Business Administration, University of Nebraska at Omaha

ECON 8300-820: Econometrics

November 27, 2023

Semester Project: Oregon Health Insurance Experiment Analysis

Project Introduction

Motivation & Purpose

The motivation of this paper is to draw insights from the Oregon Healthcare Insurance Experiment (OHIE), that can be used to improve healthcare delivery in the United States (Baicker, K., & Finkelstein, A. 2015). This experiment analyzed a group of uninsured lower-income adults in the state of Oregon who applied for a Medicaid lottery. Of the roughly 75 thousand people who applied, roughly 30 thousand were selected and were able to apply for Medicaid. The health and welfare of these individuals were then tracked over the next several years. The study presents a unique opportunity to draw statistically significant findings because the dataset is a randomized controlled trial, the gold standard for scientific studies.

This project will examine whether or not pre-existing health data can be used to accurately predict emergency department visits for insured and uninsured individuals, and if there are differences in the health outcomes for English compared to non-English speakers. We know from current research that (1) emergency department utilization can lead to improvement in healthcare outcomes and, (2) implicit bias exists against racial and ethnic minorities, and that minority populations are less likely to speak English than non-minority populations (Dennett, J. M., & Baicker, K. 2022) (Dowd B, Karmarker M, Swenson T, et al. (2014).

Many follow-up studies and conclusions have already been drawn based on the Oregon Healthcare Insurance Experiment data, and this research is added to that list. Like previous follow-up studies, our examinations can inform future healthcare research and direct resources to improve Medicaid for those enrolled in it.

Research Questions

For this research project, two questions are proposed for evaluation (1) the accuracy of predicting emergency department visits for patients who won the Medicaid lottery and (2) the health outcomes for Medicaid lottery winners in the Oregon Healthcare Study who speak English compared to those who speak a language other than English. The two questions are outlined below.

- Question 1: Are medical providers able to accurately predict if lottery winners will visit the emergency department during the survey period based on preexisting conditions?
- Question 2: Of the Medicaid lottery applicants observed in the Oregon Healthcare Study, did English speakers have greater health improvement than non-English speakers?

For these questions, we will be observing similar variables that outline the pre-existing conditions of those who applied to the Oregon Medicaid lottery. After the State of Oregon chose its lottery winners, the original investigators of the Oregon Health Insurance Experiment tracked the health of both lottery winners and losers, over the next several years. Both lottery losers and winners were observed to maintain a control group.

From the total population of applicants, the investigators drew a random sample and evaluated their health via several methods. From these data sets, we will be focusing on emergency department data, pre-existing conditions (such as hypertension, diabetes, cancer), whether these conditions were diagnosed during the Medicaid trial, spoken language, health status and improvement, and lottery status.

Research Population

The population of the Oregon Health Insurance Experiment consists of nearly 75 thousand people who applied for Medicaid when the lottery was offered in the state of Oregon in 2008 (Baicker, K., & Finkelstein, A. 2015). Of these 75 thousand applicants, about 30 thousand were randomly selected and provided Medicaid coverage. This random selection creates a natural control group and experimental group, between the lottery winners and losers. During this study, information was gathered about both the lottery winners and lottery losers through several different mechanisms at several different times. The various mechanisms are listed below:

- Application Demographic Data: the data gathered when individuals applied for the lottery, including demographic information, language spoken, age, and whether or not the individual won the lottery
- State Program data: the data gathered when individuals applied for the lottery, that details the state programs that an individual is enrolled in, such as Supplemental Nutrition Assistance Program or SNAP, Medicaid, and Temporary Assistance for Needy Families or TANF.
- Mail Surveys: surveys to collect information about the status of an individual's health were mailed out at initial Medicaid enrollment, 6 months after enrollment, and 12 months after enrollment. These surveys analyze healthcare utilization, like visits to the doctor, prescription drugs used, visits to the dentist, and whether an individual received the care that they needed.
- In-person Surveys: the in-person surveys were conducted with individuals roughly 12 months after enrollment. This survey also collected information about the health of the individual, as well as the self-assessed change in health over the last 12 months.
- Emergency Department Data: this data set contains administrative data from 12 hospitals in the Portland, OR area. It captures the emergency department visits by lottery and non-lottery winners, the reason for the visit to the emergency department, and whether the visit was emergent or non-emergent.

This paper will focus on the demographic data, the in-person surveys, and the emergency department data, to answer the questions posed in the introduction. As previously mentioned, these various data sets contain information on pre-existing conditions, emergency department visits, demographic data, health improvement, starting health, lottery status, and language spoken. These variables allow for a robust analysis of the role language plays in healthcare outcomes and the factors that impact whether or not a person will visit the emergency department in the upcoming year.

It is important to recognize the integrity and potential biases of the data from the original study, The Oregon Health Insurance Experiment (Baicker, K., & Finkelstein, A. 2015). Though this data is the gold standard for statistical analysis, with its randomized controlled design, there are some biases in the data. This data can be utilized for additional research, as many other researchers have, and the findings from additional research must highlight the limitations of the dataset.

Overall, the random sampling of the lottery winners compared to the lottery applicants does not have a statistical significance in the demographic areas of age, race/ethnicity, household income, education level, or employment status, as shown in the study "Medicaid, Health, and the Moderating Role of Neighborhood Characteristics" (Dennett, J. M., & Baicker, K., 2022). This information is illustrated in Table 1 below. Because there are no significant differences between the individuals selected for the lottery, and those not selected for the lottery, inferences drawn comparing these two groups also hold statistical significance.

Table 1

Characteristics of the 8413 respondents in the analytic sample

Demographic characteristics			
Age (Years)	40.4	40.6	0.44
	(11.5)	(11.6)	
Female (%)	56.7	56.2	0.62
Race/ethnicity			
White (%)	66.2	68.0	0.10
Black (%)	12.3	12.3	0.98
Hispanic (%)	18.0	16.7	0.21
Other (non-White) (%)	14.7	14.9	0.82
Household income	18,129	18,091	0.91
	(13,434)	(13,259)	
Education			
Less than high school (%)	21.4	20.5	0.39
High school diploma or GED (%)	47.0	46.1	0.49
Post-high school (%)	21.6	23.0	0.18
Four year degree or more (%)	10.0	10.4	0.62
Employment status			
Not currently employed (%)	50.5	51.3	0.50
Employed less than 20 h a week (%)	10.0	9.6	0.64
Employed 20 to 30 h a week (%)	11.1	11.6	0.56
Employed more than 30 h per week (%)	28.4	27.6	0.43
F-statistic for above variables	0.611		
P value	0.843		

(Dennett, J. M., & Baicker, K. 2022)

While the lottery winners do not vary from the applicants, the lottery applicants do vary from the general population, according to Census data in 2010 (U.S. Census Bureau, 2010). The applicant population in the Oregon Health Insurance Experiment is older, more racially and ethnically diverse, less educated, less wealthy, and less employed than the Census data two years later in 2010. This is stated at the outset of the original study, as the lottery was focused on a "group of uninsured low-income adults in Oregon" (Baicker, K., & Finkelstein, A. 2015).

The differences between the application population and the general population raise questions of bias. The applicant population is less likely to have insurance before the lottery than the general population, hence why they originally were motivated to apply to the lottery. Because this group is less likely to be insured, critics of universal healthcare state that the difference between private medical insurance and Medicaid is not captured.

Literature Review

Prior research provided important observations that informed our research on both emergency department visits and language and how both relate to healthcare outcomes. The study, "Unconscious (implicit) bias and health disparities: Where do we go from here?" compiles and analyzes the current research on the impact of implicit racial and ethnic bias in medicine, specifically surrounding race and ethnicity (Blair, Steiner, and Havranek, 2011). Many studies have been conducted observing implicit bias, and this study focuses on seven cornerstone studies, to provide a roadmap for clinicians to navigate bias in medicine. These studies have shown that minorities, specifically African Americans, are less likely to receive pain medication, less likely to receive hypertension medication (when blood pressure is shown to not be in control), and generally have poorer healthcare outcomes, when compared to non-minority individuals, among other findings. This is relevant to the research in this paper, as there is significantly more research about race and ethnicity, compared to language, in healthcare. However, we do know that non-English speakers are more likely to be minorities, and this study drives research toward questions on how various minority groups, including language minorities, receive differential healthcare outcomes.

There is far less current research about language's impact on healthcare outcomes than there is for race and ethnicity impact on healthcare outcomes. One of the landmark studies observing language in healthcare is " Improving the Health and Health Care of Non-English-Speaking Patients" an article in the Journal of General Internal Medicine, (Taira, D. A. 1999). This research also references that there are not many studies related to language spoken in healthcare. In contrast to research on racial and ethnic bias, this study finds that non-English speakers, the majority of whom are Latino and Asian in the United States, are healthier than their English-speaking counterparts. However, when separated by variables such as country of origin, education level, employment status, and income, the health of English compared to non-English speakers converges along these variables. The OHIE data is unique in that it controls for income level, as the lottery application was restricted to “low-income adults” (Baicker, K., & Finkelstein, A. 2015). Because non-English speakers are a non-homogenous group, it is advised by Taira to observe the group by the previously named variables, and additional research into language’s effect on health outcomes is needed.

A key study related to this investigation of healthcare outcomes for people enrolled in Medicaid, compared to those who are not, is “Medicaid, Health, and the Moderating Role of Neighborhood Characteristics” (Dennett, J. M., & Baicker, K., 2022). This study combines the Oregon Health Insurance Experiment data with neighborhood data to allow for a new perspective on the data from the original study. However, the study does not find that “neighborhood characteristics substantially affect the relationship between gaining insurance and health outcomes,” indicating that Medicaid expansion would have a positive impact across geographies and socioeconomic status. The variables observed in the “Neighborhood” study include information on land use, proximity to grocery stores, education, and race/ethnicity. However, this data did not include a variable for the primary language spoken by the neighborhood.

One of the key variables that we will be examining in this analysis of the data in the Oregon Health Insurance Experiment is emergency department visits (Finkelstein et al., 2012). Visits to the emergency department are often pointed to as an indicator of a failing healthcare system - the more often that people need to go to the emergency department, the less likely they are to get access to preventative care needed to avoid those visits. The study, “Emergency Department Utilization as a Measure of Physician Performance” (Dowd B, Karmarker M, Swenson T, et al., 2014), proposes that there is a relationship between poor physician care and an increase in emergency department visits; the less proficient the care that a patient receives from their primary physician, the more likely they are to need a future visit to the emergency department. This research also highlights that individuals who are not insured or who have recently received insurance are more likely to utilize the emergency department as their initial connection to the healthcare system, instead of primary care visits. The research in this paper will not attempt to analyze physician competence, though it may be cited as a bias in the dataset we are utilizing. However, we will review the amount of emergency department visits as an indicator of healthcare utilization. Emergency department visits are costly, and people are more likely to visit if they have the coverage for the cost of the visit.

It is important to observe the role cost plays for both individuals and communities. The primary objective of the study, “Universal Healthcare in the United States of America: A Healthy Debate” (Zieff, G., Kerr, Z. Y., Moore, J. B., & Stoner, L., 2020) was to identify the pros and cons of universal healthcare specifically in the United States. One of the main arguments that the paper discussed that oppose universal healthcare is the significant upfront costs. The argument that supports universal healthcare is that Americans will live healthier lives which would in the long term lessen the burden of healthcare costs. This article is important context to the Oregon Healthcare study because it discusses the pros and cons of universal healthcare outside of healthcare outcomes alongside the potential health benefits. The Oregon Health Insurance Experiment found significant improvement in the healthcare outcomes of the Medicaid lottery winners. It is important to keep in mind the policy implications, including cost as well as health benefits.

The article “An Assessment of the New York Health Act: A Single-Payer Option for New York State” discusses the New York Legislature’s consideration of the then-New York Health Act (NYHA) (Liu, J., White, C., Nowak, S., Wilks, A., Ryan, J., & Eibner, C. 2018). The objective of this act is to provide healthcare benefits to all residents who live in the state of New York. This act would remove all payments for medical services and even the associated fees such as copays. The legislature believes that providing universal health care upfront will lower healthcare costs in the future while also improving the quality of life for New York residents. In this study, it was identified that the funding for this act would come from higher taxes and state financing through the federal government. This article provides additional context and sentiment surrounding a state-wide universal healthcare initial, and that cost is a major concern, though research shows that it will lower the long-term costs of private insurance care.

The chapter titled “Socioeconomic Status and Heath” in the text “Social Epidemiology” (Glymour, M. M., Avendano, M., & Kawachi, I., 2014) is a publication that highlights the causal relationship between income and health. Several studies show that the more income an individual has the healthier they are. Interestingly, there are potentially some flaws with this assumption. The chapter highlighted some points such as higher income individuals could potentially prefer certain food groups or exercise levels that would lead to better health. In this case, income doesn’t necessarily cause better health or better access to healthcare, but income can allow for healthier habits than individuals with lower incomes. Other factors are more accurate indicators of health, such as zip code, and it is crucial to examine the underlying factors that influence the difference between the health of one group compared to another.

Project Methodology

Analysis

Our analysis for this project is split into two sections, to address the two questions that are posed at the outset of the project:

- Question 1: Are medical providers able to accurately predict if lottery winners will visit the emergency department during the survey period based on preexisting conditions?
- Question 2: Of the Medicaid lottery applicants observed in the Oregon Healthcare Study, did English speakers have greater health improvement than non-English speakers?

Within these analyses, we observed several pre-existing condition variables such as age, health condition before the lottery, and pain within the last four weeks as independent variables. Additionally, for the research analysis regarding language spoken, we utilized language as an independent variable along with the above named independent variables, with change in health over the last twelve months as the dependent variable.

Emergency Department Analysis

The primary objective of the Emergency Department Analysis is to determine if healthcare lottery winners will need to visit an emergency department during the duration of the study. This information can then be utilized by medical professionals to help determine staffing and care in emergency departments. To accomplish this goal, we used health condition information we obtained from the lottery survey and created a logistic regression model. Of the nearly 75,000 applicants, 29,834 were selected to be included in the study. Table #1 - Lottery Selection, shown below shows a numerical and visual representation of the population.

Table #1 - Lottery Selection

Selected	Count
No	45,088
Yes	29,834

Before the start of the trial, a survey was conducted that asked all applicants about several health-related conditions. These questions included a wide variety of health questions such as emergency department visits, chronic illness, overall health, and standard body metrics such as age and weight. Unfortunately, of the nearly 30,000 selected applicants only 5,000 completely answered the health-related questions. When developing the logistic regression model, the pool of participants only includes 5,000 individuals since NA values are not permitted in the model.

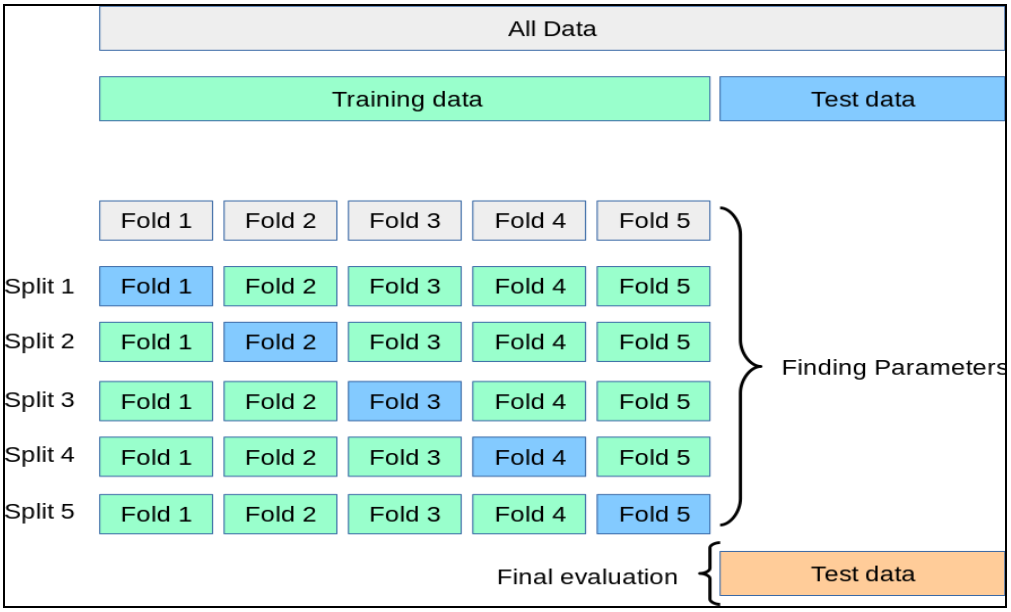
The first step of the modeling process involved establishing the testing and training datasets. We used a random function in R to split the data into training and testing sets. The variables we used within the model include: age, health during the past 12 months, level of bodily pain, emphysema, depression, asthma, diabetes, hypertension, cholesterol, heart attack, congestive heart failure, kidney failure, and cancer. The model results from this first run are displayed below:

term	estimate	std.error	statistic	p.value
(Intercept)	0.02845906	0.226301654	0.1257572	8.999241e-01
age_inp	-0.00878482	0.003354387	-2.6189046	8.821262e-03
health_last12_inp	-0.23198574	0.037532535	-6.1809237	6.372763e-10
sf4_inp	0.17245994	0.030078017	5.7337535	9.823207e-09
emp_dx_pre_lottery_inp	0.35677054	0.233953677	1.5249623	1.272685e-01
dep_dx_pre_lottery_inp	0.47842314	0.078054595	6.1293398	8.824448e-10
ast_dx_pre_lottery_inp	0.30087467	0.092298171	3.2598119	1.114861e-03
dia_dx_pre_lottery_inp	0.24548414	0.144441954	1.6995349	8.921844e-02
hbp_dx_pre_lottery_inp	0.07055793	0.103209922	0.6836352	4.942056e-01
chl_dx_pre_lottery_inp	-0.37961843	0.122833403	-3.0905147	1.998099e-03
ami_dx_pre_lottery_inp	0.69448180	0.270435753	2.5680103	1.022841e-02
kid_dx_pre_lottery_inp	0.29930270	0.258557043	1.1575886	2.470319e-01
cancer_dx_pre_lottery_inp	0.31632148	0.171996167	1.8391193	6.589764e-02

As you can see from the model, several variables are not significant such as emphysema, diabetes, and kidney disease. The model was then updated to remove the non-significant predictors and then retrained. The results of model #2 are below:

term	estimate	std.error	statistic	p.value
(Intercept)	0.018714275	0.225770382	0.08289074	9.339384e-01
age_inp	-0.007779915	0.003259599	-2.38677078	1.699709e-02
health_last12_inp	-0.238015979	0.037355199	-6.37169619	1.869489e-10
sf4_inp	0.175592306	0.030029919	5.84724543	4.997799e-09
dep_dx_pre_lottery_inp	0.486869347	0.077794376	6.25841319	3.889143e-10
ast_dx_pre_lottery_inp	0.318931222	0.091492975	3.48585477	4.905673e-04
dia_dx_pre_lottery_inp	0.259486322	0.142736244	1.81794277	6.907288e-02
chl_dx_pre_lottery_inp	-0.355030465	0.119382062	-2.97390127	2.940397e-03
ami_dx_pre_lottery_inp	0.721874521	0.268622135	2.68732330	7.202720e-03
cancer_dx_pre_lottery_inp	0.330638539	0.171080190	1.93265239	5.327903e-02

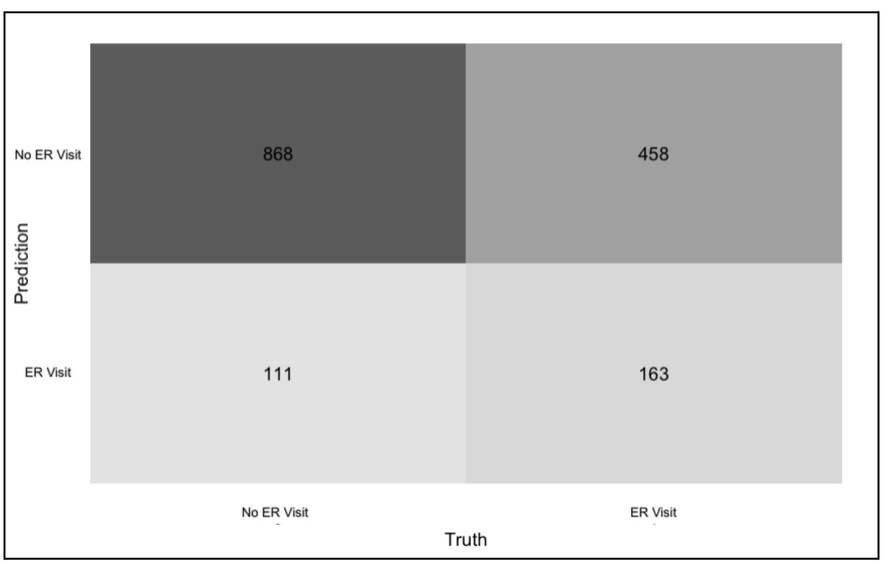
The predictor variables of the revised model are now all significant. The next step is to perform k-fold cross-validation to ensure our model is properly trained. K-fold cross-validation works by splitting the training dataset into a selected number of subgroups in which the model is trained and validated on each fold. The diagram displayed below from SKlearn provides a great example:



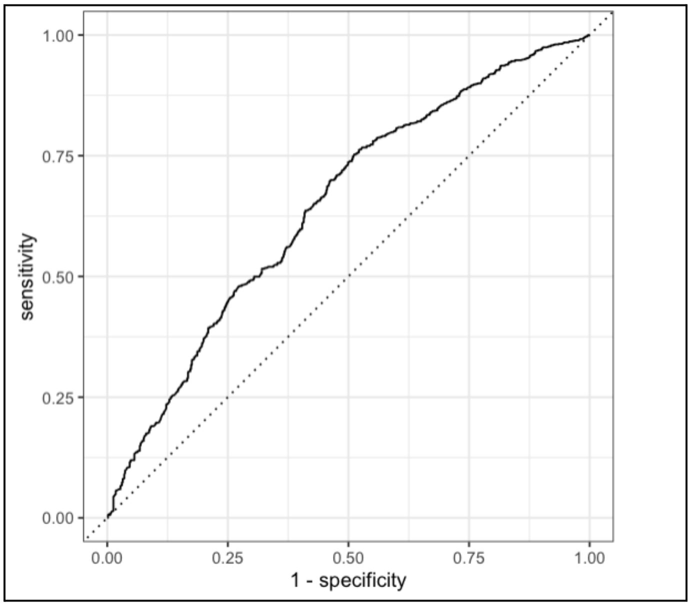
Scikit-Learn

In this example, we decided to have 10 folds used for cross-validation. The average accuracy for cross-validation for the folds is 67.6% and the area under the receiver operating characteristics curve (ROC AUC) value is 65.7. These two metrics will be referenced later to determine if our trained model can properly predict using the original testing data.

The next step is to use the trained model to predict if the lottery winners in the testing set will visit the emergency department during the survey period. Each record is run through the model and a prediction class is defined with a probability. The confusion matrix below shows the prediction results:



The accuracy is 64.4%, the sensitivity is 88.7% and the specificity is 26.2%. It's important to note that the accuracy of 64.4% is similar to the accuracy of the cross-validation results. This indicates that the model performed as expected with the testing data. Another key model performance indicator is the ROC AUC. This metric helps us identify how much better the model is than randomly guessing. In the case of this model, the receiver operating characteristic curve (ROC) is .66 which is visually represented below:



From this analysis, we were able to discover that a logistic regression model was able to predict with relatively accurate success if a lottery winner will need to visit the ER during the testing period. This information can be helpful to medical providers to keep a closer watch on those who may need an emergency department visit in the next year.

An important aspect to consider with this logistic regression analysis is the variable causality. Causality is the relationship between predictor variables. In this case, causality would include any preexisting health condition that may be correlated with another preexisting health condition. Causality analysis is essential for determining underline relationships that may exist within the model. We analyzed the correlation between all of the predictor variables and discovered that some causality exists:

Variable 1	Variable 2	Correlation
High Blood Pressure	Age	0.310
High Blood Pressure	High Cholesterol	0.348
Cancer	Age	0.113
Diabetes	High Cholesterol	0.31

Several of the predictor variables are correlated with each other. From this discovery, we can see that high blood pressure and High Cholesterol may be linked with several health conditions. It is important in logistic regression analysis and especially health studies to analyze the causality between each of the variables.

Language Analysis

The primary goal of the Language Analysis is to determine if there is a difference between the healthcare outcomes of English-speaking lottery applicants compared to non-English-speaking lottery applicants. The existing literature on racial and ethnic bias finds that minorities have worse healthcare outcomes than their non-minority peers, though data related specifically to language is limited. It is known that non-English speakers are less likely to receive treatment, but not the difference in treatment and health compared to English speakers (Taira, D. A. 1999). Since minorities are less likely to speak English, we wanted to determine if there is a gap in healthcare outcomes based on spoken language.

To perform calculations of means and regressions on the pre-existing condition variables, a "yes" response was recoded to a "1", while a "no" response was recoded to a "0". Also, for the variable measuring health from twelve months ago (health_last12_inp), the scale of "Excellent, Very Good, Good, Fair, Poor, and Very Poor," was converted to respective values of (3, 2, 1, 0, -1, and -2); responses of "don't know", "prefer not to answer" were coded as "0". Also, the values for the "Avg Perceived Change in Health vs. 12 Months Ago", were aggregated by taking the variable showing the perceived change in health (health_change_inp) and re-assigning the values of "Better, Worse, and About the Same," to "1, -1, and 0" respectively.

Before discussing the regression analysis of the independent variables of pre-existing conditions and language, and their effect on the dependent variable, the change in health over a twelve-month period, it is important to outline the differences between various groups within the population sample. Two simple t-tests, first, comparing language spoken to initial health assessment and second comparing language to change in health were conducted. These results established that there is no difference in the means of the initial health assessment but there is a significant change in health between the two groups.

The major difference between the two groups is their rates of pre-existing conditions. Non-English speakers are significantly less likely to have a diagnosis for the majority of the conditions observed in the OHIE study. The pre-existing conditions and initial self-health assessment, the means and averages for English compared to non-English speakers are shown below, alongside the four conditions that were recorded at the end of the study. As you can see below in the "Diagnosed Conditions Before Lottery" and "Diagnosed Conditions After Lottery" tables, the non-English speaking group had a greater improvement in health, driven by the non-English speaking lottery winners who saw the greatest health improvement.

Diagnosed Conditions Before Lottery			
Conditions	English_speaker	Non_English_speaker	Difference
Asthma	20.80%	6.07%	14.73%
Diabetes	7.06%	7.77%	−0.71%
Hypertension	18.94%	11.57%	7.37%
High Cholesterol	12.74%	12.06%	0.68%
Heart Attack	2.12%	0.73%	1.39%
Heart Failure	1.19%	0.40%	0.79%
Emphysema	2.38%	0.65%	1.74%
Kidney Failure	2.38%	0.65%	1.74%
Cancer	4.58%	1.54%	3.05%
Depression	36.41%	13.19%	23.23%
Overall Health	76.90%	69.42%	7.49%

Diagnosed Conditions after Lottery			
Condition	English_speaker	Non_English_speaker	Difference
Diabetes	1.56%	2.35%	−0.78%
Hypertension	5.78%	4.85%	0.92%
High Cholesterol	6.19%	7.28%	−1.10%
Depression	5.04%	5.58%	−0.54%

There is a marked difference between several percentages of pre-existing conditions in English speakers compared to non-English speakers. Asthma, hypertension, and depression are far more likely to be previously diagnosed in English speakers than in non-English speakers. The other pre-existing conditions of diabetes, cholesterol, heart attack, emphysema, congestive heart failure, kidney failure, and cancer all were similar values to one another. The reason for this may be that the latter pre-existing conditions are more emergent, requiring immediate medical attention, that will be immediately diagnosed, while the others may remain untreated. The data also shows that a year after the Medicaid lottery, the four conditions that were recorded at the end of the trial, depression, hypertension, high cholesterol, and diabetes, were essentially equal, for English speakers compared to non-English speakers (in the applicant group as a whole). This supports the idea that non-English speakers saw greater health improvement, due to undiagnosed conditions and subsequent treatment during the year post-lottery.

It is important to note that the non-English speaking group had a lower average self-evaluation of health before the study, which does not align with the above table’s finding, that non-English speakers have a lower average occurrence of pre-existing conditions. Again, this may be bias in the data from undiagnosed conditions. Essentially, though the non-English speakers who applied for the lottery have fewer listed pre-existing conditions, they have a lower level of perceived health. This lower level of beginning health may allow for the non-English speakers to have greater improvement, as the group with the highest average change in self-perceived health during the prior twelve months was the non-English-speaking lottery winners, illustrated in the “Language and Health Change” table below.

Language and Health Change		
Language	Health_Last_12	Health_Change
English-speaking	6.43%	76.90%
English-speaking, lottery winner	6.93%	80.54%
English-speaking, non-lottery winner	5.89%	72.96%
Non-English-speaking	12.14%	69.42%
Non-English-speaking, lottery winner	15.03%	68.30%
Non-English-speaking, non-lottery winner	8.69%	70.74%

The finding that the non-English-speaking lottery winners saw the greatest health improvement, matches the findings in the regression analysis we conducted. When regressing the pre-existing conditions, age, health twelve months ago, and language, against change in health we find the below results and coefficients.

OLS Regression Results: All Applicants

Characteristic	**Beta**	**95% CI**[1]	**p-value**
health_last12_inp	0.12	0.11, 0.13	<0.001
language_inp	-0.07	-0.11, -0.03	<0.001
age_inp	0.00	0.00, 0.00	<0.001
ast_dx_pre_lottery_inp	0.00	-0.03, 0.03	0.9
dia_dx_pre_lottery_inp	0.05	0.00, 0.09	0.059
hbp_dx_pre_lottery_inp	0.03	-0.01, 0.06	0.11
chl_dx_pre_lottery_inp	0.03	0.00, 0.07	0.080
ami_dx_pre_lottery_inp	0.05	-0.04, 0.13	0.3
chf_dx_pre_lottery_inp	0.00	-0.12, 0.11	>0.9
emp_dx_pre_lottery_inp	-0.04	-0.12, 0.04	0.3
kid_dx_pre_lottery_inp	-0.07	-0.15, 0.02	0.12
cancer_dx_pre_lottery_inp	-0.01	-0.07, 0.05	0.8

The differences in the previous health and change in health for the English and non-English-speaking groups are reflected in the results of the regression. Health twelve months ago, language, and age are all significant at the 0.1% level. The effect that language has on the change in health, a coefficient of -.0668305, with a p-value of less than 0.1%, means that language is a stronger indicator of the health improvement of an individual, than diabetes and cholesterol in this analysis. Because English speakers are coded as a “1” and non-English speakers had a more positive change in their health, the coefficient for “language_inp” is negative.

When the same regression is run only on lottery winners, the difference in health change based on language is even more pronounced. The coefficient of the effect of language on health improvement is of greater magnitude and significance.

OLS Regression Results: Lottery Winners Only

Characteristic	Beta	95% CI[1]	p-value
health_last12_inp	0.12	0.10, 0.13	<0.001
language_inp	-0.10	-0.15, -0.05	<0.001
age_inp	0.00	0.00, 0.00	0.021
ast_dx_pre_lottery_inp	-0.01	-0.06, 0.03	0.5
dia_dx_pre_lottery_inp	0.07	0.01, 0.14	0.025
hbp_dx_pre_lottery_inp	0.04	0.00, 0.09	0.079
chl_dx_pre_lottery_inp	0.02	-0.04, 0.07	0.6
ami_dx_pre_lottery_inp	0.04	-0.08, 0.16	0.5
chf_dx_pre_lottery_inp	-0.07	-0.22, 0.09	0.4
emp_dx_pre_lottery_inp	-0.10	-0.21, 0.01	0.072
kid_dx_pre_lottery_inp	-0.08	-0.20, 0.04	0.2
cancer_dx_pre_lottery_inp	0.04	-0.04, 0.12	0.3

Looking at these results, and seeing the significance of the independent variable language, it is important to remember the differences in the population's pre-existing conditions and beginning health. The non-English-speaking lottery winner data may be biased, in that their pre-existing conditions are not yet diagnosed, and therefore not being treated, which provides more room to improve health.

Conclusion

The data produced by the Oregon Health Insurance Experiment is the gold standard, with its randomized controlled design. Because of the quality of the data, this paper's goal is to answer two uncorrelated questions, about emergency department visits and language.

This study began with determining if a model can determine if lottery winners will need to visit an emergency department during the duration of the study. To accomplish this task, we used variables such as age, health during the past 12 months, level of bodily pain, emphysema, depression, asthma, diabetes, hypertension, cholesterol, heart attack, congestive heart failure, kidney failure, and cancer. We then created a logistic regression model to classify participants into two groups (ED Visit/No ED Visit). Multiple models were run, cross-validated, and analyzed for accuracy, sensitivity, and specificity. With this model, we were able to accurately predict if a lottery winner would visit the emergency department with 64% accuracy.

The majority of the inaccuracy of this model comes from a type II error. The model underpredicted the number of emergency department visits. The majority of the incorrect predictions come from false negatives, or individuals who visit the emergency department, but were predicted to not visit. This aligns with the findings of the study, “Emergency Department Utilization as a Measure of Physician Performance” (Dowd B, Karmarker M, Swenson T, et al., 2014), which indicates that individuals who have not previously had health insurance are more likely to visit the emergency department, as it is their most common connection to the healthcare system.

Models such as this one can be very beneficial for medical providers to determine how many patients to expect to receive in the emergency department per year, based on the current health, age, and healthcare of the population. It can also be utilized to help determine the likelihood that an individual will need to visit the emergency department, based on various current conditions.

While the results of the emergency department analysis model aligned with current research, the results of the language analysis did not. Current research surrounding racial and ethnic minority patients shows that these minority patients are more likely to have poorer health outcomes (Blair, Steiner, and Havranek, 2011). Since individuals who are racial or ethnic minorities are also more likely to speak a first language other than English, we expected to see a similar health trend between ethnic and racial minorities and non-English speakers. Research on language spoken’s effect on health also contrasted this paper’s findings, as non-English speakers had poorer health to begin with than their English-speaking counterparts (Taira, D. A. 1999). Additionally, we expected that English speakers, since English is the native language in the United States, would receive a greater benefit from receiving Medicaid, since there may be a communication barrier for non-English speakers and providers.

What this project finds is the opposite - that non-English speakers have a greater improvement in their health from Medicaid enrollment than English speakers. However, health conditions before winning the lottery may be the major driver of the difference.

There is a sizeable difference in the percentage of pre-existing conditions that are diagnosed in English compared to non-English speakers. After winning the Medicaid lottery, non-English speakers were then diagnosed with conditions, so that English and non-English speakers had similar overall group percentages diagnosed with these conditions. Since non-English speakers were more likely to have an undiagnosed condition, once they received health insurance, they were able to address their condition and saw major health improvement. For example, non-English speakers were 38% less likely to have an identified diagnosis of hypertension at the beginning of the healthcare trial, but only 16% less likely to be diagnosed with hypertension after a year of Medicaid. These undiagnosed conditions allowed for greater health improvement for non-English speakers who won the Medicaid lottery.

Because of this difference, we cannot establish a causal relationship between language and its impact on healthcare outcomes. However, language may be an indicator of health demographics.

The findings in this paper related to language show that more research is needed surrounding the impact that language spoken has on healthcare outcomes.

The contents of this paper outline two major findings: first, a method to accurately predict emergency department visits, and second, more research is needed to determine the impact of language spoken on healthcare treatment received. Hospitals can use data points similar to those in this study to replicate this model for emergency room resourcing and staffing. Language may still have a causal relationship with healthcare received, as research has shown racial and ethnic identity do, and additional research is needed to explore the relationship.

References

1. Baicker, K., & Finkelstein, A. (2015). The Oregon Health Insurance Experiment. The National Bureau of Economic Research. https://doi.org/10.1377/hpb20150716.236899
2. Blair, I. V., Steiner, J. F., & Havranek, E. P. (2011). Unconscious (implicit) bias and health disparities: Where do we go from here? The Permanente Journal, 15(2), 71–78. https://doi.org/10.7812/tpp/11.979
3. Dennett, J. M., & Baicker, K. (2022). Medicaid, Health, and the Moderating Role of Neighborhood Characteristics. *Journal of urban health: bulletin of the New York Academy of Medicine*, *99*(1), 116–133. https://doi.org/10.1007/s11524-021-00579-2
4. Dowd B, Karmarker M, Swenson T, et al. (2014) Emergency Department Utilization as a Measure of Physician Performance. American Journal of Medical Quality. 2014;29(2):135-143. doi:10.1177/1062860613487196
5. Glymour, M. M., Avendano, M., & Kawachi, I. (2014). Chapter 2 - Socioeconomic Status and Health. In *Social Epidemiology* (pp. 17–61). essay, Oxford University Press, USA.
6. Liu, J., White, C., Nowak, S., Wilks, A., Ryan, J., & Eibner, C. (2018). *An Assessment of the New York Health Act: A Single-Payer Option for New York State*. https://doi.org/10.7249/rr2424
7. Taira, D. A. (1999). Improving the Health and Health Care of Non-English-Speaking Patients. Journal of General Internal Medicine, 14(5), 322-323. https://doi.org/10.1046/j.1525-1497.1999.00345.x
8. U.S. Census Bureau QuickFacts: Oregon. (n.d.). Retrieved November 5, 2023, from https://www.census.gov/quickfacts/fact/table/OR/PST045222
9. Zieff, G., Kerr, Z. Y., Moore, J. B., & Stoner, L. (2020). Universal Healthcare in the United States of America: A Healthy Debate. *Medicina (Kaunas, Lithuania)*, *56*(11), 580. https://doi.org/10.3390/medicina56110580