ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Economics and Business Economics

Major Behavioural and Health Economics

# Does Subjective Wellbeing Affect Engaging Into Risky Health Behaviors? A Mendelian Randomization Study

Name student: Eibert Anderiesse

Student ID number: 561009

Supervisor: Elisa de Weerd

Second assessor: Lisa Voois

Date final version: 27-06-2023

# Abstract

Today's society is plagued by chronic diseases caused by risky health behaviors. Because of this, many interventions have been done with the aim of bringing this problem to a halt. However, chronic diseases are still the number one cause of death (CDC, 2021). Rather than concentrating on incentives of engaging into risky health behaviors, research has focused less on another factor influencing this behavior. Repeatedly, subjective wellbeing has been associated with risky health behaviors, but less is known about causality with this topic. After replicating these associations with the use of Ordinary Least Squares in this research, Mendelian Randomization – using polygenic scores as instruments – will be used with the aim of filling this gap in the literature. Using a sample of around 7,000 individuals from the US aged mostly 50 or older, associations found by previous research between subjective wellbeing and health behaviors smoking and physical activity have been replicated, finding an inverse and positive effect respectively. Associations between above average and good subjective wellbeing and alcohol consumption have been found, but none with heavy drinking. Moreover, causal inference has not been found between subjective wellbeing and the health behaviors, using Mendelian Randomization on a smaller part of the sample (1,500 individuals). An association has been found between subjective wellbeing and lower likelihood of smoking, but nothing can be said about causality due to the unobservable nature of some assumptions of Mendelian Randomization. Lastly, recommendations for further research and policy are given. The main improvement on this research would be data with more statistical power and based on this research, policy could focus on workplace wellbeing campaigns and the Community Reinforcement Approach.

# Table of Contents

# 1 Introduction

In today's industrialized society, risky health behaviors such as smoking, drinking, physically inactive lifestyles and unhealthy diets are indispensable as health determinants and more importantly, as causes of mortality. In the US, 7 out of 10 deaths are caused by chronic disease and this is in turn associated with some of the aforementioned preventable risky lifestyles (CDC, 2021).

Policy interventions for risky health behaviors have been widely researched. One of the most famous is the Pigouvian Tax, which imposes a special tax on externalities by taxing the consumer. This way, the incentive to buy unhealthy consumption drops, which proved to be effective for smoking and alcohol consumption (Levy et al., 2004; Wagenaar et al., 2010). A similar tax on unhealthy consumption has also been widely researched, but this tax carries higher losses than gains according to Mann (2008) and Tiffin and Arnoult (2011).

Another policy intervention that has a more positive effect on all healthy behavior could be improving one's life satisfaction or subjective wellbeing. Subjective wellbeing as defined by Stone and Mackie (2013) is the self-reported evaluation of people's lives and specific domains and activities of their lives. Wellbeing has long been a center of attention concerning its effects on a wide range of outcomes. It has been positively associated with many determinants of a healthy life and health overall (Okun et al., 1984). One determinant for a (mentally) healthy life is one's family life. Luhmann et al. (2013) found that higher life satisfaction is associated with a lower likelihood of marital separation. Earlier, it had been found by Wolfinger (1998) that marital separation results in greater likelihood of smoking and being a problem drinker, while remarriage offsets these effects partly. One could thus intuitively identify a path through which subjective wellbeing could be associated with engaging into risky health behaviors. More pathways of associations between subjective wellbeing and risky health behaviors have been found in the literature.

When it comes to preventing chronic disease, studies show that higher life satisfaction is associated with more physical activity (Melin et al., 2003). Strine et al. (2008) found the same result and furthermore that a decrease in subjective wellbeing also is associated with an increase in the prevalence of smoking, obesity and heavy drinking. This study does lack randomization, since their measurements were done using a telephone survey. Additionally, Kelloniemi et al. (2005) found that higher life satisfaction was positively associated with a healthier diet among Finnish adults. Similarly, Grant et al. (2009) found that college students with higher life satisfaction consume on average more fruit and avoid fat in their diet. They also found that higher life satisfaction was negatively associated with smoking. Other literature found the same inverse association between wellbeing and smoking (Carvajal, 2012). Less is known about alcohol consumption. Murphy et al. (2005) found a link between heavy alcohol consumption and low subjective wellbeing among female students, but a rather curvilinear relation to the social life satisfaction of males. The latter is found before by Paschall et al. (2005), who state that the association between alcohol intake and wellbeing is U-curved, which could obviously be due to environmental factors rather than through causal inference.

Clearly, there are many environmental pathways through which subjective wellbeing could affect risky health behaviors, and vice versa. An interesting turn towards causal inference could be towards genoeconomics. Genoeconomics study the interplay between behavior and genetics and identify specific genetic markers associated with phenotypic traits like decision-making, personalities and in addition, subjective wellbeing. This identification is done in a Genome-Wide Association Study (GWAS), which analyzes the entire genetic make-up of individuals to identify genetic variants correlated with a particular trait or disease. Okbay et al. (2016) found genetic variants associated with variance in subjective wellbeing, which could be used to exclude environmental factors since genetics are determined at conception.

This research will try to give a result more in line with the assumptions for causality of the effect of subjective wellbeing on risky health behaviors, by introducing a method that is not used before for this subject: Mendelian Randomization (MR). MR uses genetic variants as IV, since genetics are randomly distributed and decided at conception. This way, MR under certain assumptions can measure a causal effect whereas Ordinary Least Squares (OLS) only measures the associations between variables.

MR is an improvement compared to OLS. Firstly, confounding occurring with OLS is mitigated by MR since the variation in treatment is only decided through the genetic variants. This way, randomization within the population is unnecessary and environmental factors are ruled out. Secondly, reverse causation is controlled for if the genetic variant does not directly affect the outcome. An example study by Viinikainen et al. (2022) found that higher educated individuals engaged in less risky health activities, using MR. These researchers used this method to mitigate the limitations of OLS. The associations found by the literature between subjective wellbeing and risky health behaviors are subject to confounding and environmental factors play a great role. This paper will study the general research question:

*What is the causal effect of subjective wellbeing on engaging into risky health behaviors?*

Due to no earlier research using MR to estimate the effect of subjective wellbeing on risky health behaviors, this paper will provide a novel view for policymakers with the aim of preventing chronic disease caused by risky health behaviors. One possible policy intervention could be investing into social support programs. An example is the Community Reinforcement Approach, which is a program focused on improving wellbeing and resistance towards smoking and alcohol consumption (Meyers et al., 2011). This paper could provide valuable information for a program like this, by finding the true effect of subjective wellbeing on risky health behaviors. Additionally, this paper builds upon earlier research compared to previous research investigating smoking, alcohol consumption and physical activity and will improve on it by using a method that does not have the concerns of OLS.

This paper is structured as follows. First, it will provide the hypotheses and further literary review in the theoretic framework. Following up is a short explanation on the used GWAS by Okbay et al. (2016), the data, variables description and methodology of an OLS model and a MR model with its assumptions. The results of both models will then be reported and are followed by a discussion of the results and the limitations. Finally, suggestions for further research and policy interventions will be discussed.

# 2 Theoretic Framework

The prevalence of smoking, alcohol consumption and physical inactivity will be measured as risky health behaviors in this paper. This is mainly due to a lack of data on other health behaviors like one's diet or reckless physical behavior. This will be done separately, to give a clearer view for policy makers on each dimension. Therefore, three different hypotheses will be formed and tested.

*Smoking*

The first risky health behavior of interest, smoking, has been widely researched. Smoking causes heart disease, cancer, lung diseases and many more preventable causes of death (CDC, 2020). Even though there is a decline in smoking, a shocking 11.5 percent of US adults still engages into this risky health behavior (CDC, 2021b). New policy to improve subjective wellbeing could contribute to the decline in tobacco use. Lots of policy like a tax on tobacco or disturbing images on packages is already in place, but the literature has found another factor that is associated with smoking.

The inverse association between wellbeing and smoking found by Strine et al. (2008), Grant et al. (2009) and Carvajal (2012) was established earlier by Ashton and Stepney (1982). They found that neurotic individuals prone to higher stress levels (and therefore more likely to have a lower subjective wellbeing) are more likely to smoke. However, the association through which subjective wellbeing affects smoking is still subject to environmental factors causing stress.

As earlier mentioned, Strine et al. (2008) examined associations between life satisfaction and risky health behaviors. Among 340,575 individuals aged 18 or older from the US, they used a telephone survey questioning them about their life satisfaction and their health behavior. A limitation here is proper randomization since only people who wanted to participate over the phone are taken into account. Besides, life satisfaction was measured based on one simple question. They did control for sociodemographic characteristics and the association thus gives incentive for further research. Carvajal (2012) used longitudinal data, but the follow-up was only over a period of 18 months with a small sample size of 744 individuals. Additionally, this study found an association between global positive expectancies and less smoking, which is not the same as subjective wellbeing. Nevertheless, it could be that global positive expectancies are positively associated with subjective wellbeing. Grant et al. (2009) did investigate life satisfaction. They investigated 17,246 students aged 17 to 30 years, from a small number of universities. Again, randomization is a problem here. Obviously, their risky health behavior is different from the rest of the population and could be prone to social pressure. They did report a bidirectional relationship, which makes turning the research question around also relevant. For example, Xie et al. (2022) found that smokers have lower life satisfaction on average. Even though there is a lot of confounding and lacking representativeness, the literature incentivizes researchers to embark on the journey towards causality. Based on previous literature, the following hypothesis will be tested:

*$H_1$: Higher subjective wellbeing negatively affects the likelihood to smoke.*

*Alcohol Consumption*

The second behavior of interest is alcohol consumption. It has long been clear that alcohol consumption is risky to one's health. Some consequences may be alterations in learning, visuospatial processing, attention, memory and damage to the brain (Spear, 2018). Alcohol is sometimes seen as a luxury good and has shown to have a U-shaped curve when it comes to mood state (Paschall et al., 2005). For example: an individual could consume more alcohol when he goes through a rough and stressful period, but also when he has lots of joyful social events where alcohol is consumed. This U-shaped association is also found when income instead of mood is on the x-axis. This means that compared to middle incomes, alcohol consumption is higher among higher incomes and lower incomes (Popovici & French, 2013; Allen et al., 2017). Similarly, Paschall et al. (2005) found the same curvilinear association between alcohol consumption and mood state. They investigated the association between alcohol consumption and depressive mood among 13,892 young American adults, interviewed in 1995 and 2002. However, confounding proved to be a limitation which could mean that the curvilinear association is associated with environmental differences. MR could mitigate confounding in this matter.

When it comes to subjective wellbeing, Grant et al. (2009) did not find an effect while investigating the effect of life satisfaction on alcohol consumption. Zullig et al. (2001) found that regular alcohol use was associated with reduced life satisfaction. They investigated 5,032 high school students. Their results were however subject to confounding since they used OLS and only reported associations. Myrtveit et al. (2019) also found an inverse relationship when investigating 9,632 Norwegian students but also found something more interesting. Namely, friendships are crucial for wellbeing, but social integration in student communities is easier when consuming more alcohol. This environmental effect is again a clear barrier towards causal inference. It could also mean that there is no direct effect of subjective wellbeing on alcohol consumption. Moreover, some researchers did not find an association or something that looks more like a U-curved association. Nevertheless, some associations have been found and thus the second hypothesis will be tested using MR:

*$H_2$: Higher subjective wellbeing negatively affects alcohol consumption.*

*Physical Activity*

The third behavior of interest is physical activity. The World Health Organization (2022) found that physical activity has a positive effect on a high variety of determinants of a healthy life. It has lots of physical and mental health benefits like improving one's heart, reducing depression and enhancing cognitive skills. They also found that globally more than 80 percent of adults is insufficiently physically active and that this group has a 20 to 30 percent increased risk of mortality, mainly due to the negative effects on chronic diseases. Policy to improve physical activity is thus needed and subjective wellbeing could be an interesting target for governments. Social welfare programs could for example make sure that individuals have their basic needs met and thus improve their life satisfaction. If improved life satisfaction causes more physical activity, governments using social welfare programs could incentivize individuals to exercise.

Despite the aforementioned shocking numbers, some adults are more consistent in their physical activity than others. Research into this difference between adults is generally focused on the perceived benefits and costs of exercise, subject to discipline. Fewer researchers found that subjective wellbeing explains one's likelihood to engage into exercise. As mentioned earlier, higher life satisfaction has a positive effect on physical activity according to Strine et al. (2008). However, their study lacked proper randomization and clear definitions of life satisfaction. Melin et al. (2003) investigated the same topic. In their study using around 2500 Swedish individuals between 18 and 64 years old, only associations are found. Additionally, Schneider et al. (2009) found that greater happiness is associated with more moderate exercise. However, their sample size consisted of mostly white individuals and was very small at 98 individuals.

Reigal et al. (2014) found the same result among 2079 Spanish high school student. Interestingly, among these students there was a significant difference in the perception of self-efficacy which depended on the social context in which exercise is carried out. Doing exercise while co-operating in a team increases the number of social contacts and additionally, one's self-efficacy (Reigal et al., 2014). Sousa and Lyubomirsky (2001) found the same result and that a feeling of belonging and having discipline leads to higher satisfaction of life. One could see a 'positive vicious circle', where exercise in a social context leads to higher life satisfaction and according to previous research, again in more exercise. The pathway through which life satisfaction affects physical activity, and vice versa is thus unclear, which gives reason to test the third hypothesis:

> *$H_3$: Higher subjective wellbeing of individuals has a causal positive effect on their physical activity.*

# 3 Methodology

### 3.1 Data

The data that will be used are retrieved from the RAND HRS Longitudinal File 2020 (V1). This dataset contains longitudinal data from the Health and Retirement Study (HRS) about individuals aged 50 years and older from the US (with some exceptions being younger). Due to its national representativeness and a high number of variables on various domains like health status, demographics, socioeconomic status (SES) and more between 1992 and 2020, the file gives us valuable data over multiple time periods. Derived from the earlier versions of this file and the same respondents, is a dataset constructed by Vable et al. (2017). The Validated Measures of Childhood Socio-Economic Status 2018 (V1) file will be used to measure childhood SES. For the genetic variants explaining variance in subjective wellbeing, we use the HRS Polygenic Scores File (V4.3). This file has data on the same individuals who provided salivary DNA between 2006 and 2012 and offers a guide for the construction of genetic variants for a number of phenotypes, based on a replicated GWAS.

### 3.2 Genoeconomic Background

As mentioned before, this study will use genetic variants explaining variance in subjective wellbeing as instrumental variable. These genetic variants linked to subjective wellbeing are constructed from a GWAS conducted by the Social Science Genetic Association Consortium (SSGAC) by Okbay et al. (2016). The SSGAC is a collaboration of scientists and geneticists investigating the genetic basis of social science and behaviors, based on the Human Genome Project. They studied 298420 individuals with a European ancestry, combining approximately 9.3 million Single Nucleotide Polymorphisms (SNPs). SNPs are genetic variations that occur when a single nucleotide (A, T, C, or G) in the DNA sequence is different between individuals. The genetic variants used as IV, which also include SNPs, are denoted as Polygenic Scores (PGSs). PGSs are a numerical representation of one's genetic risk for a particular trait or disease and constructed by combining effects of multiple genetic variants across the genome. In this case, weighted sums represent those combined effects. Weights are defined by the beta estimate from the GWAS meta-analysis files associated with subjective wellbeing. PGSs are calculated by equation (1), where $W$ is the GWAS meta-analysis effect size for SNP $j$ and $G$ is the genetic make up for individual $i$ at SNP $j$.

$$PGS_i = \sum_{j=1}^{j} W_j G_{ij} \tag{1}$$

Like most genoeconomic studies, the GWAS done by Okbay et al. (2016) is not without limitations. Benjamin et al. (2012) argue that one of the pitfalls of genoeconomics is statistical power issues leading to false positives. This and other pitfalls of the used GWAS will be discussed later on in this paper.

### 3.3 Method

For this research, two different models will be tested. The first model will use OLS, to replicate existing observational correlational studies. The second model will use MR with a PGS for subjective wellbeing as IV, using Two Stage Least Squares (2SLS). With this model, causality can be found and this way the research question can be answered.

*Linear Regression Analysis*

For every risky health behavior, general equation (2) is used in this first model. $Y$ is the outcome which for $H_1$ is the prob. of currently smoking, for $H_2$ the number of days per week one consumes alcohol and for $H_3$ the frequency of physical activity. $\beta_0$ is the constant term, $\beta_1$ is the treatment effect and *SWB* is a vector containing seven levels of subjective wellbeing. Another vector $X$ is used for all control variables that will be defined in subsection 3.5. Here it is important to note that the control variables are used for all three hypotheses, and extra control variables are added when measuring physical activity. $\varepsilon$ denotes the error term.

$$Y_i = \beta_0 + \beta_1 SWB_i + \beta_2 X_i + \varepsilon_i \tag{2}$$

In order to get a robust effect, some assumptions must hold. In this research, the first concern that comes with OLS is the linearity assumption not holding. The actual effect could differ at different levels of subjective wellbeing and could thus not be linear. Secondly, the Conditional Independence Assumption (CIA) is unlikely to hold, since there can be omitted variables other than $X$ related to subjective wellbeing and risky health behaviors. Thirdly, there could be reverse causality in this case, since lots of research has shown effects of risky health behavior on wellbeing. For example, Mandolesi et al. (2018) found that physical exercise positively effects wellbeing. In order to control for these assumptions, a different approach is needed.

*Mendelian Randomization*

The second model will use MR with a PGS for subjective wellbeing as IV, with the same pattern as 2SLS following Viinikainen et al. (2022). The interest is in equation (2). *SWB* will however not be a vector of subjective wellbeing levels, but the original variable taking on values between 1 and 7. The reason for this is that the F-statistics of the 1st stage regressions of the PGS on every level of subjective wellbeing are very low. The PGS only explains variation in the continuous original variable and not the dummies. In the 1st stage denoted by equation (3), *SWB* will be regressed on the PGS for subjective wellbeing for individual $i$ and will predict $\widehat{SWB}$. $\delta_0$ is a constant and $\delta_1$ is the effect of the *PGS* on *SWB*. $X$ is a vector of childhood financial, social and human capital with effect $\delta_2$.

(3) $$\widehat{SWB}_i = \delta_0 + \delta_1 PGS_i + \delta_2 X_i$$

In the 2nd stage, $\widehat{SWB}$ will be substituted in the main equation (2), resulting in equation (4) with coefficients $\alpha_0$, $\alpha_1$ and $\alpha_2$. $u$ is the error term.

(4) $$Y_i = \alpha_0 + \alpha_1 \widehat{SWB}_i + \alpha_2 X_i + u_i$$

MR relies on the fact that one's genetics are determined at conception and thus randomly distributed. This way, the only difference between individuals is the difference in their polygenic score for subjective wellbeing. Hence, OLS-concerns are avoided if a set of assumptions hold.

### 3.4 Assumptions for Mendelian Randomization

MR can predict causality, if certain assumptions are met. This study will follow the assumptions described by von Hinke et al. (2016).

*Assumption 1: Independence*

The independence assumption holds when the PGSs are uncorrelated with the error term. This assumption cannot be validated, but some controls can be added. Firstly, population stratification might exist. This occurs when the allele distribution differs between subpopulations in a population causing spurious associations. This is also one of the pitfalls of GWAS. The GWAS used for this research used Principal Component (PC) analysis by Price et al. (2006) to solve this problem, which resulted in a European ancestry specific dataset. Secondly, genetic nurture could arise: parental genotypes may affect the fitness of their children, mediated by the environment created by the parents (Kong et al., 2018). Likewise, assortative mating could violate independence, since genes are randomly distributed, but based on the genes of the parents. To mitigate these problems, parental environmental factors during childhood are added to the second model as controls. Despite these attempts, this assumption can still not be verified.

*Assumption 2: Exclusion*

The exclusion assumption holds when the PGSs affect the outcome only through exposure to the treatment. This assumption cannot be tested. The SNPs explaining variance in subjective wellbeing could affect the outcome through biological pleiotropy. Biological pleiotropy occurs when a genetic variant influences other phenotypic traits that are most unlikely to be related with each other. If this is the case, the PGS for subjective wellbeing could directly affect risky health behaviors (Solovieff et al., 2013). Secondly, von Hinke et al. (2016) argue that there is indirect biological pleiotropy, where genetic variants influence confounders through biological pathways. If the PGS for subjective wellbeing influences certain personality traits that influence the outcome, assumption 2 is violated. This assumption cannot be verified.

*Assumption 3: Non-zero effect of the instrument on treatment*

For this assumption to hold, the PGS for subjective wellbeing is required to have an effect on subjective wellbeing. This will be met if the F-statistic in the 1st stage is above 10, which is the general rule of thumb (Angrist & Pischke, 2009). Table 3.1 contains the 1st stage regression results, using equation (3). Assumption 3 holds since the F-statistic is 19.83.

Table 3.1 1st stage 2SLS Regression Results

| | Subjective Wellbeing | |
|---|---|---|
| | $\delta_{\#}$ | F-statistic |
| PGS | 0.130*** | 19.83 |
| | (0.039) | |
| Childhood financial capital | 0.058 | |
| | (0.038) | |
| Childhood social capital | 0.185*** | |
| | (0.028) | |
| Childhood human capital | 0.091** | |
| | (0.045) | |
| Constant term | 4.854*** | |
| | (0.045) | |
| Observations | 1 536 | |

*Table 3.1 contains the results for the 1st stage regression of Mendelian Randomization. PGS stands for the polygenic score for subjective wellbeing. Robust heteroskedastic standard errors are given in brackets. *p-value<0.10, **p-value<0.05, ***p-value<0.01.*

Okbay et al. (2016) stress the limitation of their GWAS, which could also be a caveat for this assumption. They mention that the PGS for subjective wellbeing explains a low fraction of variance, being 0.9%. Therefore, they warn about low predictive statistical power which is also what Benjamin et al. (2012) mentioned as limitation.

*Assumption 4: Monotonicity*

This assumption holds when the PGS affects subjective wellbeing consistently and without contradictory effects. A non-monotonic effect could produce unmeasured confounding. Von Hinke et al. (2016) describe that gene-environment interactions may violate the monotonicity assumption. These interactions take place when the effect of a genetic variant varies depending on certain environmental factors. However, since the PGS calculated by Okbay et al. (2016) is constructed from a subset of 9.3 million SNPs, a contradictory effect on confounders is unlikely. Therefore, the chance of a gene-environment interaction violating the monotonicity assumption is low, even though such an interaction is reasonable. Nevertheless, this assumption cannot be tested due to its unobservable nature.

### 3.5 Variables

First of all, it is important to note that datasets containing both normal and genetic data are scarce. Hence, a possible limitation of this research will be to stick to the measurements of variables of one single database like the HRS, which in this case doesn't offer concise measurements for every variable. Besides, observations are lost of individuals having ethical problems with sharing genetic information. All variables are self-reported.

*Main Variables of Interest*

Subjective wellbeing from the data is measured as the respondents self-assessed quality of life. It is an averaged score on a continuous scale of 1-7, increasing in wellbeing, based on the satisfaction of their living situation, daily leisure acts, family life, financial situation, household income, health and life as a whole. This variable is transformed into a variable taking on seven values between very poor (1) and very good (7).

Smoking is measured by asking the respondents if they are currently smoking yes (1) or no (0). No data on the intensity of smoking is available.

Alcohol consumption is measured by how many days per week the respondent drinks. Additionally, heavy drinking is also measured for alcohol consumption, since there is a possibility that normal consumption has a U-curved relationship with subjective wellbeing, whereas heavy drinking is more often associated with lower life satisfaction (Strine et al., 2008; Zullig et al., 2001). Heavy drinking is measured following the guidelines offered by the National Institute on Alcohol Abuse and Alcoholism (2023). They define heavy drinking as 14 drinks per week for males, and 7 for females.

Physical activity is measured in three ways: light physical activity, vigorous physical activity and moderate physical activity. The respondent can answer: every day (1), more than once a week (2), once per week (3), one to three times a month (4) or never (5), for every level of physical activity. To account for every form of physical activity, an index is created summing up the answers for the three categories. This resulted in a variable taking on values between 3 (engaging into every level of activity every day) to 15 (never any level of activity). Thereafter, this variable is inverted and subtracted by 2. The final result is a variable taking on values between 1 (never any form of physical activity) to 13 (every form of physical activity every day).

*Control Variables*

In order to answer the research question properly and test the hypotheses, control variables will be added to the first and second model.

Control variables for the first model are age, gender, years of education, household income, religion and two of the big five personality traits: extroversion and agreeableness. Age is used since risky health behaviors and subjective wellbeing vary between different ages (Berrigan et al., 2003; Chen, 2001). Gender is controlled for, since risky health behavior differs per gender (Courtenay et al., 2002). One's years of education has also been found by de Walque (2007) to be associated with smoking and other risky health behaviors by Viinikainen et al. (2022). The same association is found on life satisfaction by Cheung and Chan (2009). Moreover, the associations between income and subjective wellbeing and alcohol consumption have already been discussed (Popovici & French, 2013; Allen et al., 2017), but associations with smoking and physical activity have also been found (Casetta et al., 2017; Shuval et al., 2017). Furthermore, risky health behaviors have been found to vary among religiosity and importance of religion by Fletcher and Kumar (2014). The same is true for variation in subjective wellbeing associated with religion (Bergan & McConatha, 2001). Lastly, a study by Raynor and Levine (2009) shows that highly extraverted individuals are more likely to engage into risky health behaviors. In addition, Vollrath et al. (1999) show that agreeableness is associated with risky health behaviors. Jovanović (2019) showed evidence for an association between both personality traits and subjective wellbeing, but ruled out that the other three traits of the big five were associated.

Specifically for testing physical activity, a set of doctor-diagnosed health problems is also controlled for. This set contains reports of high blood pressure, diabetes, cancer, lung disease and heart problems. It is clear from the literature that those with chronic diseases undertake less physical activity (Barker et al., 2019). The same holds for subjective wellbeing: Strine et al. (2008) reported higher life dissatisfaction among adults with chronic illnesses. The effects on smoking and alcohol consumption are less clear. Therefore, this control set is only added in the physical activity analysis.

Lastly, the MR model only uses three control variables retrieved from Vable et al. (2017). The first is childhood financial capital, based on the average financial resources and financial instability in childhood. The second is childhood social capital, based on the quality of the relationship with the mother and the family structure in childhood. The third is childhood human capital index, which is an index of the sum of the mother's and father's years of education. These variables are added to the model to control for genetic nurture and assortative mating.

### 3.6 Descriptive Statistics

The data used for this research consists of 7,010 individuals, with information from 2012. This year is chosen, since the F-statistic for the 1st stage regression was the best fit, compared to other time periods between 2008 and 2020. Another reason is the most responses that this year's survey provided. Table 3.2 contains the descriptive statistics for continuous variables. As expected, the median age is relatively old, being 67 years. On average, individuals drink 1.15 days per week, while being below average physically active in any form (5.45 out of 12).

Table 3.2 Descriptive Statistics for Continuous Variables

| | Obs. | Mean | Std. dev. | Min. | Max. |
|---|---|---|---|---|---|
| Age | 7 010 | 67.28 | 11.03 | 27 | 101 |
| PGS | 2 196 | -0.02 | 0.97 | -3.44 | 3.21 |
| Days/week drinking | 6 997 | 1.15 | 2.00 | 0 | 7 |
| Physical activity | 7 010 | 5.45 | 2.88 | 0 | 12 |
| Years of education | 7 010 | 12.87 | 3.00 | 0 | 17 |
| Ln (household income) | 6 947 | 10.61 | 1.02 | 2.48 | 15.11 |
| Extroversion | 7 010 | 3.17 | 0.58 | 1 | 4 |
| Agreeableness | 7 010 | 3.51 | 0.50 | 1 | 4 |
| Childhood financial capital | 2 341 | 0.00 | 1.07 | -3.10 | 2.27 |
| Childhood social capital | 2 345 | -0.19 | 1.52 | -5.64 | 1.47 |
| Childhood human capital | 2 347 | 0.31 | 1.03 | -2.57 | 2.42 |

*Table 3.2 contains the descriptive statistics for continuous variables. PGS refers to the polygenic score for subjective wellbeing, Physical activity is a scale from 0 (no form of physical activity) to 12 (three forms of physical activity every day). Extroversion and agreeableness are on a scale of 1 (not) to 4 (very). The childhood capital variables are an index.*

Table 3.3 contains the descriptive statistics for categorial variables. Interestingly, females are overrepresented with 60% to 40% males, which is obviously not representative for the national population. A possible reason for this could be the mean age being 67, since women have higher life expectancy (Rochelle et al., 2015). 14% currently smokes, being less than the national average of 18.1% in 2012 (CDC, 2014). 9% of individuals can be categorized as heavy drinkers, which is more than the national 6% (CDC, 2022). Another interesting fact to note is that only 9% of the population is not religious, compared to 29% nationally (PRC, 2021). This could be due to the high average age. On average, the individuals have decent subjective wellbeing, with good subjective wellbeing having the most observations. Very few individuals have very good subjective wellbeing (4%) or very poor subjective wellbeing (6%).

To conclude: when it comes to age, gender, religion, smoking and heavy drinking, the sample used in this research does not seem to representative for the US population. The average age of 67 and 60% being female in the sample are obviously not representative. In addition, the individuals in the sample are more religious, smoke less and drink more often heavy (PRC, 2021; CDC, 2014; CDC, 2022).

Table 3.3 Descriptive Statistics for Categorial Variables

| | Category | Obs. | Freq. |
|---|---|---|---|
| Gender | male (1) | 2 833 | 0.40 |
| | female (2) | 4 177 | 0.60 |
| Smoking | no (0) | 6 011 | 0.86 |
| | yes (1) | 961 | 0.14 |
| Heavy drinking | no (0) | 6 390 | 0.91 |
| | yes (1) | 620 | 0.09 |
| Religion | protestant (1) | 4 247 | 0.61 |
| | catholic (2) | 1 848 | 0.26 |
| | jewish (3) | 100 | 0.01 |
| | none / no pref. (4) | 642 | 0.09 |
| | other (5) | 173 | 0.02 |
| Subjective wellbeing | very poor (1) | 414 | 0.06 |
| | poor (2) | 623 | 0.09 |
| | below average (3) | 946 | 0.14 |
| | average (4) | 1 273 | 0.18 |
| | above average (5) | 1 718 | 0.24 |
| | good (6) | 1 719 | 0.25 |
| | very good (7) | 317 | 0.04 |
| High blood pressure | no (0) | 2 766 | 0.39 |
| | yes (1) | 4 244 | 0.61 |
| Diabetes | no (0) | 5 380 | 0.77 |
| | yes (1) | 1 630 | 0.23 |
| Cancer | no (0) | 5 900 | 0.84 |
| | yes (1) | 1 110 | 0.16 |
| Lung disease | no (0) | 6 275 | 0.90 |
| | yes (1) | 735 | 0.10 |
| Heart problems | no (0) | 5 259 | 0.75 |
| | yes (1) | 1 751 | 0.25 |

*Table 3.3 contains the descriptive statistics for categorial variables. Heavy drinking takes on value 1 if males drink 14 or more drinks per week, and if females drink 7 or more drinks per week. All five health problems are doctor diagnosed. Subjective wellbeing is based on survey answers on several domains that make up one's current life satisfaction. The frequency of all categories per variable equals 1, meaning that the reported frequency per category is a part of the corresponding variable.*

# 4 Results

## 4.1 Linear Regression Results

Table 4.1 shows the linear regression results, using equation (2). The reference level for subjective wellbeing is very poor subjective wellbeing, to see the direction of the relationship more clearly. To give insight in the incremental contribution of the control variables, columns (1), (3) and (5) report the results without the control variables. This resulted in lots of significant effects, but these effects cannot be interpreted due to confounding. The 5% significance level will be the criterion for significance.

When it comes to smoking in column (2), only the wellbeing levels of average and above show significant effects compared to poor subjective wellbeing, in line with $H_1$. The coefficient gets increasingly smaller compared to the reference level of very poor subjective wellbeing. Hence, associations have been found indicating a possible inverse relationship between average and higher levels of subjective wellbeing and the likelihood to smoke. This is in line with the literature mentioned in the theoretic framework. For example, based on the estimated coefficient, the likelihood to smoke of someone with very good subjective wellbeing decreases with 19.1% compared to someone with poor subjective wellbeing. Column (4) shows the results for alcohol consumption. Whereas all control variables except for age show significant associations with alcohol consumption, only above average and good subjective wellbeing show significant associations. Within the sample, having above average subjective wellbeing increases the number of days per week one drinks with 0.208 days compared to having poor subjective wellbeing. Likewise, having good subjective wellbeing increases it with 0.289 days. It is possible that these levels are associated with alcohol consumption because of alcohol consumption at joyful social events. The inverse relationship has not been found and the U-curved relationship found by Paschall et al. (2005) could not be replicated. More research did find an inverse relationship between wellbeing and heavy drinking (Witkiewitz et al., 2011; Zullig et al., 2001; Strine et al., 2008). Therefore, equation (2) is used again but this time $Y_i$ = heavy drinking. The results are reported in the appendix. The effect is not significant for any level of subjective wellbeing. Concluding, previous literature could not be replicated. Column (6) shows the results for physical activity. The positive association between subjective wellbeing and physical activity found by the aforementioned literature has also been found in this research. The results show a positive association of all levels compared to poor subjective wellbeing. In addition, every control variable except for having cancer show significant associations with physical activity as well.

Even though results from previous literature are partly replicated, the $H_0$ hypothesis of no effect cannot be rejected. Due to possible confounding of not included outcome-related variables, the Conditional Independence Assumption does not hold. To properly answer the research question and test the hypotheses, another method is needed. Subsection 4.2 will show the results of this method.

Table 4.1 Linear Regression Results

| | Smoking | | Alcohol consumption | | Physical activity | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Subjective Wellbeing | | | | | | |
| poor (2) | -0.050** | -0.026* | 0.131 | 0.071 | 0.377** | 0.352** |
| | (0.027) | (0.026) | (0.113) | (0.109) | (0.191) | (0.172) |
| below av. (3) | -0.080*** | -0.041 | 0.263** | 0.124 | 0.869*** | 0.738*** |
| | (0.025) | (0.024) | (0.108) | (0.105) | (0.177) | (0.162) |
| average (4) | -0.114*** | -0.067*** | 0.268*** | 0.116 | 1.059*** | 0.870*** |
| | (0.024) | (0.022) | (0.100) | (0.099) | (0.169) | (0.155) |
| above av. (5) | -0.133*** | -0.087*** | 0.452*** | 0.208** | 1.618*** | 1.168*** |
| | (0.023) | (0.022) | (0.100) | (0.100) | (0.164) | (0.153) |
| good (6) | -0.162*** | -0.117*** | 0.558*** | 0.289*** | 1.837*** | 1.256*** |
| | (0.023) | (0.025) | (0.100) | (0.102) | (0.163) | (0.155) |
| very good (7) | -0.191*** | -0.117*** | 0.538*** | 0.253* | 2.011*** | 1.201*** |
| | (0.026) | (0.025) | (0.149) | (0.147) | (0.212) | (0.198) |
| Gender (female) | | -0.039*** | | -0.465*** | | -0.263*** |
| | | (0.008) | | (0.052) | | (0.067) |
| Age | | -0.007*** | | 0.004 | | -0.045*** |
| | | (0.000) | | (0.002) | | (0.003) |
| Years of education | | -0.006*** | | 0.088*** | | 0.103*** |
| | | (0.001) | | (0.008) | | (0.012) |
| Ln (household income) | | -0.053*** | | 0.201*** | | 0.228*** |
| | | (0.005) | | (0.027) | | (0.036) |
| Religion | | 0.004 | | 0.129*** | | 0.105*** |
| | | (0.004) | | (0.024) | | (0.030) |
| Extroversion | | 0.003 | | 0.183*** | | 1.010*** |
| | | (0.009) | | (0.050) | | (0.066) |
| Agreeableness | | 0.003 | | -0.163*** | | -0.205** |
| | | (0.010) | | (0.058) | | (0.077) |
| High blood pressure | | | | | | -0.454*** |
| | | | | | | (0.067) |
| Diabetes | | | | | | -0.544*** |
| | | | | | | (0.076) |
| Cancer | | | | | | -0.035 |
| | | | | | | (0.088) |
| Lung disease | | | | | | -0.706*** |
| | | | | | | (0.105) |
| Heart problems | | | | | | -0.440*** |
| | | | | | | (0.077) |
| Constant term | 0.254*** | 1.373*** | 0.789*** | -2.007*** | 5.166*** | 3.163*** |
| | (0.022) | (0.076) | (0.086) | (0.386) | (0.149) | (0.531) |
| Observations | 6 909 | 6 909 | 6 934 | 6 934 | 6 947 | 6 947 |

*Table 4.1 contains linear regression results of the effect of subjective wellbeing on the likelihood of currently smoking (1) & (2), alcohol consumption in days per week one drinks (3) & (4) and the frequency of physical activity on a scale of 0 to 12 (5) & (6). Religion contains five different levels. Extroversion and agreeableness are on a scale of 1 to 4. The reference category for subjective wellbeing is very poor subjective wellbeing. Robust heteroskedastic standard errors are given in brackets. *p-value<0.10, **p-value<0.05, ***p-value<0.01.*

### 4.2 Mendelian Randomization Results

Table 4.2 reports the results for the second model, using equation (4). MR is an improvement on OLS when trying to find causality, but certain limitations arise when using this method. As earlier mentioned, the PGS for subjective wellbeing only explains 0.9% variance in subjective wellbeing. Bochud and Rousson (2010) research the usefulness of MR and state that since the variance explained by the instrument is mostly small, very large sample sizes are needed (>10,000). The sample size reported in table 4.2 is thus relatively small, also compared to the first model in subsection 4.1.

The first hypotheses $H_1$ stated that subjective wellbeing has a causal inverse relationship with the likelihood to smoke. The results reported in column (1) show an inverse association with an estimated coefficient of -0.152 that is significant at the 5% level. This means that if one's subjective wellbeing score goes up with 1 on the scale of 1 to 7, the likelihood of smoking drops with 15.2% within the sample. This relationship is however not causal, since it is unclear whether the assumptions hold. The second hypotheses $H_2$ stated that subjective wellbeing has a causal inverse relationship with alcohol consumption. The 2SLS results reported in column (2) is not significant (p>0.05), which is why the $H_0$ hypothesis of no effect cannot be rejected. The third hypotheses $H_3$ stated that subjective wellbeing has a causal positive relationship with the frequency of physical activity. Yet again, the results reported in column (3) show no significant results at the 5% level. Hence, the $H_0$ hypothesis of no effect cannot be rejected.

Table 4.2 2SLS IV Regression Results (2nd stage)

| | Smoking | Alcohol consumption | Physical activity |
|---|---|---|---|
| | (1) | (2) | (3) |
| Subjective wellbeing | -0.152** | -0.195 | -0.009 |
| | (0.072) | (0.447) | (0.567) |
| Childhood financial capital | 0.006 | -0.042 | 0.016 |
| | (0.011) | (0.063) | (0.079) |
| Childhood social capital | 0.014 | 0.050 | 0.096 |
| | (0.015) | (0.092) | (0.120) |
| Childhood human capital | 0.003 | 0.401*** | 0.872*** |
| | (0.012) | (0.084) | (0.106) |
| Constant term | 0.847** | 2.115 | 6.073** |
| | (0.350) | (2.166) | (2.753) |
| Observations | 1 536 | 1 532 | 1 530 |

*Table 4.2 contains 2SLS IV 2nd stage regression of the effect of subjective wellbeing on the likelihood of currently smoking (1), alcohol consumption in days per week one drinks (2) and the frequency of physical activity on a scale of 0 to 12 (3). The IV is a polygenic score for subjective wellbeing. Robust heteroskedastic standard errors are given in brackets. *p-value<0.10, **p-value<0.05, ***p-value<0.01.*

Even when all results were significant with the same sign as hypothesized, the three hypotheses could not be accepted since certain assumptions cannot be checked. There could be biological pleiotropy, a non-monotonic effect or no independence of the instrument. It could also very well be that there is no causal effect, and the associations are subject to other confounders that are more interesting for policy.

# 5 Conclusions and Discussion

### 5.1 Conclusions

The aim of this research was to find causal inference between subjective wellbeing and risky health behaviors. To replicate current research, a linear regression was performed and in order to find a causal effect, Mendelian Randomization was used.

The first model found several associations. Average to high subjective wellbeing was inversely associated with the likelihood of smoking and significant at the 5% level, compared to poor subjective wellbeing, indicating that higher subjective wellbeing is associated with lower likelihood of smoking. This is in line with previous research (Xie et al., 2022; Strine et al., 2008; Grant et al., 2009; Carvajal et al., 2012; Ashton & Stepney, 1982). The effect of only above average and good subjective wellbeing on alcohol consumption in days per week one drinks was significant at 5%. However, the inverse association with alcohol consumption, the inverse association with heavy drinking and the U-curved association found by previous research have not been found (Witkiewitz et al., 2011; Strine et al., 2008; Grant et al., 2009; Zullig et al., 2001; Myrtveit et al., 2019). The effect of subjective wellbeing on physical activity was significant at the 5% level on all levels of subjective wellbeing, compared to very poor subjective wellbeing. The higher the level, the greater the magnitude. This indicates a positive association between subjective wellbeing and physical activity, which is in line with previous research (Strine et al., 2008; Melin et al., 2003; Schneider et al., 2009; Reigal et al., 2014).

Previous research on the three topics has one thing in common: mere associations were found. These associations could still run through different paths where confounders come to play. The second model is used with the aim of filling this gap in the literature. This model found an inverse association between subjective wellbeing and smoking, significant at 5%. Within the sample, individuals experience a drop of 15.2% in the likelihood of smoking when their subjective wellbeing level goes up with 1 on a scale of 1 to 7. However, the MR model found no significant effects on alcohol consumption and physical activity at the 5% significance level. Therefore, nothing can be said about the sign and magnitude of the coefficients.

Since the assumptions cannot be checked, the $H_0$ hypothesis of no effect cannot be rejected for all risky health behaviors. For alcohol consumption and physical activity this can either indicate that the statistical power was not great enough, or that there is no causal effect of subjective wellbeing on the two risky health behaviors. Due to the unobservable nature of some of the assumptions, it is hard to say which it is.

### 5.2 Limitations

This study has several limitations. It is important to note that most of the limitations focus on the second model, since this model is used to test the hypotheses and answer the research question. First, the sample is not representative of the full US population. The average respondent is from the US and aged around the retirement age of 67, the individuals are only of European ancestry, data is only retrieved from 2012 and all of them gave permission to their genetic information which says something about them. Even though MR can randomize, it is within the sample which is why it lacks generalizability.

Second, all variables are self-reported. The HRS uses a survey to gain data on individuals. This is a limitation, since it cannot be checked whether an individual answers truthfully. For example, people can be ashamed in reporting their alcohol use or lie about physical activity. The way individuals interpret certain variables can also differ. A generally happy person can give life a 7 out of 10 but a person who is just happy at the moment but not in general can give life an 8 out of 10. Their reference levels differ, which could cause bias in measuring subjective wellbeing.

Third, MR used in this research has several limitations. The sample size is too small. Bochud and Rousson (2010) argue that MR requires at least 10,000 individuals and this research investigated only 1,530 individuals. This leads to statistical power issues, something which Benjamin et al. (2012) warned for. Furthermore, there could be biological pleiotropy which violates the exclusion restriction. The PGS for subjective wellbeing could affect risky health behaviors through various biological pathways, which cannot be checked. Moreover, the independence assumption could be violated since genetic nurture, population stratification and assortative mating could occur even after adding controls. In addition, the effect of the PGS on subjective wellbeing could be non-monotonic, causing spurious associations and bias.

Lastly, this study may be biased due to the definitions of the variables of interest. First of all, only three risky health behaviors are used for this research. This is a limitation for the research question, since there are more health behaviors that are considered risky which are excluded. Moreover, smoking is measured by asking the respondents whether they smoke or not. Obviously, number of cigarettes per week would have been a better measure. Further, alcohol consumption is only measured by the number of days per week one drinks and it is unclear what is defined as 'one drink'. Measurements of alcohol levels in one's blood would have been better. Moreover, the magnitude of the domains used to measure subjective wellbeing is the same for all of them. It would be very logical if an individual receives more joy from family than from leisure, but the scale used for this research does not account for this. Genoeconomics has many promises, but finding the perfect dataset stays one of its main pitfalls. This is also the reason to stick with biased definitions of variables in this research.

## 5.3 Future Research

Genoeconomics are still on the rise. With the Human Genome Project just finished, more complete datasets are becoming available. In order to find an unbiased effect, a dataset with more respondents has to be used. Say at least over 10,000 (Bochud & Rousson, 2010). Additionally, this dataset should contain the genetic makeup of the parents. Genes are randomly allocated at conception, conditional on the parents' genome. Hence, if there can be controlled for their genome, it is way more likely that the independence assumption holds.

However, the main problem of MR is yet to be solved: biological pleiotropy. Completely randomized genetic variants could still affect the outcome or confounders through biological pathways, as earlier discussed. Van Kippersluis and Rietveld (2017) found a method to control for biological pleiotropy: Pleiotropy-robust Mendelian Randomization. The logic behind this method is testing the exclusion restriction for which the first stage effect is zero, and using this subsample's obtained estimate to directly test for the effect of the genetic variants on the outcome. The two assumptions of this method do have to hold: homogeneous pleiotropic effects and random selection into the subsample of zero effect. This way, the exclusion restriction could be properly tested.

Lastly, the data on the variables of interest should be more complete. It should solve the aforementioned limitations of the outcome variables, but also give a centralized definition to life satisfaction. One suggestion is to measure life satisfaction based on several tests instead of a survey. Furthermore, MR could also work with longitudinal data. This way, time-variant confounders are controlled for. One possible way is using G2SLS by Balestra and Varadharajan-Krishnakumar (1987). This method allows for the use of IV analysis over multiple time periods and there is no reason why this wouldn't work with a genetic variant as IV.

The recommendation for further research would thus be to use a dataset containing the parents' genome, more complete measurements of the risky health behaviors and subjective wellbeing, more risky health behaviors (drug use, stress management, reckless behavior, diets etc.) and more respondents. This dataset should also be longitudinal so that the G2SLS method can be used. Lastly, Pleiotropy-robust Mendelian Randomization should be used to properly test the exclusion restriction.

## 5.4 Policy Implications

Many associations have been found on this topic, but causality could provide useful information for policy targets. To solve the problem of chronic health diseases caused by risky health behaviors, the roots of this behavior will have to be tackled. Hence, finding causal inference between subjective wellbeing and risky health behaviors, or any of the confounders in between, could be crucial in the struggle against all chronic diseases. Despite the fact that this research only found associations, policy interventions could still be based on these associations and on those of further research. For example, if an inverse association is found between smoking and subjective wellbeing, policy could target the subjective wellbeing of smokers with the aim of reducing smoking. Two are given.

The first concerns workplace wellbeing campaigns. Based on this research, employers could be encouraged to improve the subjective wellbeing of their employees in the workplace, with the aim of reducing risky health behaviors and stress levels. This could also be beneficial for employers. Bellet et al. (2023) namely found a causal link between happiness and higher productivity. These campaigns could include more recognition, better work-life integration, mental health support and better opportunities for employees to have their say.

The second direction for policy is already mentioned in the introduction. This concerns the Community Reinforcement Approach described by Meyers et al. (2011). This intervention focuses on people with substance use disorders, with the aim of deterring substance use by improving the environment or community. This approach thus improves subjective wellbeing with the intention of lowering risky health behavior in its own way. Based on the results of this research, this approach could also focus specifically on the domains of subjective wellbeing, next to improving the environment or community.

These policy implications build upon the associations discovered in this research, emphasizing the need to consider subjective wellbeing as a critical influencer of risky health behaviors, since associations have been found. By strengthening workplace wellbeing campaigns and implementing a comprehensive Community Reinforcement Approach, policymakers can address subjective wellbeing and foster healthier lifestyles among individuals. Further research in this area can provide additional insights and find causal roots for policy interventions, contributing to the ongoing efforts in reducing chronic health diseases caused by risky health behaviors.

# References

Allen, L., Townsend, N., Williams, J., Mikkelsen, B., Roberts, N., & Wickramasinghe, K. (2017). Socioeconomic status and alcohol use in low- and lower-middle income countries: A systematic review. *Alcohol*, *70*, 23–31.

Angrist, J. D., & Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ashton, H., & Stepney, R. (1983). *Smoking: Psychology and Pharmacology*. London: Tavistock Publications.

Balestra, P., & Varadharajan-Krishnakumar, J. (1987). Full Information Estimations of a System of Simultaneous Equations with Error Component Structure. *Econometric Theory*, *3*(2), 223–246.

Barker, J., Byrne, K., Doherty, A. R., Foster, C., Rahimi, K., Ramakrishnan, R., Woodward, M., & Dwyer, T. (2019). Physical activity of UK adults with chronic disease: cross sectional analysis of accelerometer measured physical activity in 96 706 UK Biobank participants. *International Journal of Epidemiology*, *48*(4), 1386.

Bellet, C. S., de Neve, J. E., Ward, G. (2023). Does Employee Happiness Have an Impact on Productivity? *Management Science*, *0*(0), 1–24.

Benjamin, D. K., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D., Harris, T. B., Launer, L. J., Purcell, S., Smith, A. V., Registry, S. T., Johannesson, M., Beauchamp, J. P., Christakis, N. A., Atwood, C. S., Hebert, B., Freese, J., Hauser, R. A., Hauser, T. S., Grankvist, A., . . . Lichtenstein, P. (2012). The Promises and Pitfalls of Genoeconomics. *Annual Review of Economics*, *4*(1), 627–662.

Bergan, A., & McConatha, J. T. (2001). Religiosity and Life Satisfaction. *Activities, Adaptation & Aging*, *24*(3), 23–34.

Berrigan, D., Dodd, K. W., Troiano, R. P., Krebs-Smith, S. M., & Barbash, R. B. (2003). Patterns of health behavior in U.S. adults. *Preventive Medicine*, *36*(5), 615–623.

Bochud, M., & Rousson, V. (2010). Usefulness of Mendelian Randomization in Observational Epidemiology. *International Journal of Environmental Research and Public Health*, *7*(3), 711 728.

Carvajal, S. C. (2012). Global positive expectancies in adolescence and health-related behaviours: Longitudinal models of latent growth and cross-lagged effects. *Psychology & Health*, *27*(8), 916–937.

Casetta, B., Videla, A. J., Bardach, A., Morello, P., Soto, N., Lee, K., Camacho, P. A., Moquillaza, R. V. H., & Ciapponi, A. (2016). Association Between Cigarette Smoking Prevalence and Income Level: A Systematic Review and Meta-Analysis. *Nicotine & Tobacco Research*, *19*(12), 1401 1407.

Centers for Disease Control and Prevention. (2014). *Current Cigarette Smoking Among Adults – United States, 2005-2012*. Retrieved from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6302a2.htm#:~:text=In%202012%2C%0an%20estimated%2018.1,for%20trend)%20(Table).

Centers for Disease Control and Prevention. (2020). *Smoking & Tobacco Use*. Retrieved from https://www.cdc.gov/tobacco/basic_information/health_effects/index.htm#:~:text=Smoking 20causes%20cancer%2C%20heart%20disease,immune%20system%2C%20including%20rheumatoid%20arthritis.

Centers for Disease Control and Prevention. (2021). *Leading Causes of Death*. Retrieved from https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

Centers for Disease Control and Prevention. (2022). *Data on Excessive Drinking*. Retrieved from https://www.cdc.gov/alcohol/data-stats.htm

Chen, C. (2001). Aging and Life Satisfaction. *Social Indicators Research*, *54*(1), 57–79.

Cheung, H. Y., & Chan, A. (2009). The Effect of Education on Life Satisfaction Across Countries. *Alberta Journal of Educational Research*, *55*(1), 124–136.

Courtenay, W. H., McCreary, D. R., & Merighi, J. R. (2002). Gender and Ethnic Differences in Health Beliefs and Behaviors. *Journal of Health Psychology*, *7*(3), 219–231.

Fletcher, J. M., & Kumar, S. (2014). Religion and risky health behaviors among U.S. adolescents and adults. *Journal of Economic Behavior and Organization*, *104*, 123–140.

Grant, N., Wardle, J., & Steptoe, A. (2009). The Relationship Between Life Satisfaction and Health Behavior: A Cross-cultural Analysis of Young Adults. *International Journal of Behavioral Medicine*, *16*(3), 259–268.

Health and Retirement Study, (RAND HRS Longitudinal File 2020 (V1)) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (March 2023).

von Hinke, S., Smith, G. D., Lawlor, D. A., Propper, C., & Windmeijer, F. (2016). Genetic markers as instrumental variables. *Journal of Health Economics*, *45*, 131–148.

Jovanović, V. (2019). Adolescent life satisfaction: The role of negative life events and the Big Five personality traits. *Personality and Individual Differences*, *151*, 109548.

Kelloniemi, H., Ek, E., & Laitinen, J. (2005). Optimism, dietary habits, body mass index and smoking among young Finnish adults. *Appetite*, *45*(2), 169–176.

van Kippersluis, H., & Rietveld, C. A. (2018). Pleiotropy-robust Mendelian randomization. *International Journal of Epidemiology*, *47*(4), 1279–1288.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjálmsson, B. J., Young, A. S., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., Gudbjartsson, D. F., Helgason, A., Bjornsdottir, G., Thorsteinsdottir, U., & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424–428.

Levy, D. E., Chaloupka, F. J., & Gitchell, J. G. (2004). The Effects of Tobacco Control Policies on Smoking Rates. *Journal of Public Health Management and Practice*, *10*(4), 338–353.

Luhmann, M., Lucas, R. E., Eid, M., & Diener, E. (2013). The Prospective Effect of Life Satisfaction on Life Events. *Social Psychological and Personality Science*, *4*(1), 39–45.

Mandolesi, L., Polverino, A., Montuori, S., Foti, F., Ferraioli, G., Sorrentino, P., & Mercuri, N. B. (2018). Effects of Physical Exercise on Cognitive Functioning and Wellbeing: Biological and Psychological Benefits. *Frontiers in Psychology*, *9*.

Mann, S. (2008). Framing Obesity in Economic Theory and Policy. *Review of Social Economy*, *66*(2), 163–179.

Melin, R., Fugl-Meyer, K. S., & Fugl-Meyer, A. R. (2003). Life Satisfaction in 18- to 64-Year-Old Swedes: in Relation to Education, Employment Situation, Health and Physical Activity. *Journal of Rehabilitation Medicine, 35*(2), 84–90.

Meyers, R. J., Roozen, H. G., & Smith, J. M. (2011). The Community Reinforcement Approach. *Alcohol Res Health, 33*(4), 380-388.

Murphy, J. D., McDevitt-Murphy, M. E., & Barnett, N. P. (2005). Drink and Be Merry? Gender, Life Satisfaction, and Alcohol Consumption Among College Students. *Psychology of Addictive Behaviors*, *19*(2), 184–191.

Myrtveit, S. M., Knapstad, M., Askeland, K. G., & Skogen, J. C. (2019). Alcohol consumption, life satisfaction and mental health among Norwegian college and university students. *Addictive Behaviors Reports, 10*, 100216.

National Institute on Alcohol Abuse and Alcoholism. (2023). *Drinking Levels Defined*. Retrieved from https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge drinking

Okbay, A., Baselmans, B. M., De Neve, J. E., Turley, P., Nivard, M. G., Fontana, M. A., ... & Gratten, J. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics, 48*(6), 624-633.

Okun, M. A., Stock, W. D., Haring, M. J., & Witter, R. C. (1984). Health and Subjective Well-Being: A Meta-Analyis. *International Journal of Aging & Human Development*, *19*(2), 111–132.

Paschall, M. J., Freisthler, B., & Lipton, R. (2005). Moderate Alcohol Use and Depression in Young Adults: Findings From a National Longitudinal Study. *American Journal of Public Health*, *95*(3), 453–457.

Pew Research Center. (2021). *About Three-in-Ten U.S. Adults Are Now Religiously Unaffiliated*. Retrieved from https://www.pewresearch.org/religion/2021/12/14/about-three-in-ten-u-s adults-are-now-religiously-unaffiliated/

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics, 38*(8), 904-909.

Popovici, I., & French, M. T. (2013). Does Unemployment Lead to Greater Alcohol Consumption? *Industrial Relations*, *52*(2), 444–466.

Raynor, D. A., & Levine, H. (2009). Associations Between the Five-Factor Model of Personality and Health Behaviors Among College Students. *Journal of American College Health*, *58*(1), 73 82.

Reigal, R., Videra, A., & Gil, J. (2014). Práctica física, autoeficacia general y satisfacción vital en la adolescencia / Physical exercise, general self-efficacy and life satisfaction in adolescence. *Revista Internacional De Medicina Y Ciencias De La Actividad Fisica Y Del Deporte*, *14*(55), 561-576.

Rochelle, T. L., Yeung, D. K., Bond, M. H., & Li, L. M. W. (2015). Predictors of the gender gap in life expectancy across 54 nations. *Psychology Health & Medicine*, *20*(2), 129–138.

Schneider, M., Graham, D. J., Grant, A. C., King, P. E., & Cooper, D. M. (2009). Regional brain activation and affective response to physical activity among healthy adolescents. *Biological Psychology*, *82*(3), 246–252.

Shuval, K., Li, Q. X., Gabriel, K. P., & Tchernis, R. (2017). Income, physical activity, sedentary behavior, and the 'weekend warrior' among U.S. adults. *Preventive Medicine*, *103*, 91–97.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics*, *14*(7), 483–495.

Sousa, L., Lyubomirsky, S. (2001). Life Satisfaction. *Encyclopedia of women and gender: Sex similarities and differences and the impact of society on gender, 2*, 667-676.

Spear, L. P. (2018). Effects of adolescent alcohol consumption on the brain and behaviour. *Nature Reviews Neuroscience*, *19*(4), 197–214.

Stone, A. A., Mackie, C. (2013). *Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Experience [Internet]*. Washington (DC): National Academies Press (US).

Strine, T. W., Chapman, D. P., Balluz, L. S., Moriarty, D. G., & Mokdad, A. H. (2008). The Associations Between Life Satisfaction and Health-related Quality of Life, Chronic Illness, and Health Behaviors among U.S. Community-dwelling Adults. *Journal of Community Health*, *33*(1), 40–50.

Tiffin, R., & Arnoult, M. H. (2011). The public health impacts of a fat tax. *European Journal of Clinical Nutrition*, *65*(4), 427–433.

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.

Viinikainen, J., Bryson, A., Böckerman, P., Kari, J. T., Lehtimäki, T., Raitakari, O. T., Viikari, J., & Pehkonen, J. (2022). Does better education mitigate risky health behavior? A mendelian randomization study. *Economics and Human Biology*, *46*, 101134.

Vollrath, M., Knoch, D., & Cassano, L. (1999). Personality, risky health behaviour, and perceived susceptibility to health risks. *European Journal of Personality*, *13*(1), 39–50.

Wagenaar, A. C., Tobler, A. L., & Komro, K. A. (2010). Effects of Alcohol Tax and Price Policies on Morbidity and Mortality: A Systematic Review. *American Journal of Public Health*, *100*(11), 2270–2278.

de Walque, D. (2007). Does education affect smoking behaviors? *Journal of Health Economics*, *26*(5), 877–895.

Ware, E. B., Gard, A. M., Schmitz, L. L., Faul, J. D. HRS Polygenic Scores – Release 4. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan; 2020.

Witkiewitz, K., Bowen, S., & Donovan, D. M. (2011). Moderating effects of a craving intervention on the relation between negative mood and heavy drinking following treatment for alcohol dependence. *Journal of Consulting and Clinical Psychology*, *79*(1), 54–63.

Wolfinger, N. H. (1998). The Effects of Parental Divorce on Adult Tobacco and Alcohol Consumption. *Journal of Health and Social Behavior*, *39*(3), 254.

World Health Organization. (2022). *Physical Activity*. Retrieved from https://www.who.int/news room/fact-sheets/detail/physical-activity

Xie, Z., Cartujano-Barrera, F., Cupertino, P., & Li, D. (2022). Cross-Sectional Associations of Self Reported Social/Emotional Support and Life Satisfaction with Smoking and Vaping Status in Adults. *International Journal of Environmental Research and Public Health*, *19*(17), 10722.

Zullig, K. J., Valois, R. F., Huebner, E. S., Oeltmann, J. E., & Drane, J. W. (2001b). Relationship between perceived life satisfaction and adolescents' substance abuse. *Journal of Adolescent Health*, *29*(4), 279–288.

# Appendix

Table A1 Linear Regression Results for Heavy Drinking

| | Heavy Drinking | |
|---|---|---|
| | (1) | (2) |
| Subjective Wellbeing | | |
| poor (2) | -0.003 | -0.040 |
| | (0.017) | (0.016) |
| below av. (3) | 0.017 | 0.012 |
| | (0.016) | (0.016) |
| average (4) | 0.004 | -0.001 |
| | (0.015) | (0.015) |
| above av. (5) | 0.018 | -0.007 |
| | (0.015) | (0.015) |
| good (6) | 0.021 | -0.010 |
| | (0.015) | (0.015) |
| very good (7) | 0.025 | -0.011 |
| | (0.021) | (0.021) |
| Constant term | 0.074*** | -0.062 |
| | (0.013) | (0.060) |
| Observations | 6 896 | 6 896 |

*Table A1 contains linear regression results of the effect of subjective wellbeing on heavy drinking. Column (1) shows the results without control variables, while column (2) shows the results controlling for gender, age, years of education, a natural logarithm of household income, extroversion, religion and agreeableness. For simplicity, these controls are left out of table A1. The reference category for subjective wellbeing is very poor subjective wellbeing. Robust heteroskedastic standard errors are given in brackets. ***p-value<0.01.*