Erasmus
School of
Economics

Thesis

What is the relation between the genetic risk of BMI and subjective life expectations?

Roel Vogel 478776

Date: 24-07-2020

Word Count: 6399

Erasmus University Rotterdam
Make it happen.

I introduction

One of the biggest health problems that the world is facing today is the problem of overweight individuals. What makes it more concerning is that this problem is still growing. In 2017, obesity was for 4.7 million individuals the cause of a premature death (Ritchie, 2017). When comparing the numbers over time, it is clear that overweight individuals is a growing problem. In 2017 8% of all deaths were caused by obesity. This is a 4.5% increase compared to the numbers of 1990. Besides that, in the United States 36% of the adults were obese in 2016 (Ritchie, 2017). The percentage of adults with obesity is lower, namely 13%. When focussing on overweight instead of obesity the numbers are even more shocking. Globally, 39% of the adults are overweight. Therefore overweight individuals are a modern world problem.

The main problem with overweight individuals is that they are more vulnerable for diseases. Having a high Body Mass Index (BMI), the measurement to determine being overweight, increases the risk of getting a heart disease, type 2 diabetes, high blood pressure, musculoskeletal disorders, gallstones, certain cancers and breathing problems (NHS website, 2018). This makes overweight individuals a huge problem, because 39% of the adults worldwide is overweight. Therefore this problem is a very interesting subject to research.

Relation between BMI and objective life expectations

There already has been a lot of research about this subject. Most of the researches cover the relation between BMI and health problems or even death. There is plenty of existing literature that have found a negative relation between BMI and the longevity. A summary of some papers about the relation between BMI and longevity are summarized below. The papers are from different time periods, different populations and used different methodologies. This shows that the findings are show a similar association between BMI and longevity, despite these differences.

Baker, Olsen and Sørensen (2007) found in a follow-up study among 276,835 Danish schoolchildren that the risk of any coronary heart disease (CHD) event, fatal or nonfatal, among adults was positively correlated with BMI at seven to thirteen years of age. This means that having a higher BMI during childhood is associated with an increased risk of CHD in adulthood. Having a higher-than-average BMI at the age of seven increases the percentage of individuals that will have a CHD by 60 years old by 1.2 percent for boys and 0.2 percent for girls. At the age of thirteen this percentage increase is 3.8 for boys and 1.2 for girls.

Månsson et al. (1996) found in a study with 5,926 respondents, male residents in Malmö, that having a high BMI increases the chance of dying from diseases of the circulatory system and

from endocrine, nutritional and metabolic diseases. The study shows that the relative risk of overweight subject was 1.3 and for the obese 2.8. These numbers are relative to the group of individuals with a normal weight. This shows an association between overweight individuals and an increased mortality from multiple diseases.

Olshansky et al. (2005) try to estimate the future longevity. The last two centuries the expected longevity showed a steady rise, so individuals are will die older, but Olshansky et al. (2005) believe that this rise could come to an end soon. One of the main reasons that this rise could come to an end, is the increasing problem of overweight individuals. The increasing amount of overweight individuals and the health risk that come with it, can lead to a decreased longevity for individuals in the future. This paper states that individuals will live longer in the absence of obesity. The paper used four different groups to test this, white males and females and black males and females. In the different groups, the effects were similar. The results in every group show negative associations between BMI and longevity.

Also Van Dam, Willett, Manson and Hu (2006) found in a study among 102,400 women that there is a negative relation between having a high BMI and the age of death among women. The reason for this effect is the higher vulnerability for disease when having a high BMI. They found that the hazard ratio for premature death was 1.18 for women with a slightly higher than normal BMI, 1.66 for women with a higher-than-average BMI and 2.79 for obese women. These ratios are all for women aged 18 and are relative to the average BMI for women at the age of 18 years old. This shows an association between BMI and premature death.

After reviewing the papers above, it is clear that BMI is negatively correlated to longevity. This means that overweight individuals have a higher mortality rate. This research will investigate the relevance and influence of the genetic risk of BMI.

Relevance of the genetic risk of BMI

One of the aspects that may cause overweight is the genetic risk of BMI. This is the aspect that this research is focussing on. Locke et al. state the 21% of the BMI variation between individuals can be accounted for by the differences between the genetics of individuals. Therefore it is possible to say that the genetic risk of BMI has a large influence on BMI.

Linnér and Koellinger (2020) state that genetic tests for common diseases are getting more precise. Recent studies show that polygenic scores, the cumulative measures of genetic liability, are similar in precision to established risk factors like the risk factors of cardiovascular disease (Abraham et al. 2019; Khera et al. 2019; Torkamani, Wineinger, and Topol 2018).

These scores have been implemented in the routine of genetic health reports that are offered by companies in the market of consumer genetics. This market has several millions of customers worldwide according to Khan and Mittelman (2018). According to Linnér and Koellinger (2020) these developments have made genetic testing affordable, what leads to the possibility for individuals to purchase estimates of their genetic liability at a young age, years before any symptoms of disease show up (Khera et al. 2019).

For medical use, the genetic risk of BMI could especially be helpful. Normally, doctors have to test their patients for their individual information. This costs a lot of time and therefore also money. If the genetics of an individual give enough the same information as the manual tests, the genetic tests can replace the manual ones. This will spare time and money, because instead of a manual test by a doctor, a tube of blood will be enough. This way, gathering individuals' information will be more efficient. Besides that, when individuals have to measure their own BMI, it is possible that individuals don't measure the right way, accidently or intentionally. In contrast to these measurement errors, the genetic risk of BMI is an objective measure. Individuals can't, accidently or intentionally, change their genetics. This leads to better measurement that are used in medical reports. The fact that the genetic risk is objective is also helpful for insurance companies. When these companies know these score they can take the risks that come with these scores into account.

The new availability of genetic risk scores may influence the expectations of individuals' health and longevity. Therefore the revelation of genetic risk scores could influence their economic behaviour (Hamermesh, 1985). Linnér and Koellinger (2020) state that when individuals have private knowledge of their polygenic scores that are not reflected in the price of insurance, the insurance can be considered as cheap or, when the other way around, expensive. Thus, when choosing to purchase some insurance this may lead to asymmetric information between the individual and the insurance company. It is in the interest for the individual to withhold or reveal their knowledge about their genetic risk when purchasing an insurance. This way the individual could influence the price of the insurance. This brings concerns to the insurance providers, because they have to deal with this asymmetric information and the adverse selection that comes with it. On the other hand, if the insurers would have insight in the polygenic scores of individuals, there would be a danger of genetic price discrimination. These developments could potentially be a problem for the affordability of insurance markets in the future (Strohmenger and Wambach, 2000; Hendren, 2013).

Linnér and Koellinger (2020) state that there is a small negative association between the genetic liability and mortality risks. This result is found for 27 common diseases. Also a

negative association is found between genetic liability and long term care insurance. These are two of the results of their study of 9,272 American participants of the Health and Retirement Study (HRS). The reason they gave for this is that individuals who observed a decline in their health status have chosen not to purchase an insurance, because they don't expect to need an long-term insurance anymore. These individuals' expectations are influenced by their declining health. For life insurances this would be the opposite. A life insurance is most of the time purchased to take care of the relatives of the individual that is dead. So it is possible that if individuals know that they will die soon, they will purchase a life insurance. To see if this is association exist, this research will also test if the subjective life expectations influence the purchase of insurance.

Mirowsky (1999) compared the actual life expectancy and the subjective life expectancy of 2,037 Americans. He found in this study that the subjective life expectancy is very similar to the actual life expectancy. This means that Americans know well where they are in their life and how long they probably have to live. The mean of the whole sample points at this conclusion, but this is not the case if the sample is separated in gender groups. Women estimate their life expectancy lower than the actual life expectancy. The opposite is true for men, they think they will live longer than the actual estimations. The mean of the whole sample is most important for insurance companies. They can take into account that the subjective life expectations doesn't differ much from the actual longevity. This is something the insurance companies can take into account. As mentioned before, some individuals don't purchase insurance if they think they won't outlive the insurance deal. Because of this, individuals with a low life expectancy will be more likely to not purchase long-term insurances.

So for the insurance market it is vital to know if there is a relation between the polygenic scores and subjective life expectations. When the customers know and the insurance providers don't know about the polygenic scores, there is a big chance that there is going to be adverse selection in the market. Mostly the individuals that believe they will die before the end of the life insurance deal will purchase it. This could increase the costs for insurance companies. To cover this, insurance companies will increase prices causing problems for the affordability of insurances. Because of this, the problem isn't only for the insurance companies, but also for the government. They have to make sure that their inhabitants are still able to afford their insurances. This makes the problem that initiated in the insurance market is important for the whole society. That is why it is very important to find out if there is a relation between the genetic risk of BMI and the subjective life expectations.

Research question

As mentioned above, most of the existing literature researched the relation between BMI and longevity. It is shown that there is a positive relation between the genetic risk of BMI an BMI itself, but the relation between the genetic risk of BMI and the subjective life expectations is not researched yet. This relation could be important for insurance companies, to have a better and more efficient way of gathering information about the individuals that purchase life insurances. By this the insurance companies can prevent losses and price rises of life insurances. This research is covering this relation and therefore this research will bring something new to the existing literature. As mentioned above, this subject is very interesting and important for insurance companies and governments. The goal of this research is to bring some new information about the importance of the genetic risk of BMI. The research question that needs to be answered to reach the goal is as follows:

Is there a relation between the genetic risk of BMI and subjective life expectations?

As mentioned before it is known that there is a negative relation between BMI and the age of death. Besides that, it is known that there is a positive relation between the genetic risk of BMI and BMI itself. It is also shown that there is a association between the subjective life expectations and the age of death. Because all of this is known, this research expects that there is a negative relation between the genetic risk of BMI and the subjective life expectations.

To formulate an answer to the research question it is, first of all, tested if the BMI and the genetic risk of BMI have a different influence on the outcome variables. This is necessary to make sure that it is useful to investigate the influence of the genetic risk of BMI instead of the BMI itself. This is tested by running all the regressions of this research twice, one time with BMI and the control variables as independent variables and one time with one extra variable, namely the genetic risk of BMI. The R^2 is checked to see if the genetic risk of BMI does increase the predictive power of the regressions.

To reach the goal of this research, multiple hypothesis will be tested. First of all the relation between the genetic risk of BMI and the self-reported probability of reaching a the age of 75 and 85 and living another ten years is tested. The existing data covers a total of thirteen different waves. Because the sample group includes different individuals in different waves, the regressions are run for every wave to see what the influence of the genetic risk of BMI is on the subjective probability of reaching a certain age. The specific hypothesis that is tested is *'there is a relation between the genetic risk of BMI and the subjective life expectations'.*

After this, it is also tested if the polygenic score of BMI is associated with the purchase of a life insurance. This is tested by running a regression for every wave, with having a life insurance as dependent variable and the polygenic score of BMI and control variables as independent variables. This regression will be a logit regression, because the dependent variable is a dummy variable.

II Methodology

Data

The dataset that is used for this research consist of two different datasets. The first dataset that is used is the RAND HRS Longitudinal File. This dataset is conducted by the Institute for Social Research at the University of Michigan. The main sponsor of this dataset is the National Institute of Aging (NIA), with additional funding of the Social Security Administration (SSA) and administered by the Institute of Social Research (ISR). The dataset consist of seven different cohorts. Namely, the Initial HRS cohort (born 1931 to 1941), AHEAD cohort (born before 1924), Children of Depression cohort (born 1924 to 1930), War Baby cohort (born 1942-1947), Early Baby Boomer cohort (born 1948 to 1953), Mid Baby Boomer cohort (born 1954 to 1959) and Late Baby Boomer cohort (born 1960 to 1965). The data covers thirteen different waves, covering the period from 1992 to 2016. Not all cohort were first interviewed in the first wave. The first interview for every cohort differs, just like the data, from 1992 to 2016. Besides that, not all the respondents have answered the exact same questions in different waves. Eventually the dataset consist of a total of 42,052 respondents. The goal of the HRS survey is to provide panel data for researches and analysis, for example, on retirement. The survey elicits information about income, demographics, health, assets, cognition, family structure and connections, housing, health care utilization and costs, job status and history, insurance and expectations.

The RAND HRS Longitudinal File is combined with a second dataset, namely the HRS Polygenic Scores (PGS). The dataset consist of polygenic scores for a variety of phenotypes for respondents of the first dataset. This is not the case for all the HRS respondents, only for those that have provided salivary DNA between 2006 and 2012. Because of this, the second dataset decreases the sample size to 10,326, with all the respondents being of European ancestry. For the construction of the PGS the researchers investigated the impact of four key decisions in the building of PGS's from published genome-wide association (GWAS) meta-analysis results: whether to use single nucleotide polymorphisms (SNP's) assessed by imputation, criteria for selecting which SNP's to include in the score, whether to account for linkage disequilibrium (LD) and if account for LD, which type of method best captures the correlation structure among SNP's. The results of these analysis made the researchers decide to provide scores that include all available SNP's in the PGS that overlap between the GWAS meta-analysis and the HRS genetic data. The formula that is used to calculate the PGS's is as follows:

Variables

The first variable that is used in this research is the polygenic score of BMI. This variable shows the genetic risk of BMI for the respondents of European ancestry. With measuring the polygenic scores for respondents it is possible to measure the association between genetic risks and individuals' lives. It is only necessary to measure this score once for every respondent, because this result won't change over time. Polygenic score of BMI has a total of 10,326 observations. The variable is standardized within ethnicity to a standard normal curve with a mean of 0 and a standard deviation of 1.

The second variable that is used is the self-reported probability of living after the age of 75. This is a variable that can change over time. Besides that, the different waves don't include the exact same respondents or the number of respondents. Test check if this leads to differences between waves, the confidence intervals of the betas of this variable are compared. When these intervals overlap there is no significant difference between the waves. The third variable is the self-reported probability of living after the age of 85. This variable is similar to the previous variable for the first four waves. From wave five to wave thirteen this variable changes into the self-reported probability of living another ten years. The self-reported probability of living after the age of 75 and 85 only have respondents that are below these ages. So there is no data of individuals above 75 who predict their chance of living after the age of 75. This is the same for the self-reported probability of living after the age of 85. The fourth variable covers the question if the respondents have a life insurance. This variable is also changeable over time and doesn't have the same respondents for every wave. This variable also is a dummy variable. This means that the variable only can be 1 or 0, in this case having a life insurance or not.

Besides the variables named above, this research makes use of some control variables. The variables that are controlled for are age, gender, household income and principal components. The principal components control for confounding from population stratification and account for any ancestry differences in genetic structures within populations that could bias estimates. Gender is a dummy variable with a value of zero if the individual is a female and the value is one if the individual is a male. The household income is implemented to control for the socioeconomic status (SES) of the individuals.
All the variables that are named above and their statistics can be found in table 1 in the appendix.

Methodology

As mentioned before, the HRS data set provided data of 42,052 respondents. In combination with the PGS data this number of respondents decreased to 10,326. After combining the datasets the data was made suitable for the regressions needed in this research. Not all respondents filled in their gender every time they took the interview and because the data doesn't provide prove of respondents that changed gender, this variable is made independent of the waves.

The other variables age and the subjective life expectations change over time, so they can be different across the different waves. Because of this, it is not possible to generalize these variables. Besides that, not every respondent have data for these variables in every wave they took the interview. Because of this the sample size variates according to the number of observations of these variables between 1,916 and 9,236.

Before it was possible to run the regressions, it was necessary to test the if the genetic risk of BMI has an additional influence on top of the influence of BMI itself. To test this, the regression with the subjective life expectations as dependent variable are run twice for every wave. The first time with the individuals BMI as independent variable and the second time with the BMI and the polygenic score of BMI as independent variables. This is necessary to see if the regression with the polygenic score of BMI increases in predictive power by comparing the R^2.

To formulate an answer to the research question, some regressions are needed. First of all the relation between the subjective life expectations and the polygenic score of BMI has to be tested. For these regressions Ordinary Least Squares (OLS) can be used. These regressions will follow the following form:

$$Y = \beta 0 + \beta 1\, X1 + \beta 2\, X2 + \beta 3\, X3 + \beta 4\, X4 + \beta 5\, X5 + \varepsilon$$

In this regression the outcome variable Y will be the subjective life expectations, so self-reported probability of living after the age of 75 and 85 and living another ten years. In these regression the X1 stands for the polygenic score of BMI. Further, the X2 and the X3 represent the control variables age and gender. The control variable household income is shown by X4. At last the X5 stands for all the principal components in the regression. The ε represents the error term. The regression for the subjective probability of living after the age of 75 is repeated for all the thirteen different waves, to see if there are significant differences across the waves. The data for the subjective probability of living after the age of 85 only consists for wave one to four. The other waves have data for the subjective probability of living another ten years. These two regressions will therefore only cover the waves where the data consists.

After the results of the previous regressions are known it is possible to test if there is a relation between the polygenic score of BMI and having a life insurance. The results of this could show the importance of the genetic risk of BMI for insurance companies. Because the outcome variable of this regression in binary, OLS can't be used. Therefore this regression will be a logistic regression. This regression will follow the following form:

$$ln\frac{\mathrm{p(ins.)}}{1-\mathrm{p(ins.)}} = \beta 0 + \beta 1\, X1 + \beta 2\, X2 + \beta 3\, X3 + \beta 4\, X4 + \beta 5\, X5 + \varepsilon$$

In this regression p(ins.) stands for the individuals' odds of having a life insurance. All the other factors stand for the same variables as in the previous regression. So X1 is the polygenic score, X2 stands for the age, X3 is the gender, X4 stands for household income and X5 are the principal components.

III Results

The first regressions of this research point out that investigating the influence of the polygenic score of BMI on subjective life expectations is useful. When comparing the R^2 of the regression with BMI as independent variable and BMI and the polygenic score of BMI, it is obvious that the R^2 is higher in the majority of the regressions with the combination of BMI and the polygenic score of BMI as independent variables. These results are shown in table 2 in the appendix. With these results it is possible to say that the polygenic score of BMI provides extra predictive power to the regressions over the BMI score itself. This means that the polygenic score of BMI has an influence on subjective life expectations on his own.

Subjective life expectations

The first regression that is run has the self-reported probability of living after a certain age as the outcome variable. As explained in the methodology, this regression is run thirteen times, one time for every wave to see if there are significant differences between the waves. This regressions have a number of observations (N) that varies between 1,916 and 4,241. In table 1 the results of these regressions are shown.

Table 1. The beta of the polygenic score of BMI on the subjective probability of living after the age of 75. Also the standard error, R^2 and number of observations (N) of the regressions.

	Statistics				
Wave	β(95% confidence interval)	Standard error	R-squared	N	Control variables
1	-0.75(-1.60;0.10)*	0.44	0.017	4,140	Yes
2	-0.79(-1.65;0.08)*	0.44	0.008	3,817	Yes
3	-1.02(-1.95;-0.10)**	0.47	0.009	3,687	Yes
4	-1.12(-1.95;-0.29)***	0.42	0.018	4,241	Yes
5	-0.81(-1.64;0.01)*	0.42	0.017	3,964	Yes
6	-1.01(-1.91;-0.11)**	0.46	0.017	3,473	Yes
7	-1.88(-2.71;-1.05)***	0.42	0.028	4,199	Yes
8	-1.97(-2.93;-1.00)***	0.49	0.028	3,316	Yes
9	-1.89(-2.90;-0.88)***	0.52	0.041	2,768	Yes
10	-1.62(-2.51;-0.73)***	0.45	0.027	3,608	Yes
11	-1.75(-2.72;-0.78)***	0.50	0.026	3,140	Yes
12	-1.59(-2.63;-0.54)***	0.53	0.025	2,550	Yes

13	-2.11(-3.35;-0.88)***	0.63	0.023	1,916	Yes

**** p<0.01, ** p<0.05, * p<0.1*

All the waves show a significant influence of the polygenic score of BMI on the subjective life expectations. This significance is for most of the waves with a p<0.01. All the other betas are significant with p<0.05 or p<0.1. The table also shows an increase in almost all the statistic as the wave number gets higher. The influence of the polygenic score becomes stronger, the R^2 increases and also the significance seems to increase along the waves. This is probably caused by the fact that the age of the respondents is higher at the thirteenth wave than at the first wave. The table also shows that all the betas are negative. This means that the polygenic score of BMI has a negative influence on the subjective life expectations. The confidence intervals overlap for all the betas, what means that there are no significant differences between the different waves.

The next thing that is noticeable is that age has a slightly positive influence. All the significant age betas are between 0.14 and 0.42. This outcome is very predictable, because it is obvious that the older you get the higher the chance is of reaching a certain age. For example, the self-reported probability of living after the age of 75, is higher for someone age 74 than for someone aged 50. It is logical that the closer you get to a certain age, you give yourself a high chance of passing that age. Because this regression doesn't include respondents that already passed the age of 75, these regression aren't biased by these respondents. The next outcome of this regression covers the gender of the respondent. The regression shows that there is a positive association between being a male and the self-reported probability of living after the age of 75. The gender beta differs between 2.47 and 6.17. These betas are all significant with p<0.01. Also the betas of the control variable household income are all positive with p<0.01, meaning that individuals with a higher income also have a higher subjective life expectation. The results of these control variables can be found in table 4 in the appendix.

The following regressions will test the relation between the genetic risk of BMI and the subjective probability of living after the age of 85. This regression is run for the first four waves. The N of these regressions varies between a minimum of 3,973 and a maximum of 4,207. Of these four waves the genetic risk only has a significant beta with p<0.01 in wave 4, namely a negative 1.71. These results can be found in table 2 below. Same as the age betas in the previous regressions, the significant betas in these regressions are both positive. One of these betas is significant with p<0.01 and the other with p<0.1. Besides that, the gender betas again all significant with p<0.01 and are all positive. The only difference is that the betas are bigger

in this regressions than in the previous regressions. They differ between 5.17 and 8.81. The results of the control variables can be found in table 5 in the appendix.

Table 2. The beta of the polygenic score of BMI on the subjective probability of living after the age of 85 (wave 1 to 4) and living another ten years (wave 5 to 13). Also the standard error, R^2 and number of observations (N) of the regressions.

	Statistics				
Wave	β(95% confidence interval)	Standard error	R-squared	N	Control variables
1	-0.74(-1.72;0.23)	0.50	0.023	4,134	Yes
2	-0.47(-1.44;0.49)	0.49	0.014	3,973	Yes
3	-0.46(-1.48;0.57)	0.52	0.018	3,975	Yes
4	-1.71(-2.65;-0.77)***	0.48	0.031	4,207	Yes
5	-0.91(-1.66;-0.16)**	0.38	0.016	6,490	Yes
6	-0.43(-1.18;0.32)	0.38	0.022	6,607	Yes
7	-1.09(-1.79;-0.39)***	0.36	0.026	8,018	Yes
8	-0.85(-1.55;-0.14)**	0.36	0.015	8,175	Yes
9	-1.20(-1.90;-0.50)***	0.36	0.024	7,919	Yes
10	-1.19(-1.85;-0.52)***	0.34	0.027	8,704	Yes
11	-0.94(-1.62;-0.27)***	0.35	0.027	8,220	Yes
12	-0.98(-1.69;-0.27)***	0.36	0.032	7,351	Yes
13	-0.90(-1.66;-0.13)**	0.39	0.036	6,337	Yes

**** p<0.01, ** p<0.05, * p<0.1*

After these four waves the regression tests the relation between the genetic risk of BMI and the subjective probability of living another ten years. N has a minimum of 6,490 and a maximum of 8,704 for these waves. Out of the nine regressions with the subjective probability of living another ten years, only wave six doesn't have a significant beta. The other significant betas vary between -0.86 and -1.28. Most of the significant betas have a $p<0.01$ and the other have a $p<0.05$. All the different betas of the genetic risk of BMI can be found in table 2. In contrast to the regressions covering the subjective probability of living after the age of 75 or 85, the age betas in these regression are all negative varying between -0.44 and -0.10. This is logical, because the older someone gets, the lower the self-reported probability of living another ten years. For example, someone age 50 will give themself a higher probability of living another ten years than someone aged 90. Besides that, the

regressions again show that there is a positive association between being a male and the subjective probability of living another ten years. The gender betas are all significant with $p<0.01$ and these betas vary between 4.57 and 6.72. The results of the control variables can be found in table 5 in the appendix.

Now the relation between the genetic risk of BMI and subjective life expectancies is known, it is possible to test if this has an effect on the fact of individuals have a life insurance or not. The logit regression of the polygenic score of BMI on having a life insurance has a N that varies between 4,269 and 9,236. Only the beta of wave six is significant with $p<0.1$. This beta is 0.05. The results of the control variables can be found in table 7 in the appendix. The age betas is significant with $p<0.01$ for every wave except wave one. All the betas are negative. The betas for the variable gender are all significant with $p<0.01$. The betas are, similar to age, all negative.

Table 3. The beta of the polygenic score of BMI on having an insurance. Also the standard error, R^2 and number of observations (N) of the logit regressions.

	Statistics				
Wave	β(95% confidence interval)	Standard error	R-squared	N	Control variables
1	-0.02(-0.10;0.06)	0.04	0.043	4,269	Yes
2	0.00(-0.07;0.07)	0.00	0.035	5,074	Yes
3	0.01(-0.06;0.08)	0.04	0.052	5,087	Yes
4	-0.01(-0.07;0.05)	0.03	0.047	6,953	Yes
5	0.02(-0.04;0.08)	0.03	0.039	6,997	Yes
6	0.05(-0.01;0.11)*	0.03	0.037	7,117	Yes
7	0.03(-0.02;0.08)	0.03	0.035	8,493	Yes
8	-0.00(-0.05;0.05)	0.03	0.032	8,585	Yes
9	0.00(-0.05;0.05)	0.03	0.033	8,430	Yes
10	0.03(-0.01;0.08)	0.02	0.027	9,236	Yes
11	0.03(-0.02;0.07)	0.02	0.028	8,796	Yes
12	0.03(-0.02;0.07)	0.02	0.026	7,937	Yes
13	0.03(-0.03;0.08)	0.03	0.023	6,918	Yes

**** p<0.01, ** p<0.05, * p<0.1*

IV Conclusion and discussion

Conclusion

The goal of this research was to find an answer to the research question *´is there a relation between the genetic risk of BMI and subjective life expectations´.* To answer this question the hypothesis *´there is a relation between the genetic risk of BMI and subjective life expectations´.* This hypothesis is tested for two different ages and the outcomes on both ages were similar. The results clearly show that the genetic risk of BMI has a negative influence on the self-reported probability of living after a certain age. The beta of the genetic risk of BMI variated between -0.75 ($p<0.1$) and -2.11 ($p<0.01$) for the self-reported probability of living after the age of 75. This means that with every increase of one standard deviation in the genetic risk score of BMI will decrease the self-reported probability of living after the age of 75 with an percentage between 0.75 and 2.11. The only significant beta for the self-reported probability of living after the age of 85 shows a similar decrease, namely 1.71 ($p<0.01$). The betas genetic risk of BMI for the self-reported probability of living another ten years are also all negative between -0.85 ($p<0.05$) and -1.20 ($p<0.01$). When comparing these effects to other papers, the effects are very similar, if not bigger, than outcomes of other researches. This means that the significant results of this research are worth mentioning. Because of this, it is possible to state that this hypothesis is not rejected, there is a relation between the genetic risk of BMI and subjective life expectations and this relation is negative.

With this hypothesis tested, it is possible to formulate an answer to the research question. The results of this research show that there is a significant negative relation between the genetic risk of BMI and subjective life expectations. As mentioned before this information could be very important for the insurance market. To test if the result of this research influence the life insurance market directly, the relation between the genetic risk of BMI and having a life insurance is also tested. Only one of the betas of the genetic risk of BMI is significant in these logit regressions. This beta is 0.05 ($p<0.1$), what means that the chance of having a insurance increases by 5% for every on standard deviation increase in the genetic risk of BMI. This results is similar to expectations, individuals with a high genetic risk of BMI estimates their subjective life expectations lower and therefore are more likely to get a life insurance.

Discussion

The results of this research were in line with the expectations beforehand. It was expected that there would be a negative relation between the genetic risk of BMI and subjective life expectations. This was the case for the hypothesis covering the subjective life expectations. With this result it is possible to say that the genetic risk of BMI is negatively related to the self-reported probability of living after a certain age. This means that a higher genetic risk of

BMI will lead to individuals giving themselves a lower probability of reaching a certain age. Within these results age has the expected effect. The older someone gets, the better someone can predict if they are going to live after the age of 75 or 85. Besides that, the research shows that the older the sample group gets, the higher the effect of the genetic risk of BMI. This may be caused by the fact that the older an individual is, the bigger the chance that their genetics worked out.

The gender variable brings remarkable results. The subjective life expectations vary between males and females. Being a male increases the subjective life probabilities with a percentage between 2.47 and 8.81. Mirowsky (1999) gives two factors that could account for these differences. The first one is that the mortality rates for men are higher than for women. Because of this, the mortality rates for men have more room for improvement. This may lead to the fact that men expect age-specific mortality rates more than women. if so, then the effect of age on the difference between subjective and actuarial life expectations can be greater for men than for women. Secondly, the current health and socioeconomic status can lead to differences between men and women. Despite the fact that women live longer, men report to feel healthier and have fewer symptoms (Ross and Bird, 1994). Reason for this can be overconfidence of men. Besides that, men have higher household income, average levels of education, personal earnings and occupational status, and lest frequent economic hardship than women. This also may lead to optimism among men about their life expectations (Ross and Bird, 1994). The second possible explanations is partly controlled for in this research. The regressions are controlled for the socioeconomic status (SES) of the individuals by implementing individuals´ income. So, the SES is not the reason for the gap between men and women in this research. On the other hand, the data didn't provide an objective health status of the respondents. Because of this, it was not possible to control for this factor and exclude this as an explanation for the differences between men and women.

With the knowledge that there is a negative relation between the genetic risk of BMI and subjective life expectations, it is possible to test if there is a relation between the genetic risk of BMI and having a life insurance. The results for this test are not as clear as the results of the previous tests. For only one wave the genetic risk of BMI showed a significant beta. The beta represent a 5% increased chance of having a life insurance for every increase of one standard deviation of genetic risk of BMI. This beta was in line with the expectations. That there is just one significant outcome makes it hard to make a clear statement about the relation between the genetic risk of BMI and having a life insurance. This research found that there may be a positive relation between the genetic risk of BMI and having a life insurance.

For insurance companies it is good to know that individuals with a high genetic risk of BMI expect to live shorter. They expect to die earlier, so they will not take the risk of buying an insurance for a couple of years and then die before that period is over. This means that only individuals that think they will live the full period of the insurance will buy one. Besides that, previous research points out that these subjective life expectations are similar to the actual life expectations. Because of this, the costs for the insurance companies will increase, because more individuals will make full use of the insurance. This may lead to an increase in the prices of insurances. Insurance companies can make sure with new policies, that they are able to use the genetic information of their customers. This will spare the companies time and money. Besides that, the information about their customers will be more objective in the future. This can prevent a rise in insurance prices. The results of this research are also interesting for the government. They have to make sure that insurances remain payable for their inhabitants. Because of this, they will also benefit from policies that allow insurance companies to use the genetics of their customers.

The dataset that is used for this research made use of random assignment, what is positive for the internal validity. On the other hand the dataset gives some external validity problems. The data only provides respondents that are American an aged 50 or older. Because of this the results of this paper only apply to individuals from the United State that are age 50 or older. This makes the external validity not so strong.

The research also has a few limitations. Firstly, as mentioned above, the external validity is not so strong. This can be fixed in further research, by researching a sample with individuals from other parts of the world and individuals aged younger than 50. Secondly, the relation between the genetic risk of BMI and having a life insurance was not as clear as expected. Further research can focus on this relation and the relation between the genetic risk of BMI and other kinds of insurances. Thirdly, this research only focuses on one genetic risk. Further research could focus on the effect of different kinds of genetic risk on subjective life expectations.

V References and Appendix

References

Abraham, G., Malik, R., Yonova-Doing, E., Salim, A., Wang, T., Danesh, J., … Dichgans, M. (2019). Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nature Communications*, *10*(1), 1. https://doi.org/10.1038/s41467-019-13848-1

Baker, J. L., Olsen, L. W., & Sørensen, T. I. A. (2007). Childhood Body-Mass Index and the Risk of Coronary Heart Disease in Adulthood. *New England Journal of Medicine*, *357*(23), 2329–2337. https://doi.org/10.1056/nejmoa072515

Hamermesh, D. S. (1985). Expectations, Life Expectancy, and Economic Behavior. *The Quarterly Journal of Economics*, *100*(2), 389. https://doi.org/10.2307/1885388

Hendren, Nathaniel. 2013. "Private Information and Insurance Rejections." Econometrica 81 (5): 1713–62. https://doi.org/10.3982/ECTA10931.

Khan, R., & Mittelman, D. (2018). Consumer genomics will change your life, whether you get tested or not. *Genome Biology*, *19*(1), 120. https://doi.org/10.1186/s13059-018-1506-1

Khera, A. V., Chaffin, M., Wade, K. H., Zahid, S., Brancale, J., Xia, R., … Kathiresan, S. (2019). Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell*, *177*(3), 587-596.e9. https://doi.org/10.1016/j.cell.2019.03.028

Linner, R. K., & Koellinger, P. D. (2020). Genetic risk scores of disease and mortality capture differences in longevity, economic behavior, and insurance outcomes. *medRxiv*.

MÅNSSON, N.-O., ERIKSSON, K.-F., ISRAELSSON, B., RANSTAM, J., MELANDER, A., & RÅSTAM, L. (1996). Body Mass Index and Disability Pension in Middle-Aged Men—Non-Linear Relations. *International Journal of Epidemiology*, *25*(1), 80–85. https://doi.org/10.1093/ije/25.1.80

Mirowsky, J. (1999). Subjective life expectancy in the US: correspondence to actuarial estimates by age, sex and race. *Social Science & Medicine*, *49*(7), 967–979. https://doi.org/10.1016/s0277-9536(99)00193-8

NHS website. (2018, 3 oktober). Being overweight, not just obese, still carries serious health

risks. Geraadpleegd op 20 juli 2020, van https://www.nhs.uk/news/obesity/being-overweight-not-just-obese-still-carries-serious-health-risks/

Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., Carnes, B. A., Brody, J., … Ludwig, D. S. (2005). A Potential Decline in Life Expectancy in the United States in the 21st Century. *New England Journal of Medicine*, *352*(11), 1138–1145. https://doi.org/10.1056/nejmsr043743

Ritchie, H. (2017, 11 augustus). *Obesity*. Geraadpleegd op 11 juni 2020, van https://ourworldindata.org/obesity#:%7E:text=13%25%20of%20adults%20in%20the,of%20energy%20intake%20and%20expenditure.

Ross, C. E., & Bird, C. E. (1994). Sex stratification and health lifestyle: consequences for men's and women's perceived health. *Journal of Health and Social Behavior*, 161-178.

Strohmenger, R., & Wambach, A. (2000). Adverse selection and categorical discrimination in the health insurance markets: the effects of genetic tests. *Journal of Health Economics*, *19*(2), 197–218. https://doi.org/10.1016/s0167-6296(99)00021-1

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, *19*(9), 581–590. https://doi.org/10.1038/s41576-018-0018-x

van Dam, R. M., Willett, W. C., Manson, J. E., & Hu, F. B. (2006). The Relationship between Overweight in Adolescence and Premature Death in Women. *Annals of Internal Medicine*, *145*(2), 91. https://doi.org/10.7326/0003-4819-145-2-200607180-00006

Appendix

Table 1. Descriptive Statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
PGS_BMI	10326	-.002	1	-3.636	3.911
gender	10326	.543	.498	0	1
age1	4309	54.504	5.639	25	75
age2	5097	58.83	7.875	27	86
age3	5107	60.782	7.912	28	88
age4	6991	61.707	8.697	30	90
age5	7036	63.599	8.769	31	92
age6	7163	65.66	8.796	33	94
age7	8534	65.129	10.108	29	96
age8	8641	67.17	10.115	31	98
age9	8496	68.907	10.087	33	98
age10	9354	67.935	11.386	30	99
age11	8870	69.224	11.155	32	100
age12	8028	70.462	10.915	34	102
age13	7011	71.816	10.502	36	103
h1icap	4309	7380.066	24474.87	0	570000
h2icap	5097	14105.5	60255.74	0	2240000
h3icap	5107	20812.89	61788.89	0	1840000
h4icap	6991	21141.66	65094.66	0	2030000
h5icap	7036	22688.93	69070.46	0	1210000
h6icap	7163	18634.22	56277.06	0	1120000
h7icap	8534	19119.03	71819.98	0	2700000
h8icap	8641	21362.7	118000	0	6520000
h9icap	8496	22678.56	80384.11	0	2010000
h10icap	9354	16873.43	56621.78	0	1310000
h11icap	8870	18837.13	81577.18	0	2840000
h12icap	8028	20200.77	101000	0	4050000
h13icap	7011	20939.83	84293.79	0	2500000
r1liv75	4140	68.307	26.424	0	100
r2liv75	3817	66.634	25.834	0	100
r3liv75	3687	69.244	27.122	0	100
r4liv75	4241	68.112	26.204	0	100
r5liv75	3964	69.273	25.183	0	100

r6liv75	3473	68.472	25.589	0	100
r7liv75	4199	67.176	26.177	0	100
r8liv75	3316	65.831	27.078	0	100
r9liv75	2768	67.504	26.232	0	100
r10liv75	3608	64.914	26.386	0	100
r11liv75	3140	63.439	26.775	0	100
r12liv75	2550	64.558	25.751	0	100
r13liv75	1916	65.039	25.994	0	100
r1liv85	4134	45.363	30.211	0	100
r2liv85	3973	42.643	29.413	0	100
r3liv85	3975	46.839	31.278	0	100
r4liv85	4207	44.257	29.87	0	100
r5liv10	6490	53.341	29.189	0	100
r6liv10	6607	51.624	29.578	0	100
r7liv10	8018	50.704	30.478	0	100
r8liv10	8175	46.415	30.58	0	100
r9liv10	7919	48.662	30.218	0	100
r10liv10	8704	44.985	30.248	0	100
r11liv10	8220	44.849	30.047	0	100
r12liv10	7351	45.569	29.919	0	100
r13liv10	6337	46.183	29.96	0	100
r1lifein	4269	.778	.415	0	1
r2lifein	5074	.788	.409	0	1
r3lifein	5087	.779	.415	0	1
r4lifein	6953	.779	.415	0	1
r5lifein	6997	.753	.431	0	1
r6lifein	7117	.732	.443	0	1
r7lifein	8493	.731	.443	0	1
r8lifein	8585	.699	.459	0	1
r9lifein	8430	.684	.465	0	1
r10lifein	9236	.659	.474	0	1
r11lifein	8796	.638	.481	0	1
r12lifein	7937	.621	.485	0	1
r13lifein	6918	.606	.489	0	1
PC1_5A	10326	0	.009	-.037	.05
PC1_5B	10326	0	.009	-.056	.016

PC1_5C	10326	0	.009	-.024	.021
PC1_5D	10326	0	.009	-.045	.034
PC1_5E	10326	0	.009	-.046	.011
PC6_10A	10326	0	.009	-.029	.035
PC6_10B	10326	0	.009	-.044	.045
PC6_10C	10326	0	.009	-.034	.037
PC6_10D	10326	0	.009	-.036	.035
PC6_10E	10326	0	.009	-.042	.03

Table 2. The difference between the R^2 when implementing BMI and BMI in combination with the polygenic score of BMI.

	Subjective probability 75		Subjective probability 85/10	
Wave	BMI R^2	BMI + PGS R^2	BMI R^2	BMI + PGS R^2
1	0.0223	0.0224	0.0268	0.0268
2	0.0102	0.0104	0.0148	0.0148
3	0.0141	0.0145	0.0200	0.0200
4	0.0245	0.0248	0.0332	0.0347
5	0.0242	0.0243	0.0192	0.0194
6	0.0269	0.0270	0.0258	0.0258
7	0.0356	0.0370	0.0283	0.0287
8	0.0370	0.0384	0.0166	0.0168
9	0.0463	0.0480	0.0235	0.0245
10	0.0337	0.0347	0.0272	0.0279
11	0.0345	0.0356	0.0268	0.0274
12	0.0347	0.0354	0.0314	0.0320
13	0.0317	0.0331	0.0374	0.0377

Table 3. The beta of the polygenic score of BMI on the subjective probability of living after the age of 75. Also the standard error, R^2 and number of observations (N) of the regressions.

	Statistics				
Wave	β(95% confidence interval)	Standard error	R-squared	N	Control variables
1	-0.75(-1.60;0.10)*	0.44	0.017	4,140	Yes
2	-0.79(-1.65;0.08)*	0.44	0.008	3,817	Yes

3	-1.02(-1.95;-0.10)**	0.47	0.009	3,687	Yes
4	-1.12(-1.95;-0.29)***	0.42	0.018	4,241	Yes
5	-0.81(-1.64;0.01)*	0.42	0.017	3,964	Yes
6	-1.01(-1.91;-0.11)**	0.46	0.017	3,473	Yes
7	-1.88(-2.71;-1.05)***	0.42	0.028	4,199	Yes
8	-1.97(-2.93;-1.00)***	0.49	0.028	3,316	Yes
9	-1.89(-2.90;-0.88)***	0.52	0.041	2,768	Yes
10	-1.62(-2.51;-0.73)***	0.45	0.027	3,608	Yes
11	-1.75(-2.72;-0.78)***	0.50	0.026	3,140	Yes
12	-1.59(-2.63;-0.54)***	0.53	0.025	2,550	Yes
13	-2.11(-3.35;-0.88)***	0.63	0.023	1,916	Yes

*** p<0.01, ** p<0.05, * p<0.1

Table 4. The betas (95% confidence intervals) of the control variables from the regressions covering the subjective probability of living after the age 75.

	Control variables		
Wave	Age	Gender	Household income
1	0.20(0.05;0.34)**	4.90(3.21;6.59)***	0.00006(0.00003;0.00010)***
2	0.17(0.00;0.33)**	2.47(0.75;4.19)***	0.00001(0.0000006;0.00002)***
3	0.18(0.01;0.35)**	2.69(0.86;4.52)***	0.00003(0.00001;0.00004)***
4	0.14(-0.01;0.28)*	4.89(3.27;6.51)***	0.00002(0.00001;0.00003)***
5	0.18(0.03;0.33)**	4.45(2.83;6.07)***	0.00002(0.00001;0.00003)***
6	0.21(0.04;0.38)**	4.57(2.79;6.34)***	0.00003(0.00002;0.00004)***
7	0.30(0.16;0.43)***	6.01(4.41;7.62)***	0.00002(0.00001;0.00003)***
8	0.28(0.10;0.46)***	5.17(3.29;7.05)***	0.00001(0.000005;0.00002)***
9	0.42(0.22;0.63)***	6.17(4.19;8.15)***	0.00003(0.00002;0.00004)***
10	0.18(0.02;0.34)**	4.70(2.97;6.43)***	0.00003(0.00001;0.00004)***
11	0.12(-0.06;0.31)	5.06(3.18;6.95)***	0.00002(0.000007;0.00003)***
12	0.20(-0.02;0.41)*	5.07(3.05;7.10)***	0.000009(0.000002;0.00002)**
13	0.26(-0.02;0.54)*	4.53(2.15;6.90)***	0.00002(0.000007;0.00003)***

*** p<0.01, ** p<0.05, * p<0.1

Table 5. The betas (95% confidence intervals) of the control variables from the regressions covering the subjective probability of living after the age 85 (wave 1 to 4) and living another ten years (wave 5 to 13).

	Control variables		
Wave	Age	Gender	Household income
1	-0.05(-0.22;0.12)	7.50(5.57;9.42)***	0.00007(0.00003;0.00010)***
2	0.15(-0.02;0.32)*	5.17(3.25;7.10)***	0.000009(-0.000004;0.00002)
3	0.33(0.16;0.51)***	5.92(3.88;7.96)***	0.00003(0.00001;0.00004)***
4	0.13(-0.03;0.29)	8.81(6.96;10.65)***	0.00002(0.000009;0.00003)***
5	-0.10(-0.18;-0.02)**	5.32(3.88;6.76)***	0.00002(0.00001;0.00003)***
6	-0.33(-0.41;-0.25)***	4.57(3.13;6.01)***	0.00003(0.00002;0.00004)***
7	-0.29(-0.35;-0.22)***	5.93(4.59;7.26)***	0.00003(0.00002;0.00004)***
8	-0.16(-0.22;-0.09)***	4.93(3.60;6.26)***	0.00001(0.000007;0.00002)***
9	-0.26(-0.32;-0.19)***	4.59(3.26;5.92)***	0.00003(0.00002;0.00004)***
10	-0.27(-0.33;-0.21)***	6.14(4.87;7.41)***	0.00003(0.00001;0.00004)***
11	-0.27(-0.33;-0.21)***	6.72(5.42;8.01)***	0.00001(0.000006;0.00002)***
12	-0.35(-0.42;-0.28)***	6.09(4.73;7.45)***	0.00001(0.000008;0.00002)**
13	-0.44(-0.52;-0.37)***	5.25(3.79;6.72)***	0.00002(0.000009;0.00003)***

**** p<0.01, ** p<0.05, * p<0.1*

Table 6. The beta of the polygenic score of BMI on the subjective probability of living after the age of 85 (wave 1 to 4) and living another ten years (wave 5 to 13). Also the standard error, R^2 and number of observations (N) of the regressions.

	Statistics				
Wave	β(95% confidence interval)	Standard error	R-squared	N	Control variables
1	-0.74(-1.72;0.23)	0.50	0.023	4,134	Yes
2	-0.47(-1.44;0.49)	0.49	0.014	3,973	Yes
3	-0.46(-1.48;0.57)	0.52	0.018	3,975	Yes
4	-1.71(-2.65;-0.77)***	0.48	0.031	4,207	Yes
5	-0.91(-1.66;-0.16)**	0.38	0.016	6,490	Yes
6	-0.43(-1.18;0.32)	0.38	0.022	6,607	Yes
7	-1.09(-1.79;-0.39)***	0.36	0.026	8,018	Yes
8	-0.85(-1.55;-0.14)**	0.36	0.015	8,175	Yes

9	-1.20(-1.90;-0.50)***	0.36	0.024	7,919	Yes
10	-1.19(-1.85;-0.52)***	0.34	0.027	8,704	Yes
11	-0.94(-1.62;-0.27)***	0.35	0.027	8,220	Yes
12	-0.98(-1.69;-0.27)***	0.36	0.032	7,351	Yes
13	-0.90(-1.66;-0.13)**	0.39	0.036	6,337	Yes

**** p<0.01, ** p<0.05, * p<0.1*

Table 7. The betas (95% confidence intervals) of the control variables from the logit regressions covering having a life insurance.

	Control variables	
Wave	Age	Gender
1	-0..00(-0.02;0.01)	-1.07(-1.25;-0.90)***
2	-0.03(-0.04;-0.02)***	-0.87(-1.02;-0.73)***
3	-0.04(-0.05;-0.03)***	-1.07(-1.22;-0.92)***
4	-0.04(-0.04;-0.03)***	-0.95(-1.08;-0.83)***
5	-0.03(-0.04;-0.03)***	-0.84(-0.95;-0.72)***
6	-0.03(-0.04;-0.03)***	-0.77(-0.89;-0.66)***
7	-0.03(-0.03;-0.02)***	-0.71(-0.81;-0.60)***
8	-0.03(-0.03;-0.02)***	-0.71(-0.80;-0.61)***
9	-0.03(-0.03;-0.02)***	-0.68(-0.78;-0.59)***
10	-0.02(-0.03;-0.02)***	-0.55(-0.64;-0.46)***
11	-0.02(-0.03;-0.02)***	-0.56(-0.65;-0.47)***
12	-0.02(-0.03;-0.02)***	-0.50(-0.59;0.40)***
13	-0.02(-0.03;-0.02)***	-0.43(-0.53;-0.33)***

**** p<0.01, ** p<0.05, * p<0.1*

Table 8. The beta of the polygenic score of BMI on having an insurance. Also the standard error, R^2 and number of observations (N) of the logit regressions.

	Statistics				
Wave	β(95% confidence interval)	Standard error	R-squared	N	Control variables
1	-0.02(-0.10;0.06)	0.04	0.043	4,269	Yes
2	0.00(-0.07;0.07)	0.00	0.035	5,074	Yes
3	0.01(-0.06;0.08)	0.04	0.052	5,087	Yes
4	-0.01(-0.07;0.05)	0.03	0.047	6,953	Yes
5	0.02(-0.04;0.08)	0.03	0.039	6,997	Yes

6	0.05(-0.01;0.11)*	0.03	0.037	7,117	Yes
7	0.03(-0.02;0.08)	0.03	0.035	8,493	Yes
8	-0.00(-0.05;0.05)	0.03	0.032	8,585	Yes
9	0.00(-0.05;0.05)	0.03	0.033	8,430	Yes
10	0.03(-0.01;0.08)	0.02	0.027	9,236	Yes
11	0.03(-0.02;0.07)	0.02	0.028	8,796	Yes
12	0.03(-0.02;0.07)	0.02	0.026	7,937	Yes
13	0.03(-0.03;0.08)	0.03	0.023	6,918	Yes

**** p<0.01, ** p<0.05, * p<0.1*

Stata commands

```
rename EA_PGS3_BMI_GIANT15 PGS_BMI
rename r1agey_m age1
rename r2agey_m age2
rename r3agey_m age3
rename r4agey_m age4
rename r5agey_m age5
rename r6agey_m age6
rename r7agey_m age7
rename r8agey_m age8
rename r9agey_m age9
rename r10agey_m age10
rename r11agey_m age11
rename r12agey_m age12
rename r13agey_m age13
rename s1gender gender1
rename s2gender gender2
rename s3gender gender3
rename s4gender gender4
rename s5gender gender5
rename s6gender gender6
rename s7gender gender7
rename s8gender gender8
rename s9gender gender9
rename s10gender gender10
rename s11gender gender11
rename s12gender gender12
rename s13gender gender13

gen gender = ((gender1==1) | (gender2==1) | (gender3==1) | (gender4==1) | (gender5==1) | (gender6==1) | (gender7==1) | (gender8==1) | (gender9==1) | (gender10==1) | (gender11==1) | (gender12==1) | (gender13==1))
```

drop if ((gender1==.u | gender1==.) & (gender2==.u | gender2==.) & (gender3==.u | gender3==.) & (gender4==.u | gender4==.) & (gender5==.u | gender5==.) & (gender6==.u | gender6==.) & (gender7==.u | gender7==.) & (gender8==.u | gender8==.) & (gender9==.u | gender9==.) & (gender10==.u | gender10==.) & (gender11==.u | gender11==.) & (gender12==.u | gender12==.) & (gender13==.u | gender13==.))

drop if PGS_BMI==.

summarize PGS_BMI gender age1 age2 age3 age4 age5 age6 age7 age8 age9 age10 age11 age12 age13 h1icap h2icap h3icap h4icap h5icap h6icap h7icap h8icap h9icap h10icap h11icap h12icap h13icap r1liv75 r2liv75 r3liv75 r4liv75 r5liv75 r6liv75 r7liv75 r8liv75 r9liv75 r10liv75 r11liv75 r12liv75 r13liv75 r1liv85 r2liv85 r3liv85 r4liv85 r5liv10 r6liv10 r7liv10 r8liv10 r9liv10 r10liv10 r11liv10 r12liv10 r13liv10 r1lifein r2lifein r3lifein r4lifein r5lifein r6lifein r7lifein r8lifein r9lifein r10lifein r11lifein r12lifein r13lifein PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r1liv75 r1bmi age1 h1icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r1liv75 r1bmi PGS_BMI age1 h1icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r2liv75 r2bmi age2 h2icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r2liv75 r2bmi PGS_BMI age2 h2icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r3liv75 r3bmi age3 h3icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r3liv75 r3bmi PGS_BMI age3 h3icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r4liv75 r4bmi age4 h4icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r4liv75 r4bmi PGS_BMI age4 h4icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r5liv75 r5bmi age5 h5icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r5liv75 r5bmi PGS_BMI age5 h5icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r6liv75 r6bmi age6 h6icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r6liv75 r6bmi PGS_BMI age6 h6icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r7liv75 r7bmi age7 h7icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r7liv75 r7bmi PGS_BMI age7 h7icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r8liv75 r8bmi age8 h8icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r8liv75 r8bmi PGS_BMI age8 h8icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r9liv75 r9bmi age9 h9icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r9liv75 r9bmi PGS_BMI age9 h9icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r10liv75 r10bmi age10 h10icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r10liv75 r10bmi PGS_BMI age10 h10icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r11liv75 r11bmi age11 h11icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r11liv75 r11bmi PGS_BMI age11 h11icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r12liv75 r12bmi age12 h12icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r12liv75 r12bmi PGS_BMI age12 h12icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r13liv75 r13bmi age13 h13icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r13liv75 r13bmi PGS_BMI age13 h13icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r1liv85 r1bmi age1 h1icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r1liv85 r1bmi PGS_BMI age1 h1icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r2liv85 r2bmi age2 h2icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r2liv85 r2bmi PGS_BMI age2 h2icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r3liv85 r3bmi age3 h3icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r3liv85 r3bmi PGS_BMI age3 h3icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r4liv85 r4bmi age4 h4icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r4liv85 r4bmi PGS_BMI age4 h4icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r5liv10 r5bmi age5 h5icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r5liv10 r5bmi PGS_BMI age5 h5icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r6liv10 r6bmi age6 h6icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r6liv10 r6bmi PGS_BMI age6 h6icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r7liv10 r7bmi age7 h7icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r7liv10 r7bmi PGS_BMI age7 h7icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r8liv10 r8bmi age8 h8icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r8liv10 r8bmi PGS_BMI age8 h8icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r9liv10 r9bmi age9 h9icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r9liv10 r9bmi PGS_BMI age9 h9icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r10liv10 r10bmi age10 h10icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r10liv10 r10bmi PGS_BMI age10 h10icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r11liv10 r11bmi age11 h11icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r11liv10 r11bmi PGS_BMI age11 h11icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r12liv10 r12bmi age12 h12icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r12liv10 r12bmi PGS_BMI age12 h12icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r13liv10 r13bmi age13 h13icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r13liv10 r13bmi PGS_BMI age13 h13icap i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r1liv75 PGS_BMI age1 i.gender h1icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r2liv75 PGS_BMI age2 i.gender h2icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r3liv75 PGS_BMI age3 i.gender h3icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r4liv75 PGS_BMI age4 i.gender h4icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r5liv75 PGS_BMI age5 i.gender h5icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r6liv75 PGS_BMI age6 i.gender h6icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r7liv75 PGS_BMI age7 i.gender h7icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r8liv75 PGS_BMI age8 i.gender h8icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r9liv75 PGS_BMI age9 i.gender h9icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r10liv75 PGS_BMI age10 i.gender h10icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r11liv75 PGS_BMI age11 i.gender h11icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r12liv75 PGS_BMI age12 i.gender h12icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r13liv75 PGS_BMI age13 i.gender h13icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

reg r1liv85 PGS_BMI age1 i.gender h1icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r2liv85 PGS_BMI age2 i.gender h2icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r3liv85 PGS_BMI age3 i.gender h3icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r4liv85 PGS_BMI age4 i.gender h4icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r5liv10 PGS_BMI age5 i.gender h5icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r6liv10 PGS_BMI age6 i.gender h6icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r7liv10 PGS_BMI age7 i.gender h7icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r8liv10 PGS_BMI age8 i.gender h8icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r9liv10 PGS_BMI age9 i.gender h9icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r10liv10 PGS_BMI age10 i.gender h10icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r11liv10 PGS_BMI age11 i.gender h11icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r12liv10 PGS_BMI age12 i.gender h12icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
reg r13liv10 PGS_BMI age13 i.gender h13icap PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

logit r1lifein PGS_BMI age1 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r2lifein PGS_BMI age2 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E

```
logit r3lifein PGS_BMI age3 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r4lifein PGS_BMI age4 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r5lifein PGS_BMI age5 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r6lifein PGS_BMI age6 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r7lifein PGS_BMI age7 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r8lifein PGS_BMI age8 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r9lifein PGS_BMI age9 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r10lifein PGS_BMI age10 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r11lifein PGS_BMI age11 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r12lifein PGS_BMI age12 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
logit r13lifein PGS_BMI age13 i.gender PC1_5A PC1_5B PC1_5C PC1_5D PC1_5E PC6_10A PC6_10B PC6_10C PC6_10D PC6_10E
```