ERASMUS UNIVERSITEIT ROTTERDAM

ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

Master Thesis in Health Economics

Gene-Environment Interplay in Shaping Health Outcomes under the COVID-19 Pandemic

ABSTRACT

This thesis investigates the extent to which the interplay of individual genetic predisposition to BMI – measured by polygenic risk scores - and the COVID-19 pandemic shaped health outcomes in U.S. elderly individuals. Exposure to the pandemic is measured by knowing someone in your social cycle who passed from COVID-19. Utilising the recently released COVID data from the Health and Retirement Study (HRS), this study is among the first to investigate how the pandemic affected associations between genetics and health-related measures. Using panel data regression analysis combined with risk stratification, the findings indicate that exposure to the pandemic as a negative shock is associated with poorer self-rated health and increased BMI scores. Exposure to the environment results in a decrease of 0.126 standard deviations in self-rated health scores. The analysis of G×E interaction revealed that individuals with lower genetic predisposition to BMI report 0.993 kg/m^2 lower BMI when exposed to the pandemic as compared to those who did not. These findings provide insight into one of the mechanisms by which individual characteristics can shape health outcomes.

"This analysis uses Early Release data from the Health and Retirement Study, (2022 HRS Core), sponsored by the National Institute on Aging (grant number NIA U01AG009740) and conducted by the University of Michigan. These data have not been cleaned and may contain errors that will be corrected in the Final Public Release version of the dataset."

Name student: K.R. van der Ploeg

Student ID + mail: 704820, 704820kp@eur.nl

Supervisor: dr. D. Muslimova

Second assessor: dr. P.L.H. Bakx

July 24, 2024

SUMMARY THESIS

Background. The COVID-19 pandemic affected healthcare systems at the global scale. Existing literature identified factors and individual characteristics that help explain observed health conditions under the pandemic. Genetics are expected to at least partly explain the observed heterogeneities. Specific genes that directly contribute to the manifestation of the COVID-19 disease have been identified. However, how exposure to the pandemic influences other health outcomes differently by underlying risk for complex diseases remains to be determined. As of today, there is no similar empirical work that studies how genetic predispositions to BMI as measured by polygenic risk scores, interact with the COVID-environment to establish health outcomes.

Main objective. This thesis aims to investigate the gene-environment association between the genetic risk for increased BMI and subjective- and objective measures of health using the COVID-19 pandemic as an environmental shock. More specifically, the focus is on differential objective- and subjective health in the U.S. elderly (50+) population, where objective measures are the presence of lung disease and BMI, and the subjective measure is self-rated health scores.

Research question and hypotheses. The research question of this thesis is ‘’Are the effects of COVID-19 on objective- and subjective health measures of the elderly U.S. population different by underlying genetic predispositions?’’. Based on existing GxE theoretical frameworks, I hypothesise that genetics do play a role in explaining the inter-individual heterogeneity in health outcomes. I hypothesise that genetic propensity for BMI moderates the effect of the pandemic on health outcomes, as such that individuals with greater genetic risk are more susceptible to worse health outcomes when exposed to the COVID environment – i.e. higher likelihood of having lung disease, lower self-perceived health scores and/or higher BMI.

Methodology. I utilize different datasets from the Health and Retirement Study to construct a panel data set that covers the survey years 2018-2022. I apply panel data regression analysis to estimate the extent to which genetic predispositions, the environment and their interaction term (as a measure for gene-environment interactions) affect the selected health outcomes. Genetic predisposition is measured by polygenic risk scores, the environment indicates knowing someone who passed from COVID-19. Lung disease is a binary indicator of lung health status. BMI is measured as absolute BMI points, whereas self-reported health is an increasing 1-5 point scale, which is standardized for the ease of interpretation

Results. The polygenic risk score for BMI is predictive of actual BMI, as higher genetic risk corresponds to higher BMI scores in this sample. A similar trend is observed for self-rated health, where greater genetic predisposition to BMI is associated with lower reported health. The association between genetic predisposition for BMI and lung health is ambiguous. The main regression analysis uncovers that the environment has a significant negative effect on self-perceived health and BMI. I find evidence for the association of G×E effects on self-rated health and BMI. Additionally, sample stratification according to genetic risk uncovers that there is heterogeneity in health outcomes between risk groups.

Conclusion and policy implications. This thesis finds that exposure to the COVID-19 pandemic, in terms of losing someone to the disease, has adversely affected health outcomes of U.S. elderly individuals. Main policy recommendations could aim to identify susceptible elderly individuals who are exposed to the environment and provide community-based support mechanisms in terms of mental and/or physical support.

Table of Contents

1. Introduction

The COVID-19 pandemic, which is caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), emerged early 2020 as one of the most pressing public health crises in recent history. Since the initial outbreak, the virus rapidly spread across continents, challenging healthcare systems and local economies worldwide, leading to rising healthcare costs (Figure 1). An important manifestation in the course of infection is the observed interindividual variability in outcomes. The disease presents multiple phenotypes with differing patterns of symptoms; ranging from asymptomatic or mild symptoms to more severe disease patterns such as respiratory distress, multi-organ complications or death.

There are a broad range of factors (e.g. individual characteristics, environmental influences, social- and economic factors) that determine or influence health outcomes after viral infection (Semenza et al., 2016). One important host risk factor for developing more severe disease complications is advanced age (Acosta and Singer, 2020). This leaves the elderly population at an increased risk for poorer clinical outcomes of COVID-19 upon infection. The elderly are considered particularly vulnerable due to a range of age-related risk factors, including a weakened immune system and the presence of comorbidities. From Figure 1 it is evident that, among people with COVID-19-related events, there are higher healthcare costs associated with the elderly population. However, such characteristics - as increased age - do not fully explain the observed heterogeneity in disease response between individuals, suggesting the involvement of additional factors.

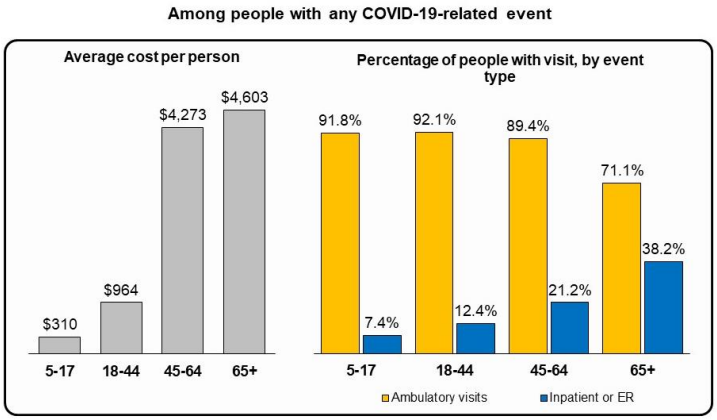

Figure 1: Healthcare costs per age category U.S. Source: Agency for Healthcare Research and Quality, 2020

Another factor hypothesized to influence disease outcomes, is varying underlying genetic predispositions (Kerner and Quintana-Murci, 2022). Increasing evidence suggests the existence of genetic variants that determine clinical outcomes following COVID-19 infection (Westerman et al., 2022). It is presumed that individual genetic variants either directly contribute to disease pathogenesis or interact with other (environmental) factors to determine disease manifestation (Virolainen et al., 2023). Gene-environment studies explore the relationship between genetic

factors and environmental influences in the progression of certain (health-related) outcomes. When the effect of environmental exposure on outcomes is different conditional on an individual's genetic makeup, this is referred to as gene-environment interactions (G×E). Previous G×E research provides evidence that certain environmental shocks can trigger underlying genetic predispositions.[1] However, little research has explored how genetic predispositions for complex health outcomes play a role in explaining later life health outcomes under the pandemic. Therefore, this thesis contributes to the growing literature on gene-environment studies.

The current thesis investigates how the recent COVID-19 shock triggered a subset of health outcomes of individuals. The health outcomes are measured by self-reported health scores, BMI and the presence of lung disease. I examine how these outcomes differ based on underlying genetic predispositions, aiming to gain novel insight into genetic variability as a mechanism for impacting health outcomes under a public health crisis. The empirical setting utilizes longitudinal survey data from the Health and Retirement Study (HRS) - a biannual study focused on the elderly (50+) U.S. population.[2] The HRS provides comprehensive data on various measures associated with ageing and moreover includes genetic data of its respondents.

To measure the variation in genetic predispositions, I use aggregated genetic data, so-called 'Polygenic Risk Scores' (PGS) to increased BMI. From literature, it is evident that increased BMI is a predictor for obesity, which is adversely associated with poor health outcomes (Tommerup et al., 2021; Djalalinia et al., 2015). An advantage of utilizing genetic risk scores over the presence of elevated BMI is that genetic predispositions are fixed from birth, and cannot be influenced by the environment thereafter (Lewis and Vassos, 2020). Existing literature has identified specific genes related to COVID-19 susceptibility; however, little is known about whether health outcomes differ by exposure to the pandemic for individuals with varying health-related polygenic risk scores (Pairo-Castineira et al., 2023; Ellinghaus et al., 2020). Therefore, this thesis adds an important dimension to the current epidemic literature on the past pandemic.

[1] For example, several studies find associations between childhood- or family history and genetic risk to explain smoking behaviour in later life (Bierut et al., 2023; Foo et al., 2022).

[2] As a condition of use, I note that the HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

The COVID-19 pandemic is a large environmental shock, which could trigger various underlying genetic susceptibilities. The pandemic is quantified by a binary indicator for knowing someone who passed from the COVID-19 disease, which aims to capture grief effects during the pandemic resulting from COVID-19-related deaths. Grief may adversely affect health outcomes through anxiety-related disorder or adverse coping mechanisms, such as binge eating, and increased alcohol and/or cigarette consumption (Prigerson et al., 1997). The impact of pandemic-related grief is expected to trigger responses beyond grief due to other cause of death (Stahl et al., 2023; Eisma et al., 2020). Compared to other COVID-19 exposures in the HRS which are endogenous[3], pandemic-related deaths are at least quasi-exogenous and independent of BMI PGS.

The aim of the thesis is threefold: 1) to explore how the COVID-19 environment affected subjective- and objective health outcomes of the elderly (50+) population in the U.S., 2) to investigate how outcomes vary across underlying genetic predispositions for complex health outcomes and 3) to examine the role of selection based on genetic predisposition (gene-environment correlation), into the COVID-19 environment as a mechanism for the potentially observed heterogeneities in Objective 2. Achieving the above objectives will help answer the main research question: "Are the effects of COVID-19 on subjective- and objective health outcomes in the elderly U.S. population different by underlying genetic predispositions?"

The results of this thesis can be summarized as follows. Genetic predisposition and exposure to the pandemic separately are predictive for self-rated health scores and established BMI. This thesis finds that there is limited evidence for gene-environment correlation as a mechanism for self-selection into the COVID-19 environment. Using panel data regression on a final sample of 3,517 elderly U.S. individuals, I find that the pandemic has significantly contributed to worse health outcomes. Individuals who have experienced the loss of a close friend or family member due to the COVID-19 shock report lower self-reported health scores and higher BMI. The G×E interaction term suggests that individuals responded differently in terms of self-reported health scores and BMI, with higher BMI amongst individuals with a lower genetic predisposition to BMI and lower self-rated health for individuals with a higher genetic predisposition.

[3] Endogeneity could partly be resolved by utilizing state data to construct exogenous measures for COVID-19 vaccination statistics. However, I do not have access to this data. Therefore, the current measure for the environment – knowing someone who passed from COVID-19 – is the most suitable.

From a societal perspective, it is important to recognize that the COVID-19 pandemic has put a large pressure on healthcare systems in many countries. The identification of individuals or groups with an increased vulnerability should be a priority for health systems for the optimal distribution of resources. A thorough understanding of the individual characteristics that underly such patterns can provide information for population risk stratification and protective measures for vulnerable groups in society. Since genetics are fixed at birth, the identification of vulnerable groups of individuals that would suffer more in terms of worsening health than an average person as a result of an environment shock, may in the future contribute to decreasing health inequalities. This might allow for better therapeutic decisions in targeted patient care when faced with a public health crisis.

The remainder of the paper will be structured as follows: Section 2 reviews the relevant theoretical background and introduces the main hypothesis of this thesis, while Section 3 discusses existing empirical studies. Section 4 elaborates on the data used for testing the hypothesis and Section 5 reports the methodology and introduces the econometric model. Section 6 reports the results and Section 7 concludes. The thesis is ended with Section 8 which discusses results, indicates relevant limitations and policy implications and suggests some avenues for further research.

2. Theoretical background

The following section presents the theoretical background. In the first part, I introduce key concepts and definitions in the field of gene-environment studies. The second part summarizes background information on molecular genetic studies that form the basis for the genetic scores used in the main analysis. Lastly, the third section elaborates on existing theoretical frameworks in the G×E literature, which links genetics and environmental factors to various outcomes, and I describe how they inform the conceptual framework.

2.1. Definitions

The molecular processes that provide the basis for human health and disease development are highly complex and only to a certain level well-understood. While some phenotypes[4] can be tracked to a single factor, most are due to a combination of various contributing factors. More specifically, it is the combination of genetic- and environmental factors that contribute to the expression of a specific trait (Biroli et al., 2022). Such interactions are called gene-environment interplay (G×E). G×E studies the effect of genetic predispositions for certain phenotypes that are either expressed or suppressed under particular environments. The environment can adopt different forms, but represent anything that is not captured by genetics: for example air pollution, social- and behavioural patterns, policy changes or some other measure of exposure or life event (Mills et al., 2020). G×E interplay is in the literature (Mills et al., 2020) described as the distinction between:

1. *Gene-Environment Interaction* (G×E): the phenotypical effect of exposure to an environmental shock varies for individuals with different underlying genotypes. For example, research has shown significant interactions between the genetic risk for major depressive disorder (MDD) combined with childhood trauma in predicting depression in later life (Mills et al., 2020).
2. *Gene-Environment Correlation* (rGE): describes how certain environments are more frequent among individuals with certain genotypes. This has implications for the interpretation of G×E. rGE may have a confounding or moderating effect, as the presence of gene-environment correlations might influence how G×E effects are manifested. For example, certain genetic traits may predispose certain individuals to

[4] A phenotype refers to the expression of certain trait (for example physical traits as eye colour), by an individual's genes (genotype) (Slavkin, 2014).

seek environmental conditions that stimulate the expression of these traits (e.g. individuals with predispositions for externalising behaviour may not comply with restricting policies and get infected more frequently). Failure to identify such relationships may mistakenly attribute observed outcomes solely to G×E interactions.

G × E studies can identify relationships between environmental exposures and genetic predispositions, however, not all individuals who are predisposed to a particular phenotype express a certain trait. For example, not all individuals who inherit genes associated with a greater risk of certain diseases, develop said disease. Studies in the G × E field aim to test whether the effect of genotype on phenotype varies across different environments, to gain more insight into the role of genetic and environmental influences.

2.2. Genetic data, Genome-Wide Association Studies and polygenic risk scores

The human genome. Deoxyribonucleic acid (DNA) is the molecule that carries hereditary material. Molecular genetics is the research area that studies the functioning and structure of DNA at the most fundamental level (Beauchamp et al., 2011). Human DNA has a total of 23 pairs of chromosomes: a maternal- and paternal copy.[5] DNA is a polymer with two complementary polynucleotide strands that coil together to form a double helix (Alberts et al., 2002). Each DNA strand is made up of four different types of subunits called nucleotides (Alberts et al., 2002).[6] The complementary strands are held together by opposing nucleotides that form base pairs through hydrogen bonds.[7] The human genome contains approximately 3 billion of such base pairs (Beauchamp et al., 2011). The nucleotide sequence and length determine genes, which hold the 'instructions' for the expression of a particular trait. In total, the entire human genome contains an estimated 19.000 protein-coding genes (Piovesan et al., 2019).

[5] 22 pairs of autosomal chromosomes and one pair of sex chromosomes.

[6] The four nucleotides of DNA are: adenine (A), thymine (T), guanine (G) and cytosine (C), (uracil (U) instead of T in RNA). DNA nucleotides are link with covalent bonds through alternating sugar- and phosphate groups to form the ''DNA backbone'' (Alberts et al., 2002).

[7] A double helix is the three-dimensional structure of DNA which arises from the structural and chemical features of the nucleotides on opposing strands. Hydrogen bonding are formed through dipole-dipole attraction between molecules (as opposed to covalent binding structure). The nucleotides (bases) reside on the inside of the helix-structure. In each case, A pairs with T, C pairs with G.

Chromosome pairs contain the same genes. These genes are often not identical but contain slight differences in the base pair present. The different variants of genes controlling for the same trait are called alleles (Taylor and Lewontin, 2017). An individual's complete set of genetic material is known as their genotype.

For the vast majority of locations across the genome (~99%), humans are identical (National Human Genome Research Institute, 2018). For the remaining locations, individuals differ in their genetic makeup. The simplest form of DNA variation are *single nucleotide polymorphisms* (SNPs). SNPs are single-point mutations where one DNA nucleotide is substituted for another. SNPs are thought to drive inter-individual genetic variation and account for the most common hereditary traits of, e.g. complex diseases, such as obesity, diabetes and psychiatric disorders (Shastry, 2009). The identification of SNPs linked to certain diseases has for this reason gained attention.

Genome-wide association studies (GWAS). SNPs identification aims to understand their role in gene functioning and affecting phenotypical outcomes (Pratt et al., 2014). Therefore, genetic association studies are increasingly used in research, as well as economics. Historically, a large number of existing research on environmental influences and genetics utilizes data on twin studies, as twins share a large fraction of their genetic material. (Friedman et al., 2021). The degree of genetic similarity between twins differs: identical twins (monozygotic) share nearly all SNPs in common, while this is approximately 50% for fraternal twins (dizygotic) (Friedman et al., 2021). Since identical twins share their genetic makeup, observed differences can be attributed to differing environmental influences. Classical twin studies use this difference (or similarity) in phenotypical outcomes between identical and fraternal twins to estimate the relative role of genetics (Hagenbeek et al., 2023). A trait that is strongly determined by genetics should result in more similar outcomes for individuals who share more of their genetic material.

GWAS emerged as a scientific method to scan the human genome to find statistically significant associations between genetic variants and phenotypes (Uffelmann et al., 2021). The expression of SNPs is known to affect the phenotype (Shastry, 2009). While most SNPs are silent, some are known to alter downstream gene expression. The identification of SNPs requires testing the genomes of numerous individuals with the same ancestry, but different phenotypes (Uffelmann et al., 2021).[8] As characteristics are often complex and involve different genetic variants,

[8] Genetic and phenotypical outcomes utilized in GWAS studies can be population-based and/or family-based designs. Population-based approach studies an unrelated, random sample the total population,

GWAS identifies many genetic variants that each modestly contribute to influence the expression of a trait. The weighted GWAS effect sizes sum to a score that predicts individual genetic risk for certain characteristics, called *polygenic scores* (PGS). Polygenic risk scores are a single value estimate for individual genetic risk to a certain trait (Choi et al., 2020). As a measure of strength, SNPs are weighted by effect size derived from GWAS studies. The HRS uses the following formula to calculate PGS (Ware et al., 2021):

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

Where i an individual (i =1,…,N), j is an SNP (j=1, …, J), W is the effect size of a certain SNP obtained via GWAS studies, G is the genotype (0,1 or 2; measured by the number of alleles present[9]) for an individual i at SNP j. Most G×E studies apply these scores from GWAS as an estimate for genetic risk (Mills et al., 2020). While GWAS studies are increasingly powerful in determining relations between genetics and outcomes. However, such studies fail to inform on the role of environmental influences on phenotypes or trait heritability.

2.3. Modelling G×E Interplay

Several models provide a theoretical framework to explain gene-environment interactions. The Diathesis stress model is the most suitable one conceptually to my research question.[10] This model was first described by Monroe and Simons (1991) and illustrated in Figure 2. According to this model, a genetic predisposition (diathesis) for a particular phenotype lies dormant, until triggered by an environmental shock. This shock is usually negative in nature. Under normal or positive circumstances, an individual with greater genetic risk resembles the resilient individual (Mills et al., 2020). The resilient individual is someone who will develop similarly under positive and negative environments. However, when faced with a

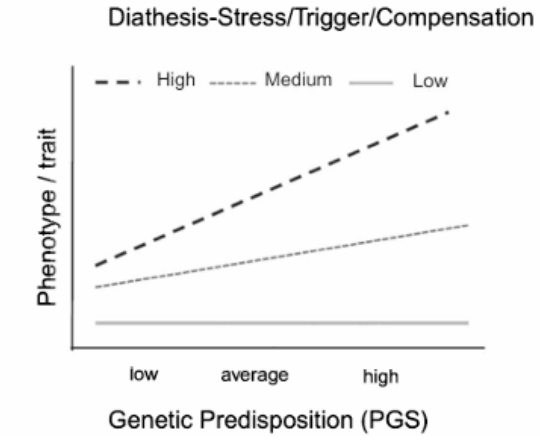

Figure 2: Diathesis stress model. The figure is adapted from Mills et al., (2020).

whereas and family-based structure considers first-degree relatives (Uffelmann et al., 2021; Veller and Coop., 2024).

[9] Individuals have two copies of the chromosomes. Therefore, the maximum number of present alleles equals 2. A allele is not present (0), present at one locus (1, heterozygous) or two loci (2, homozygotes).

[10] Other models include the social control or social push model; differential susceptibility model and the bioecological model (Mills et al., 2020). See Appendix A for a brief summary.

negative environment, an individual with greater genetic risk for a certain outcome is disproportionally affected compared to the resilient individual. This is shown by the steeply increasing dashed line at the top.

This model specifically states the presence of a negative environmental stressor, and therefore this is a suitable theoretical foundation to build the hypothesis of this thesis. In this thesis, the outcomes of interest are subjective- and objective measures of health status quantified by the presence of lung health (objective), absolute BMI (objective) and self-rated health scores (subjective), while COVID-19 is the environmental stressor. Objective measures are those measured objectively by for example a doctor, subjective measures are reported by individuals themselves. Therefore, according to the Diathesis-Stress theory, those with higher PGS to BMI are at greater risk for worse health outcomes when exposed to the pandemic.

2.4. Hypotheses

Following the Diathesis-Stress model in Section 2.3., I first hypothesize that elderly individuals with greater genetic predisposition for BMI (higher risk according to their genome), have worse health outcomes (either subjective and/ or objective outcomes[11]) under the COVID-19 pandemic, compared to individuals who did not experience the negative shock. [12] Moreover, I hypothesise that exposure to the environment and genetic predispositions have individual effects, as such that they both individually contribute to health outcomes. Worse health in general is a predictor of future hospitalization and mortality, especially in the elderly population (Zhao et al ., 2023). The genetic risk is captured by a measure that reflects different individual genetic susceptibility to increased BMI. Increased BMI is a risk factor for poorer clinical outcomes for many complex diseases and affects the risk for future hospitalization, in particular for the elderly (Ohno and Dzúrová, 2022). The empirical motivation for including this score is elaborated on in Section 3.

In the econometric model, an interaction term between the genetic score and an indicator for the environment reflects the presence and effect size of the gene-environment interaction (see Section 5.2). *Objective health measures* include 1) the presence of lung health and 2) absolute BMI scores. Worse health outcomes for these measures reflect the increased likelihood of the

[11] Objective health measures are those that can be measured and these are not dependent on individual experience. For example, they are diagnosed by a GP. Subjective health measures on the other hand are shaped by individual experience (Cleary, 1997).

[12] For a complete description and motivation for the (in)dependent variables and measure(s) to quantify the COVID-19 environment used in this study, see the empirical setting (Section 5).

presence of lung disease and an absolute higher BMI score. More specifically, this implies that I hypothesise that the COVID-19 pandemic has led to an increased likelihood of the presence of lung disease and/ or higher BMI for individuals with greater genetic predisposition to BMI. Such an effect is shown by a larger and positive interaction term (greater genetic risk and the individual experienced the negative environmental influence), to reflect worse health outcomes for high-risk individuals with the environmental stressor.

As a measure of *subjective health outcomes,* I used self-rated health scores. Worse indirect health outcomes are thus reflected by lower self-perceived health of an individual. On an increasing score scale (1 to 5), this is thus reflected by a lower score. Therefore, I hypothesise that as a consequence of a negative stressor, high-risk individuals will have lower self-rated health scores. This will be reflected by a negative gene-environment interaction term in the analysis.

Thus, I hypothesize that for the regression coefficients, the G×E interaction term is positive for the direct health measures, and negative for the indirect health measure.

3. Empirical background

The following section discusses the empirical evidence on the interplay between genetics and health outcomes. Previous empirical G×E studies identified a role for genetics in modifying lifestyle effects in developing obesity (Qi and Cho, 2012). Additionally, considering the pandemic, de Vries et al., (2021) demonstrated that general well-being under the pandemic varied, with more pronounced changes for certain subgroups. The study suggests these differences may be mediated by a genetic sensitivity to extreme changes. Nevertheless, similar empirical work linking genetic predisposition to BMI and the pandemic to general health measures is scarce. As such, the following section is split into three parts. The first part discusses empirical evidence on the impact of the COVID-19 pandemic in general on health outcomes to establish how the pandemic affected public health. The second part evaluates literature that studied the association between genetics and health outcomes in a non-pandemic setting. The last part lists existing research that identifies specific individual characteristics associated with higher risk for COVID-19.[13] These findings will help to build and motivate the conceptual framework and empirical model.

3.1. The COVID-19 pandemic and health outcomes

The most prevalent symptomatic disease manifestation of COVID-19 affects pulmonary functioning, of which *acute respiratory distress syndrome* (ARDS) is the most severe (Singh et al., 2024). Poor outcomes in patients recovering from COVID include long-term lung functioning impairment. Hooper et al. (2023) studied the effect of the pandemic on the number of days spent in poor health in a U.S. sample. The authors find an average increase of 0.14 days spend in poor health during the pandemic. This effect was even more pronounced for women, with an increase of 0.29 days on average.

Existing literature suggests that people assess their general health status more negatively during the pandemic. Lüdecke and van dem Knesebeck (2023) found that self-perceived health worsened during the COVID-19 pandemic for elderly individuals in Europe. These effects seem even stronger for individuals with lower income or education compared to the pre-pandemic era (Yun et al., 2022). The main arguments to support this observation do not only result from

[13] Understanding individual characteristics that increase the risk for poor clinical outcomes under COVID-19 allows for a more nuanced understanding of how the disease affects different populations. Insights from research on COVID-19 provides contextual relevance and can help generalize findings to other health conditions. This can help identify susceptible individuals or groups that are at higher risk for worse health outcomes in general.

fear of infection directly but also stem from other pandemic-related factors, including rising uncertainty and social isolation resulting from governmental restrictions (Oshio et al., 2022).

To curb the transmission of the virus, governments across the globe implemented policy responses aimed at restricting social contact. Such preventive measures aim to reduce infection- and mortality rates and decrease the pressure on national healthcare systems since the spread of the SARS-CoV-2 coronavirus is exponential if left unaddressed (Berger et al., 2020). Pivotal restrictions include lockdowns, social distancing and quarantine, and are collectively referred to as *non-pharmaceutical interventions* (NPIs), as they cover measures that delay disease transmission in the absence of vaccination programmes (Elbourne, 2021). Whilst such restrictive measures aim to protect the population from infection and decrease pressure on healthcare systems – and literature suggests they are effective in doing so (Ma et al., 2021) - they are also expected to have negative consequences for general health and subjective health. Reduced social contact and isolation due to NPIs seem to contribute to the psychological impact of the pandemic and contribute to lower self-rated health scores. Lloyed et al., (2023) found a higher burden of mental health issues during the pandemic. Hooper et al., (2023) establish substantial heterogeneity between subgroups in the rate of poor mental health. This effect seems more pronounced for the elderly population (Vo et al., 2021). Such NPIs also shape the pandemic as a negative environment. Therefore, the effect of COVID-19 on health outcomes might have varied depending on the pharmaceutical and non-pharmaceutical interventions.

3.2. Role of genetics on health outcomes

Previous empirical research suggests that the interplay between human genetics (nature) and environmental factors (nurture) plays an important role in the risk of expression of certain phenotypes (Meaney, 2010).[14] Identifying how genetics interact with the environment to shape health outcomes can help identify environments that mitigate or exacerbate genetic disadvantages (Barcellos et al., 2018). Genetic factors play - to different extents - a role in explaining observed interindividual differences in disease manifestation (Jackson et al., 2018).[15] Many complex diseases such as cardiovascular disease, cancers, diabetes and mental

[14] Genes affected by environmental stressors leading to disease outcomes have been identified in studies on smoking, air pollution and Alzheimer's disease (Rivera et al., 2019; Johansson et al., 2019; Dunn et al., 2019). Other studies establish a relation between smoking habits, genetic variation and family structure (Bierut et al., 2023; Foo et al., 2022).

[15] GWAS established individual genetic variations that directly affect COVID-19 infection risk and disease progression (Horowitz et al., 2022; Ishak et al., 2022)15. Niemi et al., (2022) also find that host genetics significantly influence the degree of COVID-19 severity. Examples of genes identified by GWAS, that are

health disorders are affected by both genetic factors and the environment (Nelson et al., 2006; Boardman et al., 2014). For example, Caspi et al., (2003) found that individuals with a specific genetic polymorphism are at higher risk of developing depressive disorder under stressful life events. This highlights how genetic predispositions can interact with environmental stressors. These findings suggest that genetic predispositions to certain diseases can contribute to worse disease outcomes, regardless of whether the phenotype is clinically present.

In general, it is a consensus that genetics play an important role in shaping BMI. Globally, BMI has been increasing in the past decades. Growing evidence suggests that this rise is largely driven by an individual's environment such as socioeconomic status, but genetics also play an important role (Jackson et al., 2020). In recent years, the connection between genetics in shaping BMI and obesity has been established using twin studies and, more recently, GWAS (Loos and Yeo., (2022); Jackson et al., 2020). The existing literature identified multiple genetic loci that are directly involved with realized BMI (Locke et al., 2015; Brandkvist et al., 2019). Tommerup et al., (2021) demonstrated that polygenic risk scores for BMI are associated with higher BMI and predict BMI scores in older adulthood. Additionally, GWAS established shared loci between genes predisposing to BMI as well as lung functioning (Zhu et al., 2021). While pulmonary malfunctions were long considered the results of environmental factors - such as smoking and air pollution - genetic studies have shown genes to significantly contribute as well (Sayers et al., 2024).

How genetics and environmental influences interplay in shaping subjective health remains more ambiguous. Empirical research has linked genetics to self-rated health (SRH) scores using twin studies (Romeis et al., 2000; Silventionen et al., 2007). The role of genes is supported by Franz et al., (2017), who found that genetics explain approximately 20-46% of individual differences in subjective health. More specifically, Harris et al., (2017) find that these genetic variants associated with SRH are related to physical disorders and common diseases. Moreover, other studies find higher SRH to be related to shared genetic components for underlying liability to disease severity and predisposition to high optimism (Mosing et al., 2009; Leinonen et al., 2005). Svedberg et al., (2006) found that, for individuals over the age of 45, genetic factors for physical distress contribute to subjective health. The authors suggest that the genetic effects shaping self-perceived health are mediated by the genetic predisposition to chronic diseases and

strongly associated with COVID-19 manifestation, appear to play roles in: lung functioning, immune system, cardiovascular diseases and diabetes type II (Guo et al., 2022; Ni et al., 2023; Lee et al., 2024).

functional limitations. Finkel and colleagues, (2024) support that genetic influences on subjective health were mediated through influences on physical health, including depressive symptoms. As such, this leaves a role for environmental influences to explain the remaining variance.

3.3. Individual characteristics

In the previous sections, I discussed how the COVID-19 pandemic shaped health outcomes in general and the role of genetics in shaping health. An important manifestation of COVID-19 is the heterogeneity of disease outcomes. Given the large number of cases and deaths and the discovery of new variants of the virus throughout the pandemic, understanding risk factors associated with susceptibility and severity is important for disease control. Individual characteristics can determine the risk of infection and the establishment of poor clinical outcomes. Existing literature has linked certain individual characteristics to the severity of individual outcomes of COVID-19, examples of which include: gender, age, socioeconomic status and the presence of underlying medical conditions such as increased BMI, and hypertension (Biswas et al., 2021; Niemie et al., 2022; Boutin et al., 2021). In this section, I will discuss the risk factors that are, supported by literature, most strongly associated with poor clinical outcomes.

3.2.1. Gender

De Godoy et al., (2023) found that males were more affected by COVID-19, resulting in higher mortality rates than women. This is supported by a study by Sieurin et al., (2022) who report a male sex disadvantage in COVID-19 severity, with the relative risk varying across age groups. Raja et al., (2024) also note that males have consistently higher age-adjusted COVID-19 mortality rates compared to women. It is suggested that this pattern - at least partially – stems from differences in the distribution of comorbidities across age and gender, indicating differing underlying biological mechanisms. (Alkhouli et al., 2020; Biswas et al., 2021). Women, particularly at older ages, tend to have a higher burden of comorbidities, whereas men are increasingly susceptible to critical diseases such as cardiovascular diseases, elevated blood pressure, chronic pulmonary illness and diabetes (Ahrenfeldt et al., 2019; Case and Paxon, 2005). The latter comorbidities are associated with worse progression of COVID-19, which will be discussed later.

3.2.2. Age

Early 2020, Wu et al., (2020) recognized advanced age as an important risk factor for COVID-19 mortality. Since it has been confirmed in multiple studies that age in general is an important risk factor. The elderly are most at risk of negative COVID-19 outcomes and consequences, especially among those over age 50 (Pijls et al., 2021; Fang et al., 2020; Zhang et al., 2022). According to a report from the US Centres for Disease Control and Prevention report (2020), 81% of deaths reported in the US were adults 65 years and older. Leung et al., (2020) find that the elderly have a higher mortality rate and increased risk for symptomatic infection. Factors contributing to this include the ageing process of the lungs (pulmonary ageing), but also increased risk or presence of (age-related) comorbidities (Rouatbi, 2022).

3.2.3. Comorbidities

In the previous section, I describe how genetic predispositions may affect health outcomes. However, the clinical presence of disease can also influence health outcomes. Feng et al. (2022) identified eleven overlapping genes between common comorbidities (CAD, diabetes mellites and hypertension) and COVID-19. The authors hypothesize that genetic polymorphisms of shared genes may be predictors of poor clinical outcomes. The presence of specific health-related comorbidities is associated with an increased risk of infection and more severe disease outcomes following infection. These conditions often impair one or multiple physiological functions, which can impact the progression of COVID-19 (Boutin et al., 2021). Tisminetzky et al., (2022) conducted a comprehensive literature review and found that across the world, the most prevalent underlying chronic conditions in patients with COVID-19 were hypertension, diabetes, cardiovascular disease (CVD), chronic pulmonary disease and chronic kidney disease. This is confirmed by Zhou et al., (2020) and Wu et al. (2020) who report in their study most frequent comorbidities in patients who developed acute respiratory distress syndrome, are hypertension (30%), diabetes (19%) and Coronary Artery Disease (8%).

Evidence also suggests that the presence and (or risk) of obesity is strongly associated with poor clinical outcomes of COVID-19 (Nagar et al., 2022). Dietz and Santos-Burgoa (2020) established that increased BMI is associated with decreased respiratory system responsiveness. Complex diseases such as obesity, diabetes, hypertension and cardiovascular diseases are associated with elevated risk of poor clinical outcomes (Gutierrez et al., 2020; Ejaz et al., 2020; Zhang et al., 2022). Existing literature suggests a role for shared genetics between obesity and pulmonary diseases (Hobbs et al., 2017; Zhu et al., 2020). More specifically, genetic studies

have identified genetic scores for higher BMI as a risk factor for COVID-19 severity (Leong et al., 2021). This is supported by Zhu et al., (2020), who show that individuals with greater genetic predisposition for obesity have an elevated risk of developing poor disease outcomes. However, the authors also indicate that the exact mechanism, which connects the genetic predisposition for obesity and observed obesity to severe COVID-19, is likely to be complex.[16] Nevertheless, the established relation between genetics and disease outcomes, suggests a role for obesity-related genes in shaping health outcomes under the pandemic.

[16] Increase severity for COVID-19 could stem from obesity-related changes to pulmonary functioning. Obesity decreases general lung functioning through decreased respiratory volume and effectiveness of lung muscles (Sattar et al., 2020).

4. Data

The data in this thesis is collected from the Health and Retirement Study (HRS), which is a biannual, longitudinal study initiated in 1992 by the University of Michigan. The HRS offers several publicly available datasets, including genetic scores, which are cost-free to use for research purposes upon registration at their website. The HRS is a representative sample conducted in the US and includes elderly individuals (50+) and their spouses (Bugliari et al., 2024). This dataset contains extensive baseline data at the individual level, including detailed economic and health-related information for a wide variety of categories. This thesis uses different sub-datasets: one pre-processed longitudinal file that covers the years 1992-2020; one cross-sectional dataset that offers the latest survey data available (wave 16, 2022); one 2020 COVID-specific module which was assigned during the first pandemic year (wave 15, 2020); and a dataset which contains the polygenic score data for complex phenotypes.[17] The genetic dataset contains information based on European and African ancestry. The main analysis will only include individuals classified as genetically European by the HRS, as I focus on PGS for European ancestry this reports the largest sample size.[18]

The sample used for the analysis is restricted from the main dataset. I apply sample restriction in the following manner. The total sample consists of 42,406 respondents. I followed the literature and excluded those with non-European ancestry to obtain a sample of 5,980 respondents. Individuals with missing or conflicting observations for any of the health outcomes are further excluded. Only respondents who are present in the three consecutive waves considered are included. As such, the analytic sample consists of 3,517 individuals.

4.1. Descriptive statistics

To grasp an initial understanding of the distributional changes over time in the sample, I present the descriptive statistics of the analytic sample stratified per wave in Table 1. The health outcomes are presented in the highlighted box. Firstly, note that the prevalence of lung disease followed an increasing trend during the pandemic years. Lung disease is a binary variable which indicates whether the respondent has been diagnosed with some respiratory illness or not.[19] Where 0 reports not being diagnosed and 1 corresponds to being diagnosed with lung disease.

[17] Precise details on the datasets used for data construction can be found in Appendix B.
[18] The HRS defines individuals of European ancestry as including ''all self-reported, non-Hispanic whites that had principle component (PC) loadings within ± one standard deviation of the mean for eigenvectors 1 and 2 in the PC analysis of unrelated study subjects'' (Ware et al., 2020).
[19] The variable lung disease excludes the presence of asthma.

During the pre-COVID wave, the prevalence of lung disease is 9.6%. This increased in 2020 by 0.7 percentage points to 10.3% and reached 11.1% during the late/post-COVID wave. Secondly, the self-reported health variable shows a respondent's self-reported health score on an increasing 1-5 scale with 1 corresponding to poor health and 5 excellent health. The scores remain on average rather constant over time, with a slight decreasing trend. Lastly, average BMI decreased entering the early- and mid-COVID-years, however, increased by 0.4 kg/m^2 on average following 2022.

The genetic component is captured by BMI polygenic scores. The environment is quantified by the variable 'covid deaths'. This is a binary indicator which specifies whether the respondent knows someone (friend or family member) who passed from COVID-19. In this sample, 20.5% indicate to know someone who passed from the disease.

There is an overrepresentation of women in this sample (~60%). The average age is 70-75 over the years. The gender-age distribution is compared with the U.S. gender ratio across different age groups (see Appendix B4 for details). For the age range 70-80, the gender ratio in 2021 is approximately 8 men per 10 women. From this, I can conclude that compared to the U.S. elderly population, women (men) are overrepresented (underrepresented) in this sample. The age-gender distribution is shown in Figure B4.2. in Appendix B4.

Smoking rates among U.S. elderly aged 65+ in 2021 is approximately 8.3% (CDC, 2023). This sample reports a higher percentage of smokers in 2022. Around 67% of U.S. elderly (50-80) report to occasionally consume alcohol (National Poll on Healthy Ageing, 2021). This is slightly lower in this sample. The majority of the sample reports to be either protestant (~59%) or catholic (~25%) of religion. Literature suggests that approximately 55% of U.S. elderly aged 50 and above indicate to be protestant and 49% of U.S. elderly aged 50 and above indicate to be catholic (Pew Research Center, 2014). In this sample, there thus seems to be an underrepresentation of catholic individuals.

Table 1. Descriptive statistics (N=3517): Summary statistics are presented for the entire analytic sample per wave. Presented as averages or percentages, unless stated otherwise.

	Pre-COVID (2018-2019)		Early/Mid COVID (2020-2021)		Late/ Post COVID (2022-2023)	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Health outcomes						
Lung Disease	9.6%		10.3%		11.1%	
Self-reported health	3.44	.916	3.37	0.916	3.31	0.939
BMI	28.581	5.832	28.353	5.794	28.717	7.855
BMI PGS*	-.029	.995	-.029	.995	-.029	.995
Years of education	14.042	2.267	14.042	2.267	14.042	2.267
Comorbidities†	1.044	.903	1.097	.916	1.182	.927
Age	70.548	8.857	72.549	8.883	74.549	8.883
	Count	**Share (%)**	**Count**	**Share (%)**	**Count**	**Share (%)**
Covid deaths‡	-	0	-	0	-	20.5
Smoking	-	6.7	-	6.1	-	9.8
Drinking	-	66.4	-	66.3	-	64
Male	-	40.8	-	40.8	-	40.8
Religion						
1. Protestant	2065	58.71	2065	58.71	2065	58.71
2. Catholic	871	24.77	871	24.77	871	24.77
3. Jewish	89	2.53	89	2.53	89	2.53
4. None	433	12.31	433	12.31	433	12.31
5. Other	59	1.68	59	1.68	59	1.68

Notes: PGS = polygenic risk score; SD = standard deviation

* Polygenic scores are standardized with a mean of 0 and a standard deviation of 1.

† comorbidities is a composite measure for the number of comorbidities present in the range of 0-3. Include the presence of: diabetes, coronary artery disease and hypertension.

‡ Variable is only included in the 2022 survey. Therefore, for previous years, the share equals 0.

5. Methodology

This section provides an overview of the main empirical setting for testing the hypothesis of the thesis. I start with describing the study design by introducing a directed acyclic graph (DAG) of the hypothesised associations between the different variables used in the model. The econometric model will be formally introduced in Section 5.2. Thereafter, I introduce and motivate the choice of dependent- and independent variables. Lastly, I introduce polygenic risk scores, which are a measure of genetic risk and capture the genetic component of the analysis. Lastly, I describe the COVID environment and how this is quantified to be included in the regression analysis.

5.1. Study design

Figure 3 visually illustrates the directed acyclic graph (DAG) of the hypothesised associations between the variables included in the main analysis. In this thesis, I aim to establish an association between the interaction of the genetic- and environmental variables to estimate how these affected health outcomes. Therefore, I hypothesise that the environmental exposure (E) and the genetic component (G) both have an effect on health outcomes on their own, but together they modify the effect of the other. G×E is therefore a function of both G and E. In the DAG, I also assume direct associations between G, E and SRH. Genetics can have a direct effect e.g. the genetic propensity to psychiatric and psychological disorders, which are directly associated with SHR responses (Harris et al., 2017). The same authors also establish a direct association between genetic predisposition to increased BMI and SRH scores. A direct effect of the pandemic (E) on self-rated health is supported by Hooper et al., (2023) who report a decrease in self-perceived health under the pandemic.

In Section 5.2.2. I illustrate the association between health outcomes and PGS scores (G): on average, higher BMI PGS is correlated with slightly lower SHR scores. The environment – knowing someone who passed from COVID - can affect health outcomes via for example poorer mental health, as the COVID-related death of someone can introduce stress, grief and anxiety. This may be reflected by lower SRH scores. Grief-induced stress can lead to higher BMI via lower physical activity and/or stress-induced binge eating (Rosenqvist et al., 2023).

Additionally, there might be confounding variables that affect the relationship between genetics and/or the environment on the one hand and health outcomes on the other hand. These are reflected in the DAG by C_E (confounders between the environment and health outcomes). One confounder that I can observe is socioeconomic status, which will be explored later (Section

6.4). Additionally, I hypothesize that the presence of BMI may be a mediator between genetics and lung health (LH) and SRH scores. As higher BMI affects health outcomes by e.g. increased susceptibility for certain diseases (Kivimäki et al., 2022).

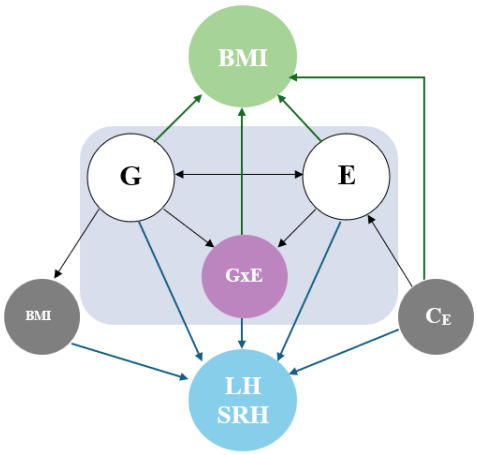

Figure 3: DAG. Visualization of associations between variables in the model. Blue and green circle correspond to the health outcomes (Y). Green arrows are associated to directly affect BMI, blue arrows lung health (LH) and self-rated health. Two-headed arrow between G and E reflect gene-environment correlation.

5.2. Econometric model

Gene-environment interaction (G×E). The main empirical setting to formally test the presence of a G×E interaction effect involves estimating Equation 1 by panel data regression. The panel data consists of repeated observations over time. This model appropriately exploits the time dimension of the dependent variables, as current values depend on their past realizations. The time period is restricted to include one pre-pandemic wave (2018) and two pandemic waves (2020/2022) and in total covers the years 2018-2022.[20]

The linear econometric model will follow the one described by Biroli and colleagues (2022). The dependent variable (Y_{it}) reflects the subjective- and objective health outcomes, measured by SRH, the presence of lung disease and BMI. G_{it} is the polygenic score for BMI, whereas E_{it} reflects the environment. The environment is quantified by a binary indicator for knowing someone who died from COVID-19 (1=yes, 0 otherwise). The interaction ($G_{it} \times E_{it}$) between the genetic index and environment is a measure of G×E effects. Equation 1 provides the regression model in the following form:

[20] Data used stems from the HRS biannual survey study. Therefore, the pre-pandemic wave corresponds to the years 2018-2019.

$$Y_{it} = \beta_0 + \beta_1 G_{it} + \beta_2 E_{it} + \beta_3 (G_{it} \times E_{it}) + \gamma X_{it} + \mu_i + u_{it} \quad \text{Equation (1).}$$

Where i denotes an individual respondent and t denotes the wave. X'_{it} is the vector of the control variables, which includes a measure for the number of existing comorbidities, age, highest level of education, a male indicator, smoking- and drinking status, and religion. μ_i reflects the random effects and u_{it} reflects the error term.

Coefficient β_1 estimates the association between polygenic score and health outcomes, whereas coefficient β_2 indicates the effect of knowing someone who passed from COVID-19. Coefficient β_3 aims to report on the existence of G×E interactions and shows how the association of the polygenic score with health outcomes is modified by COVID-19 shock. To identify the existence of a gene-environment interaction, genetic predisposition and the environment should be independently distributed (i.e. individual genetic predisposition should not influence variation in exposure to the environment) (Arold et al., 2022). This implies that there is independent variation in the genetic- and environmental components. Therefore, I also check for gene-environment correlation.

Gene-environment correlation. To test the existence of gene-environment correlation and examine how selection based on genetic predisposition may serve as a mechanism into the environment. Therefore, I regress the covid-environment on genetic scores. More formally, I employ the following panel regression analysis:

$$E_{it} = \alpha_0 + \alpha_1 G_{it} + \eta X_{it} + \theta_i + \nu_{it} \quad \text{Equation (2).}$$

Where the dependent variable E_{it} is the environmental measure for COVID-19, G_{it} the variable for genetic predisposition for BMI, X_{it} includes a vector of covariates (same as for Equation 1), θ_i the random effects and ν_{it} the error term. The coefficient α_1 serves to inform about the association between polygenic scores and the environment.

The VIF test indicates that no main variables of interest should be dropped.[21] A large VIF score (~19) is found for the covariate age, however, the effect of including or excluding age from the regression analysis is close to zero. Heteroskedasticity concerns dictate the use of robust standard errors clustered at the respondent level, which is the appropriate strategy for G×E specifications according to the literature (Biroli et al., 2022). Moreover, the analysis includes

[21] A large VIF score for the age covariate is found. However, since age is expected to affect health outcomes, the baseline is kept as described in Equation (1) and (2).

estimating the model with random effects over fixed, as genetic effects are per definition fixed over time.

5.2.1. Establishing health outcomes

As mentioned in the previous section, this thesis considers three health measures: self-rated health, BMI and the presence of lung disease. This section motivates these health outcomes. Due to the limited availability of COVID-specific outcomes, I consider subjective and more objective health measures to quantify health effects. The subjective component is captured by self-rated health. The objective measures are captured by the presence of lung disease and BMI. The section concludes by discussing covariates included in the analysis.

Self-rated health – the first outcome is self-rated health status (SRH). SRH reflects subjective individual health status at a given point in time. This measure captures a more general perception of health status under the COVID-19 pandemic.[22] One advantage of using this metric is that SRH scores are widely recognized as a convenient and simple measure of individual health (Meng et al., 2014). Therefore, they are commonly used in public health surveys. SRH has been shown to capture a wide range of health-related aspects, such as the presence of (chronic) illnesses, health behaviours and cognitive and physical limitations (Bombak, 2013). While SRH captures current health according to the respondent's definition of health, evidence indicates that it is strongly associated with other health-related measures, such as physical performance, disease diagnosis and laboratory parameters in the hospital (Wu et al., 2013). A limitation of SRH is that is it non-specific and captures health in a broad, and subjective manner. Despite its lack of specificity, low SRH scores appear predictive of longevity, especially in the elderly (Lorem et al., 2020; Stenholm et al., 2014; Dramé et al., 2023). SRH is coded as an ordinal categorical variable, with a clear ranking of distinct categories. It quantifies how a respondent would rate their current health status on a 5-point scale. It is measured using a single item question or more specifically ''*Would you say your health is excellent, very good, good, fair, or poor?*'', with 1 reflecting poor health and 5 excellent health.[23] Figure 4 shows the

[22] The main motivation for including SRH over another, more concrete measures of COVID-outcome, is data availability. Outcomes as mortality, hospitalization and covid-related events are either restricted for access, or have insufficient observations for the analysis. Therefore, SRH is chosen to capture changes in health status during the pandemic on a broader scale, as it also captures physical and mental health. However, SRH has issues regarding interpretation w.r.t. a baseline, generalizability and individual interpretation regarding ranking outcomes.

[23] Self-reported health is recoded and standardized as the variable has a different scale compared to linear predictors in the model for regression analysis.

distribution of the reported health scores per wave over the 2018-2022 sample period. The figure illustrates that there is an a shift in reported health scores when entering the early COVID-phase. There is a decrease in the number of individuals that report to be in excellent health and the number of individual that are in poor health (score 1-2) increased. During the late/ post COVID-wave, the number of individuals in poor self-rated health further increases. This suggests that during the pandemic, more respondents report to be in worse health. In the years during the pandemic, the median shifted from 4 to 3 on the 5-point scale.[24]

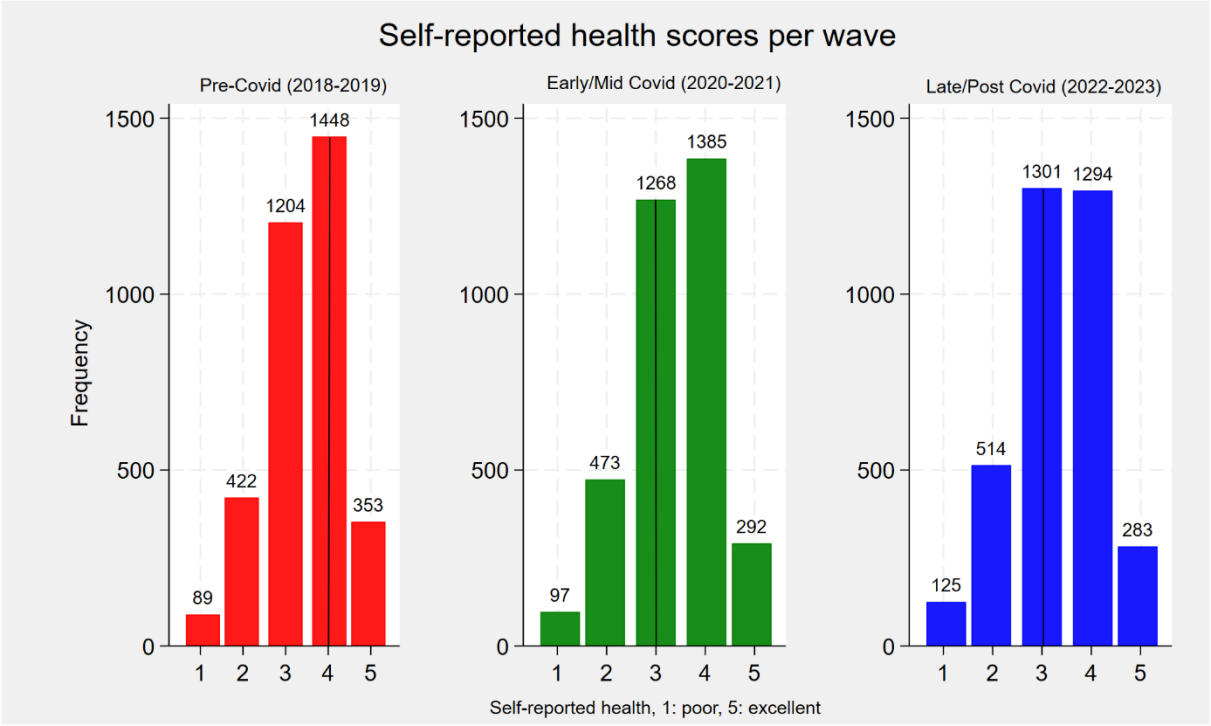

Figure 4: Self-reported health scores. The figure illustrates the distribution of self-reported health scores in the three waves included in the analysis. *Note:* the vertical black line illustrates the median score per wave.

From the figure above, there appears to be a shift in the distribution of health scores during the pandemic years, with an increasing number of individuals reporting to be in poor health. This pattern is further explored by visualising the relative change in SRH scores compared to the previous wave, to understand the relative change in scores. This helps in understanding whether most people report being in the same health status, an improvement or deterioration. Figure 5 shows the relative change in health scores compared to previous waves. It reveals that for both waves, approximately half of the sample maintained the same score. Additionally, around 800 individuals experienced a decrease in their health score by either 1 or 2 points in both waves.

[24] The COVID-19 pandemic was officially ended the 5th of may 2023 by the WHO.

Moreover, there are around 500-600 individuals who claim to be in a better health status.

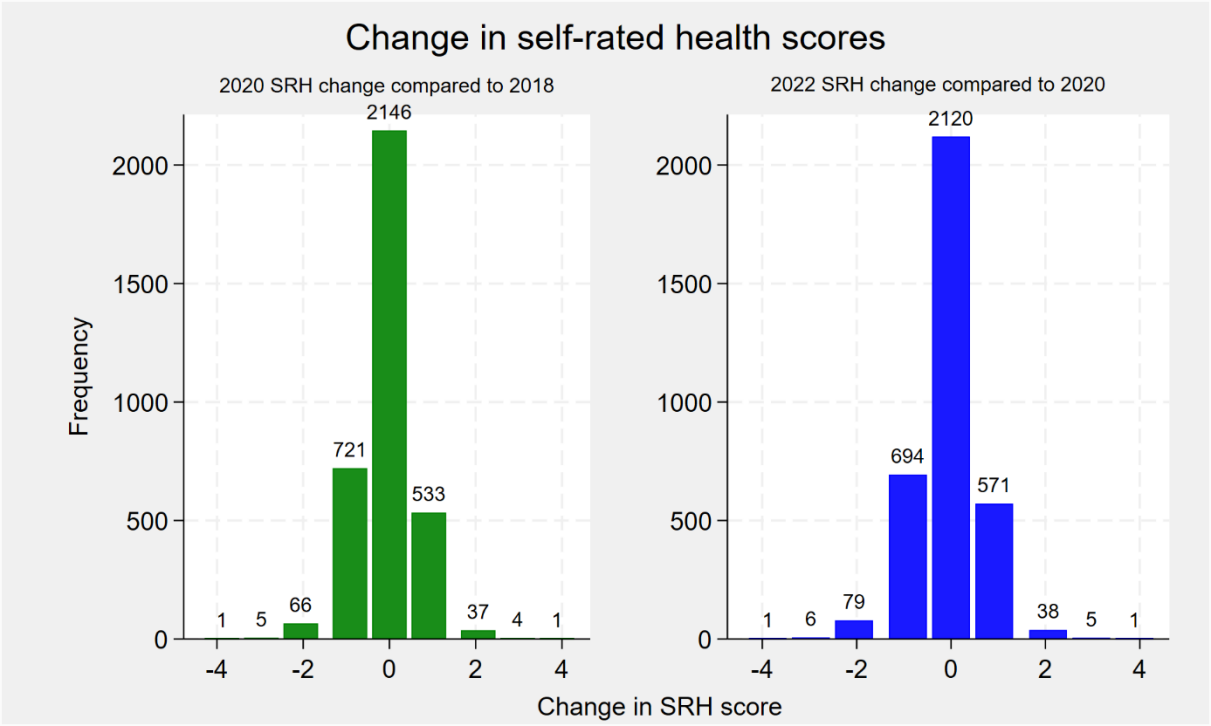

Figure 5: Relative change in SRH scores. Individual reported scores compared to the previous wave. Zero indicates individual reports the same score, whereas positive integers indicate a relative improvement in SRH scores, and negative ones a deterioration.

Lung disease - The second outcome of interest is lung disease. This variable is an objective health measure, and more directly related to COVID-19, given the impact of the disease on pulmonary functioning and the potential for long-term respiratory issues in severe cases (Chatterjee et al., 2023; Elrobaa and New, 2021). Lung health is a binary indicator and reflects the question ''*Has a doctor ever told you that you have chronic lung disease such as chronic bronchitis or emphysema?*''.[25]

Figure 6 shows the change in the prevalence of lung disease compared to the previous wave. Panel A on the left shows the reported lung health status relative to the pre-COVID wave. The majority of the respondents maintain the same lung status, indicating either no lung disease or pre-existing chronic conditions. 37 individuals do report being affected with some lung disease, which may reflect COVID-19-induced conditions. Panel B on the right depicts changes towards

[25] Emphysema and/or chronic bronchitis combined is sometimes also referred to as *chronic obstructive pulmonary disease* (COPD). Chronic bronchitis results in inflammation of the airways, whereas emphysema refers to the destruction of the lung's air sacs. Both lead to restricting airflow and breathing problems. These conditions can worsen when combined with COVID-19, leading to long-term illness and potential irreparable damage (Gerayeli et al., 2021).

the later COVID-19 years. While most respondents show no change, 47 individuals now report having lung disease, with 19 reporting no disease or recovery.

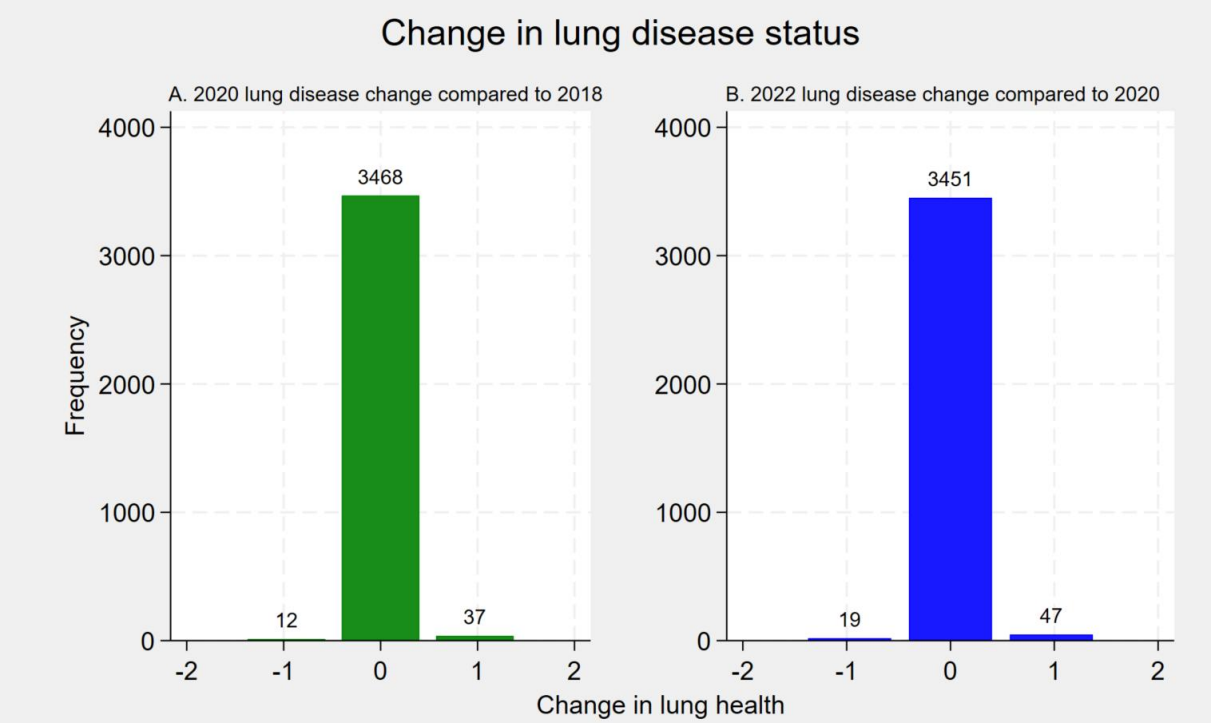

Figure 6: change in lung disease status. Change is compared to the status reported in the previous wave.

Body Mass Index (BMI) – the third health measure is BMI. Evidence suggests that the presence and (or risk) of obesity is strongly associated with poor clinical outcomes following COVID-19 (Nagar et al., 2022; Gutierrez et al., 2020; Ejaz et al., 2020; Zhang et al., 2022).[26] BMI is not directly measured in the HRS but is derived from individual height and weight measurements.[27] BMI calculations are already present in the longitudinal file, however, calculations for the most recent waves are lacking. As such, the BMI calculations for wave 16 were done manually by applying the following definition of BMI, which is in line with the HRS calculations[28] :

$$BMI = \frac{Weight\ in\ kilograms}{(Height\ in\ meters)^2}$$

Figure 7 shows the distribution of average BMI over the three survey periods. It shows that the average BMI is 28.6 kg/m^2, which corresponds to being overweight according to classification

[26] Motivation for including BMI is data availability. Nevertheless, self-rated health, lung health and BMI are other measures affected by COVID-19 (Chatterjee et al., 2023; Elrobaa and New, 2021), and an important health concern in the elderly population (Akgün et al., 2012).
[27] The HRS converts height to meters and weight to kilograms prior to BMI calculations. Weight is asked every wave, height is only asked in the first wave a new respondent enters the study. The height is carried forward for individuals that are again interviewed in later waves (Bugliari et al., 2024).
[28] For a precise description of data used for computing BMI in wave 16, see Appendix B.

by the U.S. Centers for Disease Control and Prevention (2022). This aligns closely with the average BMI of 28.2 in white elderly U.S. individuals in a study performed by Barceló et al., (2007). Table 2 reports the average BMI scores across waves.

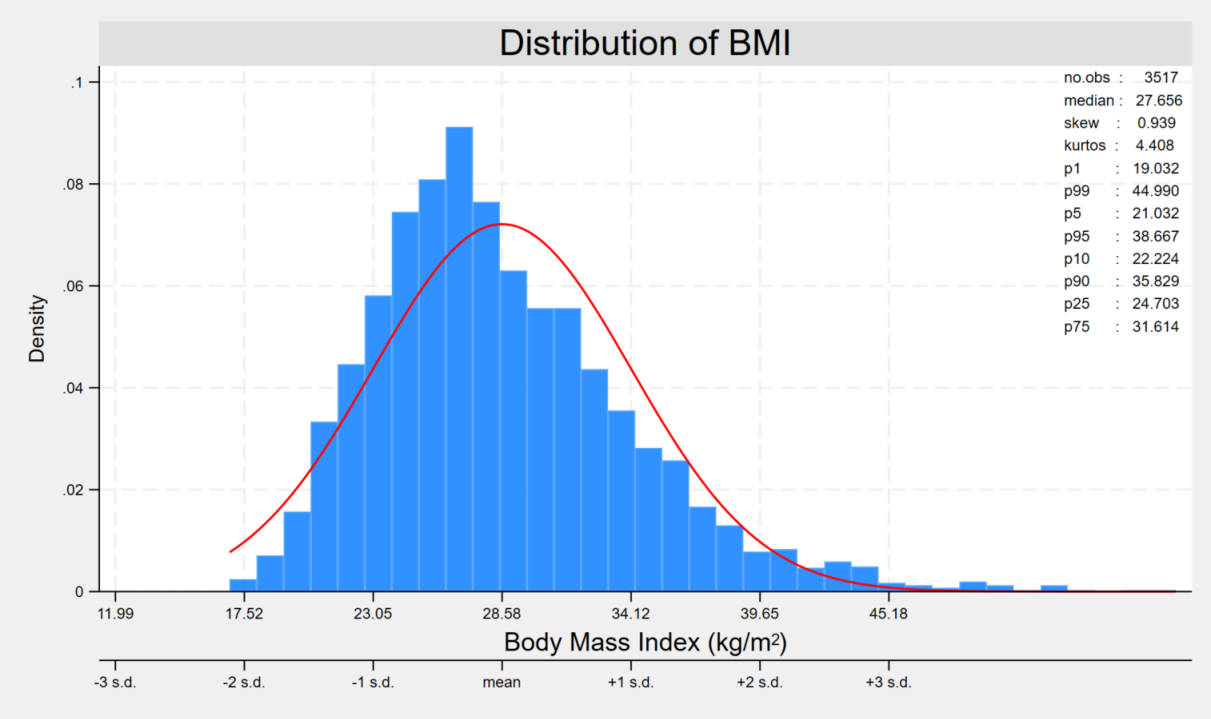

Figure 7: BMI distribution. Average BMI over the three waves

Table 2. Average BMI per wave

Variable	Wave (year)	Obs.	Mean	Std. Dev.	Min	Max
BMI	14 (2018)	3498	28.581	5.832	15	63.8
BMI	15 (2020)	3502	28.353	5.794	14.1	64.4
BMI	16 (2022)	3517	28.717	7.855	15.8	52.0

Covariates – in the baseline specification I control for different covariates that, according to the literature in Section 3, influence health status. The covariates included are age, gender, drinking- and smoking status, the number of pre-existing comorbidities and education. I additionally control for religion with self-rated health scores, as religion may affect how and if an individual reports their health status (Krause, 2010).

5.2.2. Polygenic Risk Score for BMI

The genetic variable is the polygenic risk score for BMI since the literature suggests that BMI plays an important role in establishing health outcomes. I utilize the polygenic scores made available by the HRS, which are standardized with a mean of 0 and a standard deviation of 1.

The PGS for BMI measures individual predisposition towards BMI based on the ancestral group and environmental context of the genome-wide association study, which is European ancestry in this thesis (Ware et al., 2021). HRS uses SNP weights to construct the PGS for the HRS analytic sample. The source of SNP weights is independent of the target sample and stems from GWAS meta-analysis (Ware et al., 2021).[29] Given the distribution of PGS, higher scores indicate a greater risk for a certain trait. Figure 8 illustrates the distribution of the polygenic score against the health outcomes of interest. The PGS for BMI is approximately normally distributed in the sample. Panel A represents the trend between self-rated health scores and the polygenic scores for BMI. It suggests that higher polygenic scores are associated with lower self-perceived health. This supports the hypotheses specified in Section 2. Panel B shows the polygenic score for BMI against the presence of lung disease. The association is less evident, however hints that higher genetics scores are associated with a greater likelihood of the presence of lung disease. Lastly, panel C shows the relationship between BMI PGS and BMI as a health outcome. There is a clear and positive correlation, with higher genetic scores predicting higher BMI in the sample, which is also in line with the hypothesis.

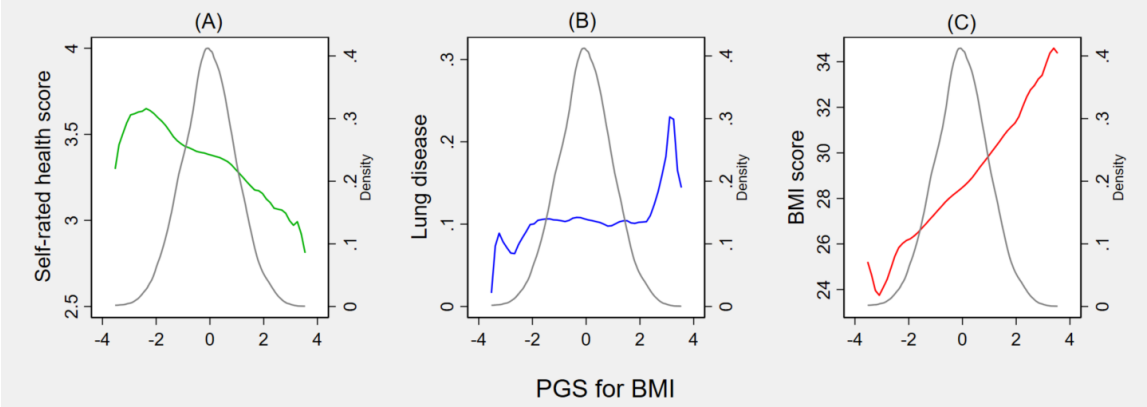

Figure 8: Trend between polygenic risk score for BMI and health outcomes. (A) self-rated health; (B) lung disease; (C) BMI scores. The horizontal axis for all figures represents the polygenic risk scores for BMI. The right vertical axis represent density. Left vertical axis reflects the scale for individual outcome.

The figure above suggests that the polygenic score is predictive of self-rated health scores and BMI. Subsequently, the predictive power of the polygenic scores is more formally assessed by regressing the outcomes on the polygenic score. Table 3 below represents the regression coefficients for different health outcomes. The results suggest that the polygenic score is predictive of worse health outcomes (i.e. lower self-rated health and higher BMI scores) for

[29] The genotyped target sample should not have been included in the original GWAS studies to predict SNP effect sizes. Otherwise, the accuracy of prediction is overestimated via *over-fitting* (e.g. overfitting occurs when the same group is used to estimate SNP effects *as well as* to make predictions) (Mills et al., 2020; Wray et al., 2013).

SRH and BMI. A one standard deviation increase in BMI PGS decreases SRH by 0.0715 standard deviations. This is statistically significant at the 1% significance level. The predictive power for lung disease is non-significant. It may be that the PGS for BMI is not sufficiently predictive of the presence of lung disease. This observation changes when considering BMI as the outcome. One standard deviation increase in the BMI PGS is associated with an increase of 1.487 points (kg/m^2) in BMI.

Table 3. Regression estimates for the effect of BMI PGS on health outcomes.

	(1) SRH	(2) Lung disease	(3) BMI
PGS for BMI	-0.0715*** (0.0135)	0.00291 (0.00496)	1.487*** (0.0901)
Observations	10,548	10,551	10,517
N	3,517	3,517	3,517

Notes: PGS for BMI is standardize to have a mean of 0 and standard deviation of 1. Regressions control for the covariates specified in section 5.2. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

In the model, I imply associations between the genetic scores and health outcomes. Polygenic scores are likely to be confounded by genetic nurture since genetics are inherited from parents (Trejo and Domingue, 2018). Genetic nurture suggests that genetic influences on individual traits can be mediated not only by an individual's genetics but also through the environment that is shaped by the parent's genetics (Kong et al., 2018). Therefore, a polygenic risk score can capture family and environmental effects, arising from for example geographical and socio-economic status effects. Therefore, to imply causality, one should control for the parental genotype (Biroli et al., 2022). However, these are unavailable from the HRS.

5.2.3. The Environment

The theoretical foundation is the diathesis stress model (Section 2.3) which postulates an environment to be negative. As such, the COVID environment is captured as a 'life event' that quantifies the pandemic as a binary indicator for knowing someone (friend and/or family member) who passed from the disease or not. The measure reflects the pandemic as a negative shock, which is in line with the theoretical model. This question is included in the 2022 survey as: '*Do you have any family members or close friends who died from COVID-19?".* Existing literature suggests that COVID-19-related deaths - combined with the pandemic circumstances - affect how individuals cope with grief (Eisma et al., 2020). While the risk of mortality after COVID-19 infection is associated with several risk factors in literature (Zhang et al., 2023), COVID-related deaths are often sudden and to a certain extent random in nature. Typically,

sudden and unexpected grief is disruptive and painful. Literature has proposed different effects of grief after the sudden death of a loved one: most individuals adapt after losing someone close, but coping mechanisms can vary significantly (Goveas and Shear, 2020; Carr et al., 2020). This can induce physical and psychological consequences, particularly in elderly individuals (Carr et al., 2001). A sudden bereavement can trigger psychiatric disorders, alcohol (mis)use, decreased life satisfaction and lower quality of life in general of an affected individual (Keyes et al., 2014; Treml et al., 2020). Sudden death of a partner in general is moreover associated with greater mortality of surviving partners in elderly individuals (Shah et al., 2013). The passing of a spouse from COVID-19 specifically is associated with greater symptoms of distress, shock and disbelief (Stahl et al., 2023). Stahl et al., (2023) find that these individuals are more likely to require clinical care for their severe pathological grief reactions. Eisma et al., (2020) also found increased grief levels after COVID-19-related bereavement for adults, compared to death due to other causes. Given the individual heterogeneity in responses to unexpected death - which worsened during the pandemic - it could be that individuals who experienced bereavement during the pandemic present different health outcomes.

Ideally, the environment for the gene-environment interaction should be exogenous to the model (Biroli et al., 2022), however endogeneity concerns arise when the environment is correlated with other (un)observed variables in the model. This measure is chosen as the environment over other candidates[30], since it is at least quasi-exogenous. Knowing someone who passed is unlikely to be shaped by the genetic measure. To test whether it is plausible to assume the exogenous nature of the environment regarding polygenic scores, I follow the strategy of Biroli et al., (2022). I test whether the polygenic scores are significantly different across the environment (rGE) via Equation 2 in the following section. Finding non-significant evidence helps support the assumption that the COVID environment is exogenous to the model and thus an appropriate choice for G×E analysis.

Secondly, individuals with certain characteristics may cluster together in the environment - e.g. elderly individuals are often surrounded by other elderly individuals, and advanced age is an important risk factor for COVID-19 mortality (Wu et al., 2020). The clustering effect may also happen for individuals with poorer underlying health, which increases the likelihood of knowing someone who passed from COVID-19. To reduce omitted variable bias, variables for age and the number of comorbidities are included as covariates in the model.

[30] For a description of the other potential COVID-environment measures, see Appendix C.

Furthermore, the risk of passing away from COVID-19 is higher in areas with low vaccination rates; higher vaccination rates are associated with reduced risk of COVID-19 mortality (Chen, 2023). Willingness to engage in COVID-related preventive behaviour, such as vaccination, depends on social ties, norms and the behaviour of friends and family (Rabb et al., 2022). However, since all the respondents in the sample are vaccinated, this factor may not play a prominent role.

A limitation of quantifying the environment as such, however, is that it does not specify who passed from COVID-19, nor when this happened. It can be expected that the death of someone closer to you may have a larger impact. Nevertheless, it still provides a general understanding of the impact of COVID-19-related deaths, regardless of the non-specificity of the deceased individual.

6. Results

6.1. Main results

Gene-environment correlation

To establish that the variation in BMI PGS is independent of the variation in exposure to COVID-19, I first explore the existence of a gene-environment correlation between knowing someone who passed from COVID-19 and the genetic predisposition for BMI, by Equation 2. The estimated effect size hovers around zero and is statistically insignificant. This suggests that individuals, based on their genetic predisposition are unlikely to select differently in the COVID-environment. Thus, knowing someone who passed from COVID-19 unlikely to be driven by gene-environment correlation. This suggests that the environment is exogenous to the BMI PGS and potential gene-environment interactions I am going to investigate further are unlikely to be driven by these correlations.

Table 4. Regression analysis Equation 2.

	Covid Environment
PGS for BMI	0.00287
	(0.00176)
Constant	-0.0123
	(0.0161)
Observations	8,044
N	3,500

Note: The table excludes coefficients for covariates (age, number of comorbidities, religion, gender, smoking- and drinking status, and highest level of education). Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The gene-environment interaction between genetic predisposition to BMI and COVID-19 environment in shaping health outcomes

Next, I investigate the extent to which the BMI PGS and exposure to the environment predict health outcomes. Additionally, I explore the G×E interaction using Equation 1 to see if the genetic predisposition to BMI moderates the effect on health outcomes when experiencing the COVID-shock.

Table 5 presents the regression results of estimating Equation 1. The columns (1-3) represent the different outcome variables. Column (1) shows the regression coefficients for self-rated health. The BMI PGS negatively predicts self-reported health scores, since individuals with higher genetic predispositions are more likely to report lower scores. More specifically, a one

standard deviation increase in BMI polygenic score is associated with a 0.0698 (≈0.066 points[31]) standard deviation lower SRH. The COVID-environment is also negatively associated with SRH. Individuals who know someone who passed from COVID-19, as opposed to those who do not, have 0.126 (≈0.119 points) standard deviation lower SRH scores. The gene-environment interaction (Covid environment × PGS for BMI) term is negative. Therefore, knowing someone who passed from COVID-19 increases the negative association of SRH and BMI PGS by ≈ 88% (0.0620/0.0698). This suggests that individuals with higher polygenic scores report lower SRH scores when exposed to the COVID-shock. However, these results are statistically insignificant.

Table 5. Panel data regression of health outcomes.

	(1) SRH	(2) Lung disease	(3) BMI
PGS for BMI	-0.0698***	-0.00317	1.506***
	(0.0136)	(0.00484)	(0.0895)
Covid environment	-0.126***	0.00339	0.871**
	(0.0423)	(0.00889)	(0.424)
Covid environment × PGS for BMI	-0.0620	0.0114	-0.993***
	(0.0379)	(0.0109)	(0.384)
Constant	-0.182	0.0561	37.87***
	(0.143)	(0.0515)	(0.923)
Observations	8,041	8,044	8,010
N	3,500	3,500	3,493

Note: The table excludes coefficients for covariates (age, number of comorbidities, religion (SRH), gender, smoking- and drinking status, highest level of education), the full table is included in Appendix D, Table D1. *Reference categories*: Covid environment: not knowing someone who passed from COVID-19. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Looking at the presence of lung disease (Column 2), both coefficients for BMI PGS and the COVID environment are insignificant and close to zero. If significant, the interaction term suggests that the likelihood of lung disease increases by 1.14 percentage points for individuals with a greater predisposition to BMI under the COVID-19 shock.

The last Column (3) shows the results for BMI. The polygenic risk score for BMI is positively associated with BMI. As such, a one standard deviation increase in BMI PGS is associated with a 1.506 points (kg/m^2) increase in BMI. The environment also suggests a positive association with BMI: those who know someone who passed from COVID-19 report 0.871 points (kg/m^2) higher BMI, compared to those who do not know someone who passed. The gene-environment interaction coefficient is negative and significant at the 1% significance level. This suggests

[31] Coefficient * SD unstandardized variable (0.946)

that the effect of a high BMI PGS on BMI is potentially moderated by exposure to the environment. More specifically, exposure to the environment reduces the positive association of the BMI PGS on BMI by approximately 66% (0.993/1.506). Respondents with greater polygenic risk scores under the COVID environment have on average 0.993 points lower BMI, compared to those who do not know someone who passed.

Figure 9 shows the predicted values of estimating Equation 1 for the three different health outcomes by the polygenic risk score for BMI by the presence of exposure to the environment. The first graph suggests that exposure to the environment is associated with more negative SRH scores with increasing BMI PGS since the line that represents exposure to the environment (green) is steeper. This is in line with the previous finding that the environment magnifies the negative association of BMI PGS on SRH. However, these results are insignificant. The second graph is suggestive of an adverse effect on the presence of lung disease for individuals with greater genetic risk, when exposed to the environment, however insignificant. The last graph shows that exposure to the environment flattens the association of the polygenic score on BMI. This suggests that knowing someone who passed from COVID-19 may dampen the genetic effect on BMI. This will be discussed in Section 8.3.

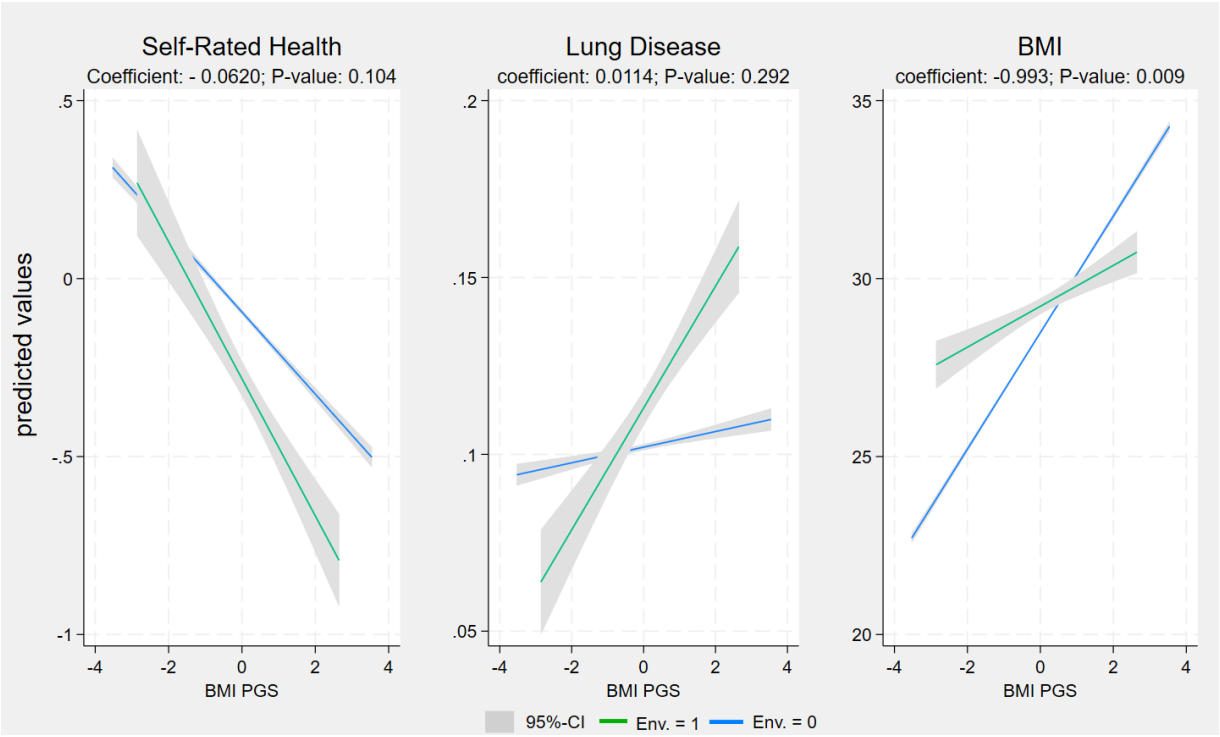

Figure 9: Predicted values for health outcomes by BMI PGS. The horizontal axis represents the polygenic risk score for BMI. The right vertical axis represents predicted values respectively. The left vertical axis reflects the scale for individual outcomes. Includes 95-percent confidence intervals for individuals that are exposed (Env. = 1; green) or not exposed (Env. = 0; blue) to the environment. Predicted values from regression including the G×E interaction.

Stratification by high- and low genetic predisposition to BMI

The main regression results suggest that individuals with different genetic predispositions to BMI respond differently when exposed to the COVID environment. Therefore, I explore whether changes in health effects are more pronounced under the environment when stratified by relative genetic risk, to allow for comparison between the groups.[32] Table 6 shows the regression results for estimating Equation 1, but excluding the G×E interaction term as stratification allows for direct interpretation of the COVID-environment coefficient. First, Columns (1) and (2) show the results for self-rated health. The effect of the environment on SRH is relatively close in absolute magnitude but slightly higher for the greater risk group. Both are significant at the 1% significance level. Individuals with high genetic risk have 0.135 (≈0.127 points) standard deviation lower SRH scores under the COVID-shock, compared to high-risk individuals who did not experience said shock. Low genetic risk individuals who know someone who passed report 0.123 (≈0.114 points) standard deviation lower SRH compared to those who did not experience said shock.

For lung disease, the environment coefficients for high- and low-genetic risk which hints that knowing someone who passed increased the likelihood of having lung disease, which is also shown in Figure 9. However, Column (3) does not show any economically significant result, as the effect size is approximately zero and non-significant. The same holds for the environment in Column (4).

Columns (5) and (6) show the stratified results for BMI. Individuals with low genetic risk have 1.589 point higher BMI when faced with the COVID-environment compared to individuals with low genetic risk who did not experience that shock. However, this effect is not seen in the high-risk group. More specifically, the coefficient is negative, which suggests that these individuals have lower BMI scores when exposed to the environment. However, the effect size is non-significant.

The results stratified by genetic risk suggest that indeed COVID-19 adversely affected health outcomes. For self-reported health scores, these effects do not seem to be mainly driven by genetic predispositions, but rather by the presence of the shock. The COVID-coefficient on BMI indicates a larger and significant effect in the low-risk group, which suggests that the

[32] As PGS are standardized with a mean of 0 and standard deviation of 1, I stratify genetic risk as follows:

$$\begin{cases} High\ genetic\ risk: \text{PGS} > 0 \\ Low\ genetic\ risk: \text{PGS} < 0 \end{cases}$$

genetic effect of BMI may be dampened when an individual experiences a COVID-19-related loss. Additionally, these effects could also partly be explained by the fact that there is a slight difference in genetic distribution under the COVID- environment (see Appendix B4 for details).

Table 6. Panel data regression of health outcomes by low and high genetic risk

	SRH		Lung disease		BMI	
	(1)	(2)	(3)	(4)	(5)	(6)
	Low risk	High risk	Low risk	High risk	Low risk	High risk
PGS for BMI	-0.0914***	-0.0799**	0.00495	-0.00885	1.352***	1.791***
	(0.0309)	(0.0320)	(0.0105)	(0.0128)	(0.176)	(0.239)
Covid	-0.123**	-0.135**	5.67e-05	0.00614	1.589**	-0.0296
environment	(0.0588)	(0.0615)	(0.0106)	(0.0155)	(0.625)	(0.576)
Constant	0.426***	0.106	0.00426	-0.180***	35.68***	39.08***
	(0.162)	(0.180)	(0.0557)	(0.0524)	(0.960)	(1.188)
Obs.	4,090	3,951	3,951	4,093	4,079	3,931
N	1,787	1,713	1,713	1,787	1,783	1,710

Note: The table excludes coefficients for covariates (age, number of comorbidities, religion (SRH), gender, smoking- and drinking status, highest level of education), the full table is included in Appendix D, Table D2. *Reference categories*: Covid environment: not knowing someone who passed from COVID-19. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

6.2. Robustness checks

In this section, I check whether the model is robust to alternative specifications of the model. Self-reported health is an ordinal categorical variable with a clear, increasing ranking of the different categories. It is not uncommon practice that such data is analysed as if it were continuous, which is the strategy in the main section of the thesis (Vermunt and Hagenaars, 2004; Robitzsch, 2020). However, this approach ignores the fact that the true distances between response levels are unknown. The literature proposes the ordinal logit model as an appropriate method to deal with such dependent variables, as it allows to predict the likelihood of falling into a specific category (Williams, 2020).

Therefore, as a robustness check, I estimate two additional models for self-rated health as the outcome variable: 1) ordinal logistic model and 2) OLS with non-standardized outcome, as in the main specification, SRH is standardized. The regression results are shown in Table 7. Column (1) reports the odds ratios after running an ordinal logistic regression. For the main outcomes, the sign of the association remains the same. A standard deviation increase in BMI PGS is associated with a decrease in the odds of having a higher SRH of 0.81, i.e. higher BMI PGS is associated with lower SRH. The COVID-environment also indicates a significant negative effect. SRH is predicted to shrink by a factor 0.83 (1-OR) when knowing someone who passed from COVID-19. The interaction term is insignificant, however, close to 1 in absolute magnitude and suggests a negative association. Since it is approximately 1, it suggests

that the combined effect does not significantly alter the odds of the outcome, which is in line with the main regression results (-.0620). Colum (2) shows the regression results of running Equation 1 with unstandardized SRH scores and thus treating the variable as if it was continuous on a 1-5 scale. The sign of the coefficients is similar, except for the G x E interaction, which changes from negative to positive. This may be explained by scale differences, as standardizing the variables centres values around zero and changes the scale. The same may hold for the coefficient of the COVID-environment, which has increased in absolute magnitude without changing the direction of the association. Therefore, the relative impact of the interaction may be different on a (non-)standardized scale. Overall, these results suggest that for self-rated health scores, the direction of the association remains stable when changing the model.

Table 7.

	SRH	
	(1)	(2)
	OR	OLS
PGS for BMI	0.8175***	-0.0471***
	(0.0448)	(0.0128)
Covid environment	0.1795**	-0.3760**
	(0.0076)	(0.1538)
Covid environment × PGS for BMI	0.9859	0.0035
	(0.0222)	(0.0050)
Observations	8,007	8,007
N	3,493	3,493

The table excludes coefficients for covariates (age, number of comorbidities, religion (SRH), gender, smoking- and drinking status, highest level of education). *Reference categories*: Covid environment: not knowing someone who passed from COVID-19. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

6.3. Sensitivity analysis

In this section, I check whether the estimated coefficients change when varying parameters in the model. For the sensitivity analysis, I estimate the baseline specification (Equation 1), but for different polygenic risk scores for other complex health outcomes to determine if the observed effects are specific to the BMI PGS. The factors for pre-existing comorbidities include coronary artery disease, type II diabetes and hypertension. Therefore, I will assess whether or not the presence of the phenotype, but also the underlying genetic risk for these diseases modifies the association with the COVID environment. The results are shown in Appendix D, Table D4. It shows that the estimated effect of the COVID-environment remains predictive of self-rated health and BMI scores. The polygenic scores for different complex diseases indicate

no significant effect, except for type 2 diabetes on BMI. Therefore, it may be that these genetic scores are insufficiently predictive for the health outcomes considered in this thesis. GWAS established genetic variants that overlap between genetic risk for BMI and type 2 diabetes, which may explain some predictive power (Basile, 2014). The coefficient for the environment remains in the same order of magnitude for different polygenic risk scores and is significant for SRH and BMI, which is consistent with the main analysis.

6.4. Mechanisms

In this section, I offer alternative explanations for the observed results from the main analysis of this thesis. I will explore different mechanisms that could explain these findings and test them accordingly.

Mediating variable – BMI

In the conceptual framework of section 5.1, I moreover allow for the existence of mediating and confounding variables, which may help explain the observed heterogeneity in health outcomes. A mediating variable for self-rated health and lung disease could be the BMI phenotype. From the literature, it is evident that higher BMI is an important risk factor for poor clinical outcomes and coupled to overall worse health status (Nagar et al., 2022; Dietz and Santos-Burgoa (2020)). Therefore, it is likely that the presence of high BMI, as opposed to whether the individual is genetically predisposed, affects health outcomes. To test for mediation, I utilize mediation analysis as proposed by Baron and Kenny (1986).[33] These results are shown in Table 9 below. All coefficients show statistically significant results for SRH. This suggests that BMI *partially mediates* SRH scores[34], since Column (3) shows that BMI remains significant after controlling for BMI PGS. whereas this is unlikely to be the case for lung disease. This suggests that BMI scores accounts partly for the relationship between polygenic risk scores and SRH.

[33] I first regress health outcomes (SRH and lung health) on polygenic scores for BMI, secondly the hypothesised mediator BMI on polygenic scores and the mediator on health outcomes on Y, and lastly a multiple regression analysis for X and M predicting Y.

[34] Partial mediation is likely to be present in the model if X and M significantly predict Y (Baron and Kenny, 1986).

Table 9.

	SRH			Lung disease		
	(1)	(2)	(3)	(4)	(5)	(6)
	SRH	BMI	SRH	Lung disease	BMI	Lung disease
PGS for BMI	-0.0715***	1.483***	-.0520***	-0.00275	1.483***	.0005
	(0.0136)	(0.0903)	(.01383)	(0.00485)	(0.0903)	(.00058)
BMI			-0.0145***			0.000534
			(0.00213)			(0.000577)
Constant	-0.156	37.97***	0.382**	0.0440	37.97***	0.0236
	(0.142)	(0.946)	(0.163)	(0.0521)	(0.946)	(0.0532)
Observations	8,041	8,010	8,007	8,044	8,010	8,010
N	3,500	3,493	3,493	3,500	3,493	3,493

Note: table excludes coefficients for covariates (age, number of comorbidities, religion (SRH), gender, smoking- and drinking status)
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Confounder– Socioeconomic status

Confounders are expected to affect the variables in the model, and not accounting for them implies that the results fail to capture the true relationship (Pourhoseingholi et al., 2012). A confounder that affects both the environment and health outcomes is socio-economic status. Lower socio-economic status is associated with an increased risk of poor health (Barakat and Konstantinidis, 2023), and higher prevalence and mortality of COVID-19 (Hawkins et al., 2020). Therefore, it is likely to assume that socio-economic status predicts individual health outcomes as well as the likelihood of exposure to the environment. According to the literature, a way to deal with confounding in G×E studies is to include the confounder–by–environment and confounder–by–gene interaction as controls (Mills et al., 2020). The variable to represent socio-economic status is a binary indicator that is 0 when an individual has an education lower than university and 1 if someone has a university degree or higher. The results are shown in Table 10 below. The main coefficients of interest are the interaction terms for the G×E effects. Firstly for SRH, compared with the interaction coefficient in the main analysis (-0.0620), Column (1) reports now a significant interaction term. This suggests that after controlling for confounding effects of SES, individuals with a greater predisposition for BMI report significantly lower SRH scores. Secondly, the interaction term for BMI remains rather unchanged and still reports statistical significance. Therefore, focussing on the main interaction effects of SRH and BMI, SES confounding does not seem to affect the main outcomes.

Table 10. panel data regressions including covariates for confounder-by-environment and confounder-by-gene as controls.

VARIABLES	(1) SRH	(2) Lung disease	(3) BMI
PGS for BMI	-0.0776***	0.000599	1.364***
	(0.0178)	(0.00739)	(0.121)
Covid environment	0.0161	-0.0621***	0.572
	(0.0672)	(0.0127)	(0.704)
Covid environment × PGS for BMI	-0.0651*	0.0114	-0.988**
	(0.0382)	(0.0107)	(0.386)
Constant	0.151	-0.0570	38.28***
	(0.125)	(0.0386)	(0.760)
Observations	8,041	8,044	8,010
N	3,500	3,500	3,493

Note: table excludes coefficients for covariates (age, number of comorbidities, religion (SRH), gender, smoking- and drinking status; confounder-by-environment and confounder-by-gene) Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

7. Conclusion

This thesis considered the interplay between genetic predisposition to increased BMI and a measure for the COVID-19 environment in explaining subjective- and objective health outcomes, using data on the elderly U.S. population from the Health and Retirement Study. The three health outcomes considered are self-rated health, the presence of lung disease and BMI. The theoretical foundation for the hypotheses is the diathesis-stress model and, supported by the empirical literature, informed the conceptual framework of the study. The main hypothesis proposed that health outcomes vary according to genetic predisposition to increased BMI during the pandemic. To gain deeper insights into the potential underlying mechanisms driving heterogeneities in health outcomes, I used panel data regression, along with stratification according to relative genetic risk. This thesis aimed to answer the following three sub-questions:

How did the COVID-19 environment affect subjective- and objective health outcomes of the elderly (50+) population in the U.S.? Exposure to the environment affects self-rated health scores and BMI adversely. This is in line with the hypothesis that the individual effect of exposure to the environment contributes to poorer health. This is in line with previous literature that reports a decline in self-rated health under the pandemic (Lüdecke and van dem Knesebeck (2023)), and an increase in BMI scores (Restrepo, 2022). More specifically, my results show that individuals who know someone who passed away from COVID-19 report 0.126 standard deviations lower SRH, and BMI scores that were 0.871 points (kg/m^2) higher. This indicates a clear association between experiencing the pandemic as a negative shock and poorer health outcomes in these areas. However, the association with the presence of lung disease remains unclear, with estimated coefficients close to zero and statistically insignificant. This may confirm that the environment captures grief effects, which do not have a direct impact on lung functioning over direct effect of infection during the pandemic.

How do these outcomes vary across underlying genetic predispositions for complex health outcomes? Whether health effects are different for individuals with greater predispositions under the pandemic remains unclear. The analysis revealed significant evidence that individuals with greater genetic predispositions report 0.993 point (kg/m^2) lower BMI under the environment, compared to those that were unexposed. When stratified by relative genetic risk, it appears that these results are more pronounced for the genetically low-risk group. Individuals with relatively low genetic predisposition to BMI have on average 1.589 points (kg/m^2) higher BMI, contradicting the hypothesis that worse health outcomes are more pronounced for those with greater genetic risk.

For self-rated health scores, both high- and low-genetic-risk groups report lower SRH when experiencing the shock, with reductions of 0.135 (0.127) and 0.123 (0.114) standard deviations, respectively. This finding indicates that the main driver of SRH heterogeneity is the experience of someone's passing.

Is there a role of selection based on genetic predisposition (gene-environment correlation), in the COVID-19 environment? In terms of self-selection effects in the COVID-environment, the evidence suggests that this is unlikely. I do not have sufficient evidence to establish any significant relationship between the polygenic risk scores for BMI on the likelihood of knowing someone who passed from COVID. Consequently, these results imply that selection into the environment based on genetic predisposition is random, with no evidence supporting gene-environment correlation. This makes the environment appropriate for the G×E analysis.

Overall, I established an association between the environment and worse self-perceived health and higher BMI. The extent to which this effect is different by genetic risk remains inconclusive. Although the analysis indicates an interaction between genes and the environment in predicting BMI outcomes, these findings do not align with the initial hypothesis. Regardless of relative genetic risk, both high- and low-risk groups are statistically significantly affected by the COVID-environment, reporting lower self-rated health scores in a similar absolute magnitude.

8. Discussion and Policy Implications

8.1. Comparison with existing literature

A growing body of literature suggests that genetics play a role in establishing complex health outcomes. Studies identified links between genes and phenotypes, often considering the role of the environment. To the best of my knowledge, this study is among the first to examine the association between health measures and polygenic risk scores for BMI in the context of the COVID-19 pandemic among elderly U.S. individuals. Previous research on genetic predisposition focused on the effect of exposure to different environments in shaping phenotypes (Meaney et al., 2010; Nelson et al., 2006; Boardman et al., 2014). The polygenic risk score for BMI has been for example been linked to increased risk for depression when faced with early-life stress (Avinun and Hariri, 2019); increased BMI under obesogenic environments for different birth cohorts (Walter et al., 2016) and BMI increases associated with job losses (Schmitz et al., 2021). Existing literature consistently shows a negative impact of increased BMI as the presence of the phenotype on overall health (Wyatt et al., 2006).

Research on the COVID-19 pandemic has typically emphasised the direct health consequences after infection or identifying factors that increase the risk of poor outcomes (Singh et al.,2024; Biswas et al., 2021; Niemie et al., 2022; Boutin et al., 2021; Alkhouli et al., 2020), rather than how the pandemic as an event, affected other health measures. The findings in this thesis align with the general perception that health outcomes under the pandemic worsened, particularly in terms of an individual's self-perceived health status and increased BMI. It suggests a role for underlying genetic predisposition for BMI in at least partly explaining the observed heterogeneity.

To contextualize the relevance of this study, it is important to consider that the main focus of this thesis is on subjective- and objective health measures as proxies for the pandemic's health effects. These measures, however, do not capture the direct effects of infection and are therefore no suitable predictors of mortality or hospitalization among elderly individuals, which is a more direct COVID measure. From an economic perspective, mortality and hospitalization are critical measures as they are directly related to hospital spending, especially for elderly adults. Nevertheless, the health outcomes considered in this thesis are, according to the literature, predictors for future hospitalization and mortality. Additionally, they are also important aspects of human capital. Understanding how exposure to the pandemic affected human capital is

another contribution of this thesis. These proxies provide valuable insight into health behaviour during the pandemic, particularly as they reflect general health conditions.

8.2. Contradictory Effects of Gene-Environment Interaction on BMI

The figures in Section 5.2.1 establish a positive relation between BMI polygenic scores and reported BMI, indicating that higher polygenic scores predict are predictive for BMI scores. This association is further validated by the regression results of BMI PGS on BMI outcomes, which confirms a positive relationship, confirming that the polygenic scores are predictive of the presence of higher BMI. This is in line with other studies using BMI PGS for predicting BMI utilizing HRS data (Walter et al., (2016); Thompson et al., (2020); Stephan et al., (2020)).

The main hypothesis is that individuals with greater genetic predisposition experiencing the negative shock have higher BMI. This would be reflected by a positive G×E interaction term. The main analysis (Table 5) establishes a negative G×E association for BMI, which is not in line with the general hypothesis and the Diathesis-Stress model that higher-risk individuals are increasingly adversely affected by the pandemic environment. Rather, these results suggest that individuals with higher BMI PGS under the negative shock report as much as 1 point (kg/m^2) lower BMI. This suggests that the genetic effect is dampened when an individual knows someone who passed from COVID-19. Since the overall effect on BMI is less than what would be expected when the presence of both factors (higher BMI PGS and knowing someone who passed from COVID-19) are added separately. This may imply some form of resilience or coping mechanism where exposure to the environment dampens the genetic impact on BMI. Experiencing the death of someone due to COVID-19 may lead to health-related changes that particularly mitigate the genetic risk for increased BMI, such as increased health awareness of changes in diet and or exercise patterns.

In terms of existing literature, this is the closest related to research performed by Schmitz et al., (2021). The authors interact the BMI PGS with business closure as a measure for unexpected job loss during the early phase of the pandemic to estimate the effect on BMI. They report that experiencing sudden job loss did not alter the genetic effect on BMI. However, they report only 375 individuals in the affected group, which has implications for the predictive power of the study.

A second explanation is that it may be that genetic predisposition to higher BMI exhibits different effects under the pandemic. Nagata et al., (2019) find that individuals with elevated genetic risk for BMI tend to engage more in weight-losing behaviour for both males and females. The authors suggest that these individuals are not only genetically prone to developing high BMI but also to potentially unhealthy weight loss behaviours. Hence, under adverse circumstances such as the pandemic, is it possible that the behaviours, mediated by genetic risk, manifest weight loss rather than increased BMI scores.

Thirdly, genes predisposing individuals to higher BMI can exhibit pleiotropy. This means that genes affect more than one phenotype (Lobo, 2014). It may be that the SNPs of certain genes strongly affect other phenotypes under the environmental conditions, dominating the effect these have on increased BMI. Evidence supports that pleiotropy exists for BMI, as multiple loci have been identified to also be associated with the risk for anorexia nervosa (Hinney et al., 2017). Nevertheless, there is limited research that explores this possible relationship between polygenic scores for BMI and weight-losing behaviour. Based on the findings of this thesis, this could be an avenue for future research.

The abovementioned explanation suggests that opposite interaction effects, compared to the hypotheses, may stem from the pleiotropic nature of the polygenic score. However, these contradictory findings may also be explained by 1) bias introduced by measurement errors (Mills et al., 2020); 2) an interaction driven by gene-environment correlations rather than their interaction, leading to a non-random distribution of risk scores across the environment (Biroli et al., 2022) or 3) inappropriate scaling of the outcome variable that may affect the direction of the interaction term (Murray et al., 2020). Firstly, BMI scores may suffer from measurement error, which may occur due to how these scores are constructed. In the HRS study, the respondent's height is only measured during the first interview and carried forward for subsequent interviews. Weight is remeasured each wave. Both measures are in U.S. metrics For BMI calculations, the HRS converts them to kilograms and meters. However, the scale used for this conversion is unspecified. Since BMI scores were not readily available for the latest wave, I relied on commonly used conversion methods. Nonetheless, the descriptive statistics for the calculated BMI scores in the latest wave align with those of previous waves, suggesting consistency.

Second, gene-environment correlation: individuals with greater genetic risk to BMI might self-select in environments where genetic risk is also higher - or phenotype is expressed- by for

example unhealthier lifestyles. These individuals may already more frequently experience the negative shock. However, the regression in 6.1. (Table 4) suggests that this is highly unlikely for this sample. Thirdly, inappropriate scaling issues. The literature suggests rescaling of the outcome variable to solve this. Therefore, I rescaled the BMI scores to an approximately normal distribution via log transformation. Running the baseline analysis and stratification by relative genetic risk, the direction of the effects remain the same in the main analysis as well as the for the stratification by high and low- risk.

8.3. General limitations

There are several general limitations to this study. Firstly, the dataset it is restricted to elderly people (50+) in the United States, which may limit the generalizability to other age groups and other countries. However, the HRS data is a comprehensive and unique dataset that combines data on individual physical health, recent COVID statistics and socioeconomic variables with genetic scores of elderly participating individuals. Therefore, the findings of this thesis can offer valuable insights into health outcomes among elderly individuals during the pandemic.

Another concern regarding the generalizability is that everyone in the analytic sample is vaccinated against COVID-19. This does not reflect the broader U.S. population, as research indicates a decline in willingness to vaccinate against COVID-19 (Koskan et al., 2023). Additionally, Daly and Robinson (2020) found that half of the Americans were undecided or unwilling to be vaccinated. Nevertheless, this does provide the potential to study health outcomes in a sample when vaccination coverage is high. This can offer valuable information for public health strategies and vaccination policy. This can be a route for future research.

Considering the polygenic scores there are also some limitations. The GWAS used for the PGS comes from European ancestry groups. While this is the main focus of the thesis, these PGS scores are not as predictive for other ancestry groups. PGS are genetic predictors for outcomes which only holds for the characteristics of the discovery sample and are thus limited in applicability and accuracy across different ancestries. This implies that the results of this thesis would be the most accurate for European ancestry populations. This reduces the external validity of the empirical findings. However, this research is still informative for the sample of interest and can be replicated in other samples once the data becomes increasingly available.

8.4. Implications for policymakers

This thesis found two key insights that are relevant to policymaking. Firstly, an association between genetic risk for BMI on self-rated health scores and actual BMI was established. This finding can inform policy aimed at identifying individuals with greater genetic risk to certain conditions, to potentially reduce adverse health outcomes and associated treatment costs. Genetic screening can be a powerful tool in predicting the individual risk of developing complex diseases. For example, preventive screening for a subset of predictive breast cancer genes (BRCA1 and BRCA2) in U.S. women led to the option for preventive risk-reducing measures in those who tested positive (Nelson et al., 2019). Specifically, carrying this genetic mutation combined with environmental exposure, such as tobacco use, has been linked to an increased risk of developing breast cancer (Kang et al., 2013). Therefore, from a policy perspective, polygenic ranking of individuals during a public health crisis to identify susceptible groups can be valuable for preventive policies. However, a potential concern is that focusing on genetic susceptibility may shift policy emphasis towards individual-level interventions, rather than addressing economic and environmental factors on a broader scale. Additionally, polygenic risk scores contain many SNPs and do not provide a diagnosis. Therefore policy based on polygenic ranking is subjected to large uncertainty (Muslimova et al., 2023).

Secondly, to address the environmental component, I established that the COVID-environment is associated with more negative health outcomes. In particular, elderly individuals who experienced the loss of someone in their social circle due to COVID-19 report lower self-rated health scores, and higher BMI. Therefore, targeted health interventions to support the elderly population especially those that are faced with sudden bereavement of someone in their social circle would be a policy recommendation. These could be mental and physical health support. This support should aim to address the adverse health effects of exposure to the environment. Secondly, physical health programmes, to mitigate the observed increases in BMI during the pandemic. Examples could be community-based physical activity and nutritional programmes tailored and accessible to elderly individuals.

Additionally, the findings of these thesis could motivate policy aimed at defining a susceptible sub-population to reduce future healthcare spending. As noted in the introduction, the pandemic led to a significant increase in COVID-related healthcare expenses, particularly for the elderly. Stratifying by genetic risk or susceptibility to certain environments could help lower financial costs.

Nevertheless, understanding the heterogeneity of environmental effects by genetic risk can be informative, as it gives insights into the underlying biological and environmental mechanisms that drive health outcomes.

8.5. Suggestions for further research

From a contextual perspective, it is important to realize that the field of GxE research is still evolving. Further research is needed to gain more insight into how genes and the environment interplay in establishing health outcomes. More specifically, one of the most prominent outcomes of COVID-19 is lung functioning. In this study, I fail to establish any significant association. It is likely that knowing someone who passed as the COVID-environment captures grief effects, over direct risk for disease manifestation under the pandemic. This is supported by the fact that no significant associations are found regarding the presence of lung disease. Additionally, the genetic predisposition for BMI may be insufficiently linked to lung functioning to explain observed heterogeneity. Future studies can utilize polygenic scores for different lung diseases to explain health outcomes. Lastly, there is evidence that lung disease in this study is chronic rather than induced by the pandemic. However, it could be that chronic lung disease worsened, but this cannot be tested with the available data. Nevertheless, if data is increasingly available, this could be studied in future research.

9. References

Ahrenfeldt, L. J., Möller, S., Thinggaard, M., Christensen, K., & Lindahl-Jacobsen, R. (2019). Sex Differences in Comorbidity and Frailty in Europe. *International Journal of Public Health*, *64*(7), 1025–1036. https://doi.org/10.1007/s00038-019-01270-9

Akgun, K. M., Crothers, K., & Pisani, M. (2012). Epidemiology and Management of Common Pulmonary Diseases in Older Persons. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *67A*(3), 276–291. https://doi.org/10.1093/gerona/glr251

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *The Structure and Function of DNA*. Nih.gov; Garland Science. https://www.ncbi.nlm.nih.gov/books/NBK26821/

Alcohol Use Among Older Adults. (2021). National Poll on Healthy Aging. https://www.healthyagingpoll.org/reports-more/report/alcohol-use-among-older-adults

Alkhouli, M., Nanjundappa, A., Annie, F., Bates, M. C., & Bhatt, D. L. (2020). Sex Differences in Case Fatality Rate of COVID-19: Insights From a Multinational Registry. *Mayo Clinic Proceedings*, *95*(8), 1613–1620. https://doi.org/10.1016/j.mayocp.2020.05.014

Arold, B. W., Hufe, P., & Stoeckli, M. (2022). Genetic Endowments, Educational Outcomes and the Mediating Influence of School Investments. *CESifo Working Paper Series*. https://ideas.repec.org/p/ces/ceswps/_9841.html

Arpey, N. C., Gaglioti, A. H., & Rosenbaum, M. E. (2017). How socioeconomic status affects patient perceptions of health care: A qualitative study. *Journal of Primary Care & Community Health*, *8*(3), 169–175. https://doi.org/10.1177/2150131917697439

Avinun, R., & Hariri, A. R. (2019). A polygenic score for body mass index is associated with depressive symptoms via early life stress: Evidence for gene-environment correlation. *Journal of Psychiatric Research*, *118*, 9–13. https://doi.org/10.1016/j.jpsychires.2019.08.008

Barakat, C., & Konstantinidis, T. (2023). A Review of the Relationship between Socioeconomic Status Change and Health. *International Journal of Environmental Research and Public Health*, *20*(13), 6249. https://doi.org/10.3390/ijerph20136249

Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences*, *115*(42), E9765–E9772. https://doi.org/10.1073/pnas.1802909115

Barceló, A., Gregg, E. W., Pastor-Valero, M., & Robles, S. C. (2007). Waist circumference, BMI and the prevalence of self-reported diabetes among the elderly of the United States and six cities of Latin America and the Caribbean. *Diabetes Research and Clinical Practice*, *78*(3), 418–427. https://doi.org/10.1016/j.diabres.2007.06.008

Barthel, M., Mitchell, A., & Holcomb, J. (2016, December 15). *Many Americans Believe Fake News Is Sowing Confusion*. Pew Research Center's Journalism Project; Pew Research Center. https://www.pewresearch.org/journalism/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology, 51(6), 1173.

Basile, K. J., Johnson, M. E., Xia, Q., & Grant, S. F. A. (2014). Genetic Susceptibility to Type 2 Diabetes and Obesity: Follow-Up of Findings from Genome-Wide Association Studies. International Journal of Endocrinology, 2014, 1–13. https://doi.org/10.1155/2014/769671

Beauchamp, J. P., Cesarini, D., Johannesson, M., van der Loos, M. J. H. M., Koellinger, P. D., Groenen, P. J. F., Fowler, J. H., Rosenquist, J. N., Thurik, A. R., & Christakis, N. A. (2011). Molecular Genetics and Economics. *Journal*

of Economic Perspectives, *25*(4), 57–82. https://doi.org/10.1257/jep.25.4.57

Berger, D., Herkenhoff, K., & Mongey, S. (2020, March 1). *An SEIR Infectious Disease Model with Testing and Conditional Quarantine*. RePEc - Econpapers. https://econpapers.repec.org/paper/nbrnberwo/26901.htm

Bierut, L., Biroli, P., Galama, T. J., & Thom, K. (2023). Challenges in studying the interplay of genes and environment. A study of childhood financial distress moderating genetic predisposition for peak smoking. *Journal of Economic Psychology*, *98*(102636), 1–14. https://doi.org/10.1016/j.joep.2023.102636

Biroli, P., Galama, T., Hinke, S. von, Kippersluis, H. van, Rietveld, C., & Thom, K. (2022). The Economics and Econometrics of Gene-Environment Interplay. *Tinbergen Institute Discussion Papers*. https://ideas.repec.org/p/tin/wpaper/20220019.html

Biswas, M., Rahaman, S., Biswas, T. K., Haque, Z., & Ibrahim, B. (2020). Association of Sex, Age, and Comorbidities with Mortality in COVID-19 Patients: A Systematic Review and Meta-Analysis. *Intervirology*, *64*(1), 1–12. https://doi.org/10.1159/000512592

Boardman, J. D., Domingue, B. W., Blalock, C. L., Haberstick, B. C., Kathleen Mullan Harris, & McQueen, M. J. (2014). Is the Gene-Environment Interaction Paradigm Relevant to Genome-Wide Studies? The Case of Education and Body Mass Index. *Demography*, *51*(1), 119–139. https://doi.org/10.1007/s13524-013-0259-4

Bombak, A. E. (2013). Self-Rated Health and Public Health: A Critical Perspective. *Frontiers in Public Health*, *1*(15). https://doi.org/10.3389/fpubh.2013.00015

Boutin, S., Hildebrand, D., Boulant, S., Kreuter, M., Rüter, J., Pallerla, S. R., Velavan, T. P., & Nurjadi, D. (2021). Host factors facilitating SARS-CoV-2 virus infection and replication in the lungs. *Cellular and Molecular Life* Sciences, 78(16), 5953–5976. https://doi.org/10.1007/s00018-021-03889-5

Brandkvist, M., Bjørngaard, J. H., Ødegård, R. A., Åsvold, B. O., Sund, E. R., & Vie, G. Å. (2019). Quantifying the impact of genes on body mass index during the obesity epidemic: longitudinal findings from the HUNT Study. bmj, 366.

Buck, D., & Frosini, F. (2012). Clustering of unhealthy behaviours over time: implications for policy and practice.

Bugliari, D., Carroll, J., Hayden, O., Hayes, J., Hurd, M., Lee, S., Main, R., Mccullough, C., Meijer, E., Pantoja, P., & Rohwedder, S. (2024). *RAND HRS Longitudinal File 2020 (V2) Documentation (Final Release) Sponsored by the Social Security Administration and the National Institute on Aging*. https://hrsdata.isr.umich.edu/sites/default/files/documentation/other/1715891203/randhrs1992_2020v2.pdf

Carr, D., Boerner, K., & Moorman, S. (2020). Bereavement in the Time of Coronavirus: Unprecedented Challenges Demand Novel Interventions. *Journal of Aging & Social Policy*, *32*(4-5), 425–431. https://doi.org/10.1080/08959420.2020.1764320

Carr, D., House, J. S., Wortman, C., Nesse, R., & Kessler, R. C. (2001). Psychological Adjustment to Sudden and Anticipated Spousal Loss Among Older Widowed Persons. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *56*(4), S237–S248. https://doi.org/10.1093/geronb/56.4.s237

Case, A., & Paxson, C. H. (2005). Sex Differences in Morbidity and Mortality. *Demography*, *42*(2), 189–214. https://doi.org/10.1353/dem.2005.0011

Caspi, A. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, *301*(5631), 386–389. https://doi.org/10.1126/science.1083968

CDC. (2020, February 11). *COVID-19 and Your Health*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html#:~:text=Older%20adults%20are%20at%20highest

CDC. (2022). *About adult BMI* . Centers for Disease Control and Prevention. https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html

CDC. (2023, May 4). *Current Cigarette Smoking Among Adults in the United States*. Smoking and Tobacco Use; CDC. https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm

Chatterjee, S., Lakshmi Vineela Nalla, Sharma, M., Sharma, N., Singh, A. A., Fehmina Mushtaque Malim, Manasi Ghatage, Mohd Mukarram, Pawar, A., Nidhi Parihar, Arya, N., & Amit Khairnar. (2023). Association of COVID-19 with Comorbidities: An Update. *ACS Pharmacology & Translational Science*, *6*(3), 334–354. https://doi.org/10.1021/acsptsci.2c00181

Chen, Y.-T. (2022). Effect of vaccination patterns and vaccination rates on the spread and mortality of COVID-19 pandemic. *Health Policy and Technology*, 100699. https://doi.org/10.1016/j.hlpt.2022.100699

Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759–2772. https://doi.org/10.1038/s41596-020-0353-1

Cleary, P. D. (1997). Subjective and Objective Measures of Health: Which is Better When? *Journal of Health Services Research & Policy*, *2*(1), 3–4. https://doi.org/10.1177/135581969700200102

Cutler, D. M., Lleras-Muney, A., & Vogl, T. (2008, September 10). *Socioeconomic Status and Health: Dimensions and Mechanisms*. Www.nber.org. https://www.nber.org/papers/w14333

Daly, M., & Robinson, E. (2021). Willingness to Vaccinate Against COVID-19 in the U.S.: Representative Longitudinal Evidence From April to October 2020. *American Journal of Preventive Medicine*. https://doi.org/10.1016/j.amepre.2021.01.008

de Godoy, M. R. P., Machado Salomão, M. L., Queiroz, F., & de Godoy, J. M. P. (2023). Prevalence of COVID-19 Among 8000 Patients at a Hospitalized and Mortality in Old Age. *Current Aging Science*. https://doi.org/10.2174/1874609816666230525124934

de Vries, L. P., van de Weijer, M. P., Pelt, D. H. M., Ligthart, L., Willemsen, G., Boomsma, D. I., de Geus, E. J. C., & Bartels, M. (2021). Gene-by-Crisis Interaction for Optimism and Meaning in Life: The Effects of the COVID-19 Pandemic. *Behavior Genetics*. https://doi.org/10.1007/s10519-021-10081-9

Dietz, W., & Santos-Burgoa, C. (2020). Obesity and its Implications for COVID-19 Mortality. *Obesity*, *28*(6). https://doi.org/10.1002/oby.22818

Djalalinia, S., Qorbani, M., Peykari, N., & Kelishadi, R. (2015). Health Impacts of Obesity. *Pakistan Journal of Medical Sciences*, *31*(1). https://doi.org/10.12669/pjms.311.7033

Dramé, M., Cantegrit, E., & Godaert, L. (2023). Self-Rated Health as a Predictor of Mortality in Older Adults: A Systematic Review. *International Journal of Environmental Research and Public Health*, *20*(5), 3813. https://doi.org/10.3390/ijerph20053813

Dunn, A. R., O'Connell, K. M. S., & Kaczorowski, C. C. (2019). Gene-by-environment interactions in Alzheimer's disease and Parkinson's disease. *Neuroscience & Biobehavioral Reviews*, *103*, 73–80. https://doi.org/10.1016/j.neubiorev.2019.06.018

Eisma, M. C., Boelen, Paul. A., & Lenferink, L. I. M. (2020). Prolonged Grief Disorder following the Coronavirus (COVID-19) Pandemic. *Psychiatry Research*, 113031. https://doi.org/10.1016/j.psychres.2020.113031

Eisma, M. C., Tamminga, A., Smid, G. E., & Boelen, P. A. (2021). Acute grief after deaths due to COVID-19, natural causes and unnatural causes: An empirical comparison. *Journal of Affective Disorders*, *278*, 54–56. https://doi.org/10.1016/j.jad.2020.09.049

Ejaz, H., Alsrhani, A., Zafar, A., Javed, H., Junaid, K., Abdalla, A. E., Abosalif, K. O. A., Ahmed, Z., & Younas, S. (2020). COVID-19 and comorbidities: Deleterious impact on infected patients. *Journal of Infection and Public Health*, *13*(12). https://doi.org/10.1016/j.jiph.2020.07.014

Elbourne, A. (2021). The Effect of the Pandemic and Non-pharmaceutical Interventions on Household Consumption. CPB Netherlands Bureau for Economic Policy Analysis.

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., Grimsrud, M. M., Milani, C., Aziz, F., Kässens, J., May, S., Wendorff, M., Wienbrandt, L., Uellendahl-Werth, F., Zheng, T., & Yi, X. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *New England Journal of Medicine*. https://doi.org/10.1056/nejmoa2020283

Elrobaa, I. H., & New, K. J. (2021). COVID-19: Pulmonary and Extra Pulmonary Manifestations. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.711616

Fang, X., Li, S., Yu, H., Wang, P., Zhang, Y., Chen, Z., Li, Y., Cheng, L., Li, W., Jia, H., & Ma, X. (2020). Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis. *Aging*, *12*(13), 12493–12503. https://doi.org/10.18632/aging.103579

Farooqi, R., Kooner, J. S., & Zhang, W. (2023). Associations between polygenic risk score and covid-19 susceptibility and severity across ethnic groups: UK Biobank analysis. *BMC Medical Genomics*, *16*(1). https://doi.org/10.1186/s12920-023-01584-x

Feng, S., Song, F., Guo, W., Tan, J., Zhang, X., Qiao, F., Guo, J., Zhang, L., & Jia, X. (2022). Potential Genes Associated with COVID-19 and Comorbidity. *International Journal of Medical Sciences*, *19*(2), 402–415. https://doi.org/10.7150/ijms.67815

Finkel, D., Gatz, M., Franz, C. E., Catts, V. S., Christensen, K., Kremen, W., Nygaard, M., Plassman, B. L., Sachdev, P. S., Whitfield, K., & Pedersen, N. L. (2024). Age and sex differences in the genetic architecture of measures of subjective health: Relationships with physical health, depressive symptoms, and episodic memory. *the Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*. https://doi.org/10.1093/geronb/gbae062

Foo, J., Streit, F., Frank, J., Zacharias, N., Zillich, L., Sirignano, L., Völker, M., Nürnberg, P., Wienker, T., Wagner, M., Nöthen, M., Nothnagel, M., Walter, H., Lenz, B., Spanagel, R., Kiefer, F., Winterer, G., Rietschel, M., & Witt, S. (2022). *Polygenic risk scores for nicotine use and family history of smoking are associated with smoking behaviou*. Europe PMC. https://europepmc.org/article/ppr/ppr586191

Franz, C. E., Finkel, D., Panizzon, M. S., Spoon, K., Christensen, K., Gatz, M., Kremen, W. S., Krueger, R., Neiderhiser, J., Reynolds, C., & Pedersen, N. L. (2016). Facets of Subjective Health From Early Adulthood to Old Age. *Journal of Aging and Health*, *29*(1), 149–171. https://doi.org/10.1177/0898264315625488

Friedman, N. P., Banich, M. T., & Keller, M. C. (2021). Twin studies to GWAS: there and back again. *Trends in Cognitive Sciences*, *25*(10), 855–869. https://doi.org/10.1016/j.tics.2021.06.007

Gerayeli, F. V., Milne, S., Cheung, C., Li, X., Yang, C. W. T., Tam, A., Choi, L. H., Bae, A., & Sin, D. D. (2021). COPD and the risk of poor outcomes in COVID-19: A systematic review and meta-analysis. *EClinicalMedicine*, *33*, 100789. https://doi.org/10.1016/j.eclinm.2021.100789

Goveas, J. S., & Shear, M. K. (2020). Grief and the COVID-19 pandemic in older adults. The American Journal of Geriatric Psychiatry, 28(10), 1119-1125.

Guerrero, L. R., & Wallace, S. P. (2021). The Impact of COVID-19 on Diverse Older Adults and Health Equity in the United States. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.661592

Guo, H., Li, T., & Wen, H. (2022). Identifying shared genetic loci between coronavirus disease 2019 and cardiovascular diseases based on cross-trait meta-analysis. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.993933

Guralnik, J. M., Balfour, J. L., & Volpato, S. (2000). The ratio of older women to men: Historical perspectives and cross-national comparisons. *Aging Clinical and Experimental Research*, *12*(2), 65–76. https://doi.org/10.1007/bf03339893

Gutierrez, J. P., & Bertozzi, S. M. (2020). Non-communicable diseases and inequalities increase risk of death among

COVID-19 patients in Mexico. *PLOS ONE*, *15*(10), e0240394. https://doi.org/10.1371/journal.pone.0240394

Hagenbeek, F. A., Hirzinger, J. S., Breunig, S., Bruins, S., Kuznetsov, D. V., Schut, K., Odintsova, V. V., & Boomsma, D. I. (2023). Maximizing the value of twin studies in health and behaviour. *Nature Human Behaviour*, 1–12. https://doi.org/10.1038/s41562-023-01609-6

Hallas, L., Hatibie, A., Koch, R., Pyarali, M., Baylor, Wood, A., & Hale, T. (2021). *Variation in US states' COVID-19 policy responses BSG Working Paper Series Providing access to the latest policy-relevant research*. https://www.bsg.ox.ac.uk/sites/default/files/2021-05/BSG-WP-2020-034-v3.pdf

Harris, S. E., Hagenaars, S. P., Davies, G., David Hill, W., Liewald, D. C. M., Ritchie, S. J., Marioni, R. E., Sudlow, C. L. M., Wardlaw, J. M., McIntosh, A. M., Gale, C. R., & Deary, I. J. (2016). Molecular genetic contributions to self-rated health. *International Journal of Epidemiology*, dyw219. https://doi.org/10.1093/ije/dyw219

Hawkins, R. B., Charles, E. J., & Mehaffey, J. H. (2020). Socio-economic status and COVID-19–related cases and fatalities. *Public Health*, *189*, 129–134. https://doi.org/10.1016/j.puhe.2020.09.016

Hinney, A., Kesselmeier, M., Jall, S., Volckmar, A-L., Föcker, M., Antel, J., Heid, I. M., Winkler, T. W., Grant, S. F. A., Guo, Y., Bergen, A. W., Kaye, W., Berrettini, W., Hakonarson, H., Herpertz-Dahlmann, B., de Zwaan, M., Herzog, W., Ehrlich, S., Zipfel, S., & Egberts, K. M. (2016). Evidence for three genetic loci involved in both anorexia nervosa risk and variation of body mass index. *Molecular Psychiatry*, *22*(2), 192–201. https://doi.org/10.1038/mp.2016.71

Hobbs, B. D., De Jong, K., Lamontagne, M., Bossé, Y., Shrine, N., Artigas, M. S., ... & International COPD Genetics Consortium. (2017). Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nature genetics, 49(3), 426-432.

Hooper, M., Reinhart, M., Dusetzina, S. B., Walsh, C. G., & Griffith, K. N. (2023). Trends in U.S. self-reported health and self-care behaviors during the COVID-19 pandemic. *PLOS ONE*, *18*(9), e0291667–e0291667. https://doi.org/10.1371/journal.pone.0291667

Horowitz, J. E., Kosmicki, J. A., Damask, A., Sharma, D., Roberts, G. H. L., Justice, A. E., Banerjee, N., Coignet, M. V., Yadav, A., Leader, J. B., Marcketta, A., Park, D. S., Lanche, R., Maxwell, E., Knight, S. C., Bai, X., Guturu, H., Sun, D., Baltzell, A., & Kury, F. S. P. (2022). Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. *Nature Genetics*, *54*(4), 382–392. https://doi.org/10.1038/s41588-021-01006-7

Huang, Q.-M., Zhang, P.-D., Li, Z.-H., Zhou, J.-M., Liu, D., Zhang, X.-R., Zhong, W.-F., Zhang, Y.-J., Shen, D., Liang, F., Song, W.-Q., Yang, S.-G., Guan, W.-J., & Mao, C. (2022). Genetic Risk and Chronic Obstructive Pulmonary Disease Independently Predict the Risk of Incident Severe COVID-19. *Annals of the American Thoracic Society*, *19*(1), 58–65. https://doi.org/10.1513/annalsats.202102-171oc

Huguet, A., Kaufman, J. H., & Diliberti, M. K. (2024, February 28). *Media Habits and Misinformation Susceptibility of Adults Aged 55 Years and Older: Findings from a RAND American Life Panel Survey*. Www.rand.org. https://www.rand.org/pubs/research_reports/RRA2909-1.html

Ishak, A., Mehendale, M., AlRawashdeh, M. M., Sestacovschi, C., Sharath, M., Pandav, K., & Marzban, S. (2022). The association of COVID-19 severity and susceptibility and genetic risk factors: A systematic review of the literature. *Gene*, *836*, 146674. https://doi.org/10.1016/j.gene.2022.146674

Jackson, M., Marks, L., May, G. H. W., & Wilson, Joanna B. (2018). The genetic basis of disease. *Essays in Biochemistry*, *62*(5), 643–723. https://doi.org/10.1042/ebc20170053

Jackson, S. E., Llewellyn, C. H., & Smith, L. (2020). The obesity epidemic – Nature via nurture: A narrative review of high-income countries. *SAGE Open Medicine*, *8*(1), 205031212091826. https://doi.org/10.1177/2050312120918265

Johansson, H., Mersha, T. B., Brandt, E. B., & Hershey, G. K. K. (2019). Interactions between environmental pollutants and genetic susceptibility in asthma risk. *Current Opinion in Immunology*, *60*, 156–162. https://doi.org/10.1016/j.coi.2019.07.010

Kang, H. J., Hong, Y. B., Yi, Y. W., Cho, C.-H., Wang, A., & Bae, I. (2013). Correlations between BRCA1 defect and environmental factors in the risk of breast cancer. *The Journal of Toxicological Sciences*, *38*(3), 355–361. https://doi.org/10.2131/jts.38.355

Kerner, G., & Quintana-Murci, L. (2022). The genetic and evolutionary determinants of COVID-19 susceptibility. *European Journal of Human Genetics*, *30*(8), 915–921. https://doi.org/10.1038/s41431-022-01141-7

Keyes, K. M., Pratt, C., Galea, S., McLaughlin, K. A., Koenen, K. C., & Shear, M. K. (2014). The burden of loss: Unexpected death of a loved one and psychiatric disorders across the life course in a national study. *American Journal of Psychiatry*, *171*(8), 864–871. https://doi.org/10.1176/appi.ajp.2014.13081132

Kivimäki, M., Strandberg, T., Pentti, J., Nyberg, S. T., Frank, P., Jokela, M., Ervasti, J., Suominen, S. B., Vahtera, J., Sipilä, P. N., Lindbohm, J. V., & Ferrie, J. E. (2022). Body-mass Index and Risk of Obesity-related Complex Multimorbidity: An Observational Multicohort Study. *The Lancet Diabetes & Endocrinology*, *10*(4). https://doi.org/10.1016/s2213-8587(22)00033-x

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., 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. https://doi.org/10.1126/science.aan6877

Koskan, A. M., Teeter, B. S., Daniel, C. L., LoCoco, I. E., Jensen, U. T., & Ayers, S. L. (2023). U.S. adults' reasons for changing their degree of willingness to vaccinate against COVID-19. *Zeitschrift Fur Gesundheitswissenschaften = Journal of Public Health*, 1–13. https://doi.org/10.1007/s10389-023-01819-2

Krause, N. (2010). God-Mediated Control and Change in Self-Rated Health. *International Journal for the Psychology of Religion*, *20*(4), 267–287. https://doi.org/10.1080/10508619.2010.507695

Lee, A., Seo, J., Park, S., Cho, Y., Kim, G., Li, J., Liang, L., Park, T., & Chung, W. (2024). Type 2 diabetes and its genetic susceptibility are associated with increased severity and mortality of COVID-19 in UK Biobank. *PubMed*, *7*(1). https://doi.org/10.1038/s42003-024-05799-1

Leinonen, R., Kaprio, J., JylhÃ¤, M., Tolvanen, A., Koskenvuo, M., Heikkinen, E., & Rantanen, T. (2005). Genetic Influences Underlying Self-Rated Health in Older Female Twins. *Journal of the American Geriatrics Society*, *53*(6), 1002–1007. https://doi.org/10.1111/j.1532-5415.2005.53319.x

Leong, A., Cole, J. B., Brenner, L. N., Meigs, J. B., Florez, J. C., & Mercader, J. M. (2021). Cardiometabolic risk factors for COVID-19 susceptibility and severity: A Mendelian randomization analysis. *PLOS Medicine*, *18*(3), e1003553. https://doi.org/10.1371/journal.pmed.1003553

Leung, C. (2020). Risk factors for predicting mortality in elderly patients with COVID-19: A review of clinical data in China. *Mechanisms of Ageing and Development*, *188*, 111255. https://doi.org/10.1016/j.mad.2020.111255

Lewis, C. M., & Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*, *12*(1). https://doi.org/10.1186/s13073-020-00742-5

Li, S. L., Pereira, R. H. M., Prete Jr, C. A., Zarebski, A. E., Emanuel, L., Alves, P. J. H., Peixoto, P. S., Braga, C. K. V., de Souza Santos, A. A., de Souza, W. M., Barbosa, R. J., Buss, L. F., Mendrone, A., de Almeida-Neto, C., Ferreira, S. C., Salles, N. A., Marcilio, I., Wu, C.-H., Gouveia, N., & Nascimento, V. H. (2021). Higher risk of death from COVID-19 in low-income and non-White populations of São Paulo, Brazil. *BMJ Global Health*, *6*(4), e004959. https://doi.org/10.1136/bmjgh-2021-004959

Lloyd, K., Schubotz, D., Roche, R., Manzi, J., & McKnight, M. (2023). A Mental Health Pandemic? Assessing the

Impact of COVID-19 on Young People's Mental Health. *International Journal of Environmental Research and Public Health*, *20*(16), 6550. https://doi.org/10.3390/ijerph20166550

Lobo, I. (2014). *Pleiotropy Is Not Polygenic Inheritance | Learn Science at Scitable*. Nature.com. https://www.nature.com/scitable/topicpage/pleiotropy-one-gene-can-affect-multiple-traits-569/

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., Croteau-Chonka, D. C., Esko, T., Fall, T., Ferreira, T., Gustafsson, S., Kutalik, Z., Luan, J., Mägi, R., Randall, J. C., & Winkler, T. W. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*(7538), 197–206. https://doi.org/10.1038/nature14177

Loos, R. J. F., & Yeo, G. S. H. (2021). The genetics of obesity: from discovery to biology. *Nature Reviews Genetics*, *23*(23), 1–14. https://doi.org/10.1038/s41576-021-00414-z

Lorem, G., Cook, S., Leon, D. A., Emaus, N., & Schirmer, H. (2020). Self-reported health as a predictor of mortality: A cohort study of its relation to other health measurements and observation time. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-61603-0

Lüdecke, D., & von dem Knesebeck, O. (2023). Worsened self-rated health in the course of the COVID-19 pandemic among older adults in Europe. *European Journal of Public Health*, *33*(6), 1148–1154. https://doi.org/10.1093/eurpub/ckad143

Magesh, S., John, D., Li, W. T., Li, Y., Mattingly-app, A., Jain, S., Chang, E. Y., & Ongkeko, W. M. (2021). Disparities in COVID-19 Outcomes by Race, Ethnicity, and Socioeconomic Status. *JAMA Network Open*, *4*(11), e2134147. https://doi.org/10.1001/jamanetworkopen.2021.34147

Meaney, M. J. (2010). Epigenetics and the Biological Definition of Gene × Environment Interactions. *Child Development*, *81*(1), 41–79. https://doi.org/10.1111/j.1467-8624.2009.01381.x

Meng, Q., Xie, Z., & Zhang, T. (2014). A Single-Item Self-Rated Health Measure Correlates with Objective Health Status in the Elderly: A Survey in Suburban Beijing. *Frontiers in Public Health*, *2*. https://doi.org/10.3389/fpubh.2014.00027

Mills, M. C., Barban, N., & Tropf, F. C. (2020). An introduction to statistical genetic data analysis. Mit Press.

Monroe, S. M., & Simons, A. D. (1991). Diathesis-stress theories in the context of life stress research: Implications for the depressive disorders. *Psychological Bulletin*, *110*(3), 406–425. https://doi.org/10.1037/0033-2909.110.3.406

Mosing, M. A., Zietsch, B. P., Shekar, S. N., Wright, M. J., & Martin, N. G. (2009). Genetic and Environmental Influences on Optimism and its Relationship to Mental and Self-Rated Health: A Study of Aging Twins. *Behavior Genetics*, *39*(6), 597–604. https://doi.org/10.1007/s10519-009-9287-7

Murray, A. L., Molenaar, D., Johnson, W., & Krueger, R. F. (2016). Dependence of Gene-by-Environment Interactions (GxE) on Scaling: Comparing the Use of Sum Scores, Transformed Sum Scores and IRT Scores for the Phenotype in Tests of GxE. *Behavior Genetics*, *46*(4), 552–572. https://doi.org/10.1007/s10519-016-9783-5

Muslimova, D., Pereira, R., Stephanie von Hinke, Hans Van Kippersluis, Rietveld, C. A., & S. Fleur W. Meddens. (2023). Rank concordance of polygenic indices: Implications for personalised intervention and gene-environment interplay. *BioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2022.05.03.490435

Nagar, M., Geevarughese, N. M., Mishra, R., Joshi, A., Galwankar, S., Yunus, M., ... & Agrawal, A. (2022). Body-mass index COVID-19 severity: A systematic review of systematic reviews. Journal of Family Medicine and Primary Care, 11(9), 5351-5360.

Nagata, J. M., Braudt, D. B., Domingue, B. W., Bibbins-Domingo, K., Garber, A. K., Griffiths, S., & Murray, S. B.

(2019). Genetic risk, body mass index, and weight control behaviors: Unlocking the triad. *International Journal of Eating Disorders*, *52*(7), 825–833. https://doi.org/10.1002/eat.23083

National Human Genome Research Institute. (2018, September 7). *Genetics vs. Genomics Fact Sheet*. Genome.gov. https://www.genome.gov/about-genomics/fact-sheets/Genetics-vs-Genomics#:~:text=All%20human%20beings%20are%2099.9

Nelson, H. D., Pappas, M., Cantor, A., Haney, E., Holmes, R., & Stillman, L. (2019). Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA1/2-Related Cancer in Women: A Systematic Review for the U.S. Preventive Services Task Force. In *PubMed*. Agency for Healthcare Research and Quality (US). https://www.ncbi.nlm.nih.gov/books/NBK545867/

Nelson, M. C., Gordon-Larsen, P., North, K. E., & Adair, L. S. (2006). Body Mass Index Gain, Fast Food, and Physical Activity: Effects of Shared Environments over Time*. *Obesity*, *14*(4), 701–709. https://doi.org/10.1038/oby.2006.80

Ni, J., Qiu, L., Yin, K.-J., Chen, G.-M., & Pan, H. (2022). Shared genetic architecture between type 2 diabetes and COVID-19 severity. *Journal of Endocrinological Investigation*, *46*(3), 501–507. https://doi.org/10.1007/s40618-022-01920-5

Niemi, M. E. K., Daly, M. J., & Ganna, A. (2022). The human genetic epidemiology of COVID-19. *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-022-00478-5

Ohno, M., & Dzúrová, D. (2022). Body Mass Index and Risk for COVID-19-Related Hospitalization in Adults Aged 50 and Older in Europe. *Nutrients*, *14*(19), 4001. https://doi.org/10.3390/nu14194001

Pairo-Castineira, E., Rawlik, K., Bretherick, A. D., Qi, T., Wu, Y., Nassiri, I., McConkey, G. A., Zechner, M., Klaric, L., Griffiths, F., Oosthuyzen, W., Kousathanas, A., Richmond, A., Millar, J., Russell, C. D., Malinauskas, T., Thwaites, R., Morrice, K., Keating, S., & Maslove, D. (2023). GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. *Nature*, 1–15. https://doi.org/10.1038/s41586-023-06034-3

Patanavanich, R., Siripoon, T., Amponnavarat, S., & Glantz, S. A. (2022). Active Smokers Are at Higher Risk of COVID-19 Death: A Systematic Review and Meta-analysis. *Nicotine & Tobacco Research*. https://doi.org/10.1093/ntr/ntac085

Pew Reserach Center. (n.d.). *Religious Landscape Study*. Pew Research Center. https://www.pewresearch.org/religious-landscape-study/database/age-distribution/

Pijls, B. G., Jolani, S., Atherley, A., Derckx, R. T., Dijkstra, J. I. R., Franssen, G. H. L., Hendriks, S., Richters, A., Venemans-Jellema, A., Zalpuri, S., & Zeegers, M. P. (2021). Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ Open*, *11*(1), e044640. https://doi.org/10.1136/bmjopen-2020-044640

Piovesan, A., Antonaros, F., Vitale, L., Strippoli, P., Pelleri, M. C., & Caracausi, M. (2019). Human protein-coding genes and gene feature statistics in 2019. *BMC Research Notes*, *12*. https://doi.org/10.1186/s13104-019-4343-8

Pourhoseingholi, M. A., Baghestani, A. R., & Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology from Bed to Bench*, *5*(2), 79–83. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4017459/

Pratt, C. W., Gill, K. J., Barrett, N. M., & Roberts, M. M. (2014, January 1). *Chapter 2 - Symptoms and Etiology of Serious Mental Illness* (C. W. Pratt, K. J. Gill, N. M. Barrett, & M. M. Roberts, Eds.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/abs/pii/B9780123870025000020

Prigerson, H. G., Bierhals, A. J., Kasl, S. V., Reynolds, C. F., Shear, M. K., Day, N., ... & Jacobs, S. (1997). Traumatic grief as a risk factor for mental and physical morbidity. American journal of psychiatry, 154, 616-623.

Rabb, N., Bowers, J., Glick, D., Wilson, K. H., & Yokum, D. (2022). The influence of social norms varies with "others" groups: Evidence from COVID-19 vaccination intentions. *Proceedings of the National Academy of Sciences*,

119(29). https://doi.org/10.1073/pnas.2118770119

Raja, A., Khan, O., Raja Subhash Sagar, Kumar, R., Parshant Dileep Bhimani, Riteeka Kumari Bhimani, Danial, M., Raja, S., Fnu Deepak, Muhammad Ashir Shafique, & Muhammmad Saqlain Mustafa. (2024). Insights into Covid-19 Mortality: A Comprehensive Study of Cardiovascular Sensitivity, Gender, Race, and Geography Trends in the United States (2020-2023). *Current Problems in Cardiology*, 102435–102435. https://doi.org/10.1016/j.cpcardiol.2024.102435

Rashidi, V., & Bourhaninejad, V. (2020). Age-based discrimination in Covid-19 patient care. *Indian Journal of Medical Ethics*, *05*(03), 255–255. https://doi.org/10.20529/ijme.2020.074

Restrepo, B. J. (2022). Obesity Prevalence Among U.S. Adults During the COVID-19 Pandemic. *American Journal of Preventive Medicine*, *63*(1). https://doi.org/10.1016/j.amepre.2022.01.012

Rivera, N. V., Patasova, K., Kullberg, S., Diaz-Gallo, L. M., Iseda, T., Bengtsson, C., Alfredsson, L., Eklund, A., Kockum, I., Grunewald, J., & Padyukov, L. (2019). A Gene–Environment Interaction Between Smoking and Gene polymorphisms Provides a High Risk of Two Subgroups of Sarcoidosis. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-54612-1

Robitzsch, A. (2020). Why Ordinal Variables Can (Almost) Always Be Treated as Continuous Variables: Clarifying Assumptions of Robust Continuous and Ordinal Factor Analysis Estimation Methods. *Frontiers in Education*. https://doi.org/10.3389/feduc.2020.589965

Romeis JC, Scherrer JF, Xian H, Eisen SA, Bucholz K, Heath AC, Goldberg J, Lyons MJ, Henderson WG, True WR (2000) Heritability of self-reported health. Health Serv Res 35:995–1010

Rosenqvist, E., Kiviruusu, O., Berg, N., & Konttinen, H. (2023). Stress-induced eating and drinking and their associations with weight among women and men during 30-year follow-up. *Psychology & Health*, 1–16. https://doi.org/10.1080/08870446.2023.2192240

Rouatbi, S. (2022). The aging lung face to COVID-19. La Tunisie Médicale, 100(2), 91.

Sattar, N., McInnes, I. B., & McMurray, J. J. (2020). Obesity is a risk factor for severe COVID-19 infection: multiple potential mechanisms. Circulation, 142(1), 4-6.

Sayers, I., John, C., Chen, J., & Hall, I. P. (2024). Genetics of chronic respiratory disease. *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-024-00695-0

Schmaling, K. B. (2022). Couples and COVID-19 vaccination: Frequency and reasons for discordance. *Vaccine*. https://doi.org/10.1016/j.vaccine.2022.02.055

Schmitz, L. L., Goodwin, J., Miao, J., Lu, Q., & Conley, D. (2021). The impact of late-career job loss and genetic risk on body mass index: Evidence from variance polygenic scores. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-86716-y

Semenza, J. C., Lindgren, E., Balkanyi, L., Espinosa, L., Almqvist, M. S., Penttinen, P., & Rocklöv, J. (2016). Determinants and Drivers of Infectious Disease Threat Events in Europe. *Emerging Infectious Diseases*, *22*(4), 581–589. https://doi.org/10.3201/eid2204.151073

Shah, S. M., Carey, I. M., Harris, T., DeWilde, S., Victor, C. R., & Cook, D. G. (2013). The Effect of Unexpected Bereavement on Mortality in Older Couples. *American Journal of Public Health*, *103*(6), 1140–1145. https://doi.org/10.2105/ajph.2012.301050

Shastry, B. S. (2009). SNPs: Impact on Gene Function and Phenotype. *Methods in Molecular Biology*, *578*, 3–22. https://doi.org/10.1007/978-1-60327-411-1_1

Shelton, J. F., Shastri, A. J., Ye, C., Weldon, C. H., Filshtein-Sonmez, T., Coker, D., Symons, A., Esparza-Gordillo, J., Aslibekyan, S., & Auton, A. (2021). Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nature Genetics*. https://doi.org/10.1038/s41588-021-00854-7

Sieurin, J., Gunnar Brandén, Magnusson, C., Maria-Pia Hergens, & Kyriaki Kosidou. (2022). A population-based cohort study of sex and risk of severe outcomes in covid-19. *European Journal of Epidemiology*, *37*(11), 1159–1169. https://doi.org/10.1007/s10654-022-00919-9

Silventoinen, K., Posthuma, D., Lahelma, E., Rose, R. J., & Kaprio, J. (2007). Genetic and environmental factors affecting self-rated health from age 16–25: A longitudinal study of Finnish twins. Behavior Genetics, 37(2), 326-333.

Singh, S., Baldwin, M. M., Daynes, E., Evans, R. A., Greening, N., R Gisli Jenkins, Lone, N., McAuley, H., Mehta, P., Newman, J. V., P. Novotný, Smith, D., Stanel, S., Toshner, M., & Brightling, C. (2023). Respiratory sequelae of COVID-19: pulmonary and extrapulmonary origins, and approaches to clinical care and rehabilitation. *The Lancet Respiratory Medicine*. https://doi.org/10.1016/s2213-2600(23)00159-5

Slavkin, H. C. (2014). From Phenotype to Genotype. *Journal of Dental Research*, *93*(7_suppl), 3S6S. https://doi.org/10.1177/0022034514533569

Stahl, S. T., Kazan, J., Lazzari, T., Krafty, R. T., Reynolds, C. F., Rollman, B. L., Smagula, S. F., & Gebara, M. A. (2023). Risk for Complicated Grief After the COVID-19 Death of a Marital Partner in Late Life. *The American Journal of Geriatric Psychiatry*. https://doi.org/10.1016/j.jagp.2023.10.019

Stenholm, S., Westerlund, H., Salo, P.,g Hyde, M., Pentti, J., Head, J., ... & Vahtera, J. (2014). Age-related trajectories of physical functioning in work and retirement: the role of sociodemographic factors, lifestyle and disease. J Epidemiol Community Health, 68(6), 503-509.

Stephan, Y., Sutin, A. R., Luchetti, M., Caille, P., & Terracciano, A. (2020). An examination of potential mediators of the relationship between polygenic scores of BMI and waist circumference and phenotypic adiposity. *Psychology & Health*, *35*(9), 1151–1161. https://doi.org/10.1080/08870446.2020.1743839

Svedberg, P., Bardage, C., Sandin, S., & L.Pedersen, N. (2006). A prospective study of health, life-style and psychosocial predictors of self-rated health. *European Journal of Epidemiology*, *21*(10), 767–776. https://doi.org/10.1007/s10654-006-9064-3

Szarpak, L., Mierzejewska, M., Jurek, J., Kochanowska, A., Gasecka, A., Truszewski, Z., Pruc, M., Blek, N., Rafique, Z., Filipiak, K. J., Denegri, A., & Jaguszewski, M. J. (2022). Effect of Coronary Artery Disease on COVID-19—Prognosis and Risk Assessment: A Systematic Review and Meta-Analysis. *Biology*, *11*(2), 221. https://doi.org/10.3390/biology11020221

Taylor, P., & Lewontin, R. (2017). *The Genotype/Phenotype Distinction (Stanford Encyclopedia of Philosophy)*. Stanford.edu. https://plato.stanford.edu/entries/genotype-phenotype/

Thompson, M. D., Pirkle, C. M., Youkhana, F., & Wu, Y. Y. (2020). Gene–obesogenic environment interactions on body mass indices for older black and white men and women from the Health and Retirement Study. International Journal of Obesity, 44(9), 1893-1905.

Tisminetzky, M., Delude, C., Hebert, T., Carr, C., Goldberg, R. J., & Gurwitz, J. H. (2020). Age, Multiple Chronic Conditions, and COVID-19: A Literature Review. *The Journals of Gerontology: Series A*. https://doi.org/10.1093/gerona/glaa320

Tommerup, K., Ajnakina, O., & Steptoe, A. (2021). Genetic propensity for obesity, socioeconomic position, and trajectories of body mass index in older adults. *Scientific Reports*, *11*(1), 20276. https://doi.org/10.1038/s41598-021-99332-7

Torres Acosta, M. A., & Singer, B. D. (2020). Pathogenesis of COVID-19-induced ARDS: implications for an ageing population. *The European Respiratory Journal*, *56*(3). https://doi.org/10.1183/13993003.02049-2020

Trejo, S., & Domingue, B. W. (2018). Genetic nature or genetic nurture? Introducing social genetic parameters to quantify bias in polygenic score analyses. *Biodemography and Social Biology*, *64*(3-4), 187–215.

https://doi.org/10.1080/19485565.2019.1681257

Treml, J., Linde, K., Engel, C., Glaesmer, H., Hinz, A., Luck, T., Riedel-Heller, S., Sander, C., & Kersting, A. (2020). Loss and grief in elderly people: Results from the LIFE-Adult-Study. *Death Studies*, *46*(7), 1–10. https://doi.org/10.1080/07481187.2020.1824203

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-021-00056-9

Uscinski, J. E., Enders, A. M., Klofstad, C., Seelig, M., Funchion, J., Everett, C., Wuchty, S., Premaratne, K., & Murthi, M. (2020). Why Do People Believe COVID-19 Conspiracy Theories? *Harvard Kennedy School Misinformation Review*, *1*(1). https://doi.org/10.37016/mr-2020-015

Vegivinti, C. T. R., Evanson, K. W., Lyons, H., Akosman, I., Barrett, A., Hardy, N., Kane, B., Keesari, P. R., Pulakurthi, Y. S., Sheffels, E., Balasubramanian, P., Chibbar, R., Chittajallu, S., Cowie, K., Karon, J., Siegel, L., Tarchand, R., Zinn, C., Gupta, N., & Kallmes, K. M. (2022). Efficacy of antiviral therapies for COVID-19: a systematic review of randomized controlled trials. *BMC Infectious Diseases*, *22*(1). https://doi.org/10.1186/s12879-022-07068-0

Veller, C., & Coop, G. M. (2024). Interpreting population- and family-based genome-wide association studies in the presence of confounding. *PLOS Biology*, *22*(4), e3002511. https://doi.org/10.1371/journal.pbio.3002511

Vermunt, J. K., & Hagenaars, J. A. P. (2004). *Ordinal Longitudinal Analysis* (R. C. Hauspi, N. Cameron, & L. Molinari, Eds.). Tilburg University Research Portal; Cambridge University Press. https://research.tilburguniversity.edu/en/publications/ordinal-longitudinal-analysis

Virolainen, S. J., VonHandorf, A., Viel, K. C. M. F., Weirauch, M. T., & Kottyan, L. C. (2022). Gene–environment interactions and their impact on human health. *Genes & Immunity*, *24*(1). https://doi.org/10.1038/s41435-022-00192-6

Vo, M. T. H., Nakamura, K., Seino, K., & Vo, T. V. (2021). Greater Risk of Negative Health Outcomes of Older Adults Living Alone in Vietnam: A Community Survey. *International Journal of Environmental Research and Public Health*, *18*(21), 11115. https://doi.org/10.3390/ijerph182111115

Walter, S., Mejía-Guevara, I., Estrada, K., Liu, S. Y., & Glymour, M. M. (2016). Association of a Genetic Risk Score With Body Mass Index Across Different Birth Cohorts. *JAMA*, *316*(1), 63. https://doi.org/10.1001/jama.2016.8729

Wang, J., & Geng, L. (2019). Effects of Socioeconomic Status on Physical and Psychological Health: Lifestyle as a Mediator. *International Journal of Environmental Research and Public Health*, *16*(2), 281. https://doi.org/10.3390/ijerph16020281

Ware, E., Gard, A., & Schmitz, L. (2021). *HRS Documentation Report HRS Polygenic Scores -Release 4.3 2006-2012 Genetic Data Report prepared by*. https://hrsdata.isr.umich.edu/sites/default/files/documentation/data-descriptions/PGNSCORES4_3DD_0.pdf

Westerman, K. E., Lin, J., Magdalena, Beza Tadess, Marchek, C., & Manning, A. K. (2022). Gene-Environment Interaction Analysis Incorporating Sex, Cardiometabolic Diseases, and Multiple Deprivation Index Reveals Novel Genetic Associations With COVID-19 Severity. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.782172

Williams, R. (2020). Ordinal Regression Models. *SAGE Research Methods Foundations*. https://doi.org/10.4135/9781526421036885901

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews. Genetics*, *14*(7), 507–515. https://doi.org/10.1038/nrg3457

Wu, C., Chen, X., Cai, Y., Xia, J., Zhou, X., Xu, S., Huang, H., Zhang, L., Zhou, X., Du, C., Zhang, Y., Song, J., Wang, S., Chao, Y., Yang, Z., Xu, J., Zhou, X., Chen, D., Xiong, W., & Xu, L. (2020). Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Internal Medicine*, *180*(7). https://doi.org/10.1001/jamainternmed.2020.0994

Wu, S., Wang, R., Zhao, Y., Ma, X., Wu, M., Yan, X., & He, J. (2013). The relationship between self-rated health and objective health status: a population-based study. *BMC Public Health*, *13*(1). https://doi.org/10.1186/1471-2458-13-320

Wyatt, S. B., Winters, K. P., & Dubbert, P. M. (2006). Overweight and Obesity: Prevalence, Consequences, and Causes of a Growing Public Health Problem. *The American Journal of the Medical Sciences*, *331*(4), 166–174. https://doi.org/10.1097/00000441-200604000-00002

Yun, J.-Y., Sim, J.-A., Lee, S., & Yun, Y. H. (2022). Stronger association of perceived health with socio-economic inequality during COVID-19 pandemic than pre-pandemic era. *BMC Public Health*, *22*(1). https://doi.org/10.1186/s12889-022-14176-8

Zhang, J., Dong, X., Liu, G., & Gao, Y. (2022). Risk and Protective Factors for COVID-19 Morbidity, Severity, and Mortality. *Clinical Reviews in Allergy & Immunology*, *64*(1). https://doi.org/10.1007/s12016-022-08921-5

Zhao, H., Ma, Q., Xie, M., Huang, Y., Liu, Y., Song, H., Gui, H., Li, M., & Wang, Q. (2023). Self-rated health as a predictor of hospitalizations in patients with bipolar disorder or major depressive disorder: A prospective cohort study of the UK Biobank. *Journal of Affective Disorders*, *331*, 200–206. https://doi.org/10.1016/j.jad.2023.02.113

Zhong, S., Huisingh-Scheetz, M., & Huang, E. S. (2022). Delayed medical care and its perceived health impact among US older adults during the COVID -19 pandemic. *Journal of the American Geriatrics Society*, *70*(6). https://doi.org/10.1111/jgs.17805

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, *395*(10229), 1054–1062. https://doi.org/10.1016/s0140-6736(20)30566-3

Zhu, Z., Hasegawa, K., Ma, B., Fujiogi, M., Camargo, C. A., & Liang, L. (2020). Association of obesity and its genetic predisposition with the risk of severe COVID-19: Analysis of population-based cohort data. *Metabolism*, *112*, 154345. https://doi.org/10.1016/j.metabol.2020.154345

Zhu, Z., Li, J., Si, J., Ma, B., Shi, H., Jun Lv, Cao, W., Guo, Y., Millwood, I. Y., Walters, R., Lin, K., Yang, L., Chen, Y., Du, H., Yu, B., Hasegawa, K., Camargo, C. A., Moffatt, M. F., Cookson, W., & Chen, J. (2021). *A large-scale genome-wide association analysis of lung function in the Chinese population identifies novel loci and highlights shared genetic aetiology with obesity*. *58*(4), 2100199–2100199. https://doi.org/10.1183/13993003.00199-2021

10. Appendix

Appendix A. Summary of G×E conceptual frameworks

Mills and colleagues (2020) have described four main theoretical frameworks that serve as a theoretical foundation to most of the G×E literature. Besides the Diathesis-Stress model, which is described in the main part of the thesis, other models include the ones summarized in Table A1 below.

Table A1. Short description of other conceptual frameworks

Theory	**Summary**
Social control/ social push model	Model that established the effect of social restrictive environments. These environments (restricted by e.g. social norms/ social control) indirectly affect the effect of genetics on phenotypical outcomes (can be either strengthened or weakened).
Differential Susceptibility model	Low- and high risk are differently susceptible to negative but also positive experiences. Arises when the most vulnerable group is disproportionally affected (influenced by genetics) in a negative environment as well as a positive environment.
Bioecological model	A theory that proposes that genetic potential is maximized in stable environments allow for positive interactions.

Appendix B. Data

B1. Data selection

The data used in this thesis originates from the Health and Retirement Study (HRS). The HRS is a nationally representative study of U.S. elderly individuals aged 50 and over. The dataset contains individual- and household level data from the time of entering the survey until death (or dropping out). Every 6 years, a novel birth cohort of participants is introduced. In total, any given wave has about 20.000 respondents with over 40.000 individuals ever interviewed from its start in 1992.

The analytic dataset used for this thesis is constructed by combining different sub-datasets from the HRS. The main source is the HRS longitudinal survey, which combines the information from all previous waves. The different sub-datasets are presented in Table B1 below. All datasets were individually adjusted and cleaned to fit the master dataset. Selection criteria for

data were 1) the number of observations and 2) the most recent estimates (2022 survey available as of February 2024). The panel dataset is restricted to cover wave 14-16, which corresponds to a biannual period of 2018-2022.

From 2006 onwards, the HRS collected genetic data from their respondents. Genetic data was obtained using saliva (DNA) samples from participants, collected during assessment from 2006-2012 (HRS documentation report). As part of the genetic data release, the HRS calculates polygenic scores (PGS) for each phenotype based on genome-wide association studies.

During the COVID-19 pandemic, the HRS conducted a COVID-19 related study amongst the participants. This was done in the form of questionnaires, which was administered randomly to 50% of the households and a non-random group of respondents who were assigned to telephone interviewing and completed those in June 2020 or later. This survey asked for the respondent's experience regarding COVID-19 by examining different variables. This midway release of the HRS is included in the dataset.

Table B1. Description of datasets

Dataset description	Source
HRS Longitudinal File 2020 (master)	HRS – HRS products [1]
HRS 2020 Tracker	HRS – 2020 HRS Core [2]
HRS 2020 COVID Survey	HRS – 2020 HRS Core [2]
HRS Polygenic Scores – Release 4	HRS – HRS Genetic Data [3]
HRS 2022 Survey data	HRS – 2022 HRS Core [4]

B2. Variable description

Table B2. below reports the name and description of the variables that were either used in the main analysis of this thesis or additional variables used other regressions or analyses. How the variables are constructed will be elaborated on in Appendix section B3.

Table B2. Descriptive statistics

Variable	Description
Lung health	Binary indicator whether respondent indicates to be diagnosed with lung disease by doctor (excluding asthma). 0=no; 1 = yes.
Self-reported health	Self-reported health score on a 1-5 scale. Increasingly recoded such that 1 (poor health) – 5 (excellent health).

Region of residence	Reports respondent region of residence. 1: Northeast; 2: Midwest; 3: South; 4: West; 5: Other.
Covid concern	Parameter to which captures the individual concern about the covid-19 pandemic in general on an increasing 1-10 scale, where 1 (not concerned) – 10 (most concerned).
Covid diagnosed	Indicates if respondent is diagnosed with covid-19. 0 = no, 1 = yes.
Covid deaths	Indicates if the respondent knows someone (friend and/or family member) who passed from covid-19. 0 = no; 1= yes.
Covid vaccine	Indicates if the respondent is ever vaccinated against covid-19. 1 = yes (all are vaccinated)
Covid variable	Combines data on year and month a respondent is vaccinated against covid-19. Increasingly coded indicating the higher, the later vaccinated. E.g. 1: January 2020; 2: February 2020 … 44: August 2023.
Covid Survey	Binary variable that represents whether a respondent was included in the 2020 covid survey. 0 = no; 1 = yes
Gender	Gender of respondent 1: male; 0: female
Race	Race of respondent. 1: white/Caucasian (only ethnicity in this sample)
Education	Categorical variable represents the highest level of education of respondent. 1: less than High-School; 2: General Educational Development (GED); 3: High-School Graduate; 4: Some College; 5: College and above
Highest Education	Categorical variable indicating the highest level of education. 0: no degree; 1: GED; 2: high-school (HS); 3: HS/ GED; 4:AA/ less than BA; 5: BA; 6: MA/ MBA; 7: Law/MD/PhD; 8: other
Education years	Reports the total number of education years ranging from 0 – 17/17+.
Religion	Religion of respondent. 1: Protestant; 2: Catholic; 3: Jewish; 4: None; 5: other
Age	Respondent age at time of the interview
Cohort	Represents the birth-cohort of a respondent which defines the years of entry into the HRS study. 0: HRS/ AHEAD overlap; 1: Study of Assets and Health Dynamics (AHEAD); 2: Children of Depression (coda); 3: HRS; 4: War Babies; 5: Early baby boomers; 6: Mid baby boomers. (7: Late baby boomers, however not in sample).
High Blood pressure	Reports if respondent has high blood pressure. 0 = no, 1 = yes
Diabetes	Reports if respondent has diabetes. 0 = no, 1 = yes
Heart condition	Reports if respondent has heart condition. 0 = no, 1 = yes
BMI	Reports BMI of the respondent
Expansion	Reports vaccine eligibility according to age and region. 14: February 2021; 15: March 2021; 16: April 2021; 17: May 2021.
Vaccination timing	Shows the time delay (in months) between vaccination eligibility (according to expansion) and self-reported time of vaccination.
Early vaccination	Indicator for individuals who report to be vaccinated before eligibility according to expansion criteria.

B3. Data construction

The primary dataset is the pre-constructed longitudinal dataset by the HRS, which has total of 42,406 respondents. The data is cleaned and irrelevant variable categories have been removed. This dataset is merged with another dataset that contains the polygenic risk scores and a separate file containing the 2020 covid-survey. Both datasets are cleaned to contain the polygenic scores and covid-variables of interest before merging to the master file. Upon merging, I checked the number of individuals for whom polygenic risk scores are present, since not all cohorts are genotyped. The respondent with available genetic data (5,980) are kept in the sample.

As the RAND HRS study is a biannual study, the newest data covering 2021/2022 was released in March of 2024. Therefore, to match the main dataset, the 2022 survey data requires to be adjusted. Consequently, the following variables were recoded: self-reported health; lung health; variables for comorbidities (high blood pressure, diabetes, heart condition) BMI; region of residence. These are constructed according to the HRS guidelines (where necessary) and renamed to match the variable names in the longitudinal file in order for them to be correctly transformed. The region of residence was not readily available for the 2022 survey release, but could be constructed using information on state data. To do this, I used the similar method as the HRS namely clustering the data to match the master. This is shown in Table B3, below.

Table B3. Grouping of state-level data to regional categories.

Value master	Region master	Value + States 2022 Survey data
1	Northeast	1. Northeast Region: New England Division (ME, NH, VT, MA, RI, CT) 2. Northeast Region: Middle Atlantic Division (NY, NJ, PA)
2	Midwest	3. Midwest Region: East North Central Division (OH, IN, IL, MI,WI) 4. Midwest Region: West North Central Division (MN, IA, MO, ND,SD, NE, KS)
3	South	5. South Region: South Atlantic Division (DE, MD, DC, VA, WV,NC, SC, GA, FL) 6. South Region: East South Central Division (KY, TN, AL, MS) 7. South Region: West South Central Division (AR, LA, OK, TX)
4	West	8. West Region: Mountain Division (MT, ID, WY, CO, NM, AZ, UT,NV) 9. West Region: Pacific Division (WA, OR, CA, AK, HI)
5	Other	11. Foreign Country: Not in a Census Division (includes U.S. territories)

The HRS derives BMI from height and weight measures. The height of the individual is measured during the first interview period and carried forward when a respondent is reinterviewed. Weight is remeasured when the respondent during subsequent interviews. Height and weight are reported in feet and pounds. I convert these by a factor 0.31 and 0.45, respectively, to meters and kilograms. I apply the general formula for calculating BMI, which is inline with the HRS calculations.

$$BMI = \frac{weight\ (kg)}{(height\ (m))^2}$$

After combining all datasets, the file is checked for duplicate respondents. The duplicates (90) are dropped from the sample. The data is reshaped from wide to long, to allow for panel data regression analysis. The total number of respondents in the analytic sample equals 3,517.

B4**.** The analytic sample

This section provides some background information about the representativeness of the analytic sample. Summary statistics per wave are presented in Section 4.1. of the main thesis. Table B4. shows the descriptive statistics of the sample for the total sample, and stratified by environment for individuals who do not know someone who passed from COVID-19 (e = 0) and individuals who do know someone who passed form COVID-19 (e = 1).

Table B4. Descriptive statistics stratified by the environment.

	Total sample (N=3517)		**e=0 (N=2796)**		**e=1 (N=721)**	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Lung Disease	10.4%		10.4%		10.4%	
Self-reported health	3.48	.923	3.39	.826	3.34	.793
BMI	28,56	5.878	28.43	5.83	29.16	6.04
BMI PGS*	-.03	0.995	-.035	.996	-.007	.991
Years of education	14.04	2.27	14.1	2.28	13.83	2.2
Comorbidities†	1.11	0.88	1.09	.879	1.146	.895
Age	72.55	8.87	72.96	8.95	70.95	8.37
	Count	**Share (%)**	**Count**	**Share (%)**	**Count**	**Share (%)**
Covid deaths		20,5	-	-	-	-
Smoking		6.23		6.1		7.1
Drinking		65.56		66.8		61
Male		40.8		41.3		39
Religion						
Protestant	2065	58.71	1644	58.80	421	58.39
Catholic	871	24.77	692	24.75	179	24.83
Jewish	89	2.53	67	2.4	22	3.05
None	433	12.31	354	12.66	79	10.96
Other	59	1.68	39	1.39	20	2.77

Notes: PGS = polygenic score; SD = standard deviation

* Polygenic scores are standardized with a mean of 0 and a standard deviation of 1.

† comorbidities is a composite measure for the number of comorbidities present in the range of 0-3. Include the presence of: diabetes, coronary artery disease and hypertension.

Gender-ratio: the gender-ratio is measured as the number of men per 100 woman (OWID, 2022). It is known that on average, females tend to live longer than men (Guralnik et al., 2000). In Figure B4.1 below, the gender ratio for different age-categories in the U.S. is shown. Consistent with literature, the gender ratio is decreasing with increasing age in the U.S., as increasingly more women tend to live to higher ages. This is the most prominent after 70/80 years of age. This is also shown in the data in this study (Figure B4.2). The data indicates a rather consistent age/gender distribution per age. Thus, given that there are more females in the

sample, the trend as visualized in the data is still representative of the general U.S. elderly population.

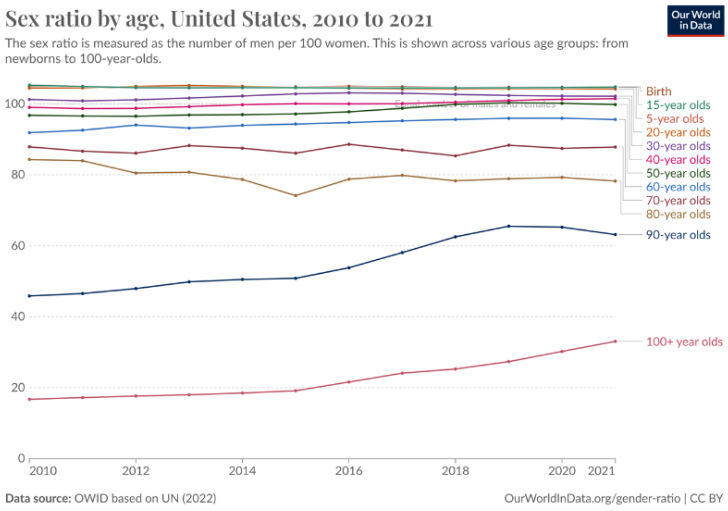

Figure B4.1. Gender-rate U.S. 2010-2021. Note: Figure retrieved from Our World in Data (2021).

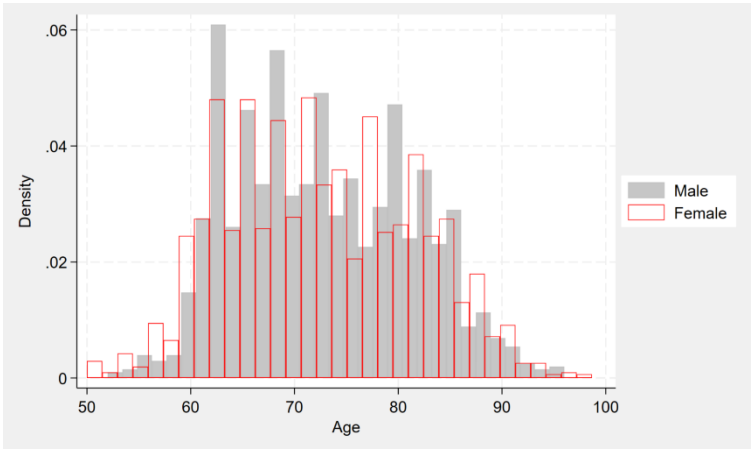

Figure B4.2. Distribution of males/females in the analytic sample by age

Appendix C. The COVID environment

The remainder of this section describes other candidate measures for the environment. In Section 5.2., I specify the pandemic as a negative shock, which is in line with the theoretical framework. However, there are different ways in which the environment can be quantified. The HRS dataset offers – together with the halfway COVID-survey[35] – a different scala on measures that capture some characteristic or individual experience of the pandemic. Therefore, in this section, I report other measures which I constructed and tested to be suitable for the environment, each with strengths and weaknesses. Ultimately, considering the strengths, weaknesses and overall suitability of these options, I choose to quantify the COVID-environment as an indicator for knowing someone (friend/ family member) who passed in the main analysis.

1. Covid testing statistics

The first measure to quantify the environment is a variable that reports numbers on individual COVID-19 testing statistics. This question was asked in the 2020 covid module. There are two options: a general question asking whether a respondent is tested for COVID-19 by: ‘’*have you been tested for the coronavirus?*‘’, and a the second question on how often someone is tested

[35] Note: the 2020 covid module was randomly administered to a subset of the respondent. Therefore, in general the number of observations for these variables tend to be lower compared to a survey question asked in the main survey.

by: ‘’*have you been tested only one, or multiple times?*’’. However, after restricting the data, the number of observations for both of these measures is low. This is partly explained by the fact that these questions stem from the midway module, which was asked to approximately half of the respondents. Additionally, it is likely that individual testing statistics (whether this is the number of times tested or being tested in general) is correlated with individual characteristics, which may be a source of endogeneity. Therefore, this measure to quantify the environment was disregarded.

2. Level of covid concern

A second candidate for the environmental measure is the level of concern for the COVID-19 pandemic. This question is also included in the 2020 COVID-module and captures the question: ‘*’Overall, on a scale from 1 to 10, where one is the least concerned and ten is the most concerned, how concerned are you about the coronavirus pandemic?*’’. After restricting the data, the number of respondents that answered this question is higher compared to the previous measure on testing statistic. The summary statistics show that on average, the respondents are quite concerned with the pandemic, with an average score as high as 7.8 points. Estimating Equation 1 with the level of concern as the environmental exposure is shown in the Table below. Higher levels of concern are associated with lower SRH scores. No significant G x E association is established.

Table C1.

	(1) SRH	(2) Lung disease	(3) BMI
PGS for BMI	-0.130*	-0.0133	1.429***
	(0.0718)	(0.0202)	(0.523)
Covid concern	-0.0172**	0.00379	0.0583
	(0.00843)	(0.00281)	(0.0525)
Covid concern × PGS for BMI	0.00347	0.00230	-0.00322
	(0.00872)	(0.00267)	(0.0620)
Constant	0.264	-0.122**	42.08***
	(0.167)	(0.0509)	(0.997)
Observations	4,607	4,610	4,595
N	1,974	1,974	1,971

Note: table excludes coefficients for covariates (age, number of comorbidities, religion, gender, smoking- and drinking status, highest level of education). Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The strength of quantifying the COVID-environment as such is that is captures the public perception of the pandemic and it is consistent with the theoretical framework that explains the shock to be negative in nature. The relative concern may reflect the severity of the pandemic at a certain point in time. However, the pressure of the pandemic is dynamic and subject to changes

over time. Nevertheless, there are some weaknesses. It is likely that the concern for the pandemic is highly correlated with other individual characteristics, such as risk aversion, education and socio-economic status. Besides, fear for the pandemic does not need to be driven by the risk for infection per se, as it may be determined by for example societal effects, such as conspiracy theories, misinformation or ‘’fake news’’. A large share of Americans (64%) experience that fake news causes confusion about the true facts of circumstances (Barthel and Holcomb, 2016). A study by Uscinski and colleagues (2020) revealed that over one third of the U.S. population believes that the coronavirus was created on purpose and spread accordingly. Nowadays, a vast majority of the misinformation is spread via online channels. Elderly individuals may have less access to the online environment which might decrease their exposure to misinformation. Nevertheless, reduced online exposure may also pose less opportunities for fact-checking. However, existing literature suggests that a large number of the elderly U.S. population is online and moreover highlight this group to be particularly vulnerable to false information (Huguet et al., 2024).

3. Level of covid concern + covid deaths

The level of covid-concern is likely to be endogenous and by itself insufficient to represent the pandemic. Therefore, I also combined this with whether the respondent knows someone (friend/ family member) who passed from COVID-19 to check if these concerns could be different according to experiencing a COVID-related death in your social circle. The descriptive statistics tell that there is a slight (0.2 point) higher level of concern among those that know someone who passed. I construct a categorical variable to reflect the composite measure of the two. The following criteria were applied:

- 0: if respondent level of covid concern < mean (<8) and does not know anyone who died from covid
- 1: if respondent level of covid concern < mean (<8) and does know anyone who died from covid
- 2: if level of concern >=mean (>=8) and does not know someone who died from covid
- 3: if level of concern >= mean (>=8) and does know someone who died from covid

However, creating the categories is arbitrary and there is inconsistent weighting of for the different variables. Additionally, there may be causality issues, as knowing someone who passed may make one more concerned for the pandemic. Therefore, this measure is not appropriate and not further explored. It did provide support to include the measure for knowing someone who passed as the environment in the main analysis.

4. Partner vaccination status

Household data is available, however there were too little observations. The strength is that partner vaccination arguably is more exogenous to the model, however on the contrary, partners are known to influence each others (health-related) behaviours (Schmaling, 2022). However, since there are insufficient observations, partner vaccination status was no option with the current dataset available.

5. Delayed care

Another consequence of the pandemic is disruptions in other hospital and/or healthcare activities. Research shows that backlogs and delayed care affected nearly one-third of U.S. elderly adults (Zhong et al., 2022). Such delays are the result of policy, which are region/state specific. Nevertheless, it is unlikely that these are correlated with for example genetic risk. Therefore, this was also explored as environmental measure. The results are shown in the table below. The regression results indicate that experiencing delayed care is predictive for lower SRH scores. Nevertheless, the sample size almost halves which reduces predictive power.

Table C2.

	(1) SRH	(2) Lung disease	(3) BMI
PGS for BMI	-0.0917***	0.00642	1.344***
	(0.0211)	(0.00809)	(0.142)
Delayed care	-0.180***	0.0192	-0.134
	(0.0411)	(0.0148)	(0.260)
Delayed care × PGS for BMI	-0.0359	-0.00446	0.209
	(0.0389)	(0.0138)	(0.264)
Constant	0.304*	-0.111**	42.56***
	(0.161)	(0.0490)	(0.963)
Observations	4,607	4,610	4,595
N	1,974	1,974	1,971

6. Timing of vaccination

In point 1 of this section, I already explored the potential for using individual testing data as the environment for this study. However, due to limited information in the number of times tested and the potentially endogenous nature of the variable, this measure is not the most suitable as the environmental indicator. The HRS released the 2022 survey in March (2024). In this survey, additional questions on the pandemic asked including vaccination status. In the 2022 survey, the respondent was asked the year and month when he or she received a vaccine against

COVID-19. I merged these variables to represent a numerical value for when an individual received their latest covid vaccine. The later an individual received their latest vaccine, the higher the number assigned.

This approach has some weaknesses. Namely, the theoretical framework explains that the environmental shock is negative in nature. One could argue that being vaccinated is a 'positive' event, as it offers protection against a disease. Therefore, the diathesis-stress model would potentially not suffice to explain any relationship. Strengths of this approach include that it utilizes the most recent data available. Additionally, it does not only represent the year an individual received their last vaccine, but also the month. Weaknesses include that there may be reverse causality which challenges exogeneity. Individuals with certain characteristics (poorer health/ lower self-perceived health status) may be more willing to be vaccinated as they are at a higher risk for covid-19 related complications. This I challenged by utilizing state-vaccination roll-out and timing of individual vaccination, since state-level roll-out is conditional on eligibility criteria. This I will discuss at the next point. However, all respondent in the analytic sample are vaccinated, so that issue does not hold. This can also introduce self-selection bias when individual who get vaccinated differ from those who do not. This may present as gene-environment correlation. Next, I aim to explore the difference between most recent individual vaccination and the vaccine availability per region (grouped by states). In the dataset I only have information regarding region of residence (not per state).

Therefore, I used state data and manually grouped this per region. Using the Figure on the right and information on how the HRS grouped individual states per region, I estimated the first month in which approximately >50% of the states in a certain region had access to the covid-19 vaccine for a specific age-group of the population, where I made the following criteria:

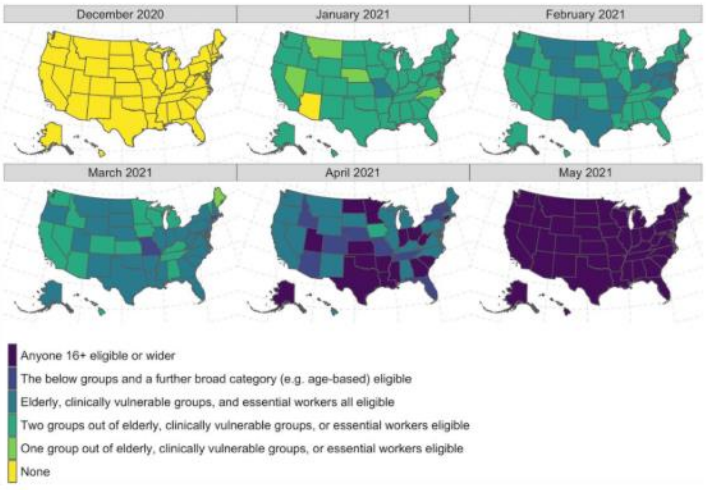

Table 4: Trends of eligible groups in state vaccine policies

State policy stage	Commonly included groups
Initial rollout	• Long-term care facility workers and residents • Healthcare workers • Common essential workers: First responders (EMS, firefighters, law enforcement), education and childcare workers (some states)
First expansion	• People aged 70+ or 75+ • Common essential workers: Education and childcare workers (some states) • People aged 16-65 with underlying health conditions (some states)
Second expansion	• People aged 60+ or 65+ • People aged 16-65 with underlying health conditions • Other frontline/essential workers as defined by state
Final expansion	• All people aged 16+

Note: figures retrieved from 'variation in US states' COVID-19 policy responses (Hallas et al., 2021)

1. first expansion: 75+ years; one or two groups out of the three categories (green colours).

2. second expansion: 65+ years, all critical groups eligible (blue)

3. final expansion: rest of the people.

Note however, that this is a quite subjective manner, as I counted the number of stated. This results in the table below. According to this method, there is a one month delay for the second expansion for the west and south region. First and final expansion are similar.

Table C3.

Region	Nr of states	First expansion (75+)	Second expansion (65+)	Final expansion (all ages)
North-East	9	February 2021	March 2021	May 2021
Midwest	12	February 2021	March 2021	May 2021
South	17	February 2021	April 2021	May 2021
West	13	February 2021	April 2021	May 2021

Using region-level eligibility criteria, this measure is more exogenous. However, this method is imprecise and may have credibility issues. It measures when an individual had their last covid-vaccine but the question does not specify whether an individual already had a shot, this is the first one or a booster. Besides the fact that it is not in line with theory, these compose one of the main reasons why it is not chosen as the environmental measure.

Using the abovementioned criteria, I constructed a time-invariant measure that depicts the difference between individual timing of vaccination and the availability of the vaccine per regio and criterium.[36] Another observation to note is that there are individuals who indicate to be vaccinated *before* availability according to state and age. This is possible, since respondents that have underlying medical conditions have earlier access to the vaccine. However, the eligibility criteria from literature not the figure does it specify what group had access. For example, the one individual that reports -16 (so vaccinated 16 months *before* availability of the vaccine. Individual states to be vaccinated in January of 2020. This is unlikely as the first covid-19 vaccine in the US was granted authorization by the FDA in December of 2020). So, I delete all individuals with observations before this date. Running the analysis however uncovers no economically meaningful interaction between genetics and the environment as all coefficients are close to zero and statistically insignificant.

[36] One issue was that since people age, some respondents became eligible for different criteria in wave 15 and/or 16. I noticed this and solved it by constructing a new variable that tracked in what wave an individual was vaccinated and their age during that wave.

Table C4.

	(1) SRH	(2) Lung disease	(3) BMI
PGS for BMI	-0.0714***	-0.00130	1.417***
	(0.0258)	(0.00967)	(0.173)
Vaccination timing	0.00421*	-3.22e-06	-0.0205
	(0.00235)	(0.000836)	(0.0148)
Delayed care × PGS for BMI	0.000273	-0.000232	0.00654
	(0.00228)	(0.000793)	(0.0140)
Observations	7,902	7,905	7,872
N	3,443	3,443	3,436

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Choosing an appropriate and exogenous measure to quantify the environment has strengths and weaknesses, as many, as shown in this section, are flawed. Therefore, I used the binary indicator of knowing whether someone passed or not.

Appendix D. Complete regression output

Table D1. Complete regression output (Table 1)

Dependent variables	(1) SRH	(2) Lung disease	(3) BMI
Independent variables			
PGS for BMI	-0.0698***	-0.00317	1.506***
	(0.0136)	(0.00484)	(0.0895)
Covid environment	-0.126***	0.00339	0.871**
	(0.0423)	(0.00889)	(0.424)
Covid environment × PGS for BMI	-0.0620	0.0114	-0.993***
	(0.0379)	(0.0109)	(0.384)
Control			
Male	0.00622	-0.0267***	1.028***
	(0.0281)	(0.00990)	(0.176)
Education			
GED	0.225**	-0.0133	0.218
	(0.101)	(0.0487)	(0.737)
High-School	0.365***	-0.119***	0.820*
	(0.0744)	(0.0344)	(0.491)
Some college	0.375***	-0.131***	0.0947
	(0.0751)	(0.0343)	(0.490)
College and above	0.566***	-0.160***	-0.626
	(0.0744)	(0.0337)	(0.484)
smoking	-0.267***	0.0321*	-1.546***
	(0.0563)	(0.0182)	(0.289)
drinking	0.150***	-0.00767	-0.270*
	(0.0247)	(0.00626)	(0.144)
age	-0.000781	0.00234***	-0.154***
	(0.00165)	(0.000537)	(0.0101)
Nr of Comorbidities			
1 comorbidity	-0.319***	0.0156**	1.593***
	(0.0281)	(0.00790)	(0.189)
2 comorbidities	-0.675***	0.0358***	2.553***
	(0.0352)	(0.0115)	(0.224)
3 comorbidities	-0.931***	0.0671***	3.423***

	(0.0562)	(0.0169)	(0.343)
Religion			
Catholic	0.0447		
	(0.0319)		
Jewish	-0.0730		
	(0.0892)		
None	-0.113**		
	(0.0458)		
Other	-0.0928		
	(0.114)		
Constant	-0.182	0.0561	37.87***
	(0.143)	(0.0515)	(0.923)
Observations	8,041	8,044	8,010
N	3,500	3,500	3,493

Note: comorbidities: composite measure indicates the number of comorbidities a person has. Comorbidities include: high blood pressure, diabetes and hearth condition. *Reference categories*: COVID deaths – not knowing someone who passed from covid-19; BMI – healthy weight; religion – protestant; education – high-school. Standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1

Table D2. Complete regression output (Table 2)

	SRH		Lung disease		BMI	
	(1) Low	(2) High	(3) Low	(4) High	(5) Low	(6) High
PGS for BMI	-0.0914***	-0.0799**	0.00495	-0.00885	1.352***	1.791***
	(0.0309)	(0.0320)	(0.0105)	(0.0128)	(0.176)	(0.239)
Covid environment	-0.123**	-0.135**	5.67e-05	0.00614	1.589**	-0.0296
	(0.0588)	(0.0615)	(0.0106)	(0.0155)	(0.625)	(0.576)
Covid environment × PGS for BMI	-0.0914***	-0.0799**	0.00495	-0.00885	1.352***	1.791***
	(0.0309)	(0.0320)	(0.0105)	(0.0128)	(0.176)	(0.239)
gender	-0.00624	0.0284	-0.0259*	-0.0264**	0.765***	1.255***
	(0.0391)	(0.0415)	(0.0149)	(0.0126)	(0.243)	(0.253)
smoking	-0.268***	-0.353***	0.0374	0.0349**	-1.583***	-1.469***
	(0.0728)	(0.0835)	(0.0257)	(0.0137)	(0.339)	(0.548)
drinking	0.179***	0.136***	-0.00744	-0.0127	-0.244	-0.335
	(0.0316)	(0.0399)	(0.00678)	(0.0124)	(0.184)	(0.233)
age	0.000809	-0.00340	0.00150**	0.00367***	-0.149***	-0.160***
	(0.00223)	(0.00250)	(0.000735)	(0.000780)	(0.0139)	(0.0144)
Nr of Comorbidities						
1 comorbidity	-0.312***	-0.326***	0.0176*	0.0146	1.681***	1.469***
	(0.0406)	(0.0388)	(0.0101)	(0.0127)	(0.282)	(0.242)
2 comorbidities	-0.671***	-0.689***	0.0602***	0.00288	2.302***	2.959***
	(0.0485)	(0.0509)	(0.0157)	(0.0167)	(0.313)	(0.316)
3 comorbidities	-0.950***	-0.917***	0.0979***	0.0140	3.365***	3.604***
	(0.0704)	(0.0952)	(0.0236)	(0.0187)	(0.437)	(0.570)
Religion						
Catholic	0.0572	0.0443				
	(0.0423)	(0.0486)				
Jewish	-0.116	-0.0299				
	(0.123)	(0.107)				
None	-0.141**	-0.0906				
	(0.0695)	(0.0611)				
Other	0.0490	-0.159				
	(0.164)	(0.156)				
Constant	0.426***	0.106	0.00426	-0.180***	35.68***	39.08***

	(0.162)	(0.180)	(0.0557)	(0.0524)	(0.960)	(1.188)
Observations	4,090	3,951	3,951	4,093	4,079	3,931
N	1,787	1,713	1,713	1,787	1,783	1,710

Note: comorbidities: composite measure indicates the number of comorbidities a person has. Comorbidities include: high blood pressure, diabetes and hearth condition. *Reference categories*: COVID deaths – not knowing someone who passed from covid-19; BMI – healthy weight; religion – protestant; education – high-school. Standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1

Other polygenic risk scores – Other polygenic risk scores of which the phenotype, according to literature, affects health outcomes under COVID-19 are: diabetes mellitus, coronary artery disfunction and hypertension. The presence of these phenotypes is in de baseline regression included as a composite measure as a covariate.

- Diabetes Mellitus: diabetic patients are increasingly susceptible to (respiratory) infections, as diabetes has been linked to impaired immune defences (dal Canto et al., 2019).
- Coronary Artery Dysfunction (CAD): Research suggests that CAD is linked to cardiovascular diseases and unfavourable results in respiratory viral infections (Szarpak et al., 2022).
- Hypertension: worldwide, there is a high prevalence of hypertension, with a peak amongst the elderly (>60%) (Zhou et al., 2021). However, given the high prevalence of hypertension in the population, this does not necessarily imply a relationship with more severe disease outcomes.

Table D3. Other polygenic risk scores.

	Self-rated health			Lung disease			BMI		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Covid environment	-0.131***	-0.129***	-0.136***	0.00389	0.00401	0.00397	0.808*	0.812*	0.813*
	(0.0426)	(0.0428)	(0.0428)	(0.00919)	(0.00886)	(0.00956)	(0.427)	(0.430)	(0.430)
PGS for T2D	0.00333			-0.00574			0.196**		
	(0.0150)			(0.00436)			(0.0891)		
Covid environment × PGS for T2D	-0.0456			0.01012			-0.242		
	(0.0381)			(0.00913)			(0.405)		
PGS for CAD		0.00240			-0.00193			0.00300	
		(0.0137)			(0.00461)			(0.0921)	
Covid environment × PGS for CAD		-0.0645			0.00696			-0.241	
		(0.0409)			(0.00924)			(0.459)	
PGS for hypertension			0.00220			-0.000940			-0.00227
			(0.0135)			(0.00466)			(0.0880)
Covid environment × PGS for hypertension			-0.02925			-0.00717			0.240
			(0.0412)			(0.00536)			(0.427)
Constant	-0.209	-0.208	-0.210	0.0563	0.0555	0.0555	38.30***	38.33***	38.33***
	(0.144)	(0.144)	(0.144)	(0.0516)	(0.0516)	(0.0516)	(0.947)	(0.950)	(0.950)
Observations	8,041	8,041	8,041	8,044	8,044	8,044	8,010	8,010	8,010
N	3,500	3,500	3,500	3,500	3,500	3,500	3,493	3,493	3,493

Note: table excludes coefficients for covariates (age, number of comorbidities, religion, gender, smoking- and drinking status, highest level of

education). Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1