Assessing Strategies for Increasing Health Checkup Participation

Steffi Chern (steffic)

5/5/2023

Introduction

(1) Maintaining good health and early detection of potential health issues require regular health exams. However, in Vietnam, the number of individuals receiving these exams annually falls short of expectations. This research paper aims to help the Vietnamese Ministry of Health understand why people are not getting annual health exams and to design effective public relations campaigns to increase the number of people receiving regular health check-ups. Specifically, we want to answer 3 questions posed by the Vietnamese Ministry of Health:

- How do people rate the value and quality of medical service, and the quality of information they receive in check-ups?
- What factors are the most important in making a person less likely to get a check-up every 12 months?
- Is quality of information received an important predictor of whether patients get check-ups, and does this differ between people with and without health insurance?

(2) The study found that people generally rated the value and quality of service to be acceptable, with higher ratings regarding medical personnel's empathy and professionalism. However, around half of the respondents felt that checkups were either unimportant or a waste of time, which was a significant predictor for checkup attendance. Beliefs about the importance of checkups, current status as a student, housewife, or having an unstable job, and not having health insurance are significant predictors of not having had a checkup in the last 12 months. Beliefs that checkups are not important and are a waste of time are

also significant predictors. However, the quality of information received during checkups was not found to be a significant factor in determining whether patients get checkups. Based on these findings, public relations campaigns should target on people don't have jobs, people who think checkups are a waste of time, people who don't have health insurance, and people who feel that regular checkups aren't important.

Exploratory Data Analysis

(1) For variables that are related to demographics, the quantitative variables are: Age, Height, Weight, and BMI. On the other hand, the categorical variables are: Sex, Job Status, and Health Insurance. To better understand these variables we have, we start with looking at the results of our univariate EDA analysis.

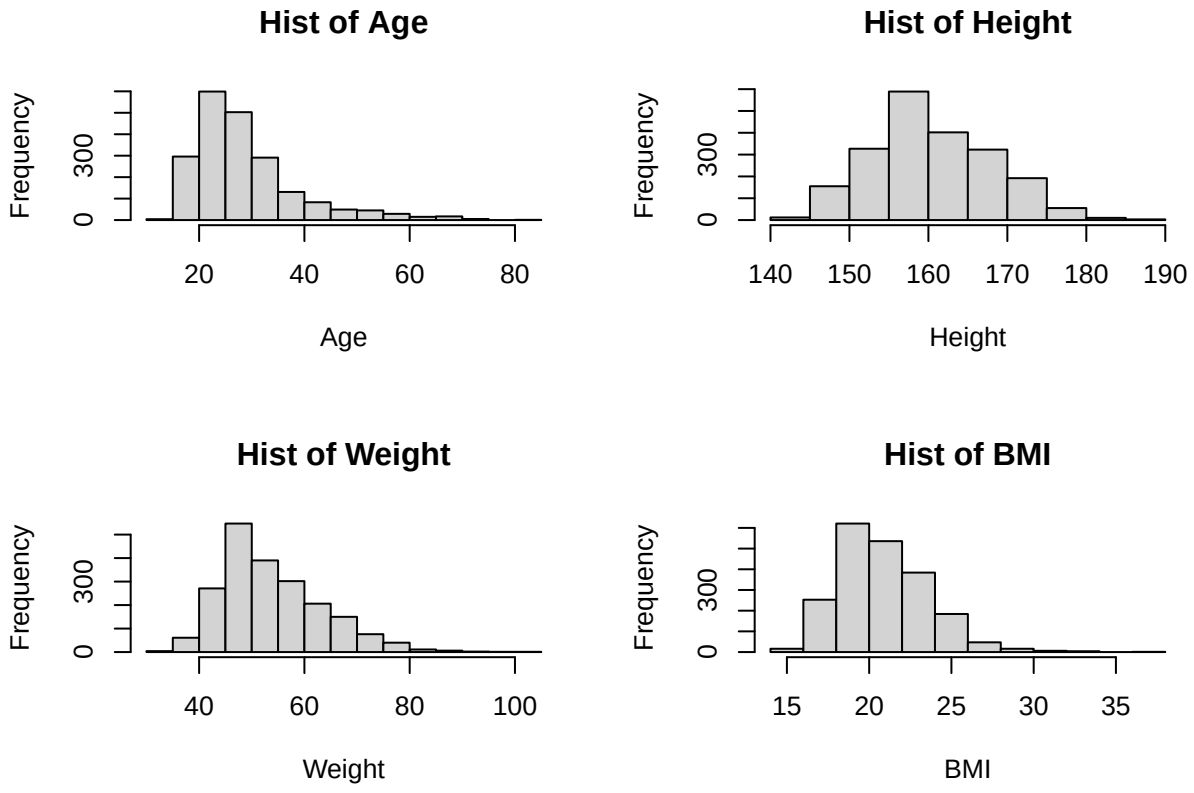


Figure 1: Histograms of Demographics Quantitative Variables

(1) From the histograms in figure 1, we observe that the quantitative variables Age, Weight, and BMI are all right skewed and unimodal, whereas Height is approximately normally

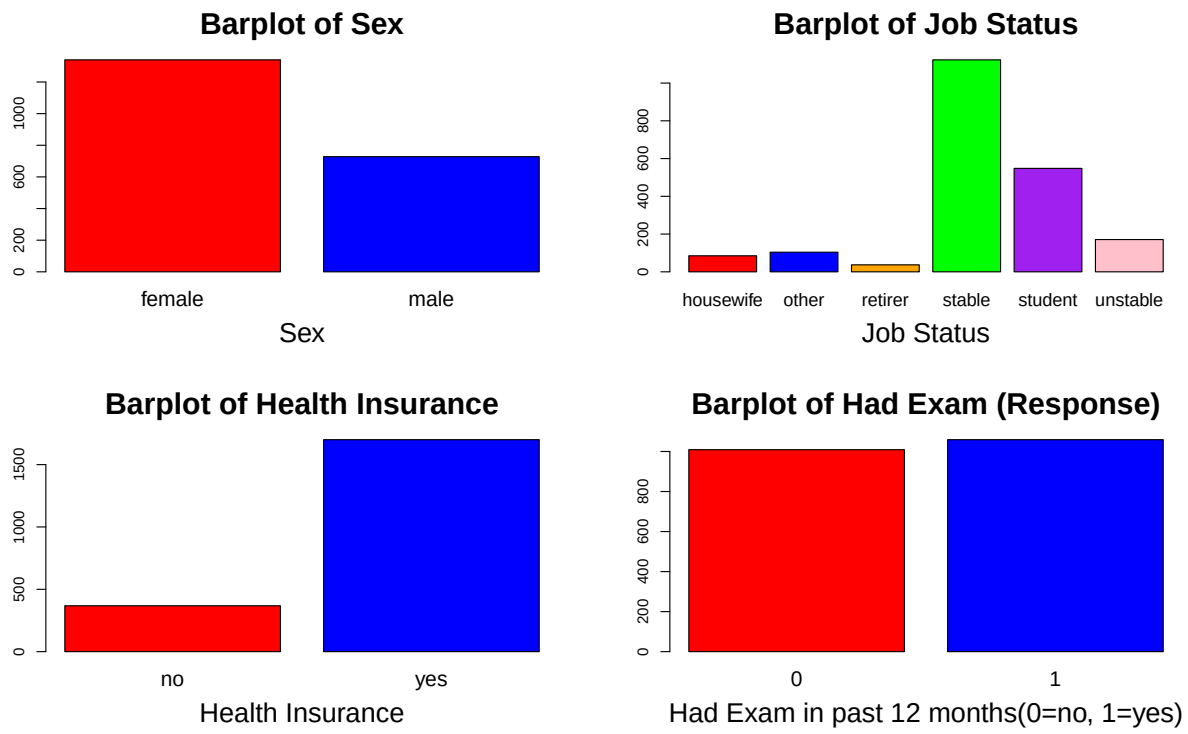


Figure 2: Barplots of Demographics Categorical Variables + Response Variable

distributed. Age has a peak around 20-30 years old, while Weight has a peak around 45-55 kg, and BMI has a peak around 18-22. Height has a mean around 160 cm.

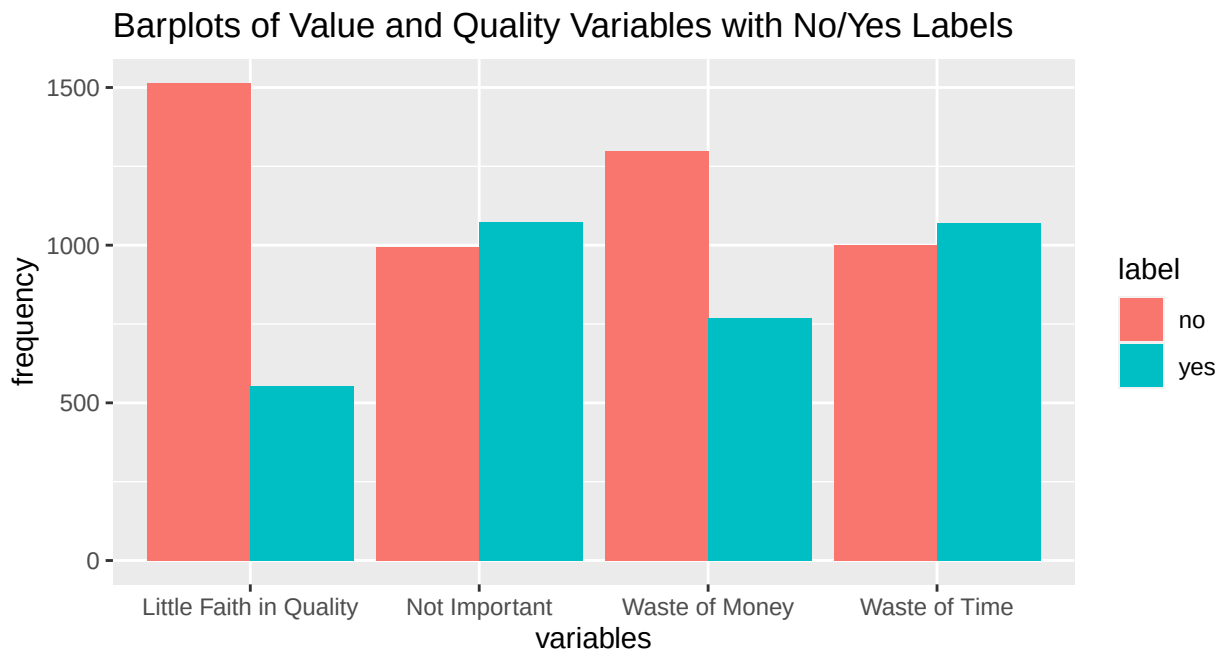
From the barplots in figure 2, we observe that there are a lot more females than males. As for the Job Status variable, we see that most people are either students or they have a stable job. We also notice that there are significantly more people who have health insurance than not having health insurance.

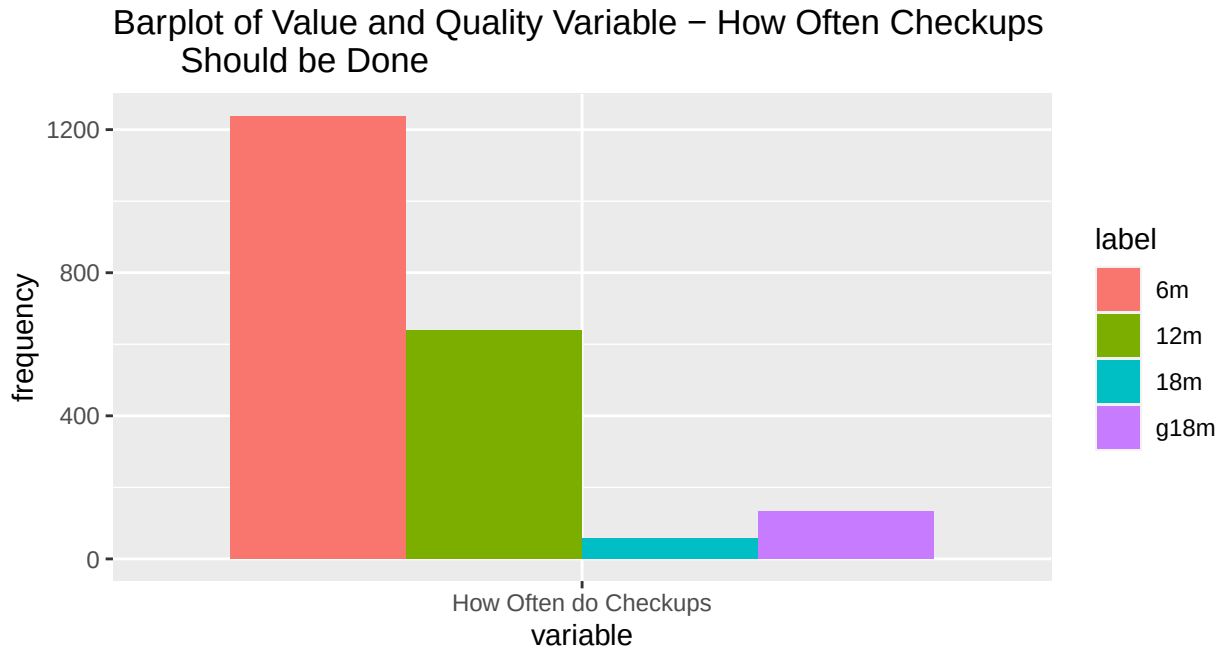
Table 1: Proportion of Respondents Had/No Exam

Had Exam	Proportion
0	0.488
1	0.512

(2) The response variable is *HadExam*, where 1 corresponds to if the respondent had a checkup in the past 12 months, while 0 indicates otherwise. Based on figure 2, we notice that the number of respondents who had a checkup in the past 12 months is really close to the number of respondents who did not have a checkup in the past 12 months. This can also

be confirmed from table 1, which showcases that around 48.8% respondents had a checkup in the past 12 months, and 51.2% respondents did not.





(3) After inspecting above, which shows the barplots of categorical variables related to value and quality of medical service, we observe that there are a lot more no's than yes's for variables Little Faith in Quality and Waste of Money. This indicates that more respondents believe that checkups are not a waste of money, nor do they have little faith in the medical services. However, there are more yes's than no's for variables Not Important and Waste of

Time, but the differences aren't that obvious (for both variables, around 48% respondents replied no while 52% respondents replied yes). This could suggest that there are slightly more people who believe checkups are a waste of time and not important. There are around 60% respondents believe that people should get checkups done every 6 months, while around 31% respondents believe checkups should be done every 12 months. Much fewer percentage of people agree that checkups should be done every 18 months or less often than every 18 months.

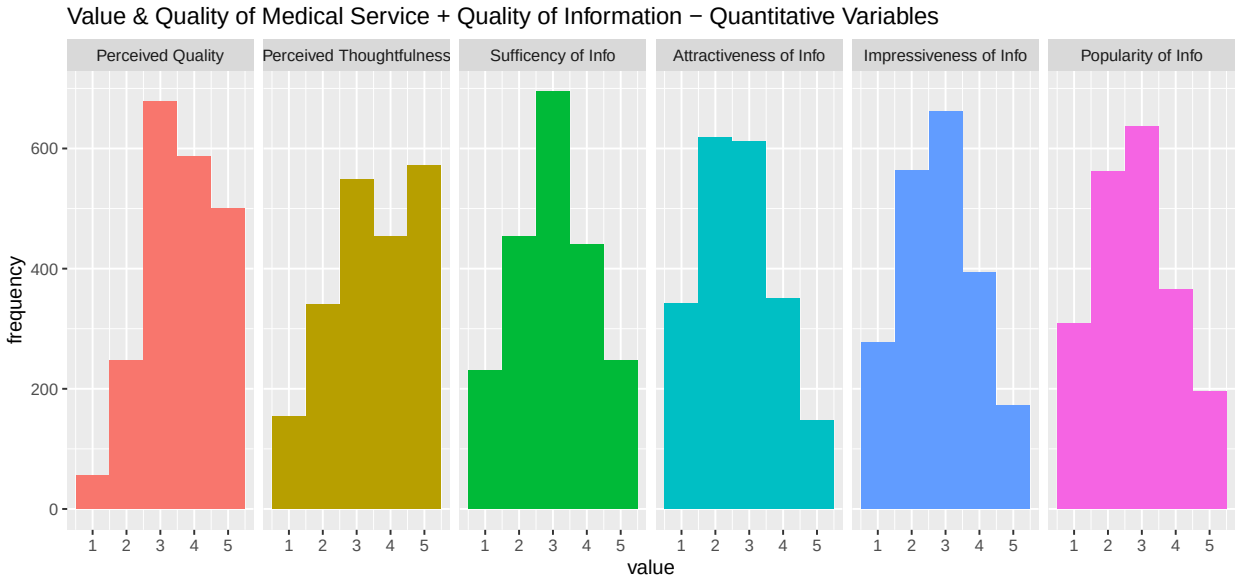


Figure 3: Barplots of Value & Quality + Quality of Info Quantitative Variables

(3) From figure 4, we discover that the variables Sufficiency of Information, Impressiveness of Information, and Popularity of Information are approximately normally distributed with mean around 3. This could suggest that most people are nearly indifferent regarding how sufficient, impressive, or popular the information they received in checkups are. We also notice that the variables Perceived Quality and Perceived Thoughtfullness are left skewed, indicating that most respondents receive high quality of medical equipment/personnel and high thoughtfulness and sense of responsibility of medical staff. Since we see that Attractiveness of Info is right skewed, this indicates that most respondents think the information they received in checkups are not that attractive.

Table 2: Had/No Exam based on Health Insurance

	no	yes
0	230	779
1	138	921

Table 3: Had/No Exam based on Waste of Time

	no	yes
0	400	609
1	599	460

Table 4: Had/No Exam based on How Often Checkups should be done

	6m	12m	18m	g18m
0	543	311	44	111
1	695	327	14	23

(4) We picked the variables Health Insurance, Waste of Time, and How Often do Checkups since they have more obvious patterns based on the univariate EDA. For people who think checkups are a waste of time, more people did not do checkups in the past 12 months. For people who has health insurance, they are more likely to have done checkups as well. We also notice that for people who believe checkups are necessary at least once a year had more people having their exams done in the past 12 months compared to those who think otherwise. Since some of the frequency differences are relatively noticeable in these variables, we expect some of predictors to be statistically significant when we try to predict whether a person had or had not taken the exam in the past 12 months.

(4) We also tried looking at the trends between quality of information variables vs the response variable. We notice that the values of Perceived Quality, Perceived Thoughtfulness, and Attractiveness of Info varied between those who took and did not take the exam over the past 12 months. This could possibly indicate that these variables will be statistically significant in predicting whether a patient has had an exam in the past 12 months or not.

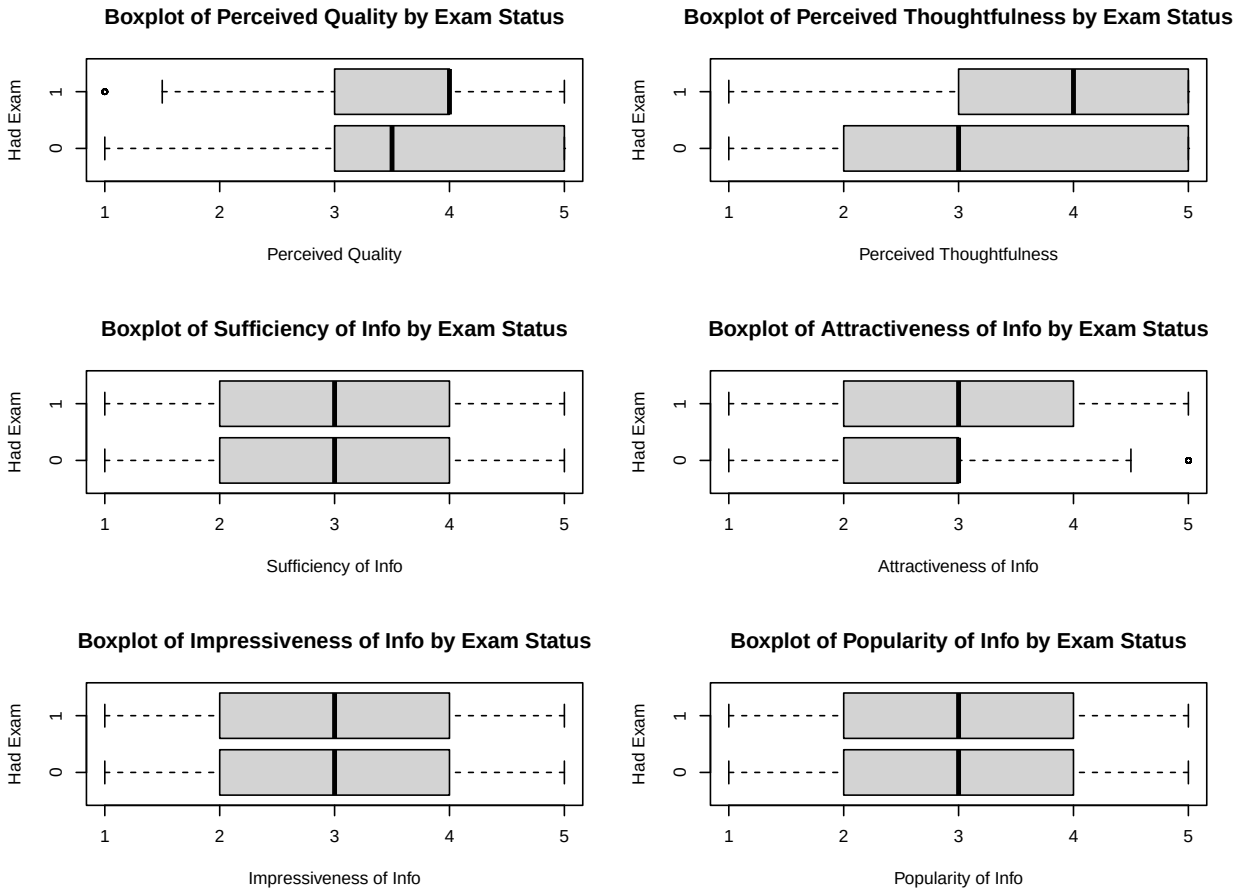


Figure 4: Bivariate EDA Between Response vs Quantitative Variables

Initial Modeling and Diagnostics

(1) To predict *HadExam* from all the demographic and value & quality predictors, we first fit a generalized linear model since the response variable is binary. We did not include the Place variable since many locations have very few respondents.

(2) Since there are many variables in the generalized linear model, we would like to determine which variables do not help with predicting the outcome. We can achieve this through a stepwise selection procedure that uses the AIC error estimate to determine which variables are not significant. From the step-wise variable selection, we found that only the variables Job Status, Waste of Time, Not Important, and How Often Respondents Think Checkups Should be Done are significant and therefore retained in the model. Most demographics related and some value & quality related variables are dropped. Specifically, the variables Age, Sex, Height, Weight, BMI, Little Faith in Quality, Perceived Quality, and Perceived Thoughtfulness are dropped from the model. Based on the AIC, not all variables are significant.

(3) We now proceed to add the health insurance and quality of information variables, as well as their interactions in our model to see if the quality of information variables have different associations between patients with and without health insurance.

(4) To determine if the model we just fitted provides a good fit to the observed data, we conduct a goodness of fit test. This is done by the likelihood ratio test, where the test statistic is equal to the difference in deviances between our fitted model and the saturated model, which follows a chi-square distribution with degrees of freedom equal to the difference in the number of parameters estimated by the two models. In this case, the null hypothesis is that our model is the true model that fits the data, while the alternative hypothesis is that the saturated model with n degrees of freedom is the true model that fits the data. We obtain a p-value less than the significant level 0.05, thus we conclude that our fitted model with the covariates included is not sufficient to explain the data.

```
## $calibration_plot
```

(5) In addition to check the goodness of fit of our fitted model, we can also check whether the model's predicted probabilities match the observed proportions of outcomes in the data by a calibration plot. From the calibration plot, we observe that most intervals lie closely to the 45-degree diagonal line, except for the predicted probability around 0.9. This suggests that most ranges of predicted probabilities are well-predicted, but may need to be aware of

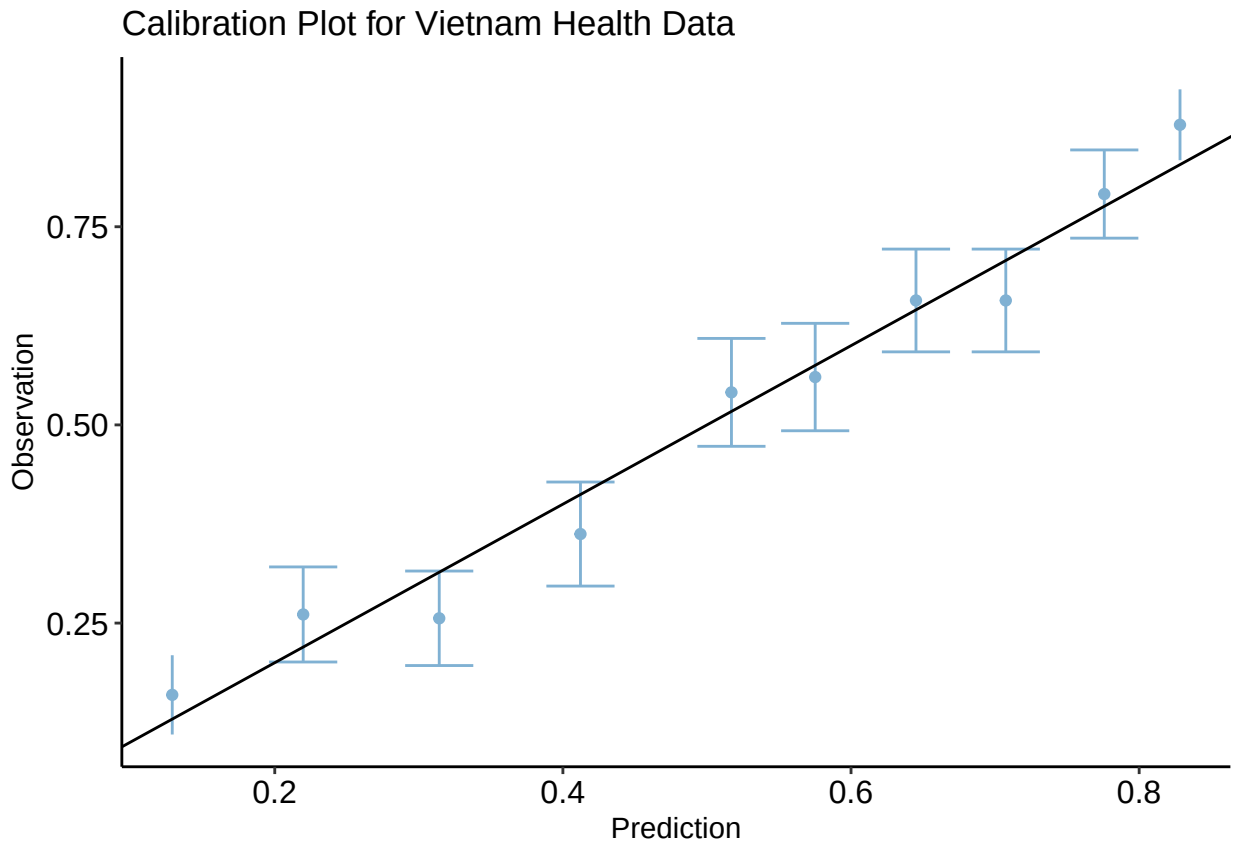


Figure 5: Calibration Plot for Vietnam Health Data

the interval that's worse calibrated. It is important to note that a model that is not perfectly calibrated can still be useful, especially when the areas of poor calibration are only limited to certain ranges of predicted probabilities and do not significantly affect the model's overall performance.

Some potential modifications to try on our fitted model is to remove non-significant predictors since there could possibly be variables that don't contribute much to the outcome. We could also try adding more interaction terms between the predictors. This allows for more complex relationships and potentially improve the accuracy of predictions.

We tried conducting one of the suggestions mentioned - removing insignificant variables through a step-wise selection procedure on our fitted model. However, too many relevant variables of interests are removed in this process, thus we will just proceed doing analysis with our original fitted model.

Model Inference and Results

(1) Previously, we added interaction terms between the health insurance and the quality of information variables in our fitted model to see if the quality of information variables have different associations between patients with and without health insurance.

To interpret the results from the interaction terms/variables, we look at the coefficients for each of them to determine if they are statistically significant or not. If an interaction variable is statistically significant (p-value < 0.05), then there is a significant relationship between that interaction variable and the log-odds of having had a checkup in the last 12 months. However, since none of the coefficients of the interaction variables are statistically significant, we cannot make any claims about the relationships between the interaction variables and the log-odds of having had a checkup in the last 12 months. In addition, there is no significant difference in the effects of these interaction variables between people with and without health insurance.

Table 5: ANOVA Test, Model with Interaction vs Model with No Interaction

Deviance	P-Value
1.2819	0.8644

(2) We conduct an ANOVA test to determine if there is a significant difference between our

fitted model with interaction terms and the same model without interaction terms. We obtain a p-value of 0.8644 (higher than the significant level), thus we do not reject the null hypothesis and conclude that the model with no interaction terms is the true model. Adding interactions between health insurance and quality of information does not significantly improve how well we predict the response variable (had exam or not in the past 12 months). We will proceed doing our analysis with the model with no interaction terms.

Table 6: Ratio between odds of having a checkup for people with most belief vs least belief in quality of info

	Odds Ratio
Ratio	1.3895

(3) To better answer the research questions in interest, we calculate the ratio between the odds of having a checkup for people with the most belief in the quality of information and the odds for those with the least belief in the quality of information. We note from above that none of the interaction terms were significant, thus the ratio does not depend on whether or not a person has health insurance.

We obtain an odds ratio of around 1.39, which indicates that higher ratings of quality of info are associated with higher odds of having a checkup. Specifically, people who rated the quality of information the highest have 1.39 times the odds of having had a checkup in the past 12 months, compared to those who rated the quality of information the lowest.

After obtaining the odds ratio, we proceed to calculate its 95% confidence interval:

Table 7: 95% Confidence Interval of Odds Ratio

Lower Bound	Upper Bound
0.934	2.066

(4) The confidence interval suggests that we are 95% confident that the true odds ratio in the population lies somewhere between 0.934 and 2.066. Since the interval does include 1, we can conclude that the odds ratio is not statistically significant at the 5% level of significance. In other words, there is no evidence to suggest that there is a significant difference in the odds of having a checkup between the people who rated highly and the people who rated poorly for the quality of information they received.

Conclusions

(1) Our study found that people generally rated the value & quality of service as acceptable, but half of the respondents consider checkups are unimportant or a waste of time, which are significant predictors in terms of checkup attendance. The factors associated with lower odds of having had a checkup in the last 12 months included beliefs that checkups are suitable once every more than 18 months or between 12 and 18 months, and checkups are a waste of time and unimpressive. Additionally, being a student, having an unstable job, or being a housewife are associated with lower odds of having a checkup, whereas having health insurance is associated with higher odds. The quality of information received was not found to be a significant factor in determining whether patients get checkups. This study suggests focusing campaigns towards people who do not classify themselves as having stable jobs, people who don't believe in annual regular checkups, people who thinks checkups are not important, and people who don't have health insurance.

(2) The reasons for some respondents to rate checkups as unimportant or a waste of time could be due to their past negative experiences or lack of knowledge about the benefits of checkups. People who don't have jobs and non-retirees may decide not to attend checkups due to other priorities in their life, such as academic studies or housework, or they do not have enough money to afford the checkups since they're not making income. Individuals with no insurance may have to pay extra for checkups, which could stop them from doing checkups. People who do not believe annual checkups are necessary may not see the urgency of the procedure, or they may not have experienced any catastrophic health events yet to make them realize the importance of doing regular checkups. These findings can be helpful for future campaigns aimed at increasing checkup attendance.

(3) There are several possible limitations for our analysis. First of all, the study only shows associations between variables and the odds of having had a checkup in the past 12 months, and does not claim causality due to the presence of some confounding variables not incorporated in the study. In addition, our logistic regression model is not perfect, especially for estimated probabilities around 0.9, which can lead to inaccuracies in inferences about the differences in odds. Lastly, the sample size may not be representative of the entire Vietnamese population, and the analysis may not apply to areas outside of Hanoi.