ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Master Thesis (Health Economics)

Should smokers, drinkers and the obese pay a higher health insurance premium?

Name: Laurens Huurman

Student number: 458097

Thesis supervisor: prof. dr. J.L.W. van Kippersluis

Second assessor: prof. dr. N.J.A. van Exel

Location: Zoetermeer

Date final version: 28-7-2021

Wordcount: 10945

Abstract

The aim of this thesis is to identify if childhood Socioeconomic Status (childhood SES) and genetic factors increase the probability of risky health behavior (smoking, alcoholism, obesity) and whether confronting people with this information change their opinions on the fairness of equal health insurance premiums. To predict the effect of childhood SES and genetic factors, Ordinary Least Squares (OLS) regressions were used on the Health and Retirement Study (HRS) datasets. A low childhood SES increases the probability to smoke and a high genetic risk increases the probability for smoking, alcoholism and obesity. Two surveys were created, one survey with this information and the other without, to see whether randomly assigned information leads to different normative viewpoints on the fairness of equal health insurance premiums. There is evidence that the respondents from the survey with information are more in favor of an equal health insurance premium, than the respondents from the survey without information.

Table of Contents

1 Introduction

Is it 'fair' that people who engage in risky health behavior such as smoking, alcoholism and obesity pay the same health insurance premium in the Dutch healthcare system? This is a question that gets a lot of attention in the media. Dutch research found that more than 40% of the population thinks that people with an unhealthy lifestyle should pay more for health insurance. However people often show solidarity with groups, they identify with in terms of lifestyle. Most of the smokers and heavy alcohol users think that their health insurance premium should remain unchanged and a substantial proportion of overweight people even think that their premium should be reduced (De Witt, 2019).

According to the Rijksinstituut voor Volksgezondheid en Milieu (RIVM), smoking, alcoholism and obesity causes 35.000 deaths and attributes to nine billion euros in health care expenses per year in the Netherlands, it is the main cause for burden of disease (Nationaal Preventieakkoord, 2018). Private health insurance is mandatory for everyone in the Netherlands and the paid premiums are equal irrespective of a person's health. Insurers set a premium where the low-risk people probably pay more than they will use medical care and high-risk people use more medical care than they pay for. Insurers must set a premium where all insured can use a sufficient level of medical care. So it is likely that people not engaging in risky health behavior compensate the costs for people that do, because they on average use more medical care. Only focusing on the current medical use, makes the equal premiums seem unfair.

However there are reasons why people might think it is fair that smokers, alcoholics and people with obesity pay the same health insurance premium. Two important reasons, that might influence people's opinion, are genetic factors and the socioeconomic status (SES) during childhood. In the literature there is extensive research on the possible effects of genetics, childhood SES and its interaction on smoking, alcoholism and obesity. People with a high genetic risk for smoking tend to smoke more, compared to people with a low genetic risk (Boomsma, Busjahn & Peltonen, 2002; Liu et al., 2019). The study from Bierut, Biroli, Galama & Thom (2018) suggests that people with a low childhood SES smoke the same amount of cigarettes or more, compared to people with a high childhood SES and that there is a clear

interaction between childhood SES and genetics. Literature shows that there is a significant effect of genetics on alcohol dependence, alcoholism and alcohol abuse (Health et al., 1997; Prescott & Kendler, 1999; Walters et al., 2018). The effect of childhood SES and the interaction is less clear (Poulten et al., 2002; Barr, Silberg, Dick & Maes, 2018). Studies show evidence of a significant effect of genetics on the probability for obesity (Maes, Neale & Eaves, 1997; Locke et al., 2015; Shungin et al., 2015). Research also shows that there is a negative association between childhood SES and BMI (Murasko, 2009; Murasko 2013; Jo, 2014)

If a person grows up in a household with both parents smoking and has a high genetic predisposition to become addicted, can you blame them? With this in mind it does not seem as unfair that everyone pays the same health insurance premium to an insurer. In the existing literature there is no clear research that considers genetics and childhood SES that might change people's opinion on the redistribution of the premiums. The aim of this research is to identify if genetics and childhood SES increase the probability of risky health behavior and whether these factors change people's opinion on the fairness of equal health insurance premiums. This leads to the following research question:

Which factors beyond an individual´s control increase the probability of risky health behavior (smoking, alcoholism, obesity) and does confronting people with this information change their opinions on the fairness of equal health insurance premiums?

This paper begins with the Theoretical Framework, which gives an extensive review of the existing literature on the topic. Research Methods will then present, the Health and Retirement Study (HRS) and survey data that is used in this research. The empirical specification of the models and how the results will be evaluated are also explained here. The results will be presented in the Results section and the research question is addressed in the Conclusion. The limitations of the research and the recommendation for future research will also be presented in this section.

2 Theoretical Framework

In the first subsection the effect of genetics, childhood SES and its interaction on risky health behavior is investigated using the existing literature. The types of risky health behavior that are focused on are smoking, alcohol use and obesity. In the second subsection the health insurance premium and the opinion of people on the fairness of equal health insurance premiums are discussed.

2.1 Genetics and childhood Socioeconomic Status

2.1.1 Smoking

The first evidence that showed a significant effect of genetics on smoking cessation, smoking initiation and the amount of cigarettes smoked, came from twin studies. The twin studies from Boomsma, Busjahn & Peltonen (2002) and Koopmans, Slutske, Heath, Neale & Boomsma (1999), found that smoking initiation is influenced by genetic factors with approximately 40%. Koopmans, et al. (1999) also found that when smoking is initiated the genetic factors influence the amount of cigarettes smoked by 86%. The twin study from Broms, Silventoinen, Madden, Heath & Kaprio (2006) suggests that genetic factors influence the amount of cigarettes smoked and the chance to quit smoking, but not the age of smoking initiation.

Advances in recent molecular genetic research on the associations between specific genetic variants and smoking, made the results on this relation more robust (Liu et al., 2019; The Tobacco and Genetics Consortium et al., 2010). In these studies Single nucleotide polymorphisms (SNPs) were used as a measure of genetic variation across individuals. SNPs are in short a variation in DNA and this causes the genetic variation (Isogg, n.d.). The samples in these type of studies are significantly larger than in the twin studies, which is an important advantage. The most recent study, from Liu et al. (2019), found that an increased genetic risk for cigarettes smoked per day is associated with an increase in additional daily smoked cigarettes and that an increased genetic risk for smoking initiation resulted in a 10-12% increased risk to become a regular smoker.

The existing literature about the effect of childhood SES on smoking initiation, smoking cessation, amount of cigarettes smoked and current smoking is less extensive, than the literature on genetics. Studies found that a low childhood SES is associated with a higher risk of current and persistent smoking for women, this effect was not significant for men (Power et al., 2005; Jefferis, Power, Graham & Manor, 2004). The latter study controlled for adult SES, the results showed a significant effect of childhood SES on smoking for men when they did not control for adult SES. Bierut et al. (2018) found that people with a high childhood SES smoked around 23 to 25 cigarettes per day at peak and people with a low childhood SES smoked around 25 to 32 cigarettes per day at peak. So this is evidence that people with a low childhood SES smoke the same or more cigarettes per day, compared to people with a high childhood SES. An important caveat is that the measure for childhood SES is different among most studies. There is no clear definition for childhood SES, this make the comparability of different studies more difficult.

The literature also shows that the genetic risk for cigarettes smoked per day at peak is moderate for people with a high childhood SES compared to people with a low childhood SES (Bierut et al., 2018). This study shows that for people with a high childhood SES there is little difference in cigarettes smoked per day at peak between a high and low genetic risk. For people with a low childhood SES the cigarettes smoked per day at peak increase significantly with a higher genetic risk. This is called the gene-by-environment (G*E) interaction. The existing literature mostly focused on the interaction between genetics and current SES. There is less literature on the interaction between genetics and childhood SES. However, the study by Bierut et al. (2018) shows that there is evidence of a G*E interaction with regard to childhood SES. This study is often used as a reference in this thesis, however there are clear differences. An important difference is that the focus of this research is not only on smoking, but also on alcohol use and obesity.

2.1.2 Alcohol use

The first evidence that showed a significant effect of genetics on alcohol dependence, alcoholism and alcohol abuse also came from twin studies (Health et al., 1997; Prescott & Kendler, 1999). They observed that a large part of the risk for alcohol dependence, alcoholism

and alcohol abuse is due to genetic factors and it showed similar results for men and women. The same as for smoking, there were advances in recent molecular genetic research on the associations between specific genetic variants and alcoholism (Liu et al., 2019; Walters et al., 2018).

Liu et al. (2019) found that an increase in the genetic risk for alcoholic drinks per week increased the amount of alcoholic drinks consumed per week. Walters et al. (2018) found that an increase in the genetic risk for alcohol dependence, increased the chance for alcohol dependence. In both studies polygenic scores (PGS) were used, this score is a person's genetic predisposition (genotype) for a particular type of risky health behavior. A higher score indicates a greater genetic liability for a specific type of risky health behavior (Ware, Gard, Schmitz & Fau, 2021). In this study the PGS for alcohol dependence and alcoholic drinks per week were used.

Then the literature on the effect of current SES on alcohol consumption. Recent studies show that people with a high SES consume the same or even greater amounts of alcohol compared to people with a low SES. However the negative effects of alcohol are more often experienced by people with a low SES (Collins, 2016; Katikireddi et al., 2017). The literature on the effect of childhood SES on alcohol consumption is less extensive. An often cited paper from Poulten et al. (2002) only showed a weak link between childhood SES and alcohol dependence. So from research there is no strong evidence of an effect of childhood SES on alcohol dependence, alcoholism and alcohol abuse.

A twin study from Jacob et al. (2003) showed evidence of an effect of the gene-by-environment interaction, the interaction between childhood SES and genetics, on alcohol dependence. They found that people that grew up in a low-risk environment (for example, the absence of parental alcoholism), compared to people that grew up in a high-risk environment, the genetic risk for alcohol dependence is probably moderated. The study from Barr, Silberg, Dick & Maes (2018) also investigates the G*E interaction, but uses PGS as a genetic measure and they found no strong evidence of a G*E interaction. For people with a high childhood SES the risk for alcohol problems increases in early adulthood and for people with a low childhood SES the risk increases in later adulthood. For women with a low childhood SES and a high genetic risk increases the risk for alcohol problems in adulthood.

2.1.3 Obesity

The same as for smoking and alcohol use, the first studies that showed a significant effect of genetics on obesity were family studies. The twin study from Price & Gottesman (1991) suggests that genes have an important influence on the accumulation of body fat. Maes, Neale & Eaves (1997) found that genetic factors influence the variance in people's body mass index (BMI) by 50-90%. In recent molecular genetic research on this effect, BMI is the most used measure for obesity. BMI is a simple calculation where you divide your weight in kilos by your height in meters squared. For example, a person of 80 kilos and a height of 1.80 meters has a BMI of $80/(1.80)^2 = 24.7$. A person with a BMI between 25 and 30 is overweight and 30 or higher is obese (Voedingscentrum, n.d.)

In molecular genetic research on the effect of specific genetic variants on obesity, SNPs were used and these were mostly associated with BMI (Locke et al., 2015; Shungin et al., 2015; Yengo et al., 2018; Frayling et al., 2007). In all of these studies they found that an increase in the genetic risk for a high BMI is accompanied with an increase in BMI. Locke et al. (2015) found that there was a difference of 1.8 kg per m^2 in mean BMI between the people with the highest genetic risk and the people with an average genetic risk.

The literature on the effect of SES on obesity is extensive (Mclaren, 2007; Sobal & Stunkard, 1989; Chapman, Fiscella, Duberstein, Coletta & Kawachi, 2009). For example Mclaren (2007) suggests that people with a low SES in highly developed countries (for example the Netherlands) have a higher probability to be obese, compared to people with a high SES. The literature on the effect of childhood SES on obesity shows a negative association between childhood SES and BMI (Murasko, 2009; Murasko, 2013; Jo, 2014). However the study from Jo (2014) found that in very low-income families this negative association does not hold. There is no significant literature on the effect of the interaction, between childhood SES and genetics, on obesity.

2.2 Health insurance premium

For Dutch inhabitants it is mandatory to take out at least basic health insurance with a health insurer. The insured pays for this every month (or another agreed period), this is the health

insurance premium. The level of this periodic amount is determined every year by the health insurers for both basic and additional health insurance. The premium for basic health insurance is the same for all insured of a particular health insurer. There is no distinction between age, sex, health, et cetera (zorgverzekering, n.d.).

Literature shows research on the opinion of Dutch inhabitants on the fairness of equal health insurance premiums over time. Since the 1980s the attitudes have not changed much, even with the growing health care costs (Kloosterman, 2011). A substantial part of the people think that smokers and heavy alcohol users should pay a higher premium. This opinion is not as substantial for people that are overweight. As mentioned before most of the smokers and heavy alcohol users think that their health insurance premium should remain unchanged and overweight people even think that their premium should be reduced. Mostly the people that do not engage in risky health behavior, think that people who do should pay higher premiums (De Witt, 2019). So these are the general attitudes towards the fairness of equal health insurance premiums.

The SES during childhood and genetics are factors that an individual does not have control over. As explained in the previous subsections these two factors have a significant effect on people engaging in risky health behavior such as smoking, alcoholism and obesity. However people might not take these factors into account when giving their opinion on the fairness of equal health insurance premiums. There is no research in the existing literature where people were first confronted with these two factors and then asked for their opinion on the fairness of equal health insurance premiums. This is the main contribution of this thesis.

The following studies used a research design where information was randomly assigned to people (Alesina, Stantcheva & Teso, 2018; Charite, Fishman & Kuziemko, 2016; Cruces, Perez-Truglia & Tetaz, 2013; Card, Mas, Moretti & Saez, 2010). In these studies some of the respondents were confronted with information and the other respondents were not. The aim of this research design is to see whether information causes different normative viewpoints between the groups with information and without information. The results showed significantly different opinions between the groups, so there is evidence that providing information change peoples' attitudes.

3 Research methods

This section is split into two parts, the first part is about the Health and Retirement Study (HRS) data and the second part is about the survey data. The first subsections of each part describe the data that is used and shows the descriptive statistics of this data. In the last subsection of each part the empirical specification of the model that is used will be explained.

3.1 Health and Retirement Study data

The Health and Retirement Study (HRS) consists of longitudinal datasets with a representative sample of 20.000 people in America (HRS Data Products, n.d.). In this research the following three datasets were used from the HRS: the 2010 HRS Core, Polygenic Score Data (PGS) and Validated Measures of Childhood Socio-Economic Status. For this study only cross-sectional data was used over the year 2010. The 2010 HRS Core dataset was chosen, because this gave the most observations and with more observations the statistical significance of the regressions will increase. Other more recent datasets, for example the 2016 HRS Core dataset, consisted of significantly less observations.

The respondents in all three datasets are of European or African ancestry. The respondents in the survey will be Dutch inhabitants, for this reason only the respondents in the HRS datasets of European ancestry are used. From the 2010 HRS Core dataset the dependent variables and gender variable were used. The Polygenic Score Data (PGS) was used for the genetic variables and the Validated Measures of Childhood Socio-Economic Status dataset was used for the childhood SES variable. The following subsections will explain the variables in depth.

3.1.1 Childhood SES variable

The childhood SES variable is a continuous variable which indicates the respondents' SES from birth to age 16. A high value means that the respondent had a high childhood SES. The average financial resources in childhood, financial instability in childhood, quality of relationship with mother, number of household adults, mother's years of education and father's years of education are the variables that were used to create the childhood SES variable in the Validated Measures of Childhood Socio-Economic Status dataset (Vable, Gilsanz, Nguyen,

Kawachi & Glymour, 2017). The childhood SES variable in this research is different from Bierut et al. (2018). More variables are used to create the childhood SES variable in this research.

3.1.2 Gender variable

The gender dummy variable takes value '1' if the respondent is a male and '0' if the respondent is a female. This is a control variable that is used, because it is possible that gender has an effect on the dependent variables and on the independent variables. For example, men are more likely to be regular smokers on average, compared to women. This is the effect of the gender variable on the dependent variable current smoker. It is also possible that there is a different effect between men and women on the childhood SES variable and the PGS variables.

3.1.3 Polygenic Score variables

In this research five PGS variables were used to indicate a respondents genetic risk for risky health behavior. For smoking there are two PGS variables, $PGS_{smoking\_initiation}$ and $PGS_{cigarettes\_per\_day}$. The $PGS_{smoking\_initiation}$ variable indicates if an individual had ever smoked regularly and the $PGS_{cigarettes\_per\_day}$ variable indicates the heaviness of smoking, either as a former or current smoker. For alcohol use the following two PGS variables were used, $PGS_{alcohol\_dependence}$ and $PGS_{drinks\_per\_week}$. The $PGS_{alcohol\_dependence}$ variable indicates the alcohol dependence of an individual. This is a form of problematic alcohol use, where a person is dependent on alcohol. Signs of alcohol dependence are drinking more than intended in a day and/or a tolerance to the effects of alcohol (Sigling, 2016). The $PGS_{drinks\_per\_week}$ variable indicates the average number of drinks consumed each week by a respondent, across all types of alcohol. For obesity the PGS_{BMI} variable was used and where a higher score means a greater genetic risk for a high BMI.

3.1.4 Interaction variables

The interaction variable indicates the interaction between the PGS variables and childhood SES variable. It might be that the genetic risk for risky health behavior is moderate for people with a high childhood SES compared to people with a low childhood SES. For example Bierut et al. (2018) found that that for people with a high childhood SES there is little difference in the amount of cigarettes smoked between a high- and low genetic risk. For people with a low

childhood SES the amount of cigarettes smoked increases significantly with a higher genetic risk. The intuition behind the interaction effect is, that people with a low childhood SES and a high genetic risk find it more difficult to resist the need for smoking, drinking or other risky health behavior than people with a high childhood SES and a high genetic risk.

3.1.5 Dependent variables

The first dependent variable is ever smoke, this is a dummy variable that takes value '1' if the respondent ever smoked and '0' if not. The second dependent variable is current smoker this is also a dummy variable that takes value '1' if the respondent is a current smoker and '0' if not. The third dependent variable is alcoholic and an individual is an alcoholic, according to the geestelijke gezondheidszorg (GGZ), when drinking 1 or more drinks per day at least 6 days per week (GGZ, n.d.). The fourth dependent variable is drinks per week, which indicates the number of alcoholic drinks consumed per week. The fifth dependent variable is binge drinker, which means that a respondent consumed 4 or more alcoholic drinks in one occasion in the past three months and takes value '1' if yes and '0' if not. The sixth dependent variable is obesity and takes value '1' if a respondents' BMI is 30 or higher and '0' if less than 30. The last dependent variable is overweight and takes value '1' if a respondents' BMI is 25 or higher and '0' if less than 25.

3.2 Descriptive statistics

The descriptive statistics in table 1 show that in this sample 56.2% of the respondents ever smoked and that 18.6% are current smokers. Followed by 17.6% of the respondents are alcoholic, 27.5% of the respondents are binge drinkers and on average 7 drinks per week is consumed per respondent. There are however in the literature a lot of different opinions on how an alcoholic should be defined. The 17.6% might change when using other criteria for an alcoholic, the mentioned definition for an alcoholic in the previous subsection is used in this research to define an alcoholic. Next, 67.8% of the respondents is overweight and 30.7% is obese. There are also more women (63.1%) than men (36.9%) in the dataset. Table 1 shows the $PGS_{smoking\_initiation}$ variable, in appendix 1 are the values for the other PGS variable. As mentioned before a higher value indicates a greater genetic liability for the specific type of

risky health behavior. A higher value for the childhood SES variable means a higher childhood SES.

Table 1. Descriptive statistics (HRS data)

Variable	Mean (SD)	Min	Max	N
$PGS_{smoking\_initiation}$	-0.007 (1.000)	-3.861	3.938	10282
Childhood SES	0.276 (0.873)	-3.322	2.809	10282
Ever smoke	0.562	0	1	9092
Current smoker	0.186	0	1	5126
Alcoholic	0.176	0	1	5035
Drinks per week	6.929 (9.394)	0	168	3514
Binge drinking	0.275	0	1	3498
Obesity	0.307	0	1	4710
Overweight	0.678	0	1	4710
Gender	0.369	0	1	10282

3.3 Model specification

To find out if there are genetic and cSES (childhood SES) factors that might have an effect on risky health behavior, Ordinary Least Squares (OLS) regressions will be used. This gave the following equation:

$$Y_r = \alpha + \beta_1 PGS_r + \beta_2 cSES_r + \beta_3 [cSES_r \times PGS_r] + \beta_4 gender_r + \varepsilon_r$$

The Y_r is the dependent variable and indicates the probability for risky health behavior of respondent *r*. The $gender_r$ variable is a control variable for the gender of respondent *r*. The $cSES_r$ variable indicates the value of the childhood socioeconomic status of respondent *r* and the PGS_r variable is the polygenic score of respondent *r*. The $cSES_r \times PGS_r$ interaction variable indicates the interaction between the polygenic score and childhood socioeconomic status of respondent *r* (Bierut et al., 2018). Lastly, the ε_r is the error term and the β are the regression coefficients. In this model only gender is used as a control variable, in Bierut et al. (2018) more control variables were used.

3.4 Survey data

Qualtrics is used to create the surveys, it is one of the leading software programs for building surveys (Qualtrics, 2021). The respondents for the surveys are Dutch adults that pay a health insurance premium and they are recruited through social media and my own network. Two surveys were created and will be assigned randomly to the respondents. In both surveys the same questions are asked, however in only one survey there is information given on the effect of childhood SES and genetics on risky health behavior. This information comes from the HRS results and will be easy to understand for the respondents. There are two different surveys to identify whether randomly assigned information leads to different normative viewpoints on the fairness of equal health insurance premiums.

The surveys start with an introduction that explains the aim of the study and contains the informed consent. Then background questions are asked, these are some of the background questions in the survey: 'What is your age?', 'how tall are you?' and 'do you smoke?'. Next, in both surveys information is given on how the health insurance system works in the Netherlands and on certain terms that are used in the surveys. Then, two questions are asked: 'Do you think that the health insurance premium for people with an unhealthy lifestyle should be higher, lower or unchanged?' and 'Do you think that the health insurance premium for people with a healthy lifestyle should be higher, lower or unchanged?'. This is the first part of both surveys and up to this point they are identical, appendix 2 contains the two surveys.

The second part for the survey with information, starts with information on childhood SES and genetic factors and explains that people do not have control over these factors. Then the effect of these factors on smoking is given and the following questions are asked: 'Do you think that the health insurance premium for smokers should be higher, lower or unchanged?' and 'Do you think that the health insurance premium for non-smokers should be higher, lower or unchanged?'. Next, information on the effect of the childhood SES and genetics on respectively alcohol use and obesity are given. The same two questions as for smoking were asked, but then for respectively heavy alcohol use and obesity. See appendix 2 and subsection 4.2.1 for the information that is given, the result section explains the effect of the childhood SES and genetics on risky health behavior in detail. The survey without information contains

the same questions. The surveys end with a question that asks for comments on the survey that respondents might have.

When evaluating the results from the survey data, it is important to have a sufficient number of respondents. This makes it possible to give valid statements on the results of the survey. To identify the minimum number of respondents that are needed, a sample size calculator was used (Qualtrics, n.d.). For the calculation the target population, the margin of error and confidence level is needed.

The target population for this research are the Dutch inhabitants that are obligated to pay health insurance premiums, which are around 13.5 million people (Zorgwijzer, 2019). Next, the margin of error that is used in the calculation is 8%. This indicates how much difference is accepted between the sample mean and the target population mean. Then, the significance level determines with how much certainty you are able to say that the target population mean falls within the margin of error. A significance level of 90% is chosen, this is an alpha of 0.1. This significance level is only used in subsection 4.2.2 and not in the empirical analysis. Finally, all these numbers were used in the calculation and it turned out that 106 respondents for each survey was the minimum sample size.

First a small pilot survey was sent to around 20 people. The most important aim of the pilot was to identify if the statistics in the survey were not too difficult to understand for the respondents. Other mistakes that were mentioned by the respondents about the surveys were also fixed after the pilot. The final surveys were set online between July 1 and July 7 2021, 212 respondents were gathered, with 106 respondents for each survey.

3.5 Descriptive statistics

The descriptive statistics in table 2 show that there is only a slight difference in age and gender between the two surveys. In the survey with information the mean age is 45 years and 33% of the respondents are male. In the survey without information the mean age is 48 years and 38% of the respondents are male. So there are more female respondents than male respondents in both surveys. In both surveys around half of the respondents has completed a study in higher education, in Dutch this means university or HBO (Hoger Beroeps Onderwijs).

There are slightly more respondents obese and overweight in the survey without information. In the survey with information 20% of the respondents smoke and only 13% of the respondents smoke in the other survey. There is almost no difference in the number of alcohol and binge drinkers between the surveys, 88% of the respondents consume alcohol and 37-40% are binge drinkers. The binge drinkers have been binge drinking on average 7.3 times in the past 3 months. Table 2 also shows p-values from simple t-tests on whether these differences are significantly different or not. The samples do not differ significantly from each other for almost all the variables. Only the mean BMI between the two surveys is significantly different at a p-value threshold of 0.1. The respondents in the survey without information have on average a higher BMI.

Table 2. Descriptive statistics (survey data)

Variable	Mean (Survey with information)	Mean (Survey without information)	**Independent samples T-test** p-value
Age	45	48	0.204
Gender	0.349	0.377	0.670
Study	0.481	0.491	0.891
BMI	24.4	25.3	0.067
Obese	0.097	0.123	0.557
Overweight	0.379	0.453	0.279
Smoke	0.198	0.132	0.197
Alcohol consumer	0.877	0.877	1.000
Binge Drinker	0.368	0.396	0.673

3.6 Model specification

To find out if there are different normative viewpoints, between the treatment group (respondents from the survey with information) and control group (respondents from the survey without information), Ordinary Least Squares (OLS) regressions will be used. This gave the following equations:

$$Y_r = \alpha + \beta_1 version_r + \beta_2 gender_r + \beta_3 age_r + \varepsilon_r \quad (1)$$

$$Y_r = \alpha + \beta_1 version_r + \beta_2 gender_r + \beta_3 age_r + \beta_4 smoker_r + \beta_5 drinker_r + \beta_6 binge\_drinker_r + \beta_7 obese_r + \beta_8 overweight_r + \beta_9 study_r + \varepsilon_r \quad (2)$$

$$Y_r = \alpha + \beta_1 version_r + \beta_2 gender_r + \beta_3 age_r + \beta_4 healthy_r + \beta_5 unhealthy_r + \varepsilon_r \quad (3)$$

$$Y_r = \alpha + \beta_1 version_r + \beta_2 gender_r + \beta_3 age_r + \beta_4 smoker_r + \beta_5 drinker_r + \beta_6 binge\_drinker_r + \beta_7 obese_r + \beta_8 overweight_r + \beta_9 study_r + \beta_{10} healthy_r + \beta_{11} unhealthy_r + \varepsilon_r \quad (4)$$

The Y_r is the dependent variable and indicates the probability that respondent *r* answers higher, lower or unchanged to the questions. It takes value '1' if respondent *r* answers higher and takes value '0' if respondent *r* answers lower or unchanged. This holds for the questions on whether unhealthy people, smokers, heavy alcohol users and people who are obese/overweight should pay higher health insurance premiums. It takes value '1' if respondent *r* answers lower and takes value '0' if respondent *r* answers higher or unchanged. This holds for the questions on whether healthy people, non-smokers, not heavy alcohol users and people who are not obese/overweight should pay lower health insurance premiums.

The $version_r$ variable is a control variable that indicates which survey respondent *r* got. The $smoker_r$, $drinker_r$, $binge\_drinker_r$, $obese_r$ and $overweight_r$ dummy variables indicate if respondent *r* belongs to one or more of these groups. The $gender_r$ and age_r variables are control variables and indicate the gender and age of respondent *r*. The $study_r$ dummy variable indicates if respondent *r* completed a study in higher education. These control variables are not necessary to take away the bias, because the two surveys were assigned randomly. They can however correct for possible remaining differences that arise by chance between the two survey groups. The coefficients of some variables, for example $gender_r$ and $study_r$ are also interesting on their own. The $healthy_r$ and $unhealthy_r$ control variables, control for the answers on whether healthy/unhealthy people should pay lower/higher health insurance premiums. The equations 3 and 4 with these two variables, control for the initial difference between the treatment and control group. Lastly, the ε_r is the error term and the β are the regression coefficients.

4 Results

This chapter will present the results from the HRS and survey data. First, the HRS data shows if genetic and childhood SES factors increase the probability of smoking, alcohol use and obesity. Secondly, the survey data provides the different normative viewpoints on the fairness of equal health insurance premiums.

4.1 Health and Retirement Study

4.1.1 Smoking

The regression results in table 3 show the effect of the cSES, PGS, interaction and gender variables on the probability to ever smoke and be a current smoker. The PGS for smoking initiation is used here, which indicates the genetic liability that a respondent has ever smoked regularly. The output from the ever smoke regression shows significant positive coefficients, at a p-value threshold of 0.01, for the variables PGS and gender. It shows significant negative coefficients for the variables cSES and interaction, at respectively a p-value threshold of 0.01 and 0.05. People with a high cSES have on average a lower probability to ever smoke compared to people with a low cSES. People with a high genetic risk have on average a higher probability to ever smoke compared to people with a low genetic risk. The interaction effect suggests that the probability to ever smoke increases faster for people with a low cSES when the genetic risk increases, than for people with a high cSES. Men have on average a higher probability to ever smoke compared to women.

The output from the current smoker regression shows a significant positive coefficient, at a p-value threshold of 0.01, for the variable PGS and a negative coefficient for the variable gender at the same p-value threshold. Results show insignificant coefficients for the variables cSES and interaction. People with a high genetic risk have on average a higher probability to be a current smoker compared to people with a low genetic risk. Men have on average a lower probability to be a current smoker compared to women.

Table 3. Results OLS regressions for smoking

Variable	Ever smoke	Current smoker
cSES	-.017***	-.008

	(.006)	(.008)
$PGS_{smoking\_initiation}$	.075*** (.005)	.024*** (.006)
Interaction	-.014** (.006)	.004 (.004)
Gender	.184*** (.010)	-.059*** (.011)
Constant	.490*** (.007)	.214*** (.008)
Observations	9092	5126

Note. Standard errors are in parentheses; * p<0.01, ** p<0.05 and *** p<0.10.

The regressions in appendix 3 are almost the same as in table 3, only the PGS is different. The PGS for cigarettes per day are used here and indicates the heaviness of smoking, either as a former or current smoker. The results from these regressions are almost identical. The signs from these regressions are the same and the magnitudes are approximately the same. The significance of the coefficients are also the same, except for the coefficient of the interaction variable which is only significant in the regression from table 3 with ever smoke as dependent variable.

Figure 1 shows that as genetic risk increases, the likelihood of smoking also increases. The intersection of the two lines indicates that when the genetic risk is higher than -1, the probability of smoking is higher for people with a low cSES (red line). If the genetic risk is lower than -1, people with a high cSES (blue line) are more likely to smoke. The probability of smoking increases faster for people with a low cSES when the genetic risk increases, than for people with a high cSES. Figure 1 and the explanation is also included in the survey for the treatment group.

Figure 1. The effect of the childhood SES and genetic risk on the probability to ever smoke.

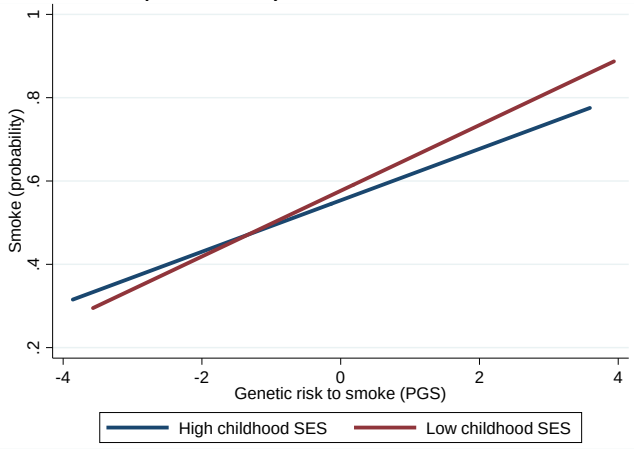

Table 1 shows the probability to ever smoke or be a current smoker in percentages, for men and women separately. The cSES(low) in table 1 indicates a low childhood SES and the PGS(high) indicates a high genetic risk for smoking initiation or cigarettes per day. The coefficients in the regressions were used to calculate the percentages and shows the magnitude of the effect. Appendix 4 shows how these calculations were done, the ever smoke regression results in table 3 are used here as example.

Table 4 shows that a low childhood SES increases the probability to ever smoke with only 5-6% (for men) and 7-8% (for women). A high PGS for smoking initiation increases the probability to ever smoke with 22% (for men) and 31% (for women). A high genetic risk increases the probability to be a current smoker with 28-31% (for men) and 20-22% (for women). Men with a high genetic risk and a low childhood SES have on average a 75-86% probability to ever smoke and men with a low genetic risk and a high childhood SES have on average a 49-60% probability to ever smoke. Women with a high genetic risk and a low childhood SES have on average a 57-67% probability to ever smoke and women with a low genetic risk and a high childhood SES have on average a 31-42% probability to ever smoke. The empty cells in table 4 and 6 mean that the coefficients for those variables were insignificant and no calculations were done.

Table 4. Effect of regression results in percentages for smoking

Variable	cSES(low)	PGS(high)	High - low
PGS Smoking Initiation			
Ever smoke (male)	5.0%	22.3%	85.8% - 49%
Current smoker (male)	-	31.0%	-
Ever smoke (female)	6.9%	30.6%	67.4% - 30.6%
Current smoker (female)	-	22.4%	-
PGS Cigarettes Per Day			
Ever smoke (male)	5.9%	5.1%	74.7% - 59.9%
Current smoker (male)	-	28%	-
Ever smoke (female)	8.1%	6.9%	56.7% - 41.9%
Current smoker (female)	-	20.3%	-

Note. The low and high in parentheses indicates respectively that the cSES variable takes value -2 and the PGS variable takes value 2.

The results in this subsection suggest that there is evidence of a significant effect of childhood SES and genetics on the probability to ever smoke and be a current smoker. A high genetic risk increases the probability to ever smoke or be a current smoker and this also holds for a low

childhood SES. There is also some evidence of an interaction effect on the probability to smoke. The probability of smoking increases faster for people with a low childhood SES when the genetic risk increases, than for people with a high childhood SES.

4.1.2 Alcohol use

The regression results in table 5 show the effect of the cSES, PGS, interaction and gender variables on the probability to be an alcoholic, binge drinker and on the number of alcoholic drinks per week. The PGS for drinks per weeks is used here, which indicates the genetic liability for the number of alcoholic drinks consumed per week. The output from the alcoholic regression shows significant positive coefficients, at a p-value threshold of 0.01, for all variables, except for the interaction variable which is insignificant. People with a high cSES have on average a higher probability to be an alcoholic compared to people with a low cSES. People with a high genetic risk have on average a higher probability to be an alcoholic compared to people with a low genetic risk. Men have on average a higher probability to be an alcoholic than women.

The output from the binge drinker regression shows the same (in)significant coefficients and the same signs, except for the coefficient of the interaction variable which has a negative sign. The magnitude of the coefficients do also not differ much, only for the gender variable. The probability to be a binge drinker is on average higher for men, compared to women. The magnitude of this effect is higher in this regression than in the alcoholic regression.

The output from the drinks per week regression shows a significant positive coefficient for the variables PGS and gender, at a p-value threshold of 0.01. People with a high genetic risk consume on average more alcoholic drinks per week compared to people with a low genetic risk. Men consume on average 3.5 alcoholic drinks more per week than women. The effect of the cSES is insignificant. Appendix 5 shows the regression results with drinks per week as dependent variable and the $PGS_{alcohol\_dependence}$ variable. These results are almost identical as the regression results from the drinks per week regression in table 5.

Table 5. Results OLS regressions for alcohol use

Variable	Alcoholic	Binge drinker	Drinks per week
cSES	.018*** (.007)	.025*** (.009)	.078 (.182)
$PGS_{drinks\_per\_week}$	.028***	.032***	.623***

	(.006)	(.009)	(.172)
Interaction	.005 (.006)	-.005 (.009)	.007 (.172)
Gender	.084*** (.011)	.179*** (.015)	3.561*** (.312)
Constant	.128*** (.007)	.171*** (.011)	5.050*** (.169)
Observations	5035	3498	3514

Note. Standard errors are in parentheses; * p<0.01, ** p<0.05 and *** p<0.10.

Table 6 shows that a high genetic risk increases the probability to be an alcoholic or a binge drinker with 18-26% (for men) and 37-44% (for women). A high genetic risk increases the consumption of alcoholic drinks with 10-15% (for men) and 16-25% (for women). Men with a high genetic risk and a low childhood SES have on average a 23% probability to be an alcoholic and men with a low genetic risk and a high childhood SES have on average a 19% probability to be an alcoholic. Women with a high genetic risk and a low childhood SES have on average a 15% probability to be an alcoholic and women with a low genetic risk and a high childhood SES have on average a 11% probability to be an alcoholic. The results suggest that a low childhood SES decreases the probability to be an alcoholic or a binge drinker.

Table 6. Effect of regression results in percentages for alcohol use

	SES (low)	PGS(high)	High - low
PGS drinks per week			
Alcoholic (male)	-17.0%	26.4%	23.2% - 19.2%
Alcoholic (female)	-28.1%	43.8%	14.8% - 10.8%
DPW (male)	-	14.5%	-
DPW (female)	-	24.7%	-
Binge (male)	-14.3%	18.3%	36.4% - 33.6%
Binge (female)	-29.2%	37.4%	18.5% - 15.7%
PGS alcohol dependence			
DPW (male)	-	9.7%	-
DPW (female)	-	16.4%	-

Note. The low and high in parentheses indicates respectively that the cSES variable takes value -2 and the PGS variable takes value 2.

The results in this subsection suggest that there is evidence of a significant effect of genetics on the probability to be an alcoholic, binge drinker and on the number of alcoholic drinks per week. A high genetic risk increases the probability to be a alcoholic or a binge drinker and also increases the number of drinks consumed per week. There is no evidence that a low childhood SES increases the probability to be an alcoholic, binge drinker and the number of alcoholic drinks per week. There is also no evidence of an interaction effect.

4.1.3 Obesity

The regression results in table 7 show the effect of the cSES, PGS, interaction and gender variables on the probability to be obese/overweight. The PGS for BMI is used here, which indicates the genetic liability that a respondent has a high BMI. The output shows significant positive coefficients, at a p-value threshold of 0.01, for the variable PGS in both regressions. Results also show a significant positive coefficient for the variable gender, at a p-value threshold of 0.01, in the overweight regression. The coefficients in both regressions for the variables cSES and interaction are insignificant. This suggests that there is no effect of the childhood SES on the probability to be obese or overweight. People with a high genetic risk have on average a higher probability to be obese or overweight compared to people with a low genetic risk. Men have on average a higher probability to be overweight than women.

Table 7. Results OLS regressions for obesity

Variable	Obese	Overweight
cSES	-.0005 (.008)	-.002 (.008)
PGS_{BMI}	.091*** (.007)	.078*** (.007)
Interaction	.007 (.007)	.005 (.008)
Gender	.008 (.013)	.129*** (.013)
Constant	.305*** (.009)	.625*** (.010)
Observations	4710	4710

Note. Standard errors are in parentheses; * p<0.01, ** p<0.05 and *** p<0.10.

Table 8 shows that a high genetic risk increases the probability to be obese with 60% and increases the probability to be overweight with 21% (for men) and 25% (for women).

Table 8. Effect of regression results in percentages for obesity

	PGS(high)
PGS BMI	
Obese	59.7%
Overweight (male)	20.7%
Overweight (female)	25.0%

Note. The high in parentheses indicates that the PGS variable takes value 2.

The results in this subsection suggest that there is evidence of a significant effect of genetics on the probability to be obese and overweight. A high genetic risk increases the probability to be obese or overweight. There is no evidence of a childhood SES and interaction effect.

4.2 Survey

4.2.1 Information survey

This subsection presents the information that is used in the survey for the treatment group and that is not already mentioned in other sections. First, there is a significant effect of childhood SES and genetics on smoking. A high genetic risk increases the probability to be a current smoker with 22-31%. The probability to ever smoke for people with a high genetic risk and a low childhood SES is 86% (for men) and 49% (for women). The probability to ever smoke for people with a low genetic risk and a high childhood SES is 67% (for men) and 31% (for women).

Next, there is a significant effect of genetics on alcohol use. A high genetic risk increases the number of alcoholic drinks consumed per week with 10-25%. The probability to be an alcoholic for people with a high genetic risk and a low childhood SES is 23% (for men) and 15% (for women). The probability to be an alcoholic for people with a low genetic risk and a high childhood SES is 19% (for men) and 11% (for women). There is a significant effect of childhood SES on alcohol use, however it seems that people with a high childhood SES consume more alcohol. Lastly, there is a significant effect of genetics on obesity. A high genetic risk increases the probability of obesity with 60% and increases the probability of being overweight with 21% (for men) and 25% (for women). The results showed no effect of childhood SES on the probability to be obese/overweight.

4.2.2 Simple comparison

First, it is determined if the respondents have different normative viewpoints between the two surveys before any information is provided. The first two questions, on whether unhealthy/healthy people should pay higher/lower health insurance premiums, are used to determine this. Figure 2 shows that in the control group a higher percentage thinks that

healthy people should pay lower health insurance premiums. The t-test results in table 9 show that this difference is significant, at a p-value threshold of 0.1. There is no significant difference in opinion between the control and treatment group on the fairness of equal health insurance premiums for unhealthy people. The same opinion on these two questions was expected between the treatment and control groups, however it seems that there is an initial difference in opinion between the two groups on whether healthy people should pay lower health insurance premiums.

Figure 2. Different viewpoints treatment- and control group

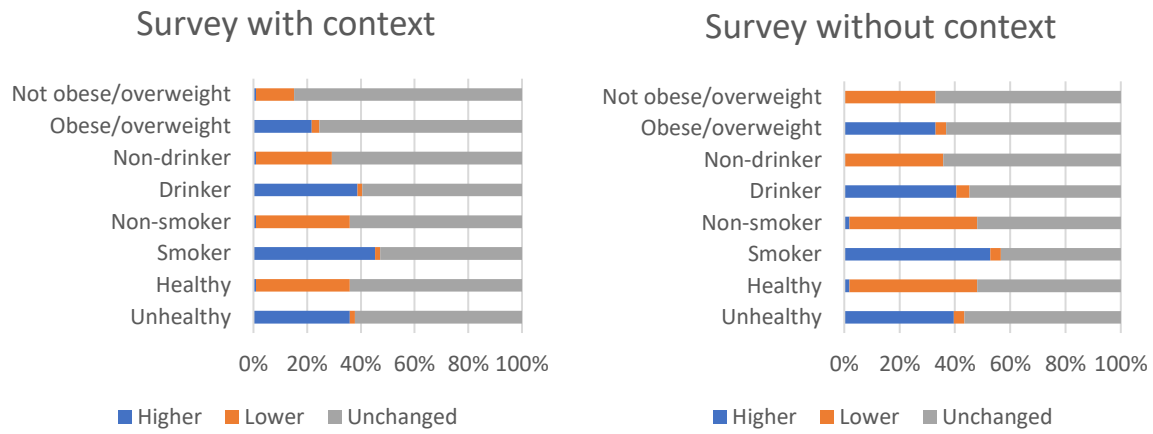

Table 9 shows that there is no significant difference in opinion between the control and treatment group on the fairness of equal health insurance premiums for smokers, heavy alcohol users and not heavy alcohol users. In the control group a higher percentage thinks that people who are obese/overweight should pay higher health insurance premiums. This difference is significant at a p-value threshold of 0.1. In the control group a higher percentage thinks that non-smokers and people who are not obese/overweight should pay lower health insurance premiums. This difference is significant at respectively a p-value threshold of 0.1 and 0.01.

Table 9. Independent samples t-test on difference between treatment-and control group

Variables	Mean difference (standard error difference)	t-value	Degrees of Freedom	p-value
Unhealthy	0.038 (0.067)	0.565	210	0.573
Healthy	0.113 (0.067)	1.682	210	0.094
Smoker	0.075 (0.069)	1.097	210	0.274

Non-smoker	0.113 (0.067)	1.682	210	0.094
Drinker	0.019 (0.067)	0.280	210	0.780
Non-drinker	0.075 (0.064)	1.175	210	0.241
Obese/overweight	0.113 (0.061)	1.855	210	0.065
Not obese/overweight	0.189 (0.057)	3.303	210	0.001

4.2.3 OLS regressions

The regression results in table 10, on whether unhealthy people should pay higher health insurance premiums, show insignificant coefficients for the variable version. There is no significant difference in opinion between the control and treatment group on the fairness of equal health insurance premiums for unhealthy people. The regression results, on whether healthy people should pay lower healthy insurance premiums, show significant negative coefficients for the variable version at a p-value threshold of 0.1. The probability that the treatment group thinks that healthy people should pay lower health insurance premiums is on average 12 percentage points lower, compared to the control group.

The regression with control variables, on whether unhealthy people should pay higher health insurance premiums, shows significant negative and positive coefficients at a p-value threshold of 0.01, for respectively the variables smoker and gender. Results show significant negative coefficients for the variables overweight and study, at respectively a p-value threshold of 0.1 and 0.05. The probability that men think that unhealthy people should pay higher health insurance premiums is on average 20 percentage points higher. The probability that smokers, people who are overweight and people that completed a study in higher education think that unhealthy people should pay higher health insurance premiums is respectively 23, 13 and 14 percentage points lower on average.

The regression with control variables, on whether healthy people should pay lower health insurance premiums, shows significant negative coefficients for the variables smoker and study, at respectively a p-value threshold of 0.1 and 0.05. The probability that smokers or

people that completed a study in higher education think that healthy people should pay lower health insurance premiums is on average 15 percentage points lower.

Table 10. Results OLS regressions for healthy- and unhealthy people

Variables	Unhealthy	Unhealthy (with control variables)	Healthy	Healthy (with control variables)
Version	-.036 (.067)	-.030 (.066)	-.119* (.067)	-.120* (.067)
Gender		.202*** (.074)		-.005 (.076)
Age		-.004 (.003)		-.002 (.003)
Smoker		-.232*** (.079)		-.148* (.084)
Drinker		-.096 (.102)		.111 (.101)
Binge drinker		-.027 (.086)		.132 (.087)
Obese		-.033 (.106)		-.032 (.111)
Overweight		-.130* (.072)		-.064 (.078)
Study		-.140** (.067)		-.150** (.068)
Constant	.396*** (.048)	.745*** (.188)	.462*** (.049)	.526*** (.186)

Note. Standard errors are in parentheses; * p<0.01, ** p<0.05 and *** p<0.10.

The rest of the regressions in this subsection will also control for the answers on whether healthy/unhealthy people should pay lower/higher health insurance premiums. As explained in subsection 3.6 the variables healthy and unhealthy control for the initial difference between the treatment and control group before any information was provided. The results for all four regressions in table 11 show insignificant coefficients for the variable version. There is no evidence that there is a significant difference in opinion between the treatment and control group on the fairness of equal health insurance premiums for smokers and non-smokers.

The regression with control variables, on whether smokers should pay higher health insurance premiums, shows significant positive coefficients at a p-value threshold of 0.05 and 0.01, for respectively the variables healthy and unhealthy. The probability that respondents think that smokers should pay higher health insurance premiums is on average 12 percentage points higher, when respondents think that healthy people should pay lower health insurance premiums. The probability that respondents think that smokers should pay higher health

insurance premiums is on average 65 percentage points higher, when respondents think that unhealthy people should pay higher health insurance premiums. The results show a significant negative coefficient for the variable smoker at a p-value threshold of 0.01 and a significant positive coefficient for the variable study at a p-value threshold of 0.1. The probability that smokers think that smokers should pay higher health insurance premiums is on average 24 percentage points lower. The probability that people that completed a study in higher education think that smokers should pay higher health insurance premiums is on average 8 percentage points higher.

The regression with control variables, on whether non-smokers should pay lower health insurance premiums, shows a significant positive coefficient for the variable healthy, at a p-value threshold of 0.01. The probability that respondents think that non-smokers should pay lower health insurance premiums is on average 78 percentage points higher, when respondents think that healthy people should pay lower health insurance premiums.

Table 11. Results OLS regressions for smokers and non-smokers

Variables	Smoker	Smoker (with control variables)	Non-smoker	Non-smoker (with control variables)
Version	-.037 (.049)	-.026 (.048)	-.023 (.041)	-.019 (.042)
Healthy	.113* (.061)	.124** (.057)	.779*** (.052)	.783*** (.051)
Unhealthy	.688*** (.054)	.650*** (.057)	.055 (.053)	.039 (.053)
Gender		.013 (.052)		.070 (.051)
Age		-.002 (.002)		.001 (.002)
Smoker		-.244*** (.066)		-.002 (.033)
Study		.077* (.047)		-.002 (.042)
Constant	.204*** (.046)	.804*** (.151)	.081** (.038)	.027 (.105)

Note. Standard errors are in parentheses; * p<0.01, ** p<0.05 and *** p<0.10.

The results for all four regressions in table 12 show insignificant coefficients for the variable version. The evidence form the HRS was the least strong for alcohol use, so this result suggests that people respond less to information that is weaker. There is no evidence that there is a significant difference in opinion between the treatment and control group on the fairness of equal health insurance premiums for heavy alcohol users and not heavy alcohol users.

The regression with control variables, on whether heavy alcohol users should pay higher health insurance premiums, shows significant positive coefficients at a p-value threshold of 0.05 and 0.01, for respectively the variables healthy and unhealthy. The probability that respondents think that heavy alcohol users should pay higher health insurance premiums is on average 15 percentage points higher, when respondents think that healthy people should pay lower health insurance premiums. The probability that respondents think that heavy alcohol users should pay higher health insurance premiums is on average 63 percentage points higher, when respondents think that unhealthy people should pay higher health insurance premiums. The results also show a significant positive coefficient for the variable age, at a p-value threshold of 0.1. The probability that respondents think that heavy alcohol users should pay higher health insurance premiums increases on average with 0,4 percentage points for each additional year of age.

The regression with control variables, on whether not heavy alcohol users should pay lower health insurance premiums, shows a significant positive coefficient for the variable healthy, at a p-value threshold of 0.01. The probability that respondents think that not heavy alcohol users should pay lower health insurance premiums is on average 66 percentage points higher, when respondents think that healthy people should pay lower health insurance premiums. Results show significant negative coefficients for the variables binge drinker and study, at respectively a p-value threshold of 0.05 and 0.1. The probability that binge drinkers and people who completed a study in higher education think that not heavy alcohol users should pay lower health insurance premiums is respectively 13 and 9 percentage points lower on average.

Table 12. Results OLS regressions for drinkers and non-drinkers

Variables	Drinker	Drinker (with control variables)	Non-drinker	Non-drinker (with control variables)
Version	.019 (.051)	.031 (.054)	-.002 (.047)	-.002 (.048)
Healthy	.134** (.068)	.147** (.068)	.641*** (.061)	.661*** (.058)
Unhealthy	.611*** (.065)	.627*** (.064)	.031 (.061)	.004 (.056)
Gender		-.018 (.055)		.068 (.055)
Age		.004* (.002)		.001 (.002)
Drinker		-.023 (.057)		.013 (.074)
Binge drinker		.000		-.134**

		(.068)		(.057)
Study		.050 (.050)		-.086* (.051)
Constant	.101** (.041)	-.094 (.133)	.050 (.036)	.082 (.134)

Note. Standard errors are in parentheses; * p<0.01, ** p<0.05 and *** p<0.10.

The regression with control variables, on whether people who are obese/overweight should pay higher health insurance premiums, shows a significant negative coefficient for the variable version, at a p-value threshold of 0.1. The probability that the treatment group thinks that people who are obese/overweight should pay higher health insurance premiums is on average 9 percentage points lower. The regression results, on whether people who are not obese/overweight should pay higher health insurance premiums, show significant negative coefficients for the variable version, at a p-value threshold of 0.01. The probability that the treatment group thinks that people who are not obese/overweight should pay lower health insurance premiums is on average 13-14 percentage points lower.

The regression with control variables, on whether people who are obese/overweight should pay higher health insurance premiums, shows significant positive coefficients at a p-value threshold of 0.1 and 0.01, for respectively the variables healthy and unhealthy. The probability that respondents think that people who are obese/overweight should pay higher health insurance premiums is on average 12 percentage points higher, when respondents think that healthy people should pay lower health insurance premiums. The probability that respondents think people who are obese/overweight should pay higher health insurance premiums is on average 47 percentage points higher, when respondents think that unhealthy people should pay higher health insurance premiums. Results show a significant positive coefficient, at a p-value threshold of 0.05, for the variable gender and a significant negative coefficient for the variable age at a p-value threshold of 0.1. The probability that men think that people who are obese/overweight should pay higher health insurance premiums is on average 14 percentage points higher. The probability that respondents think that people who are obese/overweight should pay higher health insurance premiums decreases on average with 0,4 percentage points for each additional year of age.

The regression with control variables, on whether people who are not obese/overweight should pay lower health insurance premiums, shows a significant positive coefficient for the variable healthy at a p-value threshold 0.01. The probability that respondents think that

people who are not obese/overweight should pay lower health insurance premiums is on average 42 percentage points higher, when respondents think that healthy people should pay lower health insurance premiums. Results show significant positive and negative coefficients, at a p-value threshold of 0.01, for respectively the variables gender and study. The probability that men think that people who are not obese/overweight should pay lower health insurance premiums is on average 17 percentage points higher. The probability that people who completed a study in higher education think that people who are not obese/overweight should pay lower health insurance premiums is on average 15 percentage points lower.

Table 13. Results OLS regressions for obese/overweight- and not-obese/overweight people

Variables	Obese	Obese (with all control variables)	Not-obese	Not-obese (with all control variables)
Version	-.081 (.049)	-.085* (.050)	-.134*** (.048)	-.141*** (.046)
Healthy	.117* (.063)	.121* (.065)	.453*** (.064)	.424*** (.056)
Unhealthy	.510*** (.067)	.465*** (.071)	.083 (.061)	.063 (.055)
Gender		.143** (.056)		.174*** (.049)
Age		-.004* (.002)		.002 (.002)
Obese		.102 (.065)		.106 (.079)
Overweight		-.019 (.058)		-.013 (.051)
Study		-.031 (.049)		-.146*** (.047)
Constant	.074** (.034)	.215* (.114)	.088** (.034)	.011 (.108)

Note. Standard errors are in parentheses; * p<0.01, ** p<0.05 and *** p<0.10.

5 Conclusion

The goal of this thesis was to identify if childhood socioeconomic status and genetic factors increase the probability of risky health behavior and whether confronting people with this information change their opinions on the fairness of equal health insurance premiums. The descriptive statistics from the HRS data and survey data was first looked at. Then, Ordinary Least Squares (OLS) regressions were used to identify the effect of the childhood SES and genetic factors on the probability of risky health behavior. OLS regressions were also used to identify if there are different normative viewpoints between the survey with information and without information.

The answer to the research question: '*Which factors beyond an individual´s control increase the probability of risky health behavior (smoking, alcoholism, obesity) and does confronting people with this information change their opinions on the fairness of equal health insurance premiums?*', consists of two parts.

First of all, results showed that a low childhood SES increases the probability for smoking and a high genetic risk increases the probability for smoking, alcoholism and obesity. So there is evidence of factors beyond an individual´s control that increases the probability of risky health behavior. Secondly, there is evidence that confronting people with this information change their opinions on the fairness of equal health insurance premiums. After controlling for initial differences, the fraction of respondents in the treatment group who thinks that people who are not obese/overweight should pay lower health insurance premiums is 14 percentage points lower. The fraction of respondents in the treatment group who thinks that people who are obese/overweight should pay higher health insurance premiums is 9 percentage points lower. Before controlling for initial differences, the fraction of respondents in the treatment group who thinks that non-smokers should pay lower health insurance premiums is significantly lower. There is also some evidence that women and people that completed a study in higher education are less likely to think that people who engage in risky healthy behavior should pay higher health insurance premiums.

There are several limitations to this research. First, the sample sizes of the different regressions from the HRS data are not the same. The regression with the largest sample consists of 9092 observations and the regression with the smallest sample consists of 3498 observations. The sample size of the latter regression is rather small, this increases the margin of error and reduces the power of the model. Secondly, the OLS regressions are all linear, however it might be that the relation between the indicator variables and dependent variables are non-linear. Lastly, BMI is calculated with only the weight and height of a person and this is used to indicate if people are obese or overweight. In this calculation a person's body fat percentage and muscle mass percentage are not included. These factors are also important to determine if people are obese or overweight. So the measure for BMI is not as precise as it could be.

After this research there are some implications for policy and potential future research. It would be interesting to see what other factors, beyond a person's own control, has an effect on the probability for risky health behavior and does confronting people with this information change the opinion of people on the fairness of equal health insurance premiums. Furthermore, several studies suggest that people with a lower life expectancy, because of smoking, alcoholism and obesity, have lower health care costs in total (Panhuis – Plasmans, Luijben & Hoogenveen, 2012). Including this information into the survey from this research might change the answers that the respondents gave to the questions. It would be interesting to see what this effect might be. Next, the results in this thesis suggest that confronting people with information might change their opinions on the fairness of equal health insurance premiums. It would also be interesting to see what this information contributes to potential research, on the fairness of certain payments for other types of medical expenses. Lastly, health insurers and the government often receive negative comments and media attention, because of the equal basic health insurance premium. Providing the information that there are factors beyond an individual's control that increase the probability of risky health behavior, might decrease the negative comments and media attention.

References

Alesina, A., Stantcheva, S., & Teso, E. (2018). Intergenerational Mobility and Preferences for Redistribution. American Economic Review, 108 (2), 521–554.

Barr, P.B., Silberg, J., Dick, D.M., & Maes, H. H. (2018). Childhood socioeconomic status and longitudinal patterns of alcohol problems: Variation across etiological pathways in genetic risk. Social science & medicine (1982), 209, 51–58.

Bierut, L., Biroli, P., Galama, T., & Thom, K. (2018). Childhood socioeconomic status moderates genetic predisposition for peak smoking. bioRxiv, 336834.

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. Nat. Rev. Genet., 3, 872–82.

Broms, U., Silventoinen, K., Madden, P.A.F., Heath, A.C., & Kaprio, J. (2006). Genetic architecture of smoking behavior: a study of Finnish adult twins. Twin Res Hum Genet, 9, 64–72.

Card, D., Mas, A., Moretti, E., & Saez, E. (2010). "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction". NBER Working Paper No. 16396.

Chapman, B.P., Fiscella, K., Duberstein, P., Coletta, M., & Kawachi, I. (2009). Can the influence of childhood socioeconomic status on men's and women's adult body mass be explained by adult socioeconomic status or personality? Findings from a national sample. Health Psychology, 28, 419–427.

Charite, J., Fisman, R., & Kuziemko, I. (2016). Reference points and redistributive preferences: Experiment evidence. NBER Working Paper 21009.

Collins, S.E: (2016). Associations between socioeconomic factors and alcohol outcomes. Alcohol Res, 38(1), 83-94.

Cruces, G., Perez-Truglia, R., & Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. Journal of Public Economics, 98, 100–112.

De Witt, S. (2019). Solidariteit in de gezondheidszorg; opvattingen over zorgpremie en leefstijl. CBS, Retrieved February 17, 2021, from https://www.cbs.nl/nl-nl/achtergrond/2019/40/solidariteit-in-de-gezondheidszorg

Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., … & McCarthy, M.I. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science (New York, N.Y.), 316(5826), 889–894.

GGZ. (n.d.). retrieved April 19, 2021, from https://ggzinterventie.nl/alcoholverslaving/

Heath, A.C., et al. (1997). Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. Psychol Med, 27, 1381–96.

HRS Data Products. (n.d.). Retrieved February 19, 2021, from https://hrs.isr.umich.edu/data-products

In ’t Panhuis–Plasmans, M., Luijben, G., & Hoogenveen, R. (2012). Zorgkosten van ongezond gedrag. Kosten van ziekten notities 2012-2. RIVM, Bilthoven.

Isogg (n.d.). Retrieved June 15, 2021, from https://isogg.org/wiki/Single-nucleotide_polymorphism

Jacob, T., Waterman, B., Heath, A., True, W., Bucholz, K.K., Haber, R., Scherrer, J., & Fu, Q. (2003). Genetic and environmental effects on offspring alcoholism: new insights using an offspring-of-twins design. Arch Gen Psychiatry, 60, 1256–1264.

Jefferis, B.J., Power, C., Graham, H., & Manor, O. (2004). Effects of childhood socioeconomic circumstances on persistent smoking. American Journal of Public Health, 94(2), 279–285.

Jo, Y. (2014). What money can buy: family income and childhood obesity. Econ. Hum. Biol., 15, 1–12 .

Katikireddi, S.V., Whitley, E., Lewsey, J., Gray, L., & Leyland, A.H. (2017). Socioeconomic status as an effect modifier of alcohol consumption and harm: analysis of linked cohort data. Lancet Public Health, 2 , e267-e276

Kloosterman, R. (2011). Solidariteit in de gezondheidszorg. Den Haag: Centraal Bureau voor de Statistiek.

Koopmans, J. R., Slutske, W. S., Heath, A. C., Neale, M. C., & Boomsma, D. I. (1999). The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. Behav. Genet., 29, 383–393.

Liu, M., Jiang, Y., Wedow, R., et al. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet., 51(2), 237–244.

Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., ... & Croteau-chonka, D.C. (2015). Genetic studies of body mass index yield new insights for obesity biology.Nature, 518(7538), 197-206.

Maes, H.H., Neale, M.C., & Eaves, L.J. (1997). Genetic and environmental factors in relative body weight and human adiposity. Behav. Genet., 27, 325–351.

McLaren, L. (2007). Socioeconomic status and obesity. Epidemiologic Reviews, 29(1), 29–48.

Murasko, J.E. (2009). Socioeconomic status, height, and child obesity. Econ.Human Biol., 7 (3), 376–386.

Murasko, J.E. (2013). Associations between household income, height, and BMI in contemporary US schoolchildren. Econ. Human Biol., 11 (2),185–196.

Nationaal Preventieakkoord. (2018). Retrieved February 19, 2021, from file:///C:/Users/laure/AppData/Local/Temp/nationaal-preventieakkoord.pdf

Poulton, R., Caspi, A., Milne, B.J., Thomson, W.M., Taylor, A., Sears, M.R., & Moffitt, T.E. (2002). Association between children's experience of socioeconomic disadvantage and adult health: a life-course study. Lancet (London, England), 360(9346), 1640–1645.

Power, C., Graham, H., Due, P., Hallqvist, J., Joung, I., Kuh, D., et al. (2005a). The contribution of childhood and adult socioeconomic position to adult obesity and smoking behaviour: an international comparison. International Journal of Epidemiology, 34(2), 335–344.

Prescott, C.A., & Kendler, K.S. (1999). Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. Am J Psychiatry, 156, 34-40.

Price, R. A., & Gottesman (1991). Body fat in identical twins reared apart: Roles for genes and environment. Be-hav. Genet., 21, 1-7.

Qualtrics (2021). Retrieved June 30, 2021, from https://www.qualtrics.com/about/

Qualtrics (n.d.). Retrieved July 10, 2021, from https://www.qualtrics.com/experience-management/research/determine-sample-size/

Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., ... & Workalemahu, T. (2015). New genetic loci link adipose and insulin biology to body fat distribution. Nature, 518(7538), 187-196.

Sobal, J., & Stunkard, A.J. (1989). Socioeconomic status and obesity: A review of the literature. Psychological Bulletin, 105(2), 260–275.

The Tobacco and Genetics Consortium et al. (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature Genetics, 42(5), 441–447.

Vable, A.M., Gilsanz, P., Nguyen, T.T., Kawachi, I., Glymour, M.M. (2017). Validation of a theoretically motivated approach to measuring childhood socioeconomic circumstances in the Health and Retirement Study. PLoS ONE, 12(10), e0185898.

Voedingscentrum. (n.d.). retrieved June 11, 2021, from https://www.voedingscentrum.nl/nl/bmi-meter

Walters, R.K., Polimanti, R., Johnson EC, et al. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat Neurosci., 21(12), 1656–1669.

Ware, E., Gard, A., Schmitz, L., & Fau, J. (2021). Retrieved June 15, 2021, from https://hrsdata.isr.umich.edu/sites/default/files/documentation/data-descriptions/PGNSCORES4_3DD_0.pdf

Yengo, L., Sidorenko, J., Kemper, K.E., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum Mol Genet., 27(20), 3641–3649.

Zorgverzekering (n.d.). Retrieved February 22, 2021, from https://www.zorgverzekering.org/algemene-informatie/verschillende-polissen/zorgpremie/

Zorgwijzer (2019). Retrieved July 10, 2021, from https://www.zorgwijzer.nl/zorgverzekering-2019/nederland-telt-230000-wanbetalers-zorgverzekering

Appendix

Appendix 1.

Descriptive statistics (HRS data)

Variable	Mean (SD)	Min	Max	N
$PGS_{cigarettes\_per\_day}$	-0.002 (1.002)	-4.103	3.646	10282
$PGS_{drinks\_per\_week}$	0.002 (0.999)	-3.903	3.974	10282
$PGS_{alcohol\_dependence}$	-0.005 (1.000)	-3.945	3.939	10282
PGS_{BMI}	-0.005 (0.996)	-3.636	4.078	10282

Appendix 2.

Qualtrics Survey Software.pdf

Appendix 3.

Results OLS regressions for smoking

Variable	Ever smoke	Current smoker
cSES	-.020*** (.006)	-.008 (.006)
$PGS_{cigarettes\_per\_day}$	.017*** (.005)	.022*** (.006)
Interaction	-.003 (.006)	.005 (.006)
Gender	.180*** (.010)	-.060*** (.011)
Constant	.493*** (.007)	.217*** (.008)
Observations	9092	5126

Note. Standard errors are in parentheses; * p<0.01, ** p<0.05 and *** p<0.10.

Appendix 4.

The output from the ever smoke regression in table 3 are used here as an example for the calculations. The probability that men, with a high childhood SES and low PGS, ever smoke is: 0.490 + (– 0.017 * 2) + (0.075 * -2) + 0.184= 0.490. The probability that men, with a low childhood SES and high PGS, ever smoke is: 0.490 + (– 0.017 * -2) + (0.075 * 2) + 0.184= 0.858. So the probability that men ever smoke are respectively 49.0% for low risk and 85.8% for high

risk. Keeping all variables fixed and only changing the PGS variable in the same regression gives the following probability to ever smoke for men: (0.075*2)/(0.490+0.184)=0.223. So the probability to ever smoke for men increases with 22.3% for a high genetic risk.

Appendix 5.

Results OLS regression for alcohol use

Variable	Drinks per week
cSES	.105 (.182)
$PGS_{alcohol\_dependence}$	.421** (.185)
Interaction	.015 (.229)
Gender	3.526*** (.314)
Constant	5.136*** (.169)
Observations	3514

Note. Standard errors are in parentheses; * p<0.01, ** p<0.05 and *** p<0.10.