ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics & Erasmus School of Health Policy & Management

Master Thesis Health Economics

The healthcare cost of mental health issues

Name student: Daan van Oeveren

Student ID number: 501103

Supervisor: J. L. W. van Kippersluis

Second assessor: P. H. M. van Baal

Wordcount: 8354

Date final version: 31-7-2022

In this thesis, there is an attempt to establish a causal relationship between mental health issues and out of pocket healthcare expenses. This is with an analysis based on using genetic data as an instrument in an analysis. This is called a mendelian randomization analysis, and will be expanded upon by looking at the exogeneity of the instrument. Based on the obtained results, one can claim a statistically significant positive effect between a deteriorating mental state and higher healthcare costs. This statistically significant positive effect can not be claimed for those with a strong indication of depression.

Table of Contents

1. Introduction

With the COVID-19 pandemic going on, people had to spend time in isolation, and sometimes in solitude. A decrease in mental well being, especially among students, has been observed (Savage et al., 2020) in the UK. This decrease in mental well-being could be a start of the situation worsening overtime. Think of an increase in anxiety, or a critical amount of stress. These in-time could might even lead to a depressive episode, or once these episodes start to become more common, lead to a long-term chronical depression. In a study by Satyanarayana et al. (2009), they looked at a sample of about 35000 people, within the Canadian population aged 15 and older. In this research, it discusses how chronical depression is correlated with negative physical health conditions, like high blood pressure. The treatment of these mental health conditions could lead to an increase in medical spending, due to people seeking out therapy. Besides just mental issues developing, physical health conditions also could develop which would also increase medical spending.

More genetic data is becoming available every year. With each year, it is also becoming cheaper to obtain a detailed sequence (Wetterstrand, 2021). This could open up new research options to include this genetic data. As well as any other characteristics people might have. One example where genetic predispositions could be used for is risk predictions for certain traits. These predispositions could predict the risk of how likely it is that someone could develop a mental disorder, like chronical depression. Using these genetic data, and making use of it as an instrument in statistical modelling, one could find some strong causal relationships.

Combining both the strength of genetic predispositions, and a focus on mental health, the following research question can be formulated:

What is the causal effect of common mental health outcomes, like a depressive disorder, on out of pocket healthcare expenditures?

For common mental health outcomes, both a strong indication for depression and the survey score for predicting this depressive disorder will be used. This could also bring up the sub-question on how different measurements of these mental health outcomes differ in the resulting out of pocket healthcare expenditure.

Another question one could ask if these physical risks are strongly correlated with mental health conditions. If this is the case, which physical health conditions this includes. Next to these physical risks, the cost of these risks can be used, as well as treatment costs of mental health conditions. In a study done by Egede et al. (2014), they looked a group of United States veterans

with an already existing physical medical condition of diabetes type 2. Besides this physical medical condition, they also noted any mental health diagnosis the veterans might have had. Here it is demonstrated that an increase in mental health treatments could lead to a reduction in total healthcare expenditures.

The aforementioned study provides a great basis for the hypothesis that issues related to mental health warrant an increase in healthcare expenditures. Thus it can be expected that the results obtained from the research will below will conclude an increase in out of pocket health expenditures, in relationship with a deterioration in their general mental health, such as a strong indication of depression.

Other studies have also tried to predict mental health costs, by looking at other factors which might influence these costs. One of these factors could be people their socio-economic status (Donisi et al., 2011), which shows a reduction in mental healthcare costs if the person enjoys a higher socio-economics status. This thesis would be an addition to academic literature, since it will try to establish a causal connection, instead of an association, between a deteriorating mental health and healthcare costs. This is achieved by making use of genetic data as an instrument to obtain this causal result.

With the aforementioned descriptions of these studies, which demonstrate a deterioration in mental wellbeing, and an increase in healthcare costs associated with this deterioration, this thesis is societally relevant. It will demonstrate this increase in healthcare costs for people who experience bad mental health. Finding a strong negative result would demonstrate the importance of preventative mental healthcare. Since any found increase in healthcare cost due to a bad mental health, could be minimized with this preventative healthcare.

With regards to the content, first some literature will be discussed, highlighting the reasoning of the methods used, as well as describing some of the current state of the literature with regards to mental health and healthcare costs, as well as a short introduction on genetics and the use of them in the analysis. Second, the analysis plan will be discussed and a short explanation of each method will be given. Third, the analysis results will be presented. These results will be discussed in the next section, as well as some concluding remarks will be made.

2. Literature review

First and foremost, it is important to establish a strong connection between mental health and healthcare expenditures, since this described in the first hypothesis. To dive deeper in the paper mentioned in the introduction (Egede et al., 2014), the researchers studied a group of United States veterans, who had both diabetes type 2, as well as a minimum of one mental health diagnosis. They defined two groups, one which did not attend any mental healthcare specialists at all, and those who visited at least once or more. Three different healthcare costs were looked at: inpatient, outpatient, and pharmacy costs. The research resulted that savings can be obtained for inpatient costs for those who did visit a mental health specialist, while outpatient and pharmacy increased almost 10 times the savings obtained from the inpatient cost reduction. They also found that the veterans would end up accumulating a lower cost for the treatment of diabetes, this while the costs of mental healthcare did increase.

Besides savings in healthcare costs, due to individuals visiting mental health specialist, one could also look at an estimate of increase or decrease in healthcare cost, due to depression. A research paper by Unützer et al. (2009) attempts to establish this more direct connection between having a diagnosis of a depressive disorder, or the probability of having depression, and the increase in healthcare costs compared to a group which does not show any signs of depression. They did find a difference of around 5000 USD, in a sample of about 15000 participants. They found that the difference in medical cost for mental health services was not surprisingly different between the groups of those with and without depression. The authors argue that any direct payment for mental healthcare services might be due to this form of healthcare being unaffordable for some participants. A majority of this discrepancy was due to differences in inpatient care. The authors state that this might be the case due to other illnesses developing cause of a depressive disorder, which could explain the big differences in healthcare cost.

Some of these other illnesses which could develop, or are associated with depression, can also be studied. A paper by Swendsen & Merikangas (2000) performed a family study of 226 individuals, where they examined the odds ratios of substance abuse. The odds ratios looked at in their research were alcohol abuse, and different mood disorders, like bipolar disorder, or a depressive disorder. Where it is demonstrated that there a higher odd ratio of being alcohol dependent when suffering from a depressive disorder. The study is quite wary in establishing an effect between alcoholism and a depressive disorder. They also argue reverse causality being prevalent, where excessive consumption of alcohol could be one of the factors of a depressive

disorder forming. The main caveat to discuss from this study is the link between a mental health disorder, namely depression, and substance abuse. Especially when zooming in into alcoholism, a study done by Rhem (2011), looked at the risks associated with alcohol usage. In the research report itself, he discusses many different conditions which alcohol could be described as a causal factor. An example of some of these conditions range from liver disease to diabetes. Taking into account both aforementioned studies, and an increased risk of depressive disorder having a higher risk of alcohol abuse being established, could thus lead to an increase in healthcare costs, due to other physical diseases developing. Thus leading back to the first hypothesis where a strong indication of depression does lead to a higher healthcare cost for the individual.

With a relationship between mental health issues, and an increase in healthcare costs being able to be identified, both through direct costs by using mental health experts, or indirect cost due to other physical diseases forming as a result of this mental health issue, a research paper by Richardson et al. (2017) argues reverse causality. Which would entail that an increase in healthcare costs would lead to mental health issues. In the research, a group of around 450 British university students, were asked about financial stress and mental health. It demonstrated a strong association between those who showed some amount of stress about their debt, and having higher anxiety. Thus one could argue that high medical spending, could lead people into debt. This in turn could increase their levels of anxiety, which could lead to mental health problems forming.

To combat this reverse causality, where the outcome would affect the independent variable, from plaguing the analysis results, an instrumental variable analysis can be used. This method of analysis will be discussed in-depth later on. This method requires an instrument, as is apparent from the name. One instrument available to be used are genetic predispositions in a Mendelian Randomization (Emdin, Khera & Kathiresan, 2017).

With genetic predispositions, the concept of polygenic risk scores (PGS) has to be introduced. The scores are obtained via a genome-wide association study (GWAS). In one of these studies, an attempt is made to associate genetic structures with different outcomes. These outcomes can be physical, like a determinant for height, or behavioural, like the prevalence of certain character traits. To obtain such a PGS, the following formula can be used (Dudbridge, 2013):

$$\widehat{PGS} = \sum_{i=l}^{m} \widehat{\beta_{il}}\, G_i$$

The formula shows that the score obtained is just a weight factor G, which is multiplied with the genetic score β obtained for certain genetic structures. Since multiple of these genetic structures can influence one trait, they are all summated to construct this PGS score. All these PGS scores are then standardized for the entire sample. Standardizing is the method where the mean of all the individuals in the sample for a certain variable, in this case the PGS score is set to zero, and the standard deviation is set to one. All values will then be easily comparable with one another.

This leads to the following important aspect of the analysis, the selection of the proper PGS scores. Gale et al. (2016) demonstrates a strong association between the PGS for neuroticism and several mental health disorders of which a depressive disorder is one. Yet it also demonstrates that there is a strong case of pleiotropy. Pleiotropy describes the phenomenon where one polygenic score, also affects other traits. It is demonstrated that the PGS for neuroticism has a strong correlation with anxiety, bipolar disorder, and even some physical health statuses like BMI as well. This pleiotropy will be expanded on later when discussing methods to take into account this phenomenon into the analysis. While the pleiotropy is not preferred, it does make the PGS for neuroticism usable as an instrument when looking for other mental issues besides a depressive disorder. When selecting a proper PGS as instrument for those who do have a strong indication of a depressive disorder, the PGS for depression would appear to be the right decision. A paper by Murray et al. (2021) state a valid point, where they argue that someone their PGS may portray them as a high risk for developing a depressive disorder, a relatively low percentage of these people might actually develop these mental health issues. This case is also argued for other mental health disorders, like schizophrenia or bipolar disorder. Thus, however the relationship between a high PGS for a certain mental disorder exists, it does not automatically mean that there is a strong relationship between the two, which is something important to take into account in the analysis. In the concluding arguments of their research they encourage to also include family history into the analysis when determining if someone is at high risk of depression.

The inclusion of family history brings some new prospects into the analysis. A paper by Lewis & Vassos (2020) delves deeper into the argument that other factors can influence or activate some genes, which would otherwise stay inactive. They mention factors as outside influences, like the environment one grew up in, or disease history as factors, which could influence the way PGS demonstrate and develop themselves. They also explicitly argue about the usefulness

of the genetic risk scores of those predicting a depressive disorder. Since they argue not only someone their genes, but also their environment plays a role when establishing this connection.

A dataset created by Vable et al. (2017) attempts to construct a childhood socio-economic status (SES) score. The inclusion of this variable could influence the results, by including these outside influences, by including the childhood SES as an environmental effect. This dataset thus could be used in the analysis, to take into account any of the factors which might also have an effect on the PGS. Besides this, the authors mention that outcomes later in life can partially be brought back by the circumstances someone grew up in. This childhood SES score includes childhood financial stability, social stability, and human capital, consisting of the parental education levels.

3. Methodology and Analysis

The data used for this analysis is obtained from the Health and Retirement Study (HRS) dataset. The HRS offers both a dataset on a person their characteristics, like birthyear (1992-2018 RAND HRS Longitudinal file), while another dataset offers the polygenic risk scores for certain traits of each individual person (Polygenic Score Data, 2021). The aforementioned dataset created by Vable et al. (2017), containing the childhood SES scores, is also able to be matched with the HRS dataset, thus can be used in the analysis.

When diving into the dataset with the PGS. Two datasets are available, one for people with European ancestry, and one for people with African ancestry. the PGS for European ancestry will be used. Due to most GWAS which have been performed, were focussed mostly on European heritage, which caused the sample size to be bigger compared to those of African heritage, thus the obtained PGS tend to be more reliable (Martin et al., 2017). The PGS are standardized, meaning they have a mean of 0 and a standard deviation of 1.

When diving into the RAND HRS longitudinal file, and the variables being used there, it is important to note that a clinical evaluation of people having a depressive disorder is not present. This score has to be constructed from a CES-D score. This CES-D score is obtained by a survey respondents will have to answer themselves. The regular CES-D 20 takes into account 20 different aspects of their life, which they can respond to from four different answer possibilities. For example, they might ask the respondents if their sleep was restless, if they felt sad, or felt lonely in the previous week. This will result in a score based somewhere between 0 and 60, and those with a score of 16 or more are most likely to show signs of a depressive disorder (Vilagut et al., 2016). The score obtained from the respondents was based of the CES-D 8, where only

eight different aspects of their life were asked. Next to this restricted questionnaire asked, the answer options were also simplified. Instead of answering on a scale from 0-4, they were only able to say yes or no to the questions. This leads to obtained CES-D scores ranging from 0-8. Where a cut-off score of 3 will be used to determine if someone show signs of a depressive disorder (Kozlov et al., 2020).

The out of pocket expenditures (OOP) are represented in US dollars (USD). These out of pocket expenditures consists both of direct healthcare costs, related to treatments, as well as costs for nursing home usage. If an estimate of these expenditures was given, the exact amount was estimated by factoring in the amount of direct treatments obtained (Bugliari et al., 2021). Another important factor to take into account when observing these OOP, is the influence of Medicare. Medicare is a form of health insurance for people who are 65 years of age, or older. This form of insurance consists of two parts, one part covering inpatient care, the other part covering most outpatient care, as well service provided by a doctor, as well as mental health services. Seeing as mental health is subject to the second part of this insurance plan, it is important to take in mind the details of this plan. The Original Medicare plan has a deductible of $233, after which this amount is reached, the individual will pay 20% of any medical costs made. The rest will be paid by Medicare. (Medicare, 2022). So it is the case that it will not only look at out of pocket expenditures for mental healthcare, but also other forms of case as well. In Egede et al. (2014), as mentioned before, it is expected that those with depression are also subject to higher medical expenditures due to depression inducing the development of other unhealthy habits (Swendsen & Merikangas, 2000). These unhealthy habits could cause physical health issues. These physical health issues would increase these out of pocked medical expenditures, even if covered by the Medicare plan. It is also important to note that multiple waves of data are available, and will be used for this analysis. The descriptive statistics will be displayed in table 1.1 and table 1.2.

Table 1.1 Descriptive Statistics

	Mean	Standard Deviation	Min	Max	Observations
Gender	1.596	0.491	1	2	99797
Birthyear	1937.627	9.822	1905	1979	99797
CES-D 8 score	1.200	1.778	0	8	99770
Out of pocket medical expenditures	3016.037	7869.881	0	634821	95797
Childhood SES-score	0.273	0.872	-3.322	2.809	99797
Strong indication of depression	0.171	0.377	0	1	99797

Gender is a binary value being 1 for males and 2 for females. Out of pocket medical expenditures being expressed in US dollars.

Examining the descriptive statistics in table 1.1, it can be observed that there are slightly more females represented in this sample. Another statistic that stands out is the percentage of the sample that has a strong indicator of depression. Around 17% of the sample shows this strong indication, that while in 2015, around 7% of a sample analysis has had a depressive episode (Weinberger et al., 2018).

Table 1.2 Descriptive Statistics of out of pocket medical expenditures, for those with different conditions.

	Mean	Standard Deviation	Observations
OOP (With a strong indication of depression)	4098.199	10636.880	16506
OOP (Without a strong indication of depression)	2789.369	7135.379	78803
OOP (With Medicare)	3231.131	8881.759	59822
OOP (Without Medicare)	2653.445	5757.827	35487

With OOP standing for out of pocket medical expenditures. OOP is being expressed in US dollars.

Examining the descriptive statistics in table 1.2, it can be observed that the mean OOP medical expenditures are higher for those with a strong indication of depression, as well as for those

individuals who receive Medicare. It is important to note that comparing these values is called a simple with-and-without comparison, and does not establish causality. Thus one can not say that having a strong indication of depression is the cause for these higher OOP medical expenditures.

The first model which will be estimated, will be an ordinary least squares (OLS) model. The difficulty with an OLS model is that usually plagued with biased estimates. Usually due to unobserved variables not included into the estimation. One of these unobserved variables could be attributed to financial stress. In a paper by Roberts et al. (2000), they mention that financial stress could cause could both increase the rate of depression, as well as reduce people their overall health. This could in turn lead to higher healthcare costs. Omitting this variable could lead to a positive bias, thus overestimating the actual effect of depression on healthcare costs. Due to a lot of unobservable variables, which would bias the outcome, OLS will not be the strongest method to estimate this model, thus a different model has to be introduced. On the other hand, this model is great to demonstrate biased estimators, and show how much they differ from a less biased model, as with the next model.

The new model which will be introduced, is a Mendelian Randomization (MR) analysis. This analysis is an instrumental variable (IV) analysis, which makes use of genetic predispositions as its instrument (Emdin, Khera & Kathiresan, 2017). This analysis rests on three main assumptions:

I. The genetic risk score for a certain trait has to be associated with that certain trait.
II. There should be no relationship between the risk score and the error term.
III. There should be no direct relationship between the genetic risk score and the outcome variable.

To dive deeper into these assumptions, and what they would imply for the method of analysis. The first assumption describes that when looking at the PGS for depression, it has to be associated with a strong indication of depression as well. When estimating this relationship, an F-score will be calculated as well. Depending on the size of the F-score, this relationship is considered strong or not. This F-score has to be great or equal than ten, and the results of this F-score will also be displayed in the results section. The second assumption describes that there should be no association between the instrument, in this case, the PGS used, and any unobserved variables. Thus meaning that the PGS should be exogenous, and not affected by any other

factors. The third assumption describes that, in this case, the PGS used, should not have a direct effect on the out of pocked medical expenditures (Lousdal, 2018).

Paaby & Rockman (2013) describe an issue called pleiotropy. When pleiotropy is present, this entails that the genetic risk score, besides the treatment variable, could also influence different characteristics and traits, and in turn affect the outcome variable. If this is the case, the third assumption is not satisfied, and thus the PGS would affect the outcome variable of out of pocket expenditures, via other ways than the treatment variable, it being a strong indication of depression, or CES-D score. A paper by Van Kippersluis & Rietveld (2017) tries to tackle this issue, by doing what they call a pleiotropy robust mendelian randomization (PRMR). With this method, one can study the direct effect between the instrument and the outcome variable. Thus in this case, the direct effect the PGS has on out of pocket expenditures. This is established by looking at a subgroup of the sample which does not exhibit a relationship between the PGS for depression and the treatment. Which, in this analysis, will be a group of individuals who are not portraying a strong indication of depression, but do have a high PGS for depression. They state that this analysis also rests on a couple of main assumptions. Where they are less strict satisfying the third assumption, but take into account a couple of other assumptions:

I. A subgroup is present where the first stage result is zero.
II. Similar pleiotropic effects across both the subgroup where the first stage results is zero, and the other resulting subsample.
III. The subgroup where the first stage result is zero is not selected by the outcome, nor their genetic risk score.

As an example of this first assumption, based within context, it would be necessary to establish a subgroup where a mental health outcome is not present. In this case: A subgroup where depression is not present. The result obtained will show if there is any bias present affecting the regular MR results. If the second and third assumptions are satisfied, it would satisfy that this pleiotropic effects are similar between the groups where depression is not present, and depression is present. Thus since these effects will be similar between groups, thus when including this in the PRMR analysis, it would lead to an unbiased causal effect. Which would not have been able to be obtained with regular MR, since it does not take into account these pleiotropic effects.

For the first assumption, a subgroup is selected where the first stage result is zero. In this case, it would be a subgroup where a mental health outcome is not present, as mentioned before. Yet

when looking at the third assumption, it is important that this first stage result is not selected by outcome, nor their genetic risk score. This selected subgroup, where a mental health outcome is not present, thus was selected based on outcome. This selection in turn violates the third assumption. Yet, no other subgroup in this sample exists where a guaranteed first stage results equals zero. Thus the third assumption will have to be violated, which introduces bias in the found coefficients.

To formally introduce all models, first the OLS model will be displayed, and after the MR and the PMRM models will be talked about. After both models are formally introduced, an explanation is given about the variables used.

$$1)\ OLS\colon\ OOP\ =\ \beta_0 + \beta_1 MHO + \beta_2 Covariates + \beta_3 PC + \epsilon_{OOP}$$

$$2)\ MR\ \&\ PRMR\colon \begin{cases} 1)\ OOP\ =\ \beta_0 + \beta_1 MHO + \beta_2 Covariates + \beta_3 PC + \gamma PGS + \epsilon_{OOP} \\ 2)\ MHO\ =\ \lambda_0 + \lambda_1 PGS\ + \lambda_2 Covariates + \lambda_3 PC + \epsilon_{MHO} \end{cases}$$

Within the MR & PMRM equations: 2).2) Details the first stage equation. 2).1) Details the second stage equation. Where in the equation, *OOP* stands for out of pocket healthcare expenditures. *MHO* stands for mental health outcome. The two mental health outcomes which will be looked at will be a strong indication of depressive disorder (CES-D >= 3), and CES-D score in general. When talking about covariates, birthyear, birthyear², gender and birthyear*gender will be used. The birthyear is centred, this means that the actual birthyear will be subtracted from the mean birthyear of the sample, in turn making the coefficients obtained easier to interpret.

In both models, PGS will represent the polygenic risk score. It is important to note that in MR the result found for γ is assumed to be zero. On the other hand, in PMRM γ is supposed to be non-zero, to allow for the direct effect to be displayed in the outcome. PC in both of the models will be the principal components. Including these PCs should allow for a corrected model, compared to a model which just used the PGS (Coombes et al., 2020). This correction is necessary cause of population stratification (Price et al., 2006). It is stated that a difference in genetic variation can be linked to the previous generations in a population. Population stratification is more likely to affect the results when making use of a bigger sample. The PGS for depression, and neuroticism will be used. The PGS for depression will be used for the analysis evaluating the strong indication of depressive disorder, while the PGS for neuroticism will be used for evaluating the CES-D scores.

It is important to mention that in the OLS and MR analysis, both the childhood SES score, as well as a wave variable for each different wave will be added into the model as well. The inclusions of this childhood SES score tries to take into account any environmental effects that could influence the estimation. A study done by Eley et al. (2004) shows this interaction between SES and genetic development, where a worse environment was associated with a higher rate of depressive symptoms. It is important to note that these results were only found in participating females. Next to this, the estimated standard errors will be clustered for each person, since the same individuals are present in multiple waves.

When introducing the PMRM method, the result obtained from the 2nd stage equation will be adjusted. This adjustment comes from taking into account the results obtained when doing an analysis with a subgroup where the first stage result is zero. According to the research by Van Kippersluis & Rietveld (2017), they warn about a direct effect of the PGS on the outcome variable, in this case, healthcare costs, thus violating the third assumption of a MR. Thus by estimating this direct effect in this subgroup where the first stage result is zero, this effect can be estimated, and thus be taken into account in the PMRM estimation. They also make an important note where they state the importance of a bigger sample size, to reduce any uncertainty obtained. To formally introduce this adjustment:

$$3)\ \hat{\beta} \sim N(\beta + A\mu_\gamma, V_{MR} + A\Omega_\gamma A')$$

It shows that the obtained coefficient $\hat{\beta}$ is an approximated distribution, which follows a normal distribution, indicated by N(). Where A = (MHO'PGS(PGS'PGS)$^{-1}$PGS'MHO) $^{-1}$(MHO'PGS), with MHO being the mental health outcome, thus the independent variable, and PGS being the instrument. A, V_{MR}, and β are representing the MR estimates. Due to making use of the plausibly exogenous method, and the introduction of γ in equation 2)1), which can describe non-zero values. This is displayed in equation 3) with μ_γ, being the mean of this normal distribution, and Ω_γ, being the variance of this normal distribution. This can also be written as $\gamma \sim N(\mu_\gamma, \Omega_\gamma)$.

When any prior information on the exact value of this beforementioned direct effect, where the instrument influences the outcome variable, is not available, the method described in Conley et al. (2012) can be used. They argue that the third assumption of IVs does not always hold. Thus they compute the value of direct effect that nullifies the third assumption can be used.

The method they described can be graphed, as a way to easily interpret these results. For this, Clarke & Matta (2018) offer a solution. They discuss the assumption $\gamma \sim N(\delta, \delta^2)$, which will be made use of when estimating and graphing this plausibly exogenous method. Looking at the way this formula is implemented, it portrays $\mu_\gamma = \delta$, and $\Omega_\gamma = \delta^2$, when graphing these non-zero values of γ.

4. Results

First, the results of the OLS estimation will be displayed. For both variables a model will be estimated. The first model using a strong indicator of depressive disorder as independent variable, and in the second model, the obtained CES-D scores. Both on the independent variable of OOP. These results will be interpreted and discussed later on, with the implications of these results being described.

Table 2. OLS analysis results

	(1) Out of pocket medical expenditures	(2) Out of pocket medical expenditures
Strong indication depressive disorder	1239.459*** (99.684)	
CES-D score		313.590*** (21.756)
Childhood SES-score	146.351*** (41.780)	176.608*** (41.964)
Covariates	Y	Y
PC	Y	Y
Waves	Y	Y
Observations	95309	95282
R^2	0.018	0.019

*Coefficients estimated in USD. Individual clustered standard error is reported between brackets. The entire table is reported in the appendix in table A1. With significance being displayed as: * p < 0.05, ** p < 0.01, *** p < 0.005.*

Interpreting these results, it is observable that both a strong indication of a depressive disorder, and CES-D scores have a positive association with OOP medical expenditures. Yet it is important to note that this result can be plagued with bias, due to omitted variables, as described before.

Second, the results of the regular MR analysis will be displayed. This will include both the regression analyses of the PGS of depression on a strong indication of a depressive disorder, and the PGS of neuroticism on CES-D score. *Table 3. MR analysis results*

	(1) Out of pocket medical expenditures	(2) Out of pocket medical expenditures
Strong indication depressive disorder	1679.711 (1448.137)	
CES-D score		754.427*** (240.772)
Childhood SES-score	170.891 (88.475)	316.209*** (88.793)
Covariates	Y	Y
PC	Y	Y
Waves	Y	Y
Observations	95309	95282
R^2	0.017	0.009

*Coefficients estimated in USD. Individual clustered standard error is reported between brackets. The first stage results are reported in the appendix in table A2.1 The entire table is reported in the appendix in table A2.2. With significance being displayed as: * p < 0.05, ** p < 0.01, *** p < 0.005.*

The results obtained in table 3. are obtained via the aforementioned MR analysis method, where the first stage obtained a F-score of 31.72 for the regression run with the strong indication of a

depressive disorder, and a F-score of 41.10 for the regression run with the CES-D scores. Since both of these F-scores were great or equal than ten, it means that the first assumption of MR is satisfied, and thus there is a strong connection between the PGS and resulting characteristic. What stands out from these results, is the positive and significant effect of the CES-D scores on OOP, while a strong indication of a depressive disorder does not show a significant effect. Studying these results to the OLS results in table 2, it stands out that the coefficient of the CES-D score on OOP expenditures is around 2,4 times as big in the MR estimation, compared to the OLS results. This difference could come from the difference in estimation. while OLS estimates the effect over the entire group of individuals, it is important to note that MR will only estimate this effect for the group who got a positive CES-D score because of their genetic risk, which will be called the compliers, as a result of their PGS, the so called local average treatment effect (LATE) (Frölich, 2007).

Next to this it is also important to note the difference in confidence intervals. When examining the 95% confidence interval for the OLS result, it reaches a confidence interval of [270.945, 356.236], and the interval for the MR result reaches [282.523, 1226.332]. From these intervals, one could observe that the MR results are not statistically significantly higher, on the 5% level, compared to the OLS results.

When diving deeper into the PMRM results, first the results of the pleiotropic effect will be named in table 4. This analysis is an OLS regression of the PGS on the OOP outcome, for a subgroup where the first stage result is zero. In this case, that would contain the people who do not have a strong indication for a depressive disorder, or a positive CES-D score. It is important to note that it is possible for these individuals to have a high PGS, while not demonstrating any symptoms.

Table 4. Pleiotropic Effects

	(1) Out of pocket medical expenditures	(2) Out of pocket medical expenditures
PGS Depression	20.693	
	(31.421)	
PGS Neuroticism		73.583
		(42.879)
Covariates	Y	Y
PC	Y	Y
Observations	78803	48323
R^2	0.003	0.003

*Coefficients estimated in USD. Individual clustered standard error is reported between brackets. The entire table is reported in the appendix table 3. With significance being displayed as: * p < 0.05, ** p < 0.01, *** p < 0.005.*

Observing the results for the pleiotropic effects, it is apparent that for both the PGS for depression and the PGS for neuroticism, on OOP, do not have a statistical significant effect for those individuals who did not report to have a strong indication of depression. Besides this, it is important to note that these results are for the groups not suffering from depression in the first analysis, and those with a CES-D score of 0 in the second analysis. This should demonstrate that in both groups, the exclusion restriction could be satisfied, since no statistical significant effect has been found. Due to the limiting factor that not everyone with a high PGS does demonstrate any sign of depressive symptoms, thus being selected in this subgroup, could influence the results. This could mean that people with high PGS for depression, while showing no signs of symptoms have been selected in this group in which people do not have depression. This could mean a biased result is present.

Finally, the results of the PMRM will be provided. This while keeping in mind the second assumption, being that any pleiotropic effect for the subgroup where the first stage result is 0, will be the same of the other subgroup which were not included in the first. This will estimate the results found in table 5 below.

Table 5. PRMR analysis results

	(1) Out of pocket medical expenditures	(2) Out of pocket medical expenditures
Strong indication depressive disorder	560.768 (1848.273)	
CES-D score		273.731 (329.695)
Covariates	Y	Y
PC	Y	Y

*Coefficients estimated in USD. Individual clustered standard error is reported between brackets. The entire table is reported in the appendix in table A4. With significance being displayed as: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$.*

What stands out from the results estimated above is that both results are not significant, as well a lower coefficient found compared to the results purely based on the MR analysis. One important thing to note that it was not possible to include correction for the different time periods. And however not displayed, childhood SES score was taken into account with these estimates.

The next method is the plausibly exogenous local-to-zero approach, which will be displayed graphically. In Figure 1, the results will be displayed for OOP with a strong indication of depressive disorder.

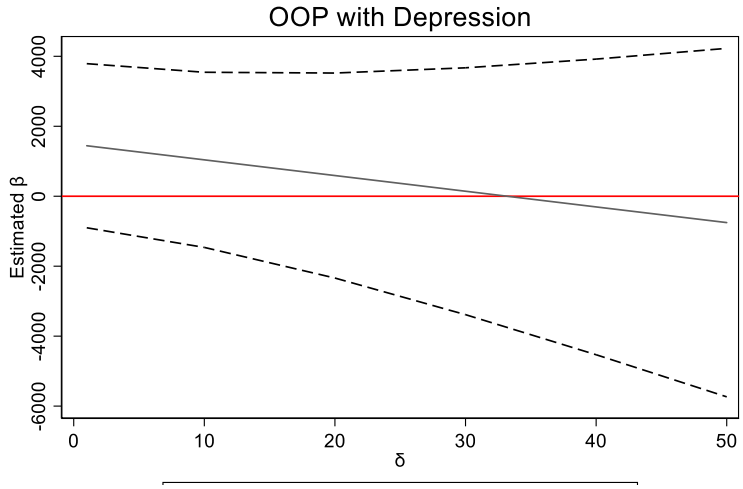

Figure 1: Plausibly Exogenous estimation for a strong indication of depression on OOP

In Figure 2, the results will be displayed for OOP for a person with a CES-D score.

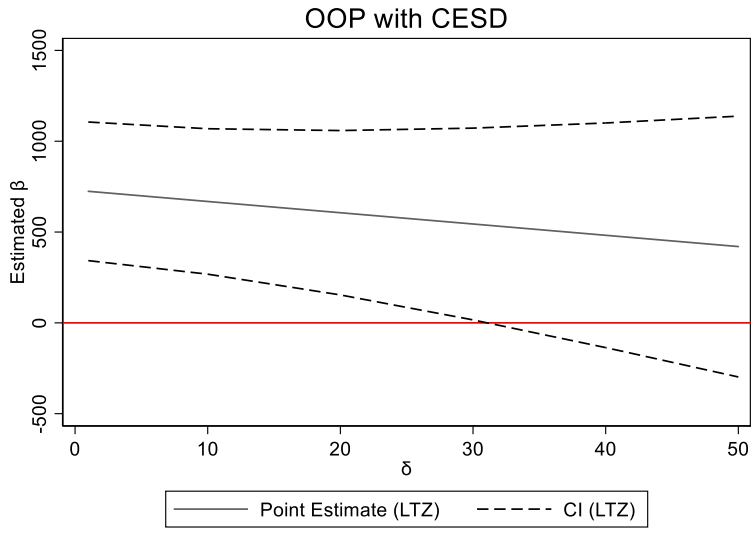

Figure 2: Plausibly Exogenous estimation for a strong indication of depression on OOP

Interpreting both these figures, looking at Figure 1, it is visible that the results are not statistically significant for any value of the estimated β. This is demonstrated by the horizontal 0-line being inside the 95% confidence interval, thus meaning that the results found are not statistically significant. This was expected, when looking back at the MR results for the effect of a strong indication of depression on OOP. How one could interpret the result Figure 2, would be looking at the horizontal 0-line, and where this line crosses the 95% confidence interval (CI). It is visible that this horizontal line crosses the 95% confidence interval line at delta (δ) = 30. As mentioned before, this delta is assumably the direct effect of the instrument on the outcome. Thus with delta being at 30, one could interpret this that as when the PGS of neuroticism has a direct effect on the outcome of out of pocket expenditures, with a coefficient of 30, then the result would be statistically insignificant. This thus demonstrates that while the result found is statistically significant at no exogeneity between the instrument and the outcome. Yet if one were to argue about the exogeneity of the instrument, and could make a valid case of this not holding, some small relationship between the instrument and the outcome variable, with the coefficient staying below that 30 threshold, would still allow for the effect found to be statistically significant. Only once the relationship between the instrument and the outcome variable would have a bigger effect than when the horizontal line crosses the confidence interval, it would no longer be statistically significant.

A thought to be had is how likely it is that the direct effect of the PGS of neuroticism is equal or bigger than 30, on the OOP medical expenditures. Looking at the results in table 4, of the pleiotropic effects of this PGS on these expenditures, it can be expected that this threshold of

30 is crossed. However the results found in table 4 were not statistically significant, a direct effect of this size is not out of the question. A study by Cuijpers et al. (2010) looked at the association between neuroticism and the economic costs. One important note with this paper is that it is not establishing a causal effect, plus it looked at the entirety of healthcare costs and non-healthcare costs. This study was also done with a sample from the Netherlands. They found an association ranging from around 5000 USD to 9000 USD, depending on the intensity of neuroticism present in the individual. When taking into account these results, it is important to note that these costs are not only out of pocket expenses, but the entire economic cost associated with neuroticism. This still could make quite a strong case that this direct effect of the PGS of neuroticism on OOP medical expenditures can be higher than this threshold of 30.

Discussion

The first thing worthy of note is the outcome of the analysis which estimated the pleiotropy effect. Seeing as these results should not be significant, it could be argued that there is no noteworthy pleiotropy inducing bias in the earlier found MR analysis results, thus one could argue that the third assumption of the MR relationship, that there should be no direct connection between the instrument and the outcome variable. Thus with this assumption being satisfied, one could argue the MR results are already unbiased. This opposes the results found from Gale et al. (2016), who did demonstrate this pleiotropy being present between neuroticism and other medical conditions. This also ties in to the result found with the plausibly exogenous method, where the PGS of neuroticism should have a smaller effect of 30 on the outcome to be statistically significant. In this paper they described health states, like anorexia nervosa, or coronary artery disease, which could have an effect on medical expenditures. Which could impose a coefficient bigger than 30 being obtained for this direct effect, thus the assumption being rejected.

Accepting this assumption of there being no connection between the PGS as instrument, and the OOP medical expenditures as outcome variable, based on the results found in the pleiotropy analysis, as well as the other assumptions for MR being satisfied, one could interpret the effect found of CES-D score on OOP expenditures as a statistically significant effect of 754.43 USD increase in OOP per positive response on the CES-D survey. On the other hand, a strong indication of a depressive disorder does show a positive effect of 1679.711 USD increase in OOP, but this estimated effect is not statistically significant.

Another observation to be made is the difference between the OLS and MR results. Where a strong indication of depressive disorder was estimated to be a statistical significant coefficient in the OLS model, while this is no longer the case when looking at the MR results. This while the variant where the CES-D score is being looked at, stays significant both in the OLS and the MR models. It is important to note that there is quite a difference between the size of the estimators. With an effect of 313.59 USD on OOP medical expenses in the OLS model and a 754.43 USD effect on OOP medical expenses in the MR model. Where the MR model is more than double the expected OOP expenses, compared to the ones obtained from the OLS model. Yet, as mentioned earlier, this doubling of the size of the effects in the MR model is not statistically significantly higher than the OLS.

To put these findings into a concrete example, the average out of pocket healthcare costs over the entire sample were 3016.04 USD. Now if one of these average people would obtain more depressive symptoms, and thus expect an increase of these OOP costs of about 754.43 USD, it would be around a 25% increase of OOP cost for this person, ceteris paribus. Comparing this to the aforementioned paper by Unützer et al. (2009), where they found an increase in 68% in total medical expenditures. This includes all types of care, and not just mental care, as well as insured and uninsured costs. A part of this difference can be explained by the difference in outcome variables looked at. In their paper, the focus was on total healthcare costs, compared to just out of pocket expenditures. Another part of this difference can also be explained by the estimation method. While in their research, they tried to establish an association between depression and healthcare costs, but might have overestimated the effects of depression on healthcare costs due to unobserved variables they did not take into account.

The found results could lead to the argument that, even though a strong indication of depression does not have a significant effect on OOP medical expenses, while observing the results from the MR model, different aspects in life measured by the CES-D questionnaire, does have an effect on OOP medical expenses. This while taking into account the CES-D questionnaire evaluates multiple facets of life. Thus it could mean that while someone might not be depressed, they are also not satisfied with their life, which could lead to the development of bad habits, deteriorating their health, and thus increasing their expected medical costs. This would be in-line with findings of Swendsen & Merikangas (2000), where they also found that substance abuse could be prevalent, it was not a direct cause of depression, nor was depression developed due to substance abuse.

On the contrary, the first important weakness of this research is the availability of mental health data in the HRS dataset. With the data obtained being subjective to the respondent itself, thus the answering of the questions asked in the survey could be answered differently, while some two respondents might be in the same health state. This subjectivity makes it more difficult to verify these results. This also includes another weakness of this data, the CES-D 8 form used is only a limited and restricted variant compared to the full CES-D 20 questionnaire. This could lead to more or less people being diagnosed with a strong indication for a depressive disorder, thus creating some sort of measurement error, introducing bias into the results. This is very apparent when comparing the percentage of people with a strong indication for a depressive disorder, which was established at around 17%. Comparing this result to the findings of the study done by Weinberger et al. (2018), where only a prevalence of around 7% was found in their sample size, which suffered from a depressive episode in the year 2015.

Another weakness based on the dataset, is from the respondents being represented in the sample. Since the sample only consists of elderly, this could influence the data collected and represented in the HRS dataset. Next to this, only the OOP expenditures is being looked at, not the expenditures made and covered by insurance, this could also influence the coefficients found, with the actual costs of healthcare being higher than the found coefficients.

When evaluating these results, one could argue about the extent this research is externally valid. With the data being focussed inside the United States, as well as only containing seniors. Thus making it difficult to apply the found results for a country in Europe, especially with regards to different costs of mental healthcare, and the difference in insurance systems in place.

5. Conclusion

Concluding, to answer the question asked at the beginning: *What is the causal effect of common mental health outcomes, like a depressive disorder, on out of pocket healthcare expenditures?* Taking into account the PMRM method, which showed the assumptions of the MR being accepted, it can be said that mental health outcomes do have an effect on these expenses. Especially when looking at CES-D scores, which measure general mental wellbeing, and how higher this score, thus how lower this mental wellbeing, how higher the expected out of pocket payments. Thus it would support the hypothesis that a worse mental wellbeing leads to higher expenses. This definite answer can not be said about people who have a strong indication of depression. This result found was not statistically significant. Thus the hypothesis which stated that there was a relationship between a deterioration in mental health and an increase in out of

pocket medical expenses can be accepted. The stronger version of this hypothesis, assuming this same increase in out of pocket medical expenses for people having a strong indication of depression has to be rejected, since the results found were not statistically significant.

While the issue of pleiotropy was not established in this sample, this while Gale et al. (2016) did find evidence of this pleiotropy existing. This could be due to the before mentioned limitation of people with a high PGS not necessarily developing depressive symptoms.

The main strengths of this research is found in its method of analysis, by being able to estimate a causal effect of these mental health outcomes, on out of pocket healthcare costs. As well as diving deeper into the assumptions behind this method of analysis, by exploring the plausible exogeneity of the instrument.

The main limitations of this research are apparent in the available data, with the CES-D 8 scores obtained being a reduced version of the complete CES-D 20 questionnaire, as well as only accounting for out of pocket expenses, and not those which people were insured for. This could lead to results distorted from reality. Even thought this will have resulted the estimated outcomes, it does establish a solid foundation for future research to be build upon. Especially if in future research more detailed dataset can be obtained, with those CES-D 20 scores, as well as the entirety of healthcare expenses.

This research would make a case to work towards making sure to keep people their minds healthy. With this focus on having a healthy mind, less mental health issues might develop, which in turn could lead to other health issues developing. Seeing as a worse mental health leads to an increase in out of pocket medical expenses, with the focus on these costs being for the individual. It also demonstrates the importance of preventative mental healthcare. Especially taking into account that these out of pocket expenses are not all the costs made, it does not include the medical expenses any insurance would have to pay. Thus this active focus on a reduction in mental health issues, as described in the CES-D 8 questionnaire, could both reduce healthcare cost for the individual, as well as the insurance provided.

6. References

Bugliari, D., Carroll, J., Hayden, O., Hayes, J., Hurd, M., Karabatakis, A., Main, R., Marks, J., McCullough. C., Meijer, E., Moldoff, M., Pantoja, P., Rohwedder, S. & St.Clair, P. (2021). RAND HRS Longitudinal File 2018 (V1) Documentation. *RAND Center for the Study of Aging.*

Clarke, D., & Matta, B. (2018). Practical considerations for questionable IVs. *The Stata Journal*, *18*(3), 663-691.

Conley, T. G., Hansen, C. B., & Rossi, P. E. (2012). Plausibly exogenous. Review of Economics and Statistics, 94(1), 260-272.

Coombes, B. J., Ploner, A., Bergen, S. E., & Biernacka, J. M. (2020). A principal component approach to improve association testing with polygenic risk scores. *Genetic epidemiology*, *44*(7), 676-686.

Cuijpers, P., Smit, F., Penninx, B. W., de Graaf, R., ten Have, M., & Beekman, A. T. (2010). Economic costs of neuroticism: a population-based study. *Archives of general psychiatry*, *67*(10), 1086-1093.

Donisi, V., Jones, J., Pertile, R., Salazzari, D., Grigoletti, L., Tansella, M., & Amaddeo, F. (2011). The difficult task of predicting the costs of community-based mental health care. A comprehensive case register study. *Epidemiology and Psychiatric Sciences*, *20*(3), 245-256.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS genetics*, *9*(3), e1003348.

Egede, L. E., Gebregziabher, M., Zhao, Y., Dismuke, C. E., Walker, R. J., Hunt, K. J., & Axon, R. N. (2014). Impact of mental health visits on healthcare cost in patients with diabetes and comorbid mental health disorders. *PLoS One*, *9*(8), e103804.

Eley, T. C., Sugden, K., Corsico, A., Gregory, A. M., Sham, P., McGuffin, P., ... & Craig, I. W. (2004). Gene–environment interaction analysis of serotonin system markers with adolescent depression. *Molecular psychiatry*, *9*(10), 908-915.

Emdin, C. A., Khera, A. V., & Kathiresan, S. (2017). Mendelian randomization. *Jama*, *318*(19), 1925-1926.

Frölich, M. (2007). Nonparametric IV estimation of local average treatment effects with covariates. *Journal of Econometrics*, *139*(1), 35-75.

Gale, C. R., Hagenaars, S. P., Davies, G., Hill, W. D., Liewald, D. C., Cullen, B., ... & Harris, S. E. (2016). Pleiotropy between neuroticism and physical and mental health: findings from 108 038 men and women in UK Biobank. *Translational psychiatry*, *6*(4), e791-e791.

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., ... & Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human Genetics*, *100*(4), 635-649.

Medicare. (2022). *Medicare & You, The official U.S. government Medicare handbook.* U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES.

Murray, G. K., Lin, T., Austin, J., McGrath, J. J., Hickie, I. B., & Wray, N. R. (2021). Could polygenic risk scores be useful in psychiatry?: A review. *JAMA psychiatry*, *78*(2), 210-219.

Paaby, A. B., & Rockman, M. V. (2013). The many faces of pleiotropy. *Trends in genetics*, *29*(2), 66-73.

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.

Rehm, J. (2011). The risks associated with alcohol use and alcoholism. *Alcohol Research & Health*, *34*(2), 135.

Richardson, T., Elliott, P., Roberts, R., & Jansen, M. (2017). A longitudinal study of financial difficulties and mental health in a national sample of British undergraduate students. *Community mental health journal*, *53*(3), 344-352.

Roberts, R., Golding, J., Towell, T., Reid, S., Woodford, S., Vetere, A., & Weinreb, I. (2000). Mental and physical health in students: the role of economic circumstances. *British Journal of Health Psychology*, *5*(3), 289-297.

Satyanarayana, S., Enns, M. W., Cox, B. J., & Sareen, J. (2009). Prevalence and correlates of chronic depression in the Canadian community health survey: mental health and well-being. *The Canadian Journal of Psychiatry*, *54*(6), 389-398.

Savage, M. J., James, R., Magistro, D., Donaldson, J., Healy, L. C., Nevill, M., & Hennis, P. J. (2020). Mental health and movement behaviour during the COVID-19 pandemic in UK university students: Prospective cohort study. *Mental Health and Physical Activity*, *19*, 100357.

Swendsen, J. D., & Merikangas, K. R. (2000). The comorbidity of depression and substance use disorders. *Clinical psychology review*, *20*(2), 173-189.

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

Kozlov, E., Dong, X., Kelley, A. S., & Ankuda, C. K. (2020). The epidemiology of depressive symptoms in the last year of life. *Journal of the American Geriatrics Society*, *68*(2), 321-328.

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome medicine*, *12*(1), 1-11.

Lousdal, M. L. (2018). An introduction to instrumental variable assumptions, validation and estimation. *Emerging themes in epidemiology*, *15*(1), 1-7.

Unützer, J., Schoenbaum, M., Katon, W. J., Fan, M. Y., Pincus, H. A., Hogan, D., & Taylor, J. (2009). Healthcare costs associated with depression in medically ill fee-for-service Medicare participants. *Journal of the American Geriatrics Society*, *57*(3), 506-510.

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.

Vilagut, G., Forero, C. G., Barbaglia, G., & Alonso, J. (2016). Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): a systematic review with meta-analysis. *PloS one*, *11*(5), e0155431.

Weinberger, A. H., Gbedemah, M., Martinez, A. M., Nash, D., Galea, S., & Goodwin, R. D. (2018). Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. *Psychological medicine*, *48*(8), 1308-1315.

Wetterstrand, K.A. (2021). The Cost of Sequencing a Human Genome/ *National Human Genome Research Institute*. Accessed on the 16th of February 2022 via: https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost

HRS datasets:

Health and Retirement Study, (Polygenic Score Data (PGS)) 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, (2022).

Health and Retirement Study, (RAND HRS Longitudinal File 2018) 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, (2022).

7. Appendix

A1. Acronyms

MR – Mendelian Randomization

OOP – Out of Pocket

PC – Principle Components

PGS – Polygenic Score

PMRM – Pleiotropy Robust Mendelian Randomization

SES – Socio-Economics Status

A2. Tables

Table A1 OLS regression

	Out of pocket medical costs	**Out of pocket medical costs**
Strong indication depressive disorder	1239.459*** (99.684)	
CES-D score		313.590*** (21.756)
Covariates		
Gender	388.142*** (63.923)	356.360*** (63.749)
Birthyear	-53.41*** (12.018)	-53.980*** (11.973)
Birthyear * Gender	9.135 (7.526)	9.143 (7.505)

Birthyear Squared	0.900*** (0.323)	0.836** (0.323)
Principal Components		
PC1A	15609.63*** (4791.074)	15497.330*** (4771.085)
PC1B	6924.189* (3270.918)	6260.903 (3259.962)
PC1C	-2084.543 (3657.294)	-2163.074 (3644.192)
PC1D	-2178.965 (3702.742)	-1837.937 (3694.21)
PC1E	-13581.04*** (4728.4)	-12904.430** (4708.136)
PC6A	-5473.838 (3471.656)	-5537.281 (3460.536)
PC6B	-9042.73** (3423.012)	-9178.194** (3411.844)
PC6C	1937.205 (3551.062)	1715.558 (3538.524)
PC6D	3743.487 (4065.306)	3663.372 (4055.817)

PC6E	-7445.135*	-7377.05*
	(3478.417)	(3474.142)
Wave		
3	1951.935***	1940.678***
	(115.503)	(115.682)
4	1990.319***	1944.821***
	(117.023)	(117.466)
5	2246.986***	2203.127***
	(123.712)	(124.074)
6	3269.449***	3233.999***
	(144.628)	(145.005)
7	4160.231***	4129.81***
	(172.276)	(172.398)
8	3431.094***	3384.475***
	(132.128)	(132.460)
9	3294.698***	3249.267***
	(133.325)	(133.689)
10	3873.42***	3836.003***
	(140.689)	(141.065)
11	3916.039***	3874.822***
	(156.623)	(156.779)
12	3764.198***	3728.059***
	(162.83)	(163.072)
13	4287.158***	4248.878***
	(241.843)	(242.213)
14	4056.277***	4017.887***
	(221.893)	(221.561)
Childhood SES	146.352***	176.608***
	(41.78)	(41.965)

Constant	-1279.686*** (185.957)	-1358.109*** (186.228)
Observations	95309	95282
R^2	0.017	0.019

*Coefficients estimated in USD. Individual clustered standard error is reported between brackets. With significance being displayed as: * p < 0.05, ** p < 0.01, *** p < 0.005.*

Table A2.1 First Stage MR

	Strong indication of depression	**CES-D**
PGS Depression	0.022***	
	(0.003)	
PGS Neuroticism		0.162***
		(0.016)
Covariates		
Gender	0.062***	0.348***
	(0.005)	(0.024)
Birthyear	0.002***	0.012***
	(0.001)	(0.004)
Birthyear * Gender	-0.001*	-0.004
	(0.000)	(0.002)
Birthyear Squared	0.000***	0.001***
	(0.000)	(0.000)
Principal Components		
PC1A	0.162	0.307
	(0.283)	(1.511)
PC1B	0.281	4.294***
	(0.270)	(1.381)
PC1C	-0.66*	-9.882***
	(0.268)	(1.71)
PC1D	-0.353	-2.182
	(0.231)	(1.214)
PC1E	-1.509***	-8.091***
	(0.284)	(1.518)
PC6A	-0.009	0.011
	(0.254)	(1.331)

PC6B	-0.336	-0.754
	(0.256)	(1.336)
PC6C	0.239	1.692
	(0.261)	(1.384)
PC6D	0.378	1.658
	(0.258)	(1.361)
PC6E	-0.628*	-3.227*
	(0.259)	(1.360)
Wave		
3	0.032***	0.160***
	(0.010)	(0.042)
4	0.073***	0.431***
	(0.010)	(0.043)
5	0.076***	0.435***
	(0.01)	(0.043)
6	0.084***	0.445***
	(0.010)	(0.043)
7	0.076***	0.403***
	(0.010)	(0.043)
8	0.092***	0.509***
	(0.010)	(0.043)
9	0.083***	0.47***
	(0.010)	(0.043)
10	0.083***	0.447***
	(0.010)	(0.044)
11	0.089***	0.487***
	(0.010)	(0.044)
12	0.088***	0.461***
	(0.010)	(0.045)
13	0.077***	0.426***
	(0.011)	(0.046)

14	0.081***	0.440***
	(0.011)	(0.047)
Childhood SES	-0.054***	-0.307***
	(0.003)	(0.015)
Constant	0.003	0.258***
	(0.012)	(0.057)
Observations	95309	95282
R^2	0.030	0.044

*Coefficients estimated in USD. Individual clustered standard error is reported between brackets. With significance being displayed as: * p < 0.05, ** p < 0.01, *** p < 0.005.*

Table A2.2 Second stage MR-results

	Out of pocket medical costs	**Out of pocket medical costs**
Strong indication depressive disorder	1679.711	
	(1448.137)	
CES-D score		754.428***
		(240.772)
Covariates		
Gender	361.100***	204.415*
	(110.271)	(99.291)
Birthyear	-54.582***	-59.575***
	(12.379)	(12.152)
Birthyear * Gender	9.620	11.164
	(7.654)	(7.510)
Birthyear Squared	0.860*	0.589
	(0.350)	(0.343)
Principal Components		
PC1A	15504.050***	14837.210***
	(4825.199)	(4827.878)
PC1B	6789.596*	4796.602
	(3301.86)	(3319.470)
PC1C	-2092.316	-2304.551
	(3649.611)	(3661.963)
PC1D	-1961.728	-544.816
	(3783.676)	(3779.939)
PC1E	-12858.930*	-9080.939
	(5370.445)	(5371.340)
PC6A	-5483.560	-5665.762
	(3467.714)	(3488.475)

PC6B	-8915.282**	-8839.700**
	(3457.667)	(3440.838)
PC6C	1892.374	1210.557
	(3534.809)	(3583.183)
PC6D	3593.819	2931.380
	(4130.436)	(4117.093)
PC6E	-7306.242*	-6705.508
	(3494.998)	(3520.868)
Wave		
3	1938.035***	1869.924***
	(127.031)	(121.964)
4	1958.392***	1754.623***
	(161.904)	(153.709)
5	2213.701***	2010.994***
	(169.135)	(155.825)
6	3232.465***	3037.686***
	(199.498)	(184.416)
7	4127.023***	3951.941***
	(210.628)	(206.234)
8	3390.752***	3159.971***
	(195.769)	(177.897)
9	3258.507***	3041.961***
	(184.702)	(171.425)
10	3836.942***	3639.209***
	(192.863)	(174.855)
11	3876.890***	3660.73***
	(213.513)	(189.088)
12	3725.837***	3525.166***
	(215.188)	(192.733)
13	4253.356***	4061.682***
	(271.654)	(252.131)

14	4021.084***	3824.895***
	(256.863)	(229.811)
Childhood SES	170.891	316.209***
	(88.475)	(88.793)
Constant	-1281.445***	-1475.136***
	(185.709)	(205.415)
Observations	95309	95282
R^2	0.017	0.009

*Coefficients estimated in USD. Individual clustered standard error is reported between brackets. With significance being displayed as: * p < 0.05, ** p < 0.01, *** p < 0.005.*

Table A3 First Stage PMRM

	Strong indication of depression	**CES-D**
PGS Depression	20.693	
	(31.421)	
PGS Neuroticism		73.583
		(42.879)
Covariates		
Gender	355.636***	278.458***
	(61.820)	(71.825)
Birthyear	-21.258	-16.730
	(11.408)	(13.602)
Birthyear * Gender	2.907	0.165
	(7.426)	(9.623)
Birthyear Squared	0.616	0.637
	(0.317)	(0.485)
Principal Components		
PC1A	15878.060***	14011.750*
	(5469.883)	(6618.428)
PC1B	7262.320*	6758.390*
	(3085.041)	(3360.526)
PC1C	-1736.510	-9269.590
	(3746.837)	(5331.386)
PC1D	-2594.543	-4199.826
	(3725.156)	(3957.290)
PC1E	-14450.900**	-17856.720**
	(5386.529)	(6690.979)
PC6A	-5283.908	-5416.820
	(3428.933)	(3621.142)

PC6B	-7057.255*	-8468.225*
	(3401.288)	(4002.528)
PC6C	493.91	4045.106
	(3453.121)	(3506.044)
PC6D	2417.627	-1002.812
	(4051.516)	(4963.287)
PC6E	-7028.196*	-9697.311*
	(3470.638)	(3889.434)
Childhood SES	136.049***	96.785
	(42.174)	(49.735)
Constant	2127.634***	2061.045***
	(103.848)	(127.66)
Observations	78803	48323
R^2	0.003	0.003

*Coefficients estimated in USD. Individual clustered standard error is reported between brackets. With significance being displayed as: * p < 0.05, ** p < 0.01, *** p < 0.005.*

Table A4 Second Stage PMRM

	Out of pocket medical costs	**Out of pocket medical costs**
Strong indication depressive disorder	560.769	
	(1848.273)	
CES-D score		273.731
		(329.695)
Covariates		
Gender	384.639***	327.501**
	(123.463)	(124.428)
Birthyear	-26.560*	-28.580**
	(10.649)	(10.351)
Birthyear * Gender	6.181	6.665
	(5.702)	(5.575)
Birthyear Squared	0.818***	0.735**
	(0.256)	(0.268)
Principal Components		
PC1A	15948.580***	15246.450***
	(2818.142)	(2822.669)
PC1B	6948.520*	6714.441*
	(2835.617)	(3096.618)
PC1C	-2558.668	-6658.472
	(2951.375)	(3844.020)
PC1D	-1610.737	-857.503
	(2876.400)	(2895.648)
PC1E	-14900.37***	-13328.45***
	(4022.604)	(3923.841)

PC6A	-6093.245*	-6279.875*
	(2779.119)	(2789.100)
PC6B	-9141.792***	-9077.601***
	(2843.031)	(2806.639)
PC6C	2403.788	2240.556
	(2826.845)	(2856.471)
PC6D	3616.498	3279.628
	(2868.524)	(2845.316)
PC6E	-7848.157**	-8103.465**
	(2943.576)	(2940.028)
Childhood SES	119.834	178.722
	(105.705)	(106.274)
Constant	2189.178***	2034.966***
	(175.850)	(244.295)
Observations	95309	95282

*Coefficients estimated in USD. Individual clustered standard error is reported between brackets. With significance being displayed as: * p < 0.05, ** p < 0.01, *** p < 0.005.*