

An Investigation of How Personal Perception of One's Health and Their Perception of Healthcare affects the effectiveness of Treatments to Link Patients To Primary Medical Care.

Arm and Guy

May 14, 2021

```
urlfile="https://raw.githubusercontent.com/gthampak/Arm_Guy_MATH150_Project/main/HELpdata.csv"
HELpdata <- read_csv(url(urlfile))
```

#3# Introduction

Connecting with medical care is a choice. Many different factors could be involved when deciding whether or not to connect with medical care. Perception of your own personal health influences your decision to connect with medical care for non-major medical emergencies. Although we are able to do some basic self diagnosis with help of the internet, there is a high chance of a misdiagnosis given the various issues with access to useful and accurate medical information. Since we do not deem it a necessity if our health is not in a critical condition, we may decide not to connect with medical care. Apart from the personal health perceptions, there are also perceptions of health care system, which could include the ability to pay for health treatment. Lastly, one may not be in the optimal mental space to decide whether or not to connect with medical care, which could be due to poor mental health or substance abuse.

There have been many studies and surveys which try to gauge the accuracy of an individual personal health perception and their views towards healthcare. We attempt to widen the scope of personal perception of general health to investigate factors which could influence an individual's view and opinions on their own health as well as primary healthcare. Better understanding the influences on one's decision to connect with medical care can help focus efforts in specific sectors to give care to people who require care but don't know it themselves.

There are many factors that could lead to differences in health perception. A study shows that people with higher education better assess their health status (more consistent with medical professionals) (Souto et al.). Additionally, a study shows that men and women do not perceive their health differently, but have different views on what constitutes a healthy lifestyle (Mayo Clinic). It has also been shown that there are higher correlations between overall health and abstract health-related variables such as feelings of general health, pain, and worry than for concrete and observable variables, such as "bed days" and activity limitations (Leite et al.). Furthermore, substance users, specifically alcohol, tend to underestimate the negative effect of alcohol consumption on health and its contribution to injuries ((Sanchez-Ramirez et al.)).

As described, there has been many studies on variables that affect health perception and self-evaluation. Although many previous literatures explore factors which could impact perception of health, there haven't been many who investigate the relationship of these factors with health care perception. A study by Samet et al. explores alcohol dependency as a factor that affects linkage to primary health care as a secondary variable, with the primary independent variable being treatment groups to encourage linking with primary health care. The study concluded that the treatment was similarly effective for subjects with alcohol and illicit drug problems.

We hypothesize that if factors affect perception of health, they will also have an effect of the individuals

willingness to seek medical treatment. If you perceived yourself as healthy, the time to link with primary healthcare will be longer than those of perceive their health to be poor.

We investigate significant factors of perception of personal health, to find if they are significant in predicting the time to link with primary healthcare. We begin by identifying factors which affect personal health perception, such as age and education. Using a coxph model, we find significant factors in predicting times to link to primary healthcare.

Aim/Hypothesis

Our primary goal was to assess how personal perception of someone's health and their perception of health care affects the effectiveness of novel multi-disciplinary clinic for linking patients in a residential detoxification program to primary medical care.

Methods

First, we looked at variables that are related to health and health perception and separated them into different categories.

General/ASI-Composite scores

- ANY_UTIL Any recent health utilization (0=No, 1=Yes)
- CES_D CES-D score, measure of depressive symptoms, high scores are worse
- CNTRL InDUC-2L-Control score
- C_AU ASI-Composite score for alcohol use
- C_DU ASI-Composite score for drug use
- C_MS ASI-Composite medical status

Questions related to opinion on and habits towards healthcare

- D5_REC Medical trtmnt is important y/n (0=No, 1=Yes)
- D5 How import is trtmnt for these med probs (0=Not at all, 1=Slightly, 2= Moderately, 3= Considerably, 4= Extremely)
- E10A have you been to med clinic-last 6 months (0=No, 1=Yes)
- E15C2 I did not know where to go for help (0=No, 1=Yes)
- E8A1 Saw MD/H care worker regarding alcohol/drugs-last 6 months (0=No, 1=Yes)
- FRML_SAT Formal substance abuse treatment y/n (0=No, 1=Yes)

SF-36 Scores

- GH SF-36 general health perceptions (0-100)
- HT Raw SF-36 health transition item
- MH SF-36 mental health index (0-100)
- PF SF-36 physical functioning (0-100)
- RE SF-36 role-emotional (0-100)
- RP SF-36 role physical (0-100)

Drugs-related variables

- H14 According to interviewer, which substance is main problem (0=No problem, 1=Alcohol, 2=Alcohol to intox, 3=Heroin 4=Methadone, 5=Other opiate/analg, 6=Barbituates, 7=Sed/hyp/tranq, 8=Cocaine, 9=Amphetamines, 10=Marij/cannabis, 15=Alcohol and one or more drug, 16=More than one drug)
- PRIM_SUB First drug of choice (0=None, 1=Alcohol, 2=Cocaine, 3=Heroin, 4=Barbituates, 5=Benzos, 6=Marijuana, 7=Methadone, 8=Opiates)

Drug and Healthcare related variables

- H19A How import is treatment for alcohol problems now (0=Not at all, 1=Slightly, 2=Moderately, 3=Considerably, 4=Extremely)
- H19B How important is trtmnt for drug probs now (0=Not at all, 1=Slightly, 2=Moderately, 3=Considerably, 4=Extremely)

Interview's perspective on Patient

- S1A At interview pt obviously depressed/withdrawn (0=No, 1=Yes)
- S1B at interview pt obviously hostile (0=No, 1=Yes)
- S1C At interview patient obviously anxious/nervous (0=No, 1

Demographics and Education-related variables

- age
- HS_GRAD High school graduate (0=No, 1=Yes)
- A9 Years of education completed

Model Building

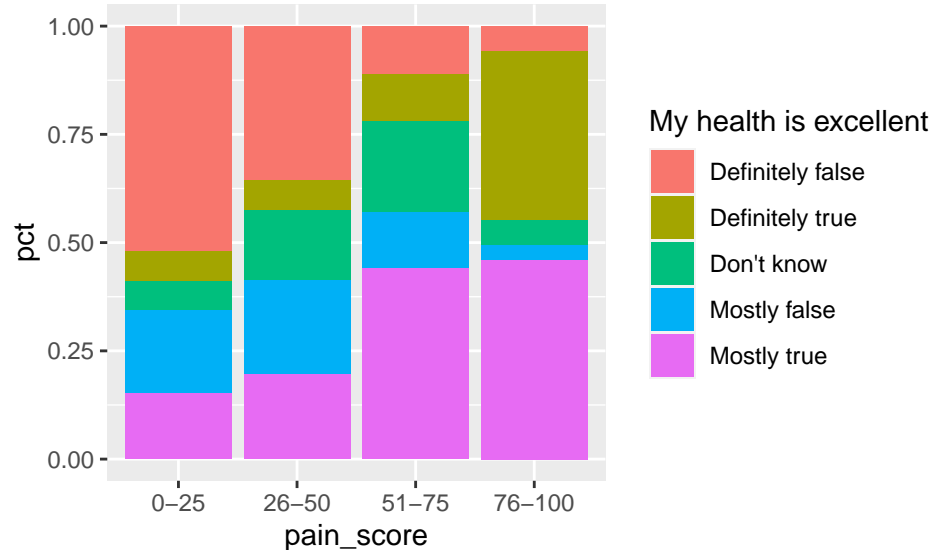
The functions we used for our primary data analyses are `coxph(Surv())` to test for significance of Hazard Ratio coefficients and `survfit(Sruv())` to plot survival curves. First we put all the variables above into a single model and removed the ones with highest significance one at a time. We also ran models with variables from a single category (above), and removed variables with highest significance. We did this to prevent putting highly correlated variables from the same category into the main model. We also tried looking for interaction between variables, but none were significant.

Results

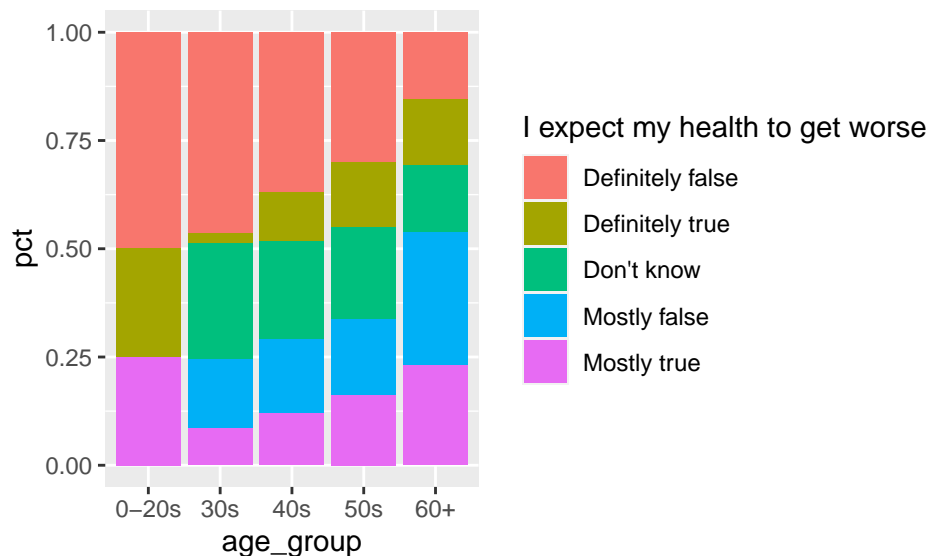
Exploratory Data Analyses

First, we wanted to see whether a higher pain score translate to a lower self rating of health. We hypothesize that it does.

From this visualization, we see that as pain score increases, the people who rated their health to be excellent increases, which makes sense.

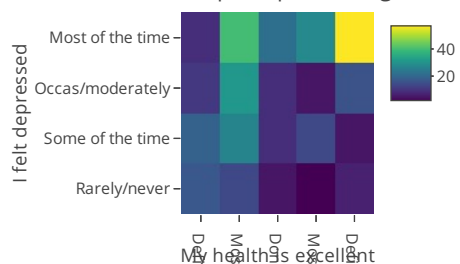


Next, we were interested in the correlation between age and perception of future health. We hypothesis that the people of older age will more likely believe that their health will get worse. This is mostly true as the percentage of people who believe their health will get worse increases as age increases, as shown below.



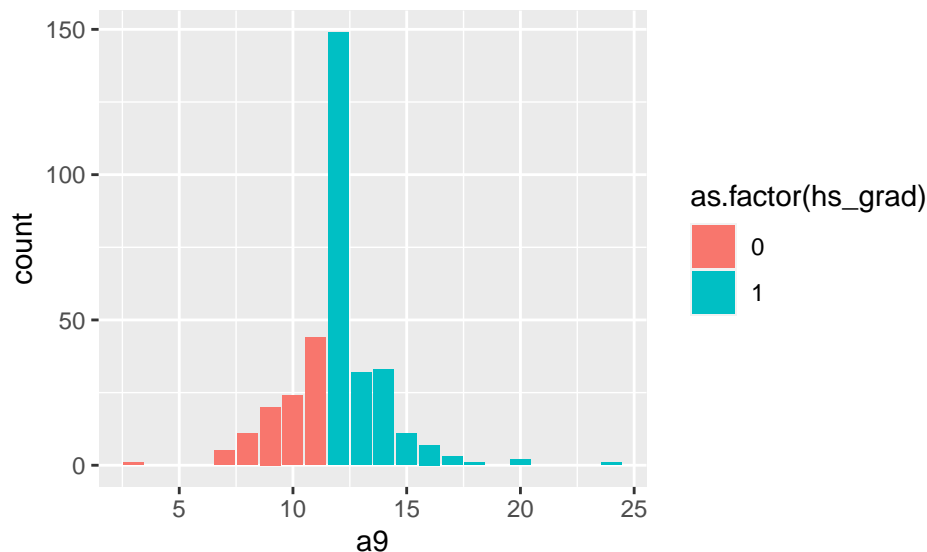
Next, we looked for patterns between perception of mental health and general health. This visualization is an effort to answer the question, do individuals incorporate their mental health to how they rate their general health. It seems that the majority of people who think their general health is bad also are suffering from poor mental health, which means that mental health is also considered in overall general health.

Personal mental health perception and general health



Now, we investigate variables associated with education as education may affect the individuals perception of health care, their knowledge of the health care system and their ability to self diagnose themselves. We plot a quick histogram to see whether years of education (a9) correspond with the high school variable (hs_grad) and it does. 12 years of formal education is when high school finishes, and this is the case here.

Here, we note that a lot of the patients in the study are high school graduates.

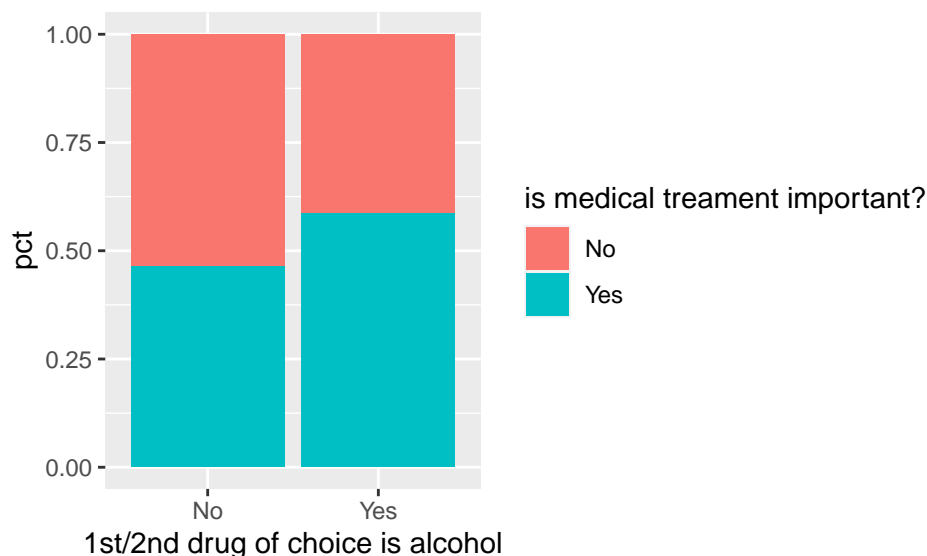


T-tests

With that, we wanted to know whether high school graduates and non-high school graduates view the importance of medical treatment differently. We performed a t-test between `hs_grad` and `d5_rec`, which asks patients whether Medical treatment is important (0=No, 1=Yes). We get a p-value of 0.7741, and cannot reject the null hypothesis that the proportion of patients who think medical treatment is important are the same between high school graduates and non-graduates.

```
##
## Welch Two Sample t-test
##
## data: d5_rec by hs_grad
## t = 0.28744, df = 198.75, p-value = 0.7741
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.09854964 0.13218183
## sample estimates:
## mean in group 0 mean in group 1
## 0.5523810 0.5355649
```

Another factor we thought could potentially affect patients view on the importance of medical treatment are those who use substances. We chose to do a t-test to see whether the proportion of people who view medical treatment as important is different between patients who have alcohol as their primary substance and those who do not. The p-value for this t-test is 0.03, which is a significant difference and we reject the null hypothesis. The test shows that the patients with alcohol as their primary substance is more likely to view medical treatment as important.



```
##
## Welch Two Sample t-test
##
## data: d5_rec by alcohol
## t = -2.1824, df = 254.19, p-value = 0.02999
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.23130020 -0.01187097
## sample estimates:
## mean in group 0 mean in group 1
## 0.4640000 0.5855856
```

We also wanted to see whether the proportion of high school graduates and non-high school graduates who view medical treatment as important is the same. According to the t-test below, the difference in proportion between the two groups is not significant.

```
##
## Welch Two Sample t-test
##
## data: d5_rec by hs_grad
## t = 0.28744, df = 198.75, p-value = 0.7741
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.09854964 0.13218183
## sample estimates:
## mean in group 0 mean in group 1
## 0.5523810 0.5355649

##
## Welch Two Sample t-test
##
## data: d5_rec by a1
## t = -1.0227, df = 138.71, p-value = 0.3082
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.18727736 0.05958477
## sample estimates:
```

```
## mean in group 1 mean in group 2
##      0.5265152      0.5903614
```

The above t-test shows that we cannot reject the null hypothesis that the proportion of patients who view medical treatment as important is the same for both male and female patients.

Significant Variables after preliminary models * group * alcohol * coc_her * hs_grad * a9 (education) * pf * any_util

Survival Analysis Model

```
coxph(Surv(dayslink, linkstatus) ~ group + alcohol + hs_grad + any_util + age, data=HELPdata) %>% tidy()
```

```
## # A tibble: 5 x 5
##   term      estimate std.error statistic  p.value
##   <chr>      <dbl>      <dbl>      <dbl>    <dbl>
## 1 group      1.74        0.227        7.68 1.62e-14
## 2 alcohol    0.494        0.204        2.43 1.53e- 2
## 3 hs_grad   -0.543        0.184       -2.95 3.19e- 3
## 4 any_util  -0.409        0.206       -1.98 4.76e- 2
## 5 age        0.0239       0.0119        2.00 4.55e- 2
```

Our final model includes the following variables: * group * alcohol * hs_grad * any_util * age

From this coxph model, **group**'s coefficient estimate is 1.73914222 which means that patients in treatment group 1 are $e^{1.73914222} = 5.692458$ times as likely to receive primary care at any given time that patients from treatment group 0. p-value for this coefficient is 1.621803e-14, which means we reject the null hypothesis that treatment group has no effect on 'risk' of seeking primary healthcare (with all else constant).

alcohol's coefficient estimate is 0.49420991 which means that patients with alcohol as their primary substance are $e^{0.49420991} = 1.639203$ times as likely to receive primary care at any given time that patients who do not have alcohol as their primary substance. p-value for this coefficient is 1.527822e-02, which means we reject the null hypothesis that alcohol as primary substance does not correlate with 'risk' of seeking primary healthcare (with all else constant).

hs_grad's coefficient estimate is -0.54287790 which means that patients with who are high school graduates are $e^{-0.54287790} = 0.5810736$ times as likely to receive primary care at any given time that patients who are not high school graduates. p-value for this coefficient is 3.189917e-03, which means we reject the null hypothesis that high school graduation does not correlate with 'risk' of seeking primary healthcare (with all else constant).

any_util's coefficient estimate is -0.40873422 which means that patients with recent health utilization are $e^{-0.40873422} = 0.6644908$ times as likely to receive primary care at any given time that patients who have no recent health utilization. p-value for this coefficient is 4.755411e-02, which means we reject the null hypothesis that recent health utilization does not correlate with 'risk' of seeking primary healthcare (with all else constant).

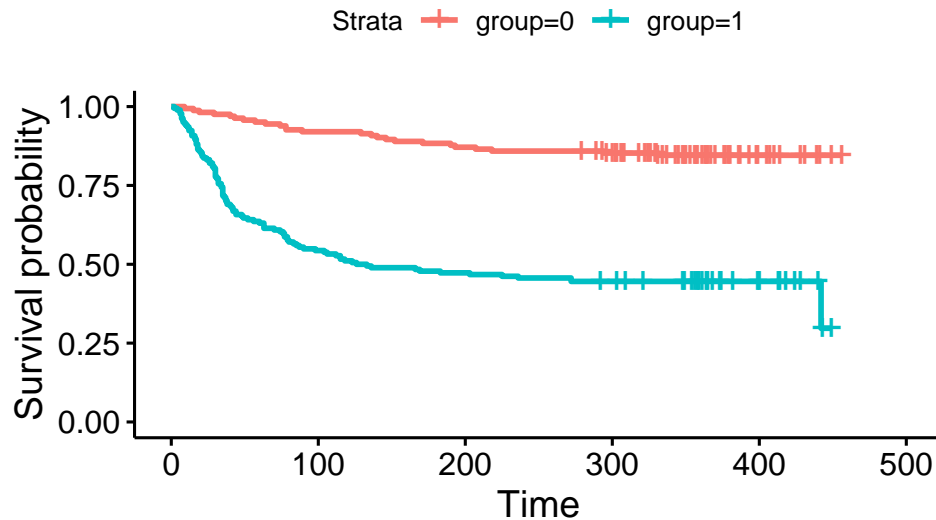
age's coefficient estimate is 0.02386218 which means that a one year increase in age corresponds to a $e^{0.02386218} = 1.024149$ multiplicative factor increase in 'risk' to link with primary care at any given time. p-value for this coefficient is 4.548318e-02, which means we reject the null hypothesis that age does not correlate with 'risk' of seeking primary healthcare (with all else constant).

CoxPH Model Assumptions

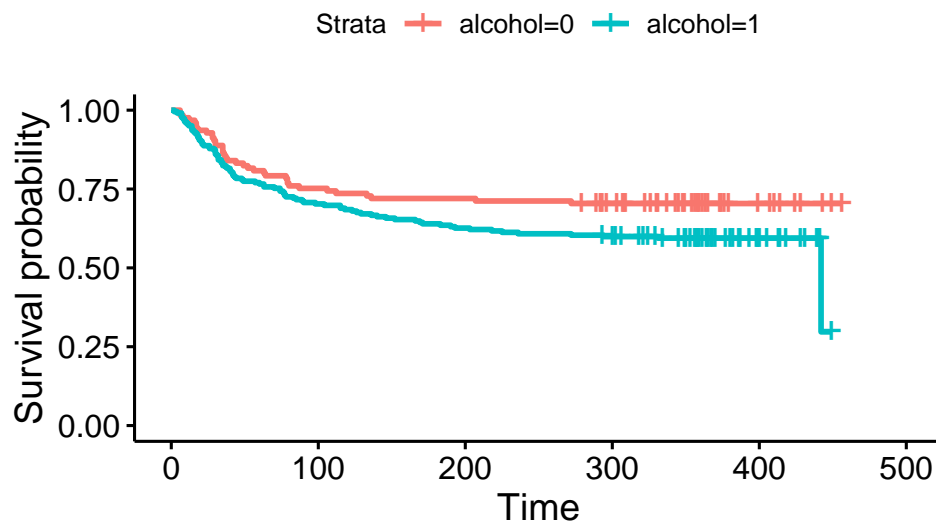
Because our primary method of survival analysis is the Cox proportional hazard model, we have to see whether our model follows Cox PH model assumptions to determine whether a fitted Cox regression model is able to describe the data appropriately. The Cox PH assumption is that the ratio of the hazards for any two individuals is constant over time, with the hazard ratios determined by the coefficients of the model, i.e., the hazards are proportional. This indicates that the hazard ratio is independent of time.

First, as a preliminary check, we can plot the associated KM curves for the variable individually and observe whether they intersect. If they do, then the hazard ratio is not constant can the variable varies with time.

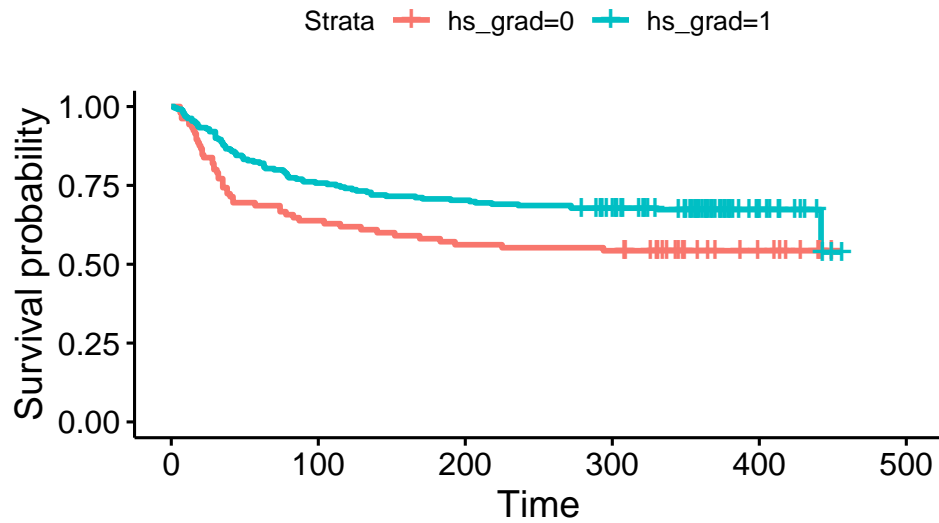
```
HELPdata_survfit <- survfit(Surv(dayslink, linkstatus) ~ group, data=HELPdata)
ggsurvplot(HELPdata_survfit, conf.type = "TRUE")
```



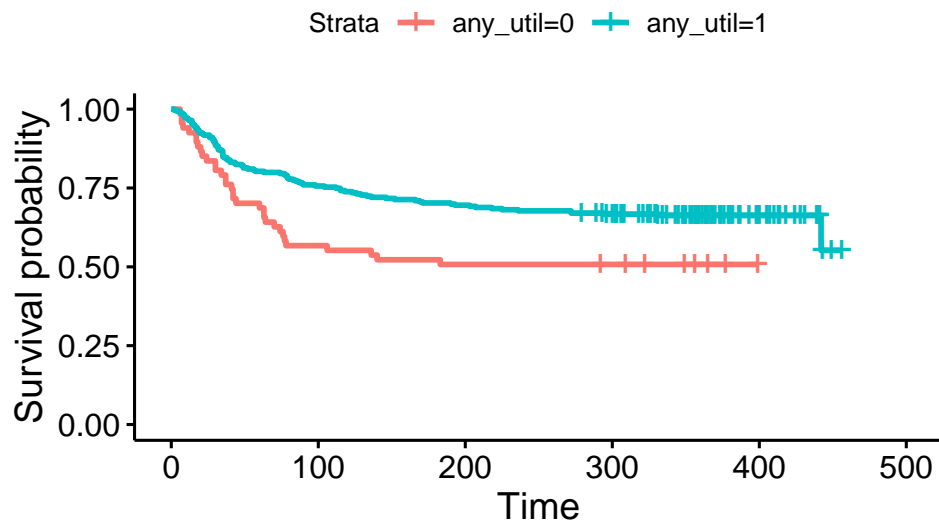
```
HELPdata_survfit <- survfit(Surv(dayslink, linkstatus) ~ alcohol, data=HELPdata)
ggsurvplot(HELPdata_survfit, conf.type = "TRUE")
```



```
HELPdata_survfit <- survfit(Surv(dayslink, linkstatus) ~ hs_grad, data=HELPdata)
ggsurvplot(HELPdata_survfit, conf.type = "TRUE")
```

```
HELPdata_survfit <- survfit(Surv(dayslink, linkstatus) ~ any_util, data=HELPdata)
ggsurvplot(HELPdata_survfit, conf.type = "TRUE")
```



As shown in the above plots, the KM curves of observations in the two groups of binary variables in the model does not intersect, thus the assumption appear to hold. Additionally, each pair of curve seem to have the same shape.

We can also use the `cox.zph` function which test for independence between residuals and time and performs a global test for the model. If coefficients are

For each covariate, the function `cox.zph()` correlates the corresponding set of scaled Schoenfeld residuals with time, to test for independence between residuals and time. Additionally, it performs a global test for the model as a whole. If the there is a non-significant relationship between residuals and time, then the assumption is held by the model (STHDA).

```
cox.zph(coxph(Surv(dayslink, linkstatus) ~ group + alcohol + hs_grad + any_util + age, data=HELPdata))
```

```
##          chisq df      p
## group    6.7776  1 0.0092
## alcohol   1.4511  1 0.2284
## hs_grad   1.0620  1 0.3028
## any_util  0.3541  1 0.5518
## age       0.0278  1 0.8675
## GLOBAL    9.2460  5 0.0996
```

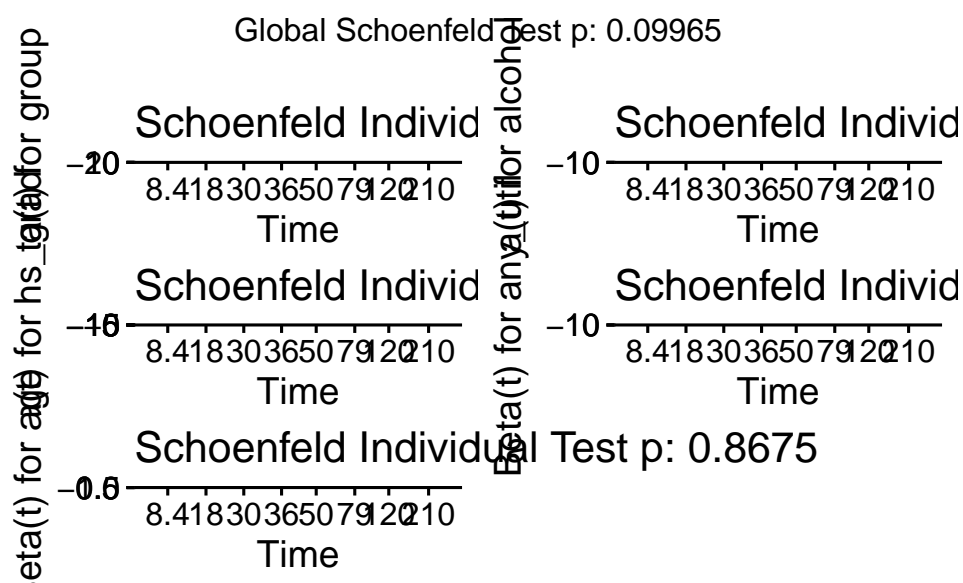
For our model, the Cox PH assumptions hold as none of the p-values for variables (other than group/treatment) has p-value below $\alpha = 0.05$. The global p-value is also above $\alpha = 0.05$. We suspect that the p-value for group variable is low despite having a nice KM curve because there may be other variables and/or interaction with other variables we did not account for in our model.

Schoenfeld residuals Plots Schoenfeld residuals are defined the Cox proportional hazard model in order to verify the proportional hazards assumption of the cox ph model. Schoenfeld showed in (reference paper) that if the survival data follows the proportional hazards assumption, then the Schoenfeld residuals are uncorrelated with each other. This is a useful and easy way to verify the assumption of the cox ph model, by calculating the Schoenfeld residuals, plotting a residual plot, and observing the plot for no correlation.

In order to understand the calculations of the Schoenfeld residuals, we observe residuals used in linear regression. Residuals are calculated by seeing how far the actual value is from the predicted value, where the y predicted value is calculated from a given x , explanatory variable. Schoenfeld residuals are calculated in the opposite way, where we fixed time (the response variable in the cox ph model) in order to predict the explanatory variables. The process for calculating the predicted values requires all the observations that are at risk at time T . Then, we calculate the expected value by summing across all the observations, multiplied by the probability that the event occurs using the β coefficients from the cox ph model. This expected value is then the value of the predicted variable at time T . Then, we subtract the real value from the predicted value and that becomes our Schoenfeld residuals at the given time T .

We can then plot the Schoenfeld residuals for our model, where there is one plot for each of the predictor variables in the model. We observe that the Schoenfeld residuals are not correlated, which means that we have satisfied the proportional hazards assumption for our cox ph model, which is the same result as the methods used to check the proportional hazards assumption above.

```
model <- cox.zph(coxph(Surv(dayslink, linkstatus) ~ group + alcohol + hs_grad + any_util + age, data=HEA))
ggcoxzph(model)
```



Discussion

There are factors related to perception of personal health and healthcare that affect an individual's likelihood to link with primary healthcare. In the model above, we show five such factors, four of which are more general (not only applicable to this study).

Our null hypothesis is that factors affecting health perception and perception on health care are unrelated to the likelihood of a person linking with primary healthcare at any given time, with all else being constant. Therefore, with several significant variables, we reject the null hypothesis and conclude variables affect health perception also correlates with likelihood of primary care linkage.

The significance of the **group** variable shows that treatment 1 increases the likelihood of a patient linking up with primary healthcare.

The significance of the **alcohol** variable shows that people who acknowledge their high consumption and dependence on alcohol are more likely (1.639203 times as likely) to link up to primary healthcare, especially if they have sought help on their drinking problems before. We see from our exploratory data analysis that people with alcohol as their primary substance are more likely to view medical treatment as important (difference in proportion 95% CI: (-0.23130020, -0.01187097)).

The significance of the **hs_grad** variable shows that high school graduates are less willing to link up with primary healthcare (0.5810736 as likely) compared to non high school graduates. This may be because with minor problems, high graduates may be more confident in self-diagnosing and looking for solutions themselves before seeing the need to link up with primary healthcare. From our preliminary data analysis, we found that the proportion of high school graduates and non-high school graduates who view medical treatment as important are not significantly different (difference in proportion 95% CI= (-0.09854964, 0.13218183)). Thus, we believe that a possible reason high school graduates are less likely to link up with primary health care is because they are less likely to seek medical treatment under less severe circumstances/sickness.

We feel this is consistent with and related to the study by Souto et al. which shows that people with higher education better assess their health status more consistent with medical professionals. We can infer that people with higher education, are more confident with self-diagnosis and self-treatment and thus take longer, with all else constant, to link with primary care.

The significance of the **any_util** variable in the model shows that patients who have recently received healthcare are less likely to link up with primary healthcare (0.6644908 times as likely). A possible reason is because healthcare is expensive and people may be less likely to visit primary healthcare successively, they may have had an unpleasant visit, a recent confirmation from a healthcare professional may increase their confidence in regards to their health, so they are less likely to think linking with primary healthcare is necessary.

The significance of the **age** variable in the model shows that the older you are, the more likely you are to link up with primary healthcare (1.024149 as likely for every one year increase). This directly relates to our initial exploratory data analysis where we saw that people in older age groups have more negative outlooks on their future health, which could explain their increased likelihood of linking with primary healthcare (with all else constant).

More generally, our exploratory data analysis results are largely consistent with previous studies discussed in the introduction. There is no significant difference in proportion of patients who view medical treatment as important for males and females (95% CI: (-0.18727736, 0.05958477)), which is consistent with the study by Mayo clinic on gender and health perception. However, there are significant differences between patients with and without alcohol as primary substances, similar to the study done by Samet et al. on primary care linkage and Sanchez-Ramirez et al. on health perceptions of alcohol dependents. One difference that we had was high school graduation as a factor in determining views on importance or medical treatment, as the proportion of those who view it as important is the same in both groups (95% CI: (-0.09854964, 0.13218183)). However, this is not an inconsistency with the study by Souto et al. on education and self-diagnosis assessment as view on medical treatment importance is different than self-diagnosis.

Interesting Note

When building our model, Variables that are significant tend to be binary variables with an even distribution of yes/no answers (or even more lob-sided towards yes (1) response). We suspect this is because more observations in each group directly translates to higher power, a lower absolute difference can result in lower (and potentially significant) p-values.

References