Factors Associated with Regular Mammogram Testing

Ella Hagopian & Kaitlin Olson

Introduction

Breast cancer is the second most common type of cancer in women and one of the most common causes of death in women each year (CDC, 2018). Regular mammogram testing is highly effective for catching cases early enough to be able to treat the cancer (Healthy People, 2019). Mammograms are the most effective when they are done regularly. Although many already adhere to a regular schedule, the United States hopes to improve the rates of mammogram testing by 10 percent by 2020 (Healthy People, 2019). It is therefore important to evaluate why some people are less likely to get regular testing in order to find solutions to make mammogram testing easier and more accessible to everyone.

Our project aims to explore possible reasons for why people may not get mammograms as often as they should and demographics of those who are likely to get tested on a regular basis. Knowing why patients are less likely to choose to get screened for breast cancer can lend information about how doctors and practices can change techniques in order to widen accessibility to screening. There has been some previous research on potential factors that impact people's decisions about screening. Resnick (2003) studied older populations and concluded that age, discussions with primary physicians, inconvenience and discomfort, and prior mastectomies were all associated with people's decisions to get a regular mammogram. Fayanju et al. (2014) expands that list with fears of cost and receiving bad news as additional factors. We will be expanding on the results of these studies by examining factors such as income, if they have insurance, history of breast problems, doctor recommendations, age, and a family history. By modeling the number of mammograms based on these factors, we wish to identify which are the most significant indicators of the number of mammograms people choose to get. Based on prior research, we hypothesize that younger women with low income, no insurance, no lump history, no history of familial breast cancer, and no doctor recommendation should be the ones with the lowest number of mammograms.

Materials and Methods

We are using compiled survey data from IPUMS Health Surveys which is a combination of National Health Interview Survey microdata (1963-present) and Medical Expenditure Panel Survey data (1996-present). The survey data provides individual-level information on demographics plus answers to health-related survey questions from a wide range of years. We chose the most recent year available with all of our variables of interest; using 2015 data, we selected columns pertaining to a person's age, income, insurance status, history of breast

problems, family history of breast cancer, and doctor recommendations for mammograms in addition to the number of mammograms they have had over the past six years (Table 1).

We manually separated the originally provided income levels into low, middle, or high income as described in Table 1. We also considered any values from any of the variables coded as "Unknown-refused", "Unknown-not ascertained", "Unknown-don't know", or "NIU" as NA and removed those with missing values from our final dataset. Because many of the levels counted as NA in the variable corresponding to the number of mammograms were from those too young to have had a mammogram, we effectively ended up with a sample aged 30 and older and went from 103,789 observations to 2,249 observations after removing NAs.

We started with exploratory data analysis to better understand our variables and the relationships between them. From there, we decided to remove the variables associated with if they had ever had a mammogram and the reason for their latest mammogram because everyone had had at least one mammogram and their reason for their last mammogram seemed complicated because we would be predicting past behavior from current indicators. The number of mammograms from the past 6 years is our primary response variable and ranges from 0 to 13+ mammograms. We expected the data to roughly follow a poisson distribution because it is count data; however, due to the nature of the variable, there was an unexpected spike in the response numbers corresponding to 6 mammograms. This made intuitive sense because those with a value of 6 mammograms are those who go regularly and follow the recommendation of going annually. Nevertheless, we wanted to find a way of filtering out some of them in order to have the data better fit a poisson distribution and better understand the individuals with a low probability of getting regular mammogram testing. We randomly selected 10% of the observations with a value of 6 and removed the other 90% so that the remaining 10% fell between the number of observations at 5 and the number at 7. Ultimately, we decided on quasipoisson over poisson because there was more variance than is assumed under a poisson regression ($\varphi = 1.76$). After checking for overdispersion, we fit a full model including an interaction between doctor recommendation status and income and decided to keep that interaction based on a nested F test. Finally, we removed maternal history because it was insignificant at the 5% level.

We also wanted to evaluate the demographics of "true sixes", or those with a high probability of having 6 mammograms in the past 6 years and getting regularly tested. To do this, we went back to the pre-filtered data that contained all of the sixes. We created a new indicator variable to predict those with 6 mammograms versus those with any value other than 6 and fit a logistic regression with all the same variables used as in our quasipoisson final model mentioned above. We then only examined those with a probability of greater than 40% (the median probability) of having 6 mammograms by performing summary statistics.

Results

Random Sampling Results

The dataset using a random 10% of the sixes was evaluated with a quasipoisson model that uses income, insurance, history of lump removal, age, and doctor recommendation status to predict the number of mammograms a person has had in the past six years. Income was a strong predictor for the number of mammograms. The average number of mammograms in the past six years for low, middle, and high income individuals are 2.64 (sd=2.40), 3.12 (sd=2.32), and 3.48 (sd=2.67) respectively. Compared to those with a low income, there is a 168% increase in the average number of mammograms for those classified as middle income (t=2.64, p<0.01) and a 78% increase in the average number of mammograms for those classified as high income (t=1.12, p=0.27), holding insurance, lump history, and age constant for those who have not visited a doctor (Figure 1).

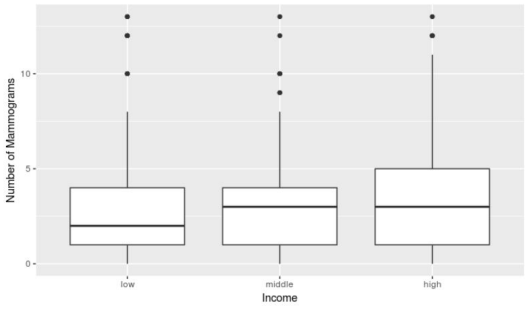

Figure 1. Number of mammograms for low, middle, and high-income individuals.

Similarly, insurance status was associated with the number of mammograms. Those with no insurance have an average of 1.51 (sd=1.26) mammograms per six years, while those with insurance have an average of 2.99 (sd=2.48). Compared to those with no insurance, there is a 57% increase in the average number of mammograms for those with insurance (t=3.308, p<0.001), holding income, lump history, age, and doctor visitation status constant (Figure 2).

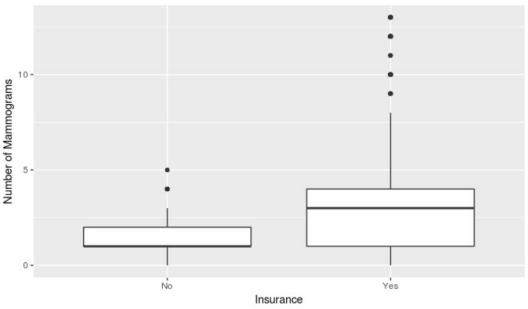

Figure 2. Number of mammograms for individuals with and without insurance.

An important predictor for getting a consistent number of mammograms is whether or not a patient has had a lump in their breast previously. Interestingly enough, the cancerous/benign condition of the lump does not change the amount of average mammograms tests these patients have done. In comparison to those who have never had a lump before, patients that have had a non-cancerous lump got 3.87 mammograms over six years (sd=2.92) or 33% more mammograms on average (t=6.217 p<0.001) while those that have had a cancerous lump got 3.75

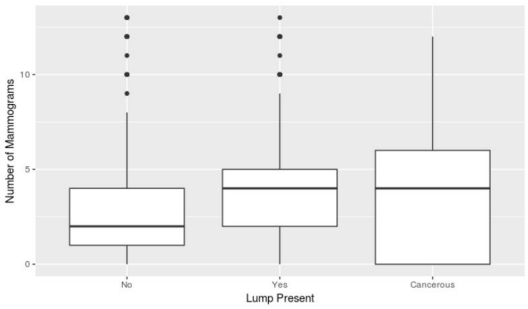

Figure 3. Number of mammograms by an individual's lump history.

mammograms over six years (sd=3.55) or 31% more mammograms on average (t=1.928, p=0.054), holding income, insurance, doctor recommendation, and age constant (Figure 3).

Visiting the doctor and whether or not they gave a recommendation to get tested is a strong predictor of number of mammograms done, as one would expect. Compared to individuals that had not visited the doctor recently, low income patients that did go to the doctor but did not get a recommendation got 2 mammograms over six years (sd=2.01) or 95% more mammograms on average (t=2.362, p=0.0183), holding insurance, lump history, and age constant. Furthermore, in comparison to those who have not gone in to the doctor, low income patients that got a specific recommendation to get a mammogram from a doctor had an average of 3.56 mammograms over six years (sd=2.52) or 295% more mammograms on average (t=4.931, p<0.001), holding insurance, lump history, and age constant (Figure 4).

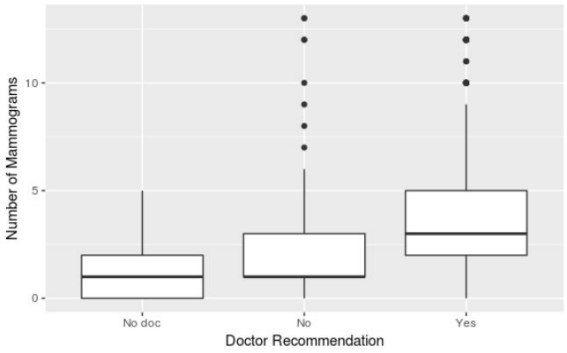

Figure 4. Relationship between doctor visit/recommendation and number of mammograms.

Age also is associated with number of mammograms a patient does. For every additional year increase in age, the average number of mammogram tests an individual has done increases by 0.4% (t=2.575, p=0.010), holding insurance, income, doctor recommendation and lump history constant.

Additionally, our model indicates that the relationship between income and doctor recommendation status is important for predicting the number of mammograms (Figure 5). Those who have seen a doctor but did not get a recommendation for a mammogram in the middle income group get an average of 167% more mammograms (t=-1.737, p=0.082591) and those in the high income group get an average of 217% more mammograms (t=-0.18, p=0.859) holding age, insurance, and lump history constant.. Further, the middle income group gets an average of 324% more mammograms (t=-2.426, p=0.015) when they have a doctor recommendation and the high income group gets an average of 366%

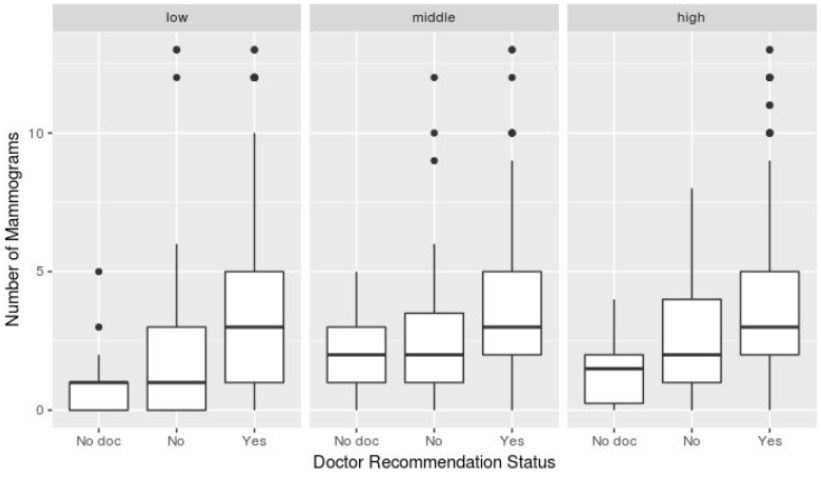

Figure 5. The relationship between number of mammograms and the interaction between doctor visit/recommendation and income.

more mammograms (t=-0.790, p=0.429) when they get a recommendation for a mammogram from their doctor holding age, insurance, and lump history constant.

Demographics of "True Six" Observations

The summary statistics of the "true sixes" taken from the predictions from the logistic model tell us about the demographics of those who are more likely to get regularly tested for mammograms (Table 2). For instance, everyone in the true six data has insurance. In the original data, 3% have no insurance. Everyone in the filtered data has also seen a doctor, and only 5% did not receive a recommendation for a mammogram from their doctor. Comparatively, 2% of the original data sample had not seen a doctor in the past year and of those who had, 29% did not have a recommendation. Additionally, the filtered sample is comparatively older with an age range of 41 to 85 and median of 64 as compared to the original data's range of 30 to 85 and a median of 61. There is a small shift towards higher income as well (28% vs. 21% high income, 49% vs. 52% low income). Although we did not consider maternal history of breast cancer in the quasipoisson final model, there is a shift in percentages from 28% who mentioned a maternal breast cancer history in the original data to 31% who mentioned a maternal breast cancer history in the filtered data. Finally, there is a higher percentage of those who had a lump removed (3% vs. 2% had cancerous lump removed, 32% vs. 22% had benign lump removed).

Discussion

Our results both align with results of previous studies and confirm our hypothesis that younger women with low income, no insurance, no lump history, and no doctor recommendation are the ones with the lowest average number of mammograms. Contrary to our original hypothesis but aligning with results from another study saying that women with a family history of breast cancer are just as likely to get a mammogram before the age of 40 as those who do not have a family history, an individual's maternal history with cancer does not influence patients to get more regularly tested, as we had expected it to (Qin 2017). Additionally, contrary to our hypothesis, the magnitude of change for each additional year of age was relatively small, so age, although significant in our model, does not seem to be an incredibly important factor in deciding to get regular testing. All of these results together could be used to lend useful information to healthcare providers about what subgroups are more or less likely to get tested for breast cancer through the means of a mammogram. Knowing that the fear of high costs is likely the reason many individuals do not get tested, it would be beneficial for healthcare providers that do offer free or reduced priced testing to continue advertise their services to these women. This may be difficult as many of these women also do not frequently visit the doctor. However, the fact that there are sufficient programs suggests that healthcare providers know and are attempting to reach women who have not previously gotten tested at a consistent rate.

Because of the way we designed our study, we must be cautious about a few things. The first is that when we removed the NA group from our dataset, we lost a significant amount of responses. Although the sample size we were left with was still robust, this has limited our generalizability quite a bit. We are also only able to generalize this women from the United States that are at least 30 years of age, as this was the sample from which the data was collected and relevant. Additionally, when we randomly selected individuals from the group that had 6 mammograms, we may have lost information that could have shifted the results. However, in the process of creating our final model we were cautious to adjust for overdispersion by using a quasipoisson model type. Furthermore, we tested significance of multiple nested models to find the final model that best displayed the results from our dataset.

In the future, we could further examine the impact of randomly removing those sixes by exploring the demographics of the 90% we did not choose. For alternative studies, it would be interesting to investigate other predictors that are tied to low testing for other gynecological cancers that are less at the forefront of public discussion. This could lend further information to healthcare providers about how they could improve access and/or exposure to other women's health issues.

Additional Tables

Table 1. All variables considered during initial phase of project. Includes if the variable is an explanatory or response variable, the type of variable, definitions, and values associated with each level of the variable.

Name	Variable Role	Definition	Values	Units
Income	Explanatory	Total combined family income	Categorical: $0-$49,999 = low $50,000-$99,999 = middle Over $100,000 = high	$
Insurance	Explanatory (Potential Confounder)	Has health insurance or not	Categorical: Yes / No	--
Mamever	Explanatory	Has ever had a mammogram	Categorical: Yes / No	--
Lumpever	Explanatory	Has ever had non-cancerous lump removed	Categorical: Yes / No /Yes, lump was cancerous	--

DrRec	Explanatory	Doctor recommended mammogram in the past 12 months	Categorical: Yes / No / No, no doctor in past year	--
Age	Explanatory (Potential Confounder)	Individual's age	Numeric: 30-85	Years
NumMam	Response	Number of mammograms from the last 6 years	Numeric: 0-13+	Number of mammograms
Reason	Explanatory	Reason for most recent mammogram	Categorical: Part of routine physical exam/screening, because of specific breast problem, follow-up to previously identified breast problem, baseline or initial mammogram, family history, other	--
Mom	Explanatory	Biological mother had breast cancer	Categorical: Mentioned a history / not mentioned	--

Table 2. Comparison of demographics from original dataset and dataset only containing those with a probability of greater than 40% of having 6 mammograms in the past 6 years. For each variable besides age, there is the percentage of observations at each level for both the original and filtered data and the change in percentage from the original to the filtered. For age, provided is the range and median.

Variable	**Original**	**True 6 Group**	**Difference**
Income	52% low income 27% middle income 21% high income	49% low income 22% middle income 29% high income	-3% low income -5% middle income +8% high income
Insurance	3% no insurance 97% insurance	0% no insurance 100% insurance	-3% no insurance +3% insurance
Lumpever	76% no lump 22% benign lump 2% cancerous lump	64% no lump 33% benign lump 3% cancerous lump	-12% no lump +11% benign lump +1% cancerous lump

DrRec	2% no doctor 29% no recommendation 51% recommendation	0% no doctor 5% no recommendation 95% recommendation	-2% no doctor -24% no recommendation +44% recommendation
Age	(30-85), median=61	(41-85), median=64	+3 median
Mom	28% mentioned 72% not mentioned	31% mentioned 69% not mentioned	+3% mentioned -3% not mentioned

Works Cited

Cancer (2019, May). Retreived from https://www.healthypeople.gov/2020/topics-objectives/topic/cancer

Division of Cancer Prevention and Control (2018, May 29). Breast Cancer. Retrieved from https://www.cdc.gov/cancer/breast/statistics/index.htm

Fayanju, O. M., Kraenzle, S., Drake, B. F., Oka, M., & Goodman, M. S. (2014, September). Perceived barriers to mammography among underserved women in a Breast Health Center Outreach Program. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4135000/

Qin, J., White, M. C., Sabatino, S. A., & Febo-Vázquez, I. (2017). Mammography use among women aged 18-39 years in the United States. 687-693. doi:10.1007/s10549-017-4625-6

Resnick, B. (2003), Health Promotion Practices of Older Adults: Model Testing. Public Health Nursing, 20: 2-12. doi:10.1046/j.1525-1446.2003.20102.