Tracing the Genetic Footprints of the UK National Health Service

Nicolau Martin-Bassols[1], Elisabetta De Cao[1], Pietro Biroli[1], Massimo Anelli[2], Silvia Mendolia[3]

[1]University of Bologna, [2]University of Bocconi, [3]University of Turin

Abstract

Selection biases can distort causal estimates and undermine policy evaluations, yet detecting and quantifying selection remains empirically challenging. This paper introduces a novel method for detecting and quantifying selective survival using genetic data. Specifically, Polygenic Indices (PGIs) are predetermined at birth and should be equally distributed between treatment and control groups in the absence of selection. Any systematic differences in PGIs after an intervention suggest the presence of selection or survival effects. Our primary case study focuses on the sharp introduction of the National Health Service (NHS) in the UK in July 1948, which provided free and universal healthcare. Following the NHS implementation, hospital births increased, leading to improved neonatal health. First, we use a regression discontinuity design (RDD) and analyze administrative data from the Registrar General's Statistical Review to find that the NHS reduced infant mortality rates (IMR) by 5.48 per 1,000 births, a 16.1% decline relative to the mean. This sharp decline left a genetic imprint on affected cohorts. NHS exposure increased PGI values predictive of detrimental medical conditions, such as chronic obstructive pulmonary disease (COPD) and depression, while reducing those linked to beneficial traits, such as life satisfaction and self-rated health. These shifts are substantial, with effect sizes reaching 10% of a standard deviation (SD) in the UK Biobank (UKB) and up to 40% in the English Longitudinal Study of Ageing (ELSA). These results remain robust when the UKB is restricted to sibling pairs born before and after the NHS implementation, and when family fixed effects are included. Moving on to the estimation of long-term effects, we leverage the availability of genetic data to partially control for selection effects and bound the estimated treatment effect. In the nationally representative ELSA dataset, NHS exposure is associated with improved self-reported health and lower body mass index (BMI) among individuals aged 45 to 80. Controlling for PGIs increases estimated effect sizes by up to 43%, suggesting that selection bias attenuated previous estimates. However, these effects are not replicated in UKB, where no significant long-term health impacts are observed. Our findings highlight the importance of accounting for survival bias in policy evaluations and demonstrate how genetic data can improve causal inference by detecting and adjusting for selection effects.

Keywords: Survival bias, Genetics, National Health System
JEL classification: I10, I38, C21

PRELIMINARY DRAFT – PLEASE DO NOT CIRCULATE OR CITE

1 Introduction

The fetal origins hypothesis has demonstrated that environmental factors experienced during fetal development can have long-lasting effects on an individual's health and economic outcomes throughout their life course (Almond & Currie, 2011; Currie et al., 2014). According to this theory, exposures in the womb, such as maternal nutrition, stress, and environmental toxins, can influence the development of key biological systems, which in turn shape an individual's susceptibility to various diseases and socioeconomic conditions later in life (Godfrey & Barker, 2001). The hypothesis emphasizes that the critical period of fetal development extends beyond birth, with early-life experiences playing a pivotal role in determining developmental pathways (Currie & Rossin-Slater, 2015). In particular, the moment of birth itself, encompassing factors such as the delivery method (natural birth or cesarean), the presence of birth complications, and immediate postnatal care, significantly influences the trajectory of physical and cognitive development (Dominguez-Bello et al., 2016; Penders et al., 2006). These early-life factors may interact with genetic predispositions and later life circumstances, underscoring the importance of understanding the interplay between prenatal and postnatal environments in shaping long-term health and economic outcomes (Gluckman & Hanson, 2006).

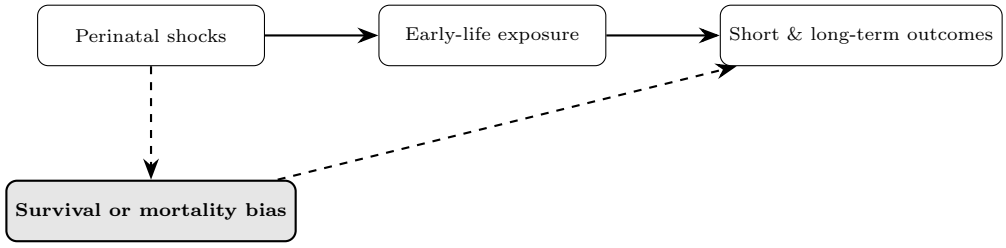
This line of research traces its origins to Barker's seminal study on the relation between infant mortality rates and later life morbidity and mortality across regions in the UK (Barker, 1990, 1995; Baker et al., n.d.) as well as the long-term effects of the Dutch Hunger Winter, a period of extreme famine in the Netherlands during the winter of 1944 to 1945 (Stein et al., 2004; Scholte et al., 2015). These findings highlighted the lasting consequences of in-utero and early-life deprivation, revealing that individuals exposed to poor perinatal conditions are at a higher risk of developing chronic conditions such as cardiovascular disease, diabetes, and hypertension later in life. Since then, the scope of the fetal origins hypothesis has broadened to encompass a range of prenatal shocks, including environmental, human-made, and policy-induced events, all of which have been shown to have profound and lasting effects on those exposed during critical periods of early development. For example, environmental shocks such as earthquakes (Menclova & Stillman, 2020), bushfires (Jayachandran, 2009), and air pollution (von Hinke & Sørensen, 2023; Martin-Bassols et al., 2024) have been linked to significant health and socioeconomic challenges in adulthood. Similarly, human-made shocks, including industrial accidents like Chernobyl (Almond, 2006), nuclear testing (Black et al., 2019), and the September 11th terrorist attacks (Currie & Schwandt, 2016), have also shown profound impacts on those in-utero at the time. In addition to these events, policy changes such as adjustments to maternal leave, shifts in healthcare access (Currie & Gruber, 2001), and

the introduction of food stamps (Hoynes et al., 2016) have been studied for their effects on fetal development and subsequent health and economic outcomes.

Crucially, these studies have considered a wide variety of both economic and health outcomes, measured at different stages of life. Early-life outcomes typically include indicators such as birthweight (Barker, 1990; Gluckman, 2004; Menclova & Stillman, 2020), gestational age (Menclova & Stillman, 2020), and early-life mortality (Jayachandran, 2009). As individuals age, the research expands to encompass later-life outcomes, such as academic performance (e.g., school grades, IQ) (Almond et al., 2009; Currie & Thomas, 1995), which in turn influence labor market outcomes, including employment and earnings throughout the life cycle (Currie & Thomas, 1995; Case et al., 2002). Furthermore, these studies extend into older age, examining health outcomes such as hospitalizations and the prevalence of chronic conditions (von Hinke & Sørensen, 2023; Martin-Bassols et al., 2024), providing a comprehensive view of how early-life shocks can shape individuals' trajectories well into later adulthood.

One challenge faced by all these studies is selective survival or mortality. This refers to the events under consideration selectively influencing individuals' survival or mortality, leading to compositional changes in the samples of individuals exposed before and after them (Nobles & Hamoudi, 2019). This phenomenon is illustrated in the directed acyclic graph (DAG) in Figure 1. Perinatal shocks, such as those discussed earlier, can affect individuals during critical early-life stages, including gestation, birth, and infancy, potentially influencing both short- and long-term outcomes. However, unobservable selective survival or mortality can bias causal estimates. These biases do not necessarily operate in the same direction to the treatment effect. If selective mortality is substantial, it can inflate estimated effects, leading to an overestimation. Conversely, it may attenuate effects or, in extreme cases, even reverse the sign in the estimated effects. A similar mechanism applies to fertility: if an event alters reproductive patterns, the resulting shifts in sample composition may further confound causal estimates. We do not consider this type of bias.

Figure 1: DAG of the Fetal Origins Hypothesis and Survival or Mortality Bias



Note: Own elaboration.

In this paper, we propose a novel method for detecting and quantifying selective survival or

mortality using genetic data. Specifically, we leverage polygenic index (PGI) data to explore how the distribution of these indexes in affected populations changes after significant events. Our primary focus is on the introduction of the National Health Service (NHS) in the UK in 1948, which provided universal healthcare access, free at the point of use. We exploit the fact that, following the NHS's implementation, pregnant mothers became more likely to deliver in hospital settings, which we show had important health benefits for the children born. We analyze data from three key sources: the Registrar General's Statistical Review, the English Longitudinal Study of Ageing (ELSA), and the UK Biobank (UKB). Using a regression discontinuity design (RDD) and the Registrar General's Statistical Review database, we estimate the short-term effects of the NHS on infant mortality rates (IMR), showing a decrease of 5.48 deaths per 1,000 births, or a 16.1% reduction compared to the mean. Building on the substantial effects on IMRs, we next explore whether the NHS left a genetic footprint on the affected cohorts. Across ELSA and UKB, we consistently find that the NHS implementation increased values in PGIs associated with qualitatively detrimental phenotypes, such as depression, COPD, and schizophrenia, while reducing those linked to beneficial traits, such as life satisfaction, self-rated health, and IQ. For both, increases and decreases, effect sizes are as large as 10% of a standard deviation (SD) for UKB estimates, and up to 40% for ELSA. Finally, we attempt to correct for this survival bias when exploring long-term effects. For this purpose, we examine long-term outcomes such as self-reported health, and BMI among individuals aged 45 to 80, while controlling for the same PGIs that were shown to be largely affected by the NHS. For the nationally representative ELSA database, we find that the NHS led to improvements in these health measures, and, more importantly, the inclusion of PGIs as controls seems to partly correct for the survival bias, increasing the estimated effects by up to 43%. However, these results cannot be replicated with the UKB, for which no health effects are observed following the NHS implementation.

Our work makes several contributions to the field. Most importantly, it proposes an empirical approach to test for selective survival or mortality. This method can be applied not only to quasi-natural experiments but also to randomized controlled trials (RCTs). Whenever treatment effects can influence sample composition, our approach provides a way to assess whether this poses a substantial threat or if selection is negligible. As discussed in greater detail in the following section, the literature has explored various methods to testing for this type of bias. Existing methods include variance reductions (or expansions) in outcome measures and deviations in high-resolution birth count data, both of which can provide evidence of selective survival or mortality (Nobles & Hamoudi, 2019). The former captures shifts in trait distributions among surviving births, while the latter detects changes in birth rates following exposure. However, due to the

requirement for historical high-resolution and large databases, these methods have seen limited empirical application. Another widely studied method is Secondary Sex Ratio (SSR) analysis. Since male fetuses are more vulnerable to prenatal mortality, an increase in the female-to-male birth ratio can indicate selective fetal loss (James & Grech, 2017). However, there is ongoing debate about whether SSR is a reliable measure of selective survival or mortality, given the complex, gender-specific dynamics of survival during gestation (Orzack et al., 2015). Our findings offer an additional way to detect selection effects, making them a valuable tool for studying survival or mortality bias across different settings. With the decreasing cost of genetic data collection (Muir et al., 2016), this approach will become more feasible in the future, allowing for more precise assessments of selection effects in a broader range of contexts.

Another contribution is our attempt to mitigate survival bias when exploring long-term effects. Our results on survival bias, which reveal shifts in genetic profiles after the implementation of the NHS, suggest that such interventions can alter genetic profiles across exposed population groups. By controlling for individuals' PGIs, we attempt to rebalance the treatment and control groups, enabling a cleaner comparison and yielding a more precise and unbiased treatment effect. While our results are promising when using the nationally representative ELSA, they are not robust in the UKB. This discrepancy could stem from the UKB's lack of national representativeness and its overrepresentation of high-SES individuals (Swanson, 2012), who may responded differently to the NHS expansion compared to the average population. We attempt to address this by incorporating sample-representative weights (van Alten et al., 2024), but this does not lead to a clearer picture. We interpret this as evidence that the UKB cannot be directly compared to the ELSA, as it captures different dynamics of the British population. These findings provide the first suggestive evidence that PGIs could play a key role in mitigating survival bias. This presents a novel direction for future research, emphasizing the need for deeper exploration to better understand and address selective survival and mortality.

Finally, our work provides novel insights into both the short- and long-term effects of the NHS implementation. While only one study has previously examined the effects of the NHS, focusing on its intergenerational dynamics (Jeffrey, 2023), our research contributes new evidence on its short-term effects, particularly its impact on IMR. Additionally, in line with Jeffrey (2023), we find suggestive evidence that the NHS may have also influenced long-term health outcomes. These results offer empirical support for the societal benefits of improved access to healthcare services, especially at the time of delivery and during the early stages of life.

The rest of the paper is organized as follows. The next section provides a literature review, outlining previous work on selective survival or mortality and research on policy evaluations similar

to the NHS. This is followed by a section on the institutional context, which describes the NHS and the healthcare system in the UK prior to its implementation. We then provide an introduction to genetics, including a discussion of Genome-Wide Association Studies (GWAS) and PGIs. The subsequent sections detail the data and methodology used in the analysis, followed by a presentation of the results. Given the nature of this study, which relies on multiple restricted-access data sources, a comprehensive 'Next Steps' section is included, based on the data access we are currently awaiting. Finally, the paper concludes with a discussion of the findings and their implications for both research and policy.

2 Literature Review

2.1 Survival or Mortality Bias

A theoretical framework for selective survival or mortality is outlined by Nobles & Hamoudi (2019). Drawing on medical evidence, the survival rate from fertilized egg to live birth may be as low as one-third, suggesting the presence of a health threshold for survival. Approximately 30% of fertilized eggs fail to implant, while 30 to 40% of pregnancies end due to detectable chromosomal abnormalities. Even among chromosomally viable pregnancies, up to 50% may still result in spontaneous loss (Hardy & Hardy, 2015; Zhang et al., 2018). Pregnancy survival is further influenced by maternal exposure to biological and psychosocial stress, infections, and the management of chronic conditions (Kawwass et al., 2016; Bailey & Berry, 2005; Arck et al., 2008). This process can be understood as a function of health endowment values surpassing a survival threshold shaped by contextual factors. A shock during gestation can either improve or worsen the health of the exposed cohort, with additional members having health values that surpass or fail to meet the survival threshold. Consequently, summary measures of cohort health or economic outcomes, which assess the difference in average outcomes between treated and untreated cohorts, may lead to biased fetal origin effects due to survival-related changes. Through simulations, Nobles & Hamoudi (2019) demonstrate that these biases can be economically meaningful, leading to substantial bias.

Various methods have been proposed to test for this bias. One approach involves assessing variance reductions or expansions in outcome measures to indicate changes in pregnancy loss, or using high-resolution time series birth count data to detect deviations, particularly when the exposure is discrete in time or space (Nobles & Hamoudi, 2019). However, detecting meaningful shifts in pregnancy survival requires large datasets with high-resolution data, typically at a finer temporal scale than annual data, which is often unavailable. When only small datasets are accessible, the limited sample size may lead to false negatives due to insufficient statistical power to detect true

changes.

Another approach involves exploring SSR analysis. Male fetuses are believed to be more vulnerable to prenatal mortality due to their larger size, which makes them more exposed to prenatal shocks and stressors (James & Grech, 2017). As such, an increase in the female-to-male birth ratio can indicate selective fetal loss, with male fetuses being more likely to be lost under adverse prenatal conditions. This method has been most widespread used by this literature so far to test for selective survival or mortality (Song, 2015; James & Grech, 2018; Sanders & Stoecker, 2015; Torche & Kleinhaus, 2012; James & Grech, 2017). However, Orzack et al. (2015) shows that contrary to previous beliefs of a male-biased conception, sex ratios are balanced at conception. The sex ratio initially decreases due to excess male mortality in the first week, increases during the first 10 to 15 weeks due to excess female mortality, and levels off around 20 weeks. After 28 weeks, the ratio declines slightly due to excess male mortality, and overall female mortality during pregnancy exceeds male mortality. These findings challenge the idea of persistent male frailty, showing that male and female biases vary across gestation rather than following a fixed pattern, limiting the capacity of SSR to test for survival or mortality bias.

This leads to the final approach for testing such bias, namely exploring changes in genetic profiles before and after the implementation of early-life shocks. So far, no paper has directly employed this method; however, a couple of studies have used similar approaches to address related research questions. Using the UKB database, Furuya et al. (2024) find that in-utero exposure to a higher area-level IMRs is associated with higher values in PGIs associated to better educational attainment. Their results suggest genetic selection in-utero, with more vulnerable individuals failing to surpass the survival threshold in areas with higher IMR, which is reflected in changes in the educational attainment PGI. However, their work is associational, and parents of individuals with different genetic endowments may select into different IMR areas. More similar to our approach and building on the classic study by (Stein et al., 2004; Scholte et al., 2015), Zhou et al. (2024) explore whether the Dutch Hunger Winter causally affected the prevalence of PGIs related to BMI. They find close-to-zero effects, which could be driven by their focus on the BMI PGI, which may not be relevant to resilience in-utero, the dimension that would be affected by selective mortality.

2.2 Universal Health Coverage Policies

Relatively few studies have examined the long-term effects of the NHS implementation. The most comprehensive work to date is the thesis by (Jeffrey, 2023), which investigates both the long-term health and intergenerational impacts of the NHS. The results show that individuals exposed to the NHS at birth report better self-reported health in adulthood and have a lower

incidence of cancer diagnoses later in life. Our long-term findings using ELSA partially align with these results, showing improvements in self-reported health as well. However, these patterns may not be replicated with the UKB data. The most novel aspect of Jeffrey (2023)'s work is the intergenerational analysis, which suggests that the effects of the NHS could extend up to three generations. Interestingly, only limited evidence is found regarding short-term effects on IMR. Our results provide a clearer picture on this issue.

Although research on the historical effects of the implementation of the NHS is limited, extensive studies have examined the impacts of early childhood healthcare interventions, with significant insights emerging from analyses of similar programs in other contexts. Research on the US Medicaid expansions and their effects on mortality reductions shows that these benefits persist into childhood and adolescence, as well as producing long-term improvements in educational attainment and lower welfare dependency (Currie et al., 2008; Brown et al., 2015; Wherry & Meyer, 2016). Furthermore, the Well-Child Programs in Scandinavian countries, introduced in the 1930s, offer valuable long-term insights. These universal programs targeted infants within their first year of life, leading to significant improvements in infant survival rates, particularly through home visiting programs in Norway and Denmark, which reduced mortality (Hjort et al., 2017; Bütikofer et al., 2019). In Sweden, the programs helped reduce infant mortality due to prematurity (Bhalotra & Venkataramani, 2011). These early interventions were linked to enduring health benefits, including reduced cardiovascular disease incidence and lower age-specific mortality later in life.

3 Background

3.1 History NHS

Before the establishment of the NHS in 1948, healthcare in the UK was largely private, requiring patients to pay for medical services, with limited free care available through various fragmented systems. Voluntary hospitals, funded by charitable donations, provided some treatment, while local authority hospitals, remnants of the Poor Law system, catered mainly to the indigent. The 1944 Education Act introduced free medical services for schoolchildren, and public health initiatives, such as tuberculosis sanatoria and vaccination programs, aimed to control infectious diseases. Additionally, the National Insurance Act of 1911 created a system of 6,000 Approved Societies that offered basic healthcare to insured workers, covering about 25% to 33% of the population but excluding hospital treatment, medications, and care for dependants, leaving many without adequate medical support (Rivett, 1998).

On July five, 1948, the UK government introduced the National Health Service (NHS), revolu-

tionizing healthcare by making it universally accessible and free at the point of use, funded through general taxation. Built on three core principles, free provision of healthcare, access based on clinical need rather than financial means, and the equalization of medical services, the NHS aimed to eliminate disparities in care. The response was immediate and overwhelming; by July 6, 84% of the population had already registered with a doctor, rising to 91% by the end of the month. By the close of 1948, an astonishing 96% of Britons were enlisted with a general practitioner, marking a swift and widespread embrace of the new system (Rivett, 1998).

There were important supply-side changes, including the nationalization of hospitals under the Ministry of Health, integrating 2,835 hospitals and 388,000 beds into the system. General practitioners (GPs) transitioned to independent contractors, receiving fixed fees per treatment as determined by Executive Councils. By September 1948, 18,165 of the 21,000 GPs had enrolled as NHS contractors. These structural changes facilitated an immediate surge in healthcare utilization, with a 13% increase in consultations and a doubling of prescriptions for proprietary medicines, reflecting the rapid expansion of access to medical services (Rivett, 1998).

3.2 NHS Impact on Birth Health: Mechanisms

The introduction of the NHS allowed pregnant mothers to deliver in a hospital setting free of charge. However, there is limited evidence regarding the proportion of mothers directly affected by this policy. We can only infer that its impact was substantial based on the reduction in IMR, as presented in this paper. Health improvements for children affected by the policy at the time of birth could be driven by three main channels. First, the shift to hospital-based childbirth has been associated with significant reductions in maternal mortality, as access to skilled medical professionals, sterile environments, and emergency interventions greatly improve survival rates (Symonds et al., 2023). This reduced maternal mortality would have benefited children throughout their development. Second, infants born in hospital settings experience lower rates of infections, reduced complications such as neonatal hypoxia, and improved outcomes for preterm births due to improved neonatal care (Backes & Scrimshaw, 2020). Finally, the introduction of universal healthcare mitigates the financial burden on families, reducing out-of-pocket expenses related to childbirth and allowing for broader access to essential maternal and infant care services, which in turn supports long-term economic stability and well-being (Bufe et al., 2021). We won't be able to distinguish which of these channels was the primary driver of our results. Nevertheless, all these mechanisms influenced individuals' health in the same direction, and it is likely that it was a combination of them.

3.3 Genetics, GWAS, and PGIs

Genetics is the study of heredity and the variation of inherited characteristics. Deoxyribonucleic acid (DNA), the fundamental molecule found in all living cells, consists of two long chains of nucleotides, each made up of three components: a phosphate group, a sugar molecule, and a nitrogenous base. The four types of nitrogenous bases in DNA are adenine (A), guanine (G), cytosine (C), and thymine (T). These nucleotides form the genetic code, which encodes genes that instruct cells to produce amino acids, the building blocks of proteins. Just as instructions are organized in books, DNA is organized into 22 pairs of chromosomes, one set inherited from the mother and one from the father, along with two sex chromosomes that determine biological sex (Alberts et al., 2015).

While 99% of human DNA is identical across all people, Single Nucleotide Polymorphisms (SNPs) are small genetic variations that occur in the DNA sequence of individuals. SNPs represent the 1% of genetic differences, consisting of single nucleotide changes, either A, T, C, or G. Unlike genes, which are long sequences of DNA that encode the instructions for building proteins and regulating cellular functions, SNPs are single-point variations that can occur within or between genes. While genes provide the blueprint for proteins, SNPs can influence how those proteins are produced or function, affecting traits such as physical appearance, disease susceptibility, and drug response. These variations can occur in non-coding regions or within the coding regions of genes, impacting gene expression or protein function, thereby contributing to phenotypic differences among individuals (Alberts et al., 2015).

Genome-Wide Association Studies (GWAS) are a powerful, hypothesis-free study design that scans the entire genome to identify genetic variants associated with complex traits or diseases. By analyzing large samples and often performing meta-analysis to increase statistical power, GWAS aims to uncover SNPs linked to the phenotypes of interest. The findings from these studies have led to the identification of specific, fixed positions on chromosomes, known as loci, where particular genes or genetic markers are located. These loci are associated with various diseases and traits, offering valuable insights into human phenotypes. However, while GWAS can highlight associations between genetic variants and traits, the underlying biological mechanisms remain not fully understood, and these studies do not establish causal relationships (Benjamin et al., 2024; Biroli et al., 2022).

PGIs are a genetic tool constructed by aggregating the weighted average of effect sizes derived from GWAS. In this indexes, larger effect variants contribute more significantly, reflecting the cumulative genetic contribution to specific phenotypes. PGIs integrate information from multiple

loci, offering a comprehensive assessment of an individual's genetic predisposition to traits such as disease risk, educational attainment, or behavior. Each locus (singular of loci) can have different versions of a gene, known as alleles. The major allele is the most common variant in a population, while the minor allele is the less frequent one. An individual's genotype at a given locus, denoted as G_{ij}, can take values of 0, 1, or 2, representing the number of minor alleles they carry (e.g., 0 for no copies, 1 for one copy, and 2 for two copies). The formula for calculating a PGI is represented as:

$$PGI_i = \sum_{j=1}^{n} W_j G_{ij}$$

where W_j represents the effect size of the j-th variant, and G_{ij} denotes the genotype of the i-th individual at the j-th locus (Benjamin et al., 2024; Biroli et al., 2022).

Some of the main limitations of PGIs include their predictive power, as they often explain only a small portion of the variation in complex traits or diseases, with environmental factors and gene-environment interactions playing significant roles that PGIs cannot capture. Additionally, PGIs do not establish causality; they are associated with genetic variants linked to traits, but many SNPs included may be in linkage disequilibrium with other variants that truly influence the phenotype, making direct causality difficult to infer. Population stratification is another limitation, as PGIs developed using datasets from specific populations may lose predictive accuracy when applied to individuals from different populations, reducing their generalizability. Environmental confounding is also a concern, as PGIs may overestimate the genetic contribution to a trait while neglecting complex gene-environment interactions. Finally, ethnic and ancestral bias can affect the performance of PGIs, particularly when they are applied to underrepresented populations, such as non-European groups, leading to biased predictions and limited utility across diverse populations (Biroli et al., 2022).

4 Data

4.1 Registrar General's Statistical Review

The Registrar General's Statistical Review was an annual publication produced by the Registrar General of the United Kingdom, which summarized and analyzed vital statistics related to population, health, mortality, births, and other demographic trends. It provided insights into the distribution, size, and characteristics of the population, including information on marriage, death, and disease rates. The review was a key resource for understanding the social and health trends

within the UK during its time of publication. The Registrar General was responsible for overseeing the national census and the registration of births, deaths, and marriages, and the Statistical Review was one of the key ways in which the data collected by these activities was disseminated (London & Region, 1952; Carrier, 1960).

We take district-year-level data from the Registrar General's Statistical Review for 1,472 districts from the years 1938 to 1958. We combine data on the number of live births and mortality before age one. This allows us to create a measure for the IMR at a year-district level, which we use to explore whether this measure was affected by the introduction of the NHS.

$$\text{IMR} = \frac{\text{Number of deaths of infants under one year}}{\text{Number of live births}}$$

For our main regressions, considering the years 1946 to 1950, the sample consists of 1,472 district-observation units, with 2,924 observations pre-NHS implementation and 4,410 observations post-NHS implementation. The panel is not fully balanced due to missing observations in some time periods.

4.2 ELSA

The ELSA is a large, ongoing study that collects data on the health, economic, and social circumstances of over 18,000 individuals aged 50 and older in England. The study, which began in 2002, follows participants over time to examine various factors affecting aging, including physical and mental health, income, employment, and quality of life. ELSA provides valuable insights into the aging process, helping to inform policies related to health, social care, and pensions. Crucially, this database is nationally representative of the English population.

The restricted data version of ELSA includes information on the month of birth of individuals, which we use to identify those born before or after the implementation of the NHS. Furthermore, ELSA includes self-reported health data (rated on a scale from one to five, with higher values indicating worse health status) and BMI measurements taken later in life. This information is used to explore the long-term effects of the NHS.

In 2013 and 2014, genome-wide genotyping was conducted at University College London (UCL) on 7,597 ELSA participants of European ancestry using Illumina HumanOmni2.5 BeadChips, which measure approximately 2.5 million genetic markers with a 2.5% minor allele frequency (MAF). PGIs were constructed using externally valid SNP weights from GWAS. Whenever possible, meta-analyses that excluded ELSA in the discovery analysis were selected to ensure data independence. SNP weights were obtained from consortium webpages, requested from consortium authors, sourced

from dbGap, or extracted from published supplemental material. This approach enabled the creation of a broad set of PGIs capturing a wide range of phenotypes (Ajnakina & Steptoe, 2022).

The initial ELSA sample consists of 21,344 individuals. For our main specification, after restricting the sample to a bandwidth of one and a half years pre- and post-NHS implementation, the number of individuals decreases to 1,110. This further declines to 705 individuals when restricting the sample to those with genotyped information, which constitutes our final sample. Importantly, and as is common practice in the literature (Benjamin et al., 2024), this sample is restricted to individuals of European ancestry, primarily of white descent.

4.3 UKB

The UKB is a large-scale biomedical database and research resource containing in-depth genetic, health, and lifestyle information from approximately 500,000 UK participants aged 40 to 69 at recruitment (2006 to 2010). It integrates genetic data, physical and cognitive measures, biochemical assays, and extensive longitudinal health records, making it a valuable resource for studying complex disease etiology and risk factors. With its linkage to national health registries, UKB supports a wide range of research in genetics, epidemiology, and public health, enabling discoveries on disease prevention, treatment, and policy implications.

The UKB recruited participants through 22 assessment centers across the UK, inviting individuals via mailed letters. Approximately 9.2 million invitations were sent, yielding a response rate of around 5%. This low response rate resulted in a sample that is not nationally representative, as it is biased toward individuals with higher education, higher socioeconomic status (SES), and better overall health. Consequently, while UKB provides valuable insights into genetic and health-related research, findings must be interpreted with caution (Swanson, 2012). Recent efforts have aimed to improve national representativeness in the UKB by creating sample weights based on the UK census (van Alten et al., 2024).

The UKB has information on the month of birth of individuals, which allows us to identify if individuals where born before or after the implementation of the NHS. It also collects self-reported health measures (rated on a scale from one to six, with higher values indicating worse health status) and measures for BMI and Body Fat Percentage (BFP), all collected later in life.

The UKB collected blood samples from approximately 500,000 participants, and genotyping was performed using the UKB Axiom Array, designed to capture genome-wide genetic variation, including SNPs and short insertions and deletions (indels). DNA was extracted from the stored blood samples for this genotyping process. In addition, the UKB undertook whole exome sequencing for approximately 470,000 participants and whole genome sequencing for all 500,000

participants (Bycroft et al., 2018).

Given the widespread use of the UKB in most GWAS studies to date, a key challenge in constructing PGIs using UKB data is avoiding overfitting. Using the UKB for both GWAS discovery and main analysis risks circularity. To mitigate this, the UKB has been systematically partitioned into three subsets, where GWAS weights are derived from two-thirds of the sample and applied to the remaining third (Becker et al., 2021). This approach has led to the creation of three distinct UKB PGI repositories, which can be used independently but should not be combined.

The initial UKB sample consists of 502,336 individuals. Our main analysis focuses on the first PGI repository, which includes siblings. Although within-family analysis is not central to our work, it is used as a robustness check. This reduces our sample to 148,593 individuals. After narrowing the sample to the relevant time window of 18 months pre- and post-NHS, the sample shrinks to 22,748. Finally, when excluding individuals with missing month of birth or outcomes, the final sample consists of 21,344 individuals. Again, this sample is restricted to individuals of European ancestry.

5 Methodology

Our main identification strategy is an RDD, specifically a sharp RDD (Cattaneo et al., 2023; Cattaneo & Titiunik, 2022). This method relies on the assumption that the NHS introduced a discontinuity in access to the healthcare system in the UK. From one day to the next, the fifth of July 1948, pregnant mothers became eligible to deliver in a hospital setting free of cost (Rivett, 1998). Since pregnant mothers cannot precisely choose their delivery dates, some randomness was introduced around the cutoff for NHS eligibility. Below, we discuss the feasibility of this assumption and potential threats to identification.

The NHS scenario is best suited for a fuzzy RDD, as it did not change access to healthcare for everyone but rather affected the probability of pregnant mothers delivering at a hospital (Cattaneo & Titiunik, 2022). As such, the most appropriate approach would be to estimate a two stage least squares (TSLS), with the first stage estimating the effect of the NHS on hospital deliveries, and the second stage estimating the effects of increased hospital deliveries on our outcomes of interest. However, we do not have information on hospital deliveries, which makes it impossible to estimate a first stage. Therefore, we focus on the intention to treat (ITT), assuming that the NHS overall increased in-hospital deliveries, though we do not have data on the exact proportion. As reflected in the results, the use of a sharp RDD wont be a major problem as we still obtain precisely estimated effects.

Our main specification is given by:

$$Y_i = f(\text{month_birth}_i, X') + \beta T_i + \epsilon_i$$

where Y represents the outcome variables of interest, including IMR for the short-term effects, PGIs for the survival bias, and self-reported health and BMI (BFP) for the long-term effects.

The function $f(\text{month_birth}, X')$ is a linear specification of the birth month and a set of control variables, such as gender, age, quarter of birth fixed effects, and 10 ancestry specific principal components (PCs). For the long-term effects estimation, we also include PGIs as control variables. We choose a linear functional form for our main specification since our ELSA sample is small, and we want to avoid overfitting the data. Alternative functional forms are presented in the robustness checks.

T is the treatment indicator, where zero indicates being born just before the implementation of the NHS, and one indicates being born immediately after. β is the treatment estimator, capturing the difference in the outcome at the cutoff between the function plotted for observations just before and just after the NHS implementation.

ϵ represents the error term.

The baseline specification uses a bandwidth of two years for regressions based on the Registrar General's Statistical Review and 18 months for those using ELSA and UKB. Since the Registrar General's Statistical Review data is yearly, some estimations rely on very short bandwidths (two years pre- and post-NHS). To ensure convergence, we cluster at the individual level and use uniform kernels. As a robustness check, we extend the bandwidth and apply cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest-neighbor variance estimation and triangular kernels, our preferred specification. In ELSA, standard errors are clustered at the local authority district (LAD) level, with triangular kernels applied. This stricter clustering is imposed due to ELSA's small sample size, to prevent overfitting. For the UKB, we implement cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest-neighbor variance estimation and triangular kernels. Alternative specifications demonstrate the robustness of our results.

The UKB makes it possible to identify siblings, which will allow us to extend these models by including family fixed effects (Amin et al., 2022). This will explore sibling differences in PGIs before and after the implementation of the NHS, based on the assumption that the distribution of PGIs within families should remain constant over time. We will run a similar model to the one previously estimated, with PGIs as the outcome variable, treatment being post-NHS birth, date of birth (to control for cohort effects), demographics of the respondents (gender and date of

interview, capturing age), and 10 ancestry-specific principal components (PCs). Additionally, we will include demographics of the siblings (gender and date of interview), number of male, female, and adopted siblings, and the number of older siblings to control for family composition, as well as family-specific fixed effects. Unlike a RDD, this model will not include the interaction between the running variable and the NHS treatment because, by definition, we won't be capturing the exact difference at the cutoff of the NHS implementation, as siblings are born at different times. As such, we will capture the overall difference between pre- and post-NHS. Although this model will lose the precision of the NHS implementation date, it will allow us to compare siblings growing up in a similar environment and household, which should have been affected by the NHS in a comparable manner.

As in any RDD setting, there are several threats to the identification. To begin with, there is the possibility that individuals may have anticipated the implementation of the NHS. In this case, individuals could have postponed the decision to have children in order to be eligible to deliver in a hospital setting. However, the announcement of the NHS implementation was made in January 1948, and during the six months leading up to its implementation, it faced resistance in Parliament, with its final approval remaining uncertain (Webster, 2002). Therefore, individuals delivering at the cutoff time of the NHS implementation would have already been pregnant at the time of its announcement. There may have been individuals who actively avoided pregnancy because they could not afford hospital costs, and who decided to have children after the NHS was announced. However, such individuals would not have affected the dynamics at the cutoff of the NHS implementation. Furthermore, the existence of a large group of such individuals seems implausible, given the limited availability of contraceptive methods at the time, with significant breakthroughs only occurring in the 1960s when the contraceptive pill became widely accessible (Murphy, 1993). Unfortunately, we only have yearly birth data form the Registrar General's Statistic Review, making it impossible to provide further evidence on the size of these demographic shifts. However, based on the arguments above, we assume that such reactions were marginal or non-existent.

A further thread is the occurrence of events parallel to the introduction of the NHS, which could potentially confound its effects. For instance, bread rationing was introduced on 21 July 1946 and ended on 24 July 1948, just weeks after the NHS was established. However, rationing during the post-war period was intricate, and focusing on a single item would be an oversimplification. This is evident in the fluctuations of rationed items throughout this time. Key events included reductions in bacon, fat, and soap rations in 1945; the introduction of bread and potato rationing due to poor harvests in 1946 to 1947; and strikes that disrupted food imports. Meanwhile, petrol

rationing underwent repeated adjustments, with restrictions easing in 1948 and ending in 1950. Clothes rationing concluded in 1949, and the gradual dismantling of food rationing culminated with its complete removal in July 1954 (Zweiniger-Bargielowska, 2000; Kynaston, 2008; Zweiniger-Bargielowska, 1994). To make sense of these dynamics, we retrieved data from The National Food Survey 1945–1949, focusing on aggregated calorie consumption from 1945 to 1950 (of Food, 1951). Daily calorie intake in 1945, the last year of war, was 2,375; it decreased slightly to 2,307 in 1946 and 2,308 in 1947; increased to 2,387 in 1948; and reached 2,425 in 1949. These data reveal that calorie intake in 1948 closely resembled levels in 1945 and showed only minor variation compared to 1946 and 1947. This stability suggests that changes in food rationing are unlikely to account for the significant effects reported in this paper.

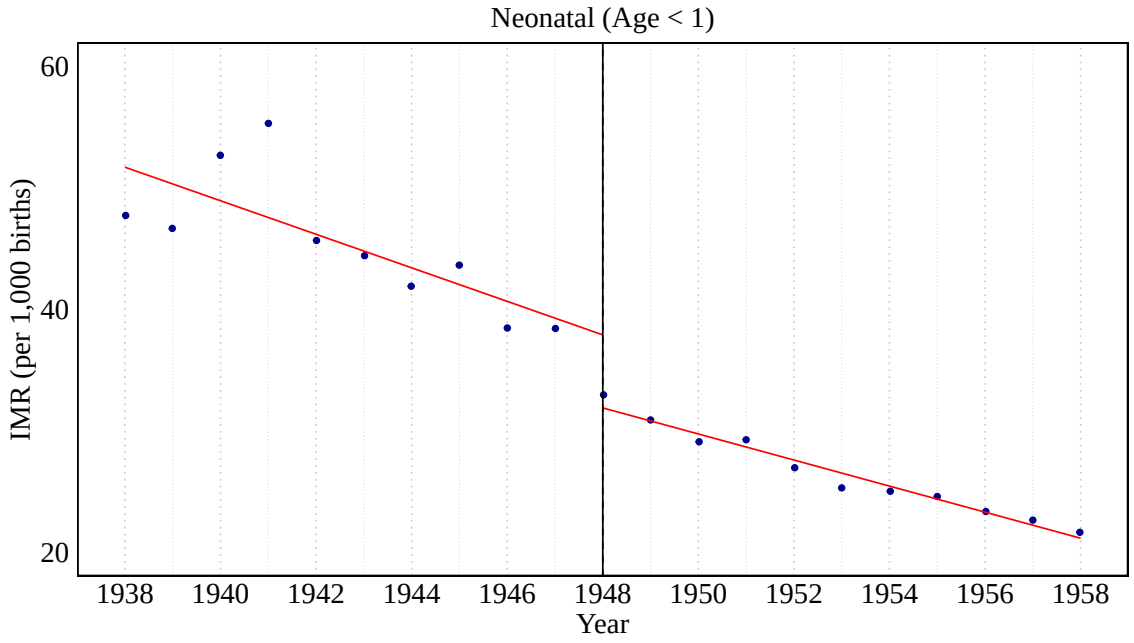
A further limitation arises from the fact that the NHS was introduced on a specific day, the 5th of July, while the available data only captures individuals' month of birth. This creates a problem because the exact timing of the NHS implementation is lost, making it difficult to identify which individuals are treated and which are not, as we can only group them by their month of birth. This aggregation reduces the precision required for a sharp RDD, where treatment assignment depends on a clear and immediate cutoff. To address this challenge, as a robustness test we use RDD under local randomization as a robustness check (Branson & Mealli, 2018). Local randomization approximates a quasi-experimental setting by focusing on individuals close to the cutoff date and assuming that they are similar in all respects except for the treatment assignment. We find no significant differences in the results, further confirming the robustness of our findings despite this data limitation.

6 Results

6.1 Short-term Effects

6.1.1 Main Results

To begin, we explore the short-term effects of the NHS on IMR. These results are crucial, as the rest of the paper builds upon them. Only if we find substantial effects on IMR, suggesting significant demographic changes, does it make sense to trace genetic footprints in the population. These results are based on data from the Registrar General's Statistical Review. Figure 2 presents compelling evidence of a significant decrease in IMR following the implementation of the NHS. What makes this figure particularly convincing is that it descriptively aligns well with several other historical events during that period.



	Bandwidth two years	Bandwidth six years
Coefficient	−5.48	−3.90
SE	1.04	0.61
p−value	0.00	0.00
Observations pre−NHS	2,942	7,352
Observations post−NHS	4,410	8,822
Average	33.96	34.76

Figure 2: Short-Term Effects of the NHS on Infant Mortality Rates
Note: Source Registrar General's Statistical Review.

To begin with, this graph reflects the start of WWII. Unsurprisingly, in 1940 and 1941, IMR increased by 7.34 deaths per 1,000 births, or 15.7% compared to 1939. This rise aligns not only with the onset of WWII but, more importantly, with the start of the Blitz, a period of sustained bombing raids on British cities between 1940 and 1941 that caused widespread destruction and loss of life. The resulting strain on the healthcare system and living conditions likely contributed to the sharp increase in IMR during this time (Gardiner, 2010).

Interestingly, a major drop in IMR occurred in 1942, in the midst of WWII, decreasing by 8.33 deaths per 1,000 births or 15.4% compared to the previous year. This is consistent with the start of Operation Barbarossa, which involved Nazi Germany's invasion of the Soviet Union beginning on June 22, 1941. It marked the largest military operation in history, with millions of soldiers advancing toward key Soviet territories, including Ukraine and the Caucasus region, leading to a significant shift in the course of the war. This forced the Nazis to divert resources from the Blitz, contributing to the end of that campaign (Erickson, 2019).

A further decrease in IMR can be observed in 1946, after the end of WWII, with a reduction by

5.18 deaths per 1,000 births or 11.8% compared to the previous year. A final discontinuous drop can be seen in 1948, with the implementation of the NHS. This drop is substantial in magnitude, involving a reduction in IMR by 5.47 deaths per 1,000 births or 14.2% compared to the previous year. Crucially, subsequent to the NHS implementation, no further discontinuities are observed. On the contrary, a very linear and smooth decreasing trend in IMR can be followed, which is expected with improved economic conditions, sanitation, and public policies (Marshall et al., 2019). In summary, this figure strongly aligns with our expectations of how different historical events should have affected IMR, providing confidence in the quality of the Registrar General's Statistical Review data and its ability to capture historical trends.

Focusing on the implementation of the NHS and its causal impact on IMR, the table at the bottom of Figure 2 presents estimates, standard errors, and p-values from a sharp RDD model exploring the causal effect of the NHS on IMR. When considering a strict bandwidth of two years pre- and post-NHS implementation, the effect of the NHS was a reduction of 5.48 deaths per 1,000 births, or a reduction of 16.1% compared to the mean (33.96). The effect is statistically significant at the 99% confidence level. Importantly, the two-year bandwidth pre- and post-NHS avoids including WWII time periods, preventing contamination of the pre-NHS period. When including a more generous time window of six years pre- and post-NHS implementation, the results maintain the same trend, with a reduction of 4.38 deaths per 1,000 births, or 12.6% compared to the mean (34.76), which is also precisely estimated.

These results point to a major effect of the NHS on IMR, which is crucial for justifying the existence of survival bias. Based on the theoretical framework of (Nobles & Hamoudi, 2019), the NHS had a health effect that enabled a significant group of individuals to surpass the threshold for survival at birth. This created important differences between the individuals born before and after that event, differences that could be captured in individuals' genetics. In the next section, we attempt to capture these differences, tracing the genetic footprint of the NHS.

6.1.2 Robustness Checks

We run the following robustness checks. To begin with, we add a quadratic fit to the function imposed pre- and post-NHS implementation. Table A.1 shows that this alternative functional form more than doubles the estimated effects while remaining precisely estimated. Nevertheless, the descriptive Figure 2 suggests that a linear fit suits the data best, and as such, the previous specification remains our preferred choice.

In Table A.2, we explore the effects of the NHS on total mortality. The results indicate a substantial reduction in total mortality following the implementation of the NHS, further rein-

forcing the notion that this was a highly effective policy. We also explore potential differences in mortality between females and males. As suggested in Section 3.2, one possible effect of the NHS could be a reduction in maternal mortality at childbirth, which might result in higher mortality rates for females compared to males. However, our results do not support this hypothesis, as male mortality remains 23.8% higher than female mortality. This discrepancy is likely due to the fact that mortality at childbirth constitutes only a small share of total mortality, causing any effects in that direction to be diluted within the overall mortality patterns (Say et al., 2014).

6.2 Survival Bias

6.2.1 Main Results

In this section, we explore the effects of the NHS on different PGIs observed in the population. Genetics are determined at birth and as such PGIs are exogenous to the implementation of the NHS, which should have no effect on them. However, the NHS could have influenced survival rates. If a sufficiently large number of individuals with distinct genetic traits survived after the implementation of the NHS, this could have left a genetic footprint, reflected in changes in the PGIs observed pre- and post-NHS. This way, although the NHS would have not directly affected the genetics of individuals, it would have affected the genetic distribution at a population level, and this is what we explore with the models presented in this section.

We analyze data from ELSA and UKB, as both databases provide large sets of PGIs. A challenge arises when deciding which PGIs to explore. Examining all PGIs could be problematic, as it would require running numerous regressions for a large number of outcomes. This would not only make interpreting the results challenging but also increase the risk of spuriously identifying effects on different PGIs and mistakenly attributing them to the causal effect of the NHS when, in reality, they may simply result from multiple hypothesis testing (Benjamini & Hochberg, 1995; Dunn, 1961). On the other hand, selecting only specific PGIs poses the risk of consciously or unconsciously introducing publication bias by choosing PGIs where significant effects are found and justifying their selection post hoc (Sterling, 1959).

To mitigate this threat, we take advantage of both ELSA and UKB categorizing PGIs into meaningful conceptual groups. We focus on groups associated with physical conditions that could contribute to resilience in-utero, as well as groups that include the most widely used PGIs, such as educational attainment. To further reduce the number of outcomes, in cases where multiple PGIs are available, either from different GWASs or derived from single versus multiple imputation methods, we prioritize PGIs from the most recent GWAS and focus on estimates that include

multiple imputations. For ELSA, we choose PGIs in the categories of socio-economic traits, adult mental health and well-being, physical health, anthropometric traits, and intelligence (Ajnakina & Steptoe, 2022). For UKB, we select PGIs in the categories of anthropometric measures, cognition and education, health and health behaviors, and personality and well-being (Becker et al., 2021). However, as a robustness check, we also present results including all available PGIs and apply multiple hypothesis testing adjustments using different methods. Both approaches confirm the robustness and consistency of our findings.

We begin with the results from the nationally representative ELSA database. In Figure 3, we present a forest plot of the different effects of the NHS on various PGIs. Each dot represents the causal effect of a separate regression with a different PGI as the outcome, along with its 95% confidence interval. The results show a clear trend: PGIs associated with more detrimental conditions, such as schizophrenia, or migraine, take on higher values after the implementation of the NHS, while PGIs linked to more beneficial conditions, such as IQ or educational attainment, show lower values. For many of the PGIs at the middle range, it remains unclear if their values increased or decreased after the implementation of the NHS. This could be driven by some PGIs having limited predictive power (Benjamin et al., 2024). On the other hand, the relationship between certain PGIs and resilience at birth remains unclear in many instances. For example, while a higher BMI later in life is a known risk factor (Sowers, 2003), higher birth weight can be considered a protective factor in early life (Diabelková et al., 2022). Nonetheless, focusing on the extremes of the figure, a clear pattern emerges. Importantly, the largest effect sizes range up to 40% of a SD. However, these very substantial effects could be misleading, as these estimates are noisy due to the small sample size in ELSA. The confidence intervals for the highest (schizophrenia) and lowest (IQ) estimated effects span nearly 60% of a SD. Therefore, while the effect of schizophrenia is reported to be 40% of a SD, it could, with 95% confidence, be as low as 10% of a SD.

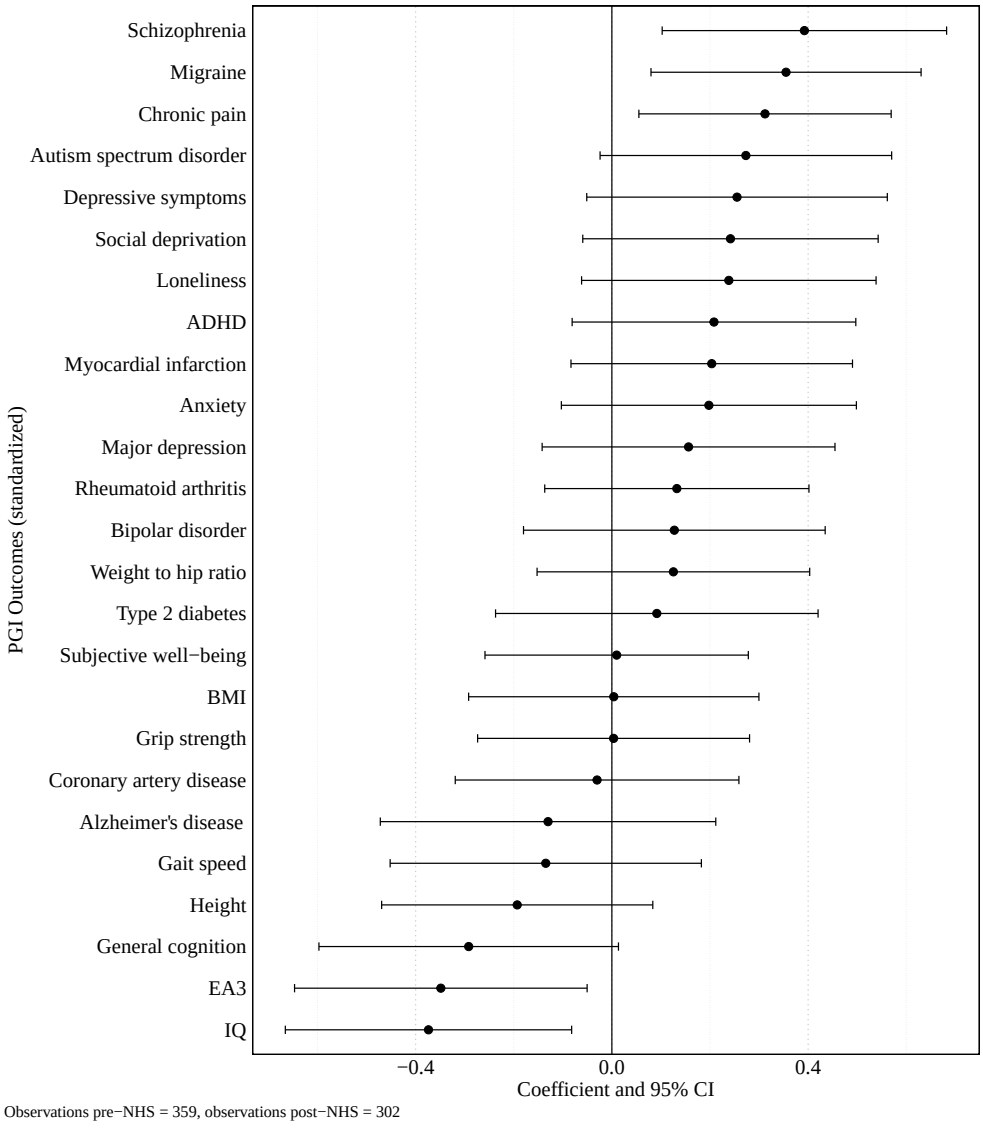


Figure 3: Survival Effects of the NHS on PGIs using ELSA

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Standard errors clustered at the LAD level. The bandwidth is 18 months per- and post-NHS implementation. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.

Next, we repeat the analysis using the UKB, which provides a much larger sample size. However, as discussed in section 4.3, it is biased towards higher SES individuals (Swanson, 2012). In Figure 4, we again present a forest plot of the NHS effect on various PGIs. The previously reported pattern reappears, with PGIs associated with more detrimental conditions, such as depression or COPD, exhibiting higher values after the implementation of the NHS, while PGIs linked to more beneficial conditions, such as life satisfaction with finances or self-rated health, show lower values. These results are estimated with much greater precision, likely due to the much larger sample size of the UKB. The effects on the upper and lower ranges of the table now span about 10% of

a standard deviation, which aligns with the lowest end estimations from Figure 3. Overall, the results present a consistent and robust pattern. This is particularly convincing given the differing nature of both databases: one has a small sample size but is nationally representative (ELSA), while the other has a much larger sample size but is biased toward higher SES individuals (UKB).

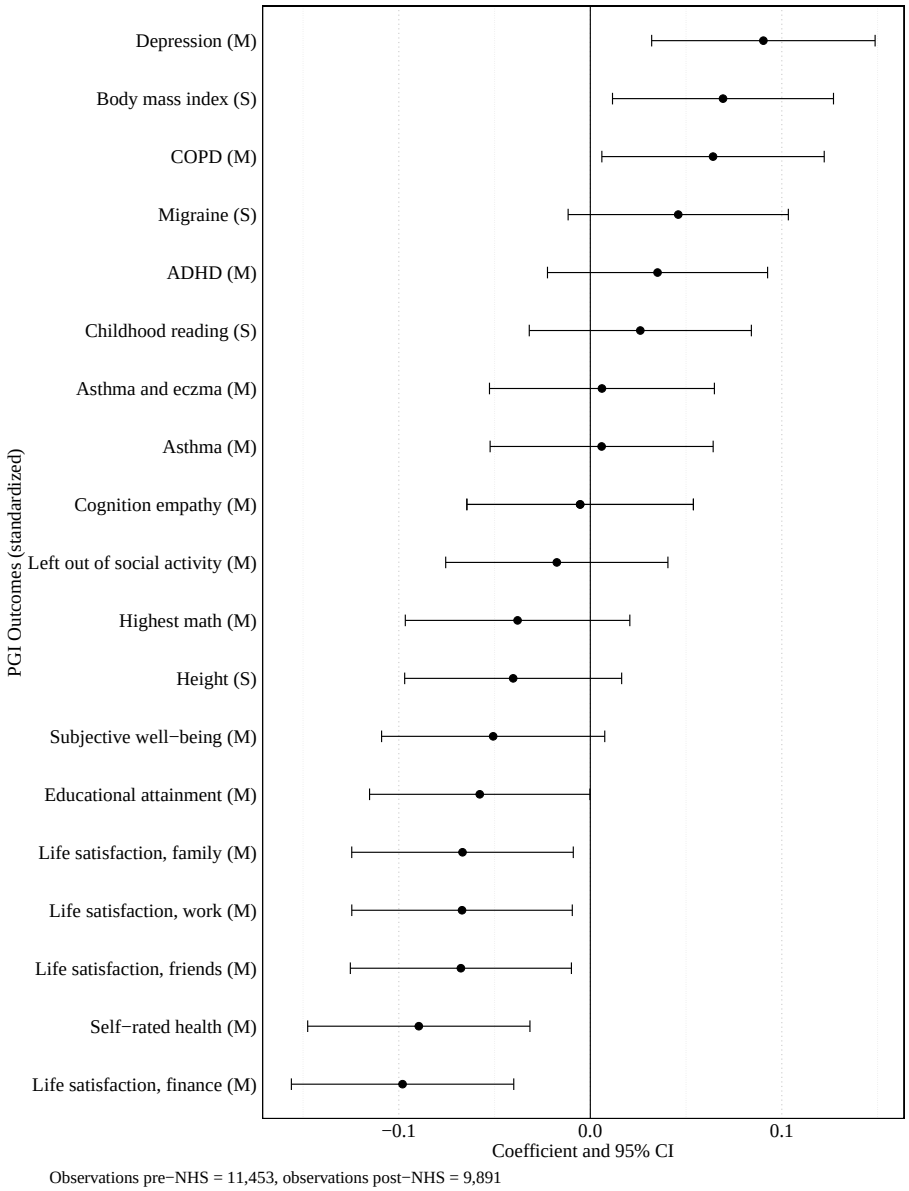


Figure 4: Survival Effects of the NHS on PGIs using UKB

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. The bandwidth is 18 months per- and post-NHS implementation. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

Finally, we run an analysis based on the UKB exploring differences in PGIs within sibling pairs, controlling for family fixed effects (no siblings can be identified in ELSA). Figure 5 shows that although the coefficients are estimated with much more noise and most confidence intervals overlap the zero line, the previously observed pattern is still clearly visible. The interpretation of these results is that, within families, siblings born after the implementation of the NHS have higher PGI values associated with qualitatively detrimental conditions, while conditions associated with qualitatively beneficial outcomes show lower values. The estimates are large in magnitude, reaching up to 30% of the SD. The larger and noisier estimates are driven by the much smaller siblings sample, consisting of 595 families and 1,190 sibling pairs. When extending the analysis to a bandwidth of 36 months (Figure B.1), with a sample size of 2,373. families and 4,746. sibling pairs, the results are more precisely estimated, while the same pattern remains.

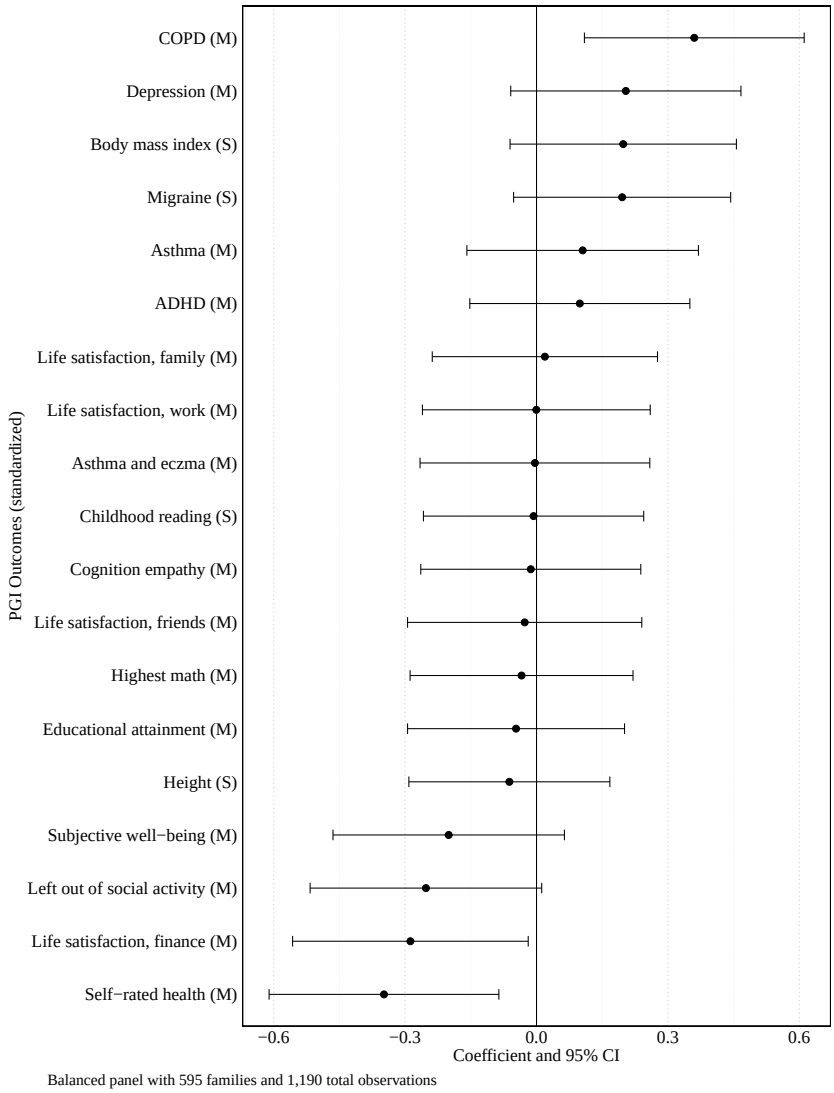


Figure 5: Effect of the NHS on PGIs using UKB: Sample of Siblings and Family Fixed Effects

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. The bandwidth is 18 months per- and post-NHS implementation. Controls include year-month of birth, interview date (age), gender, quarter of birth fixed effects, sibling characteristics including year-month of birth, gender, number of siblings (males, females, adopted males, adopted females), and number of older siblings, 10 ancestry-specific PCs, and family fixed effects. Source: UKB.

A key limitation of our analysis is that survival at birth should be linked to resilience in early life, which we aim to capture using various PGIs. However, this approach assumes that certain SNPs prevalent in the examined PGIs are indeed relevant to resilience. A more direct approach would involve focusing on more relevant phenotypes to survival at birth, such as birth weight or other related proxies. In the 'Next Steps' section, we explain how we plan to address this limitation. That said, even without knowing the underlying biological effects of these PGIs, the observed patterns are unlikely to be driven by chance. Moreover, the consistency across the various

robustness checks presented in the next section, suggests that we are capturing some real patterns.

6.2.2 Robustness Checks

For both the ELSA and UKB datasets, we perform multiple robustness checks. To begin with, we replicate the results using different bandwidths, kernels, and functional forms. Results for ELSA, using different bandwidths and kernels, are shown in Figure B.2, and with a squared functional form in Figure B.3. The results are highly consistent, with the main differences observed when reducing the bandwidth to 12 months. Effect sizes largely increase and become noisier, likely due to the small sample size of ELSA, but the results remain qualitatively similar. For the UKB, results using different bandwidths can be found in Figures B.4 and B.5, results using different kernels in Figures B.6 and B.7, and results using a quadratic fit in Figure B.8. Results for the UKB are highly consistent across specifications, which is expected due to its much larger sample size. The quadratic fit does increase the treatment effects for both ELSA and UKB, which could signal overfitting observations with extreme values. For this reason, we prefer to take the more conservative linear fit as our preferred specification.

Second, we estimate the sharp RDD model under local randomization. This approach addresses the limitation of our monthly-level data, given that the NHS implementation occurred on a daily level. Local randomization approximates a quasi-experimental setting by focusing on individuals near the cutoff date and assuming they are similar in all respects except for treatment assignment (Branson & Mealli, 2018). Figures B.9 for ELSA and B.10 for UKB show that applying this technique does not lead to major differences in the results.

Third, we adjust our results for multiple hypothesis testing using the Benjamini-Hochberg (Benjamini & Hochberg, 1995) and Bonferroni corrections (Dunn, 1961). Figures B.11 for ELSA and B.12 for the UKB show that regardless of the method we use, the p-values for the PGIs at the extremes of the forest plots in Figures 3 and 4 remain statistically significant after adjusting for multiple hypothesis testing.

Fourth, we perform a permutation test to assess whether our results could arise by chance. Specifically, we randomly reassign treatment to different months while preserving the structure of the data and estimate the treatment effect for each permutation. By repeating this process 3,000 times, we generate a distribution of placebo treatment effects under the null hypothesis of no true effect. We then compare the observed treatment effect to this empirical distribution to determine whether it is statistically distinguishable from chance. Figures B.13 for ELSA and B.14 for the UKB show that, for the PGIs with the largest observed effects, the differences between the placebo and the treatment effects are statistically significant.

Fifth, we extend our analysis to the full set of PGIs available in ELSA and UKB, which confirms the pattern observed in the selected subset of PGIs. Figures B.15 for ELSA and B.16 for UKB illustrate this consistency. As expected, the expanded set includes more noisy estimates, but the overall pattern remain robust: PGIs associated with qualitatively more detrimental phenotypes tend to increase, while those linked to more favorable traits tend to decline. Interestingly, several "older generation PGIs" in ELSA capture the strongest effects of the NHS.

Finally, we conduct a heterogeneity analysis by gender. Previous research on SSR (Nobles & Hamoudi, 2019) suggests that males are more vulnerable in-utero and should, therefore, have been more affected by the implementation of the NHS. We do not observe significant gender differences in the ELSA data, as shown in Figure B.17. However, the small sample size in ELSA likely limits the ability to detect such heterogeneity. This is further supported by the UKB data, illustrated in Figures B.18 and B.19, where effect sizes for males are nearly twice as large as for females. These findings are consistent with SSR and further support the robustness of our results.

6.3 Long-term Effects

6.3.1 Main Results

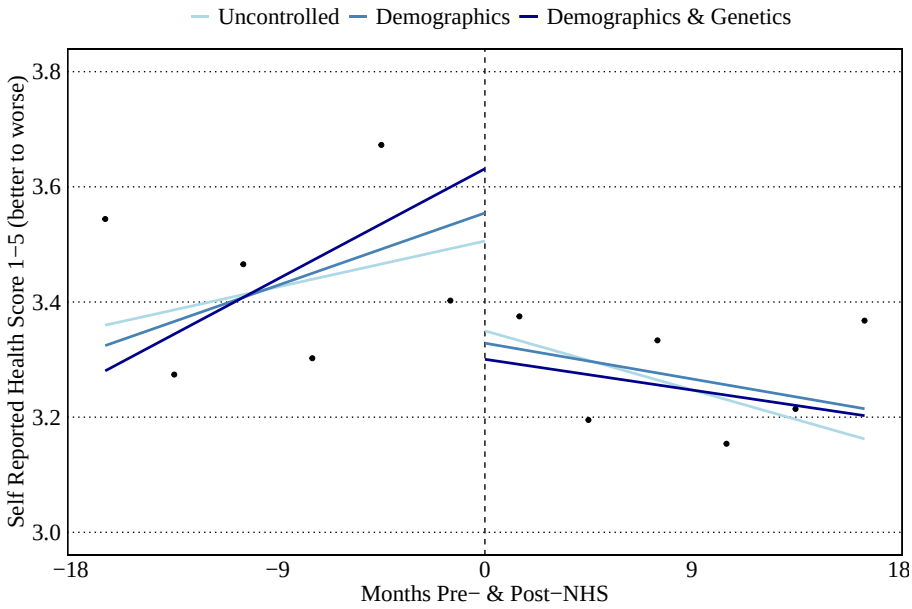
In the previous section, we presented robust evidence of how survival bias can be captured using genetic data. Referring back to the DAG diagram in Figure 1, the proposed method examines whether a substantial bias exists, as indicated by the vertical dashed arrow connecting "Perinatal shocks" with "Survival or mortality bias" But what happens if significant bias is identified? Can short- and long-term outcomes still be explored? Is it possible to reduce the size of this bias? In this section, we undertake an exercise to attempt reducing survival bias and estimate more accurate short- and long-term effects.

Our approach involves adding PGIs as control variables when exploring long-term effects. We use the same information that helped test for survival bias to correct for it, thereby closing the path from "Survival or mortality bias" to "Short & long-term outcomes" in Figure 1. This may seem problematic at first glance, as we have demonstrated an effect of the NHS on the PGIs, and now we want to include this seemingly endogenous variables as a controls. However, as discussed earlier, the NHS did not directly alter individuals' genetics but rather changed the genetic distribution in the population. Genetics are pre-imposed at birth and exogenous to the NHS. Therefore, by including PGIs as controls, we rebalance the sample pre- and post-NHS implementation, and attempt to address the survival bias.

To perform this exercise, we choose outcome variables that are available in both the ELSA

and UKB datasets and have been shown to be sensitive to shocks during gestation. The first variable is self-reported health, as Jeffrey (2023) reported improvements in this outcome after the implementation of the NHS. The second variable is BMI (BFP), as the other paper most related to our work (Zhou et al., 2024) also focuses on that outcome.

Again, we begin by exploring the results obtained from the ELSA. Figure 6 presents an RDD plot with self-reported health before and after the implementation of the NHS as an outcome. The graph reveals a discontinuity following the NHS implementation, with improvements in self-reported health, aligning with the findings of Jeffrey (2023). However, the graph is noisy, likely due to the small sample size of ELSA. There are three fitted lines with varying shades of blue: a light blue line representing an entirely uncontrolled model, a medium blue controlling for basic demographics, and a dark blue incorporating both basic demographics and genetic information (PGIs). Notably, the fitted lines shift with the inclusion of different controls, with the most pronounced change occurring when PGIs are added. For greater precision, the table below the figure provides RDD estimates, standard errors, and p-values capturing the causal effect of the NHS on self-reported health. The fully uncontrolled model indicates an improvement in self-reported health of 0.16 following the NHS implementation, representing a 4.7% relative increase compared to the mean. This coefficient increases in magnitude when basic demographics are included. However, the most substantial change arises when PGIs are added as controls, raising the estimated effect on self-reported health to 0.33, or 9.8% relative to the mean. This larger effect size helps mitigate the limitations of ELSA's small sample size, and the results become statistically significant at the 95% confidence level. We interpret this increase in the coefficients as an adjustment for survival bias.

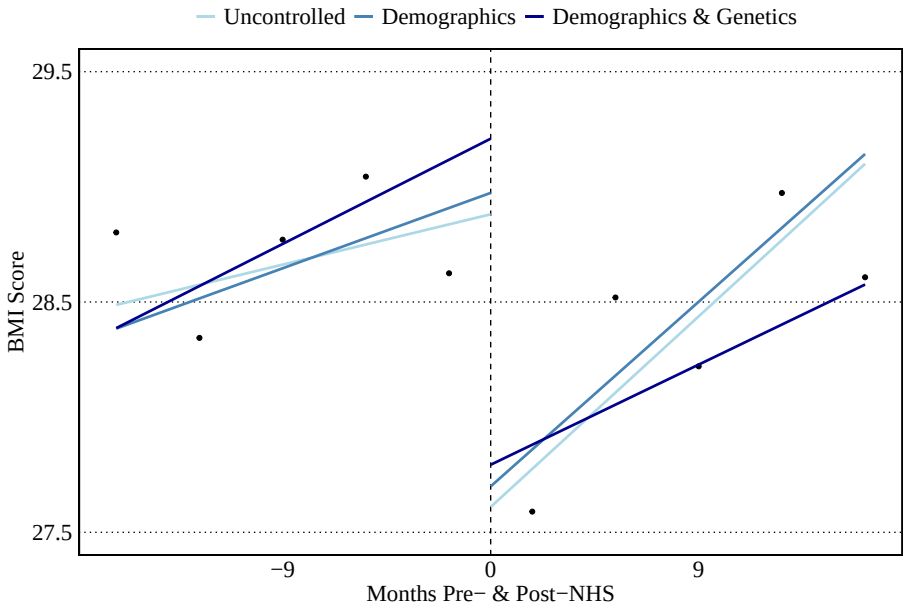


	Uncontrolled	Demographics	Demographics and Genetics
Coefficient	−0.16	−0.23	−0.33
SE	0.16	0.15	0.14
p−value	0.33	0.14	0.02
Observations pre−NHS	359	359	359
Observations post−NHS	302	302	302
Average	3.37	3.37	3.37

Figure 6: Long-term effects of the NHS on Self-reported Health using ELSA

Note: Standard errors clustered at the LAD level. The bandwidth is 18 months per- and post-NHS implementation. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.

These results remain consistent when we repeat the analysis using BMI as an outcome in Figure 7. The discontinuity in this graph is much clearer, indicating a decrease in BMI after the implementation of the NHS, which can be considered a beneficial health effect given the overall high BMI levels observed in this sample (Sowers, 2003). This beneficial effect of an early life intervention on BMI aligns with the work of Barker (1990, 1995). There is an overall increasing trend in BMI over time, which is expected if younger cohorts have higher BMI levels. Crucially, the table below the figure shows that when adding basic demographics as control variables, the main coefficients remain unchanged. However, the inclusion of PGIs as controls leads to a 10.6% increase in the treatment effect. Once more, this larger effect size compensates for the small ELSA sample, and the effect now becomes significant at the 90% confidence level.



	Uncontrolled	Demographics	Demographics and Genetics
Coefficient	−1.27	−1.27	−1.42
SE	0.81	0.79	0.72
p−value	0.12	0.11	0.05
Observations pre−NHS	357	357	357
Observations post−NHS	299	299	299
Average	28.53	28.53	28.53

Figure 7: Long-term effects of the NHS on BMI using ELSA

Note: Standard errors clustered at the LAD level. The bandwidth is 18 months per- and post-NHS implementation. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.

Unfortunately, these results cannot be replicated with the UKB, leaving them inconclusive. Table 1 repeats the RDD analysis using the UKB, focusing on self-reported health, BMI, and BFP as outcomes. Surprisingly, in Panel A, where no controls are added, the results are the opposite of those observed with ELSA. They suggest that the NHS had a detrimental effect on individuals born after its implementation, worsening self-reported health and increasing BMI and BFP. Crucially, these results are very imprecisely estimated. This remains unchanged with the inclusion of basic demographic controls and PGIs in Panel B. This pattern could be explained by the fact that the UKB is biased toward higher-SES individuals (Swanson, 2012), who may have been affected by the NHS differently than the rest of the population. For instance, it is likely that high-SES individuals already had access to healthcare services before the NHS was established. Since the NHS primarily increased hospital capacity without introducing new services (Jeffrey, 2023), this may have worsened healthcare access for high-SES individuals. To test this hypothesis, we incorporate sample-representative weights in Panel C (van Alten et al., 2024). However, the results are mixed, with the inclusion of weights increasing the effects on self-reported health and BFP while reducing the effect on BMI. A key difference from the ELSA results is that the inclusion of PGIs does not appear to affect the estimates.

Table 1: Long-term Effects of the NHS on Self-reported Health, BMI, and BFP

	Self-reported health	BMI	BFP
Panel A: No controls	0.025	0.147	0.339
	0.247	0.3	0.067
Panel B: Controls	0.024	0.145	0.336
	0.252	0.309	0.069
Panel C: Controls & Weights	0.048	0.056	0.393
	0.282	0.818	0.153
Observations pre-NHS	10,939	10,806	10,803
Observations post-NHS	9,390	9,287	9,286

Note: Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation, standard errors in parentheses, *** p<0.001; ** p<0.01; * p<0.05. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. The bandwidth is 18 months per- and post-NHS implementation. Linear fit and triangular kernels. Source: UKB.

These are somewhat unsatisfying result, as the promising pattern observed in ELSA could not be replicated in the UKB. Our interpretation of these results is that bias in the UKB is so substantial (Swanson, 2012) that, even after applying sample-representative weights (van Alten et al., 2024), this database cannot be directly compared to ELSA. Nevertheless, we believe these findings remain valuable, as they provide insights into potential ways to address survival bias. Further research should continue to assess whether this approach could effectively minimize this type of bias.

6.3.2 Robustness Checks

In this section, we conduct a series of robustness checks on the results derived from ELSA. While the analysis of long-term outcomes remains inconclusive and these checks do not alter the UKB results, it is important to emphasize that the findings from the ELSA analysis are robust across various tests and are not merely spurious.

To begin with, we test whether these results hold when considering other closely related outcomes, namely the prevalence of obesity, defined as BMI levels over 30, and ever being diagnosed with heart problems. Tables C.1, C.2, and C.3 further present results using different bandwidths, kernels, and a quadratic fit. Overall, results with different bandwidths and kernels show a consistent trend, with effect sizes remaining similar. However, results using the quadratic fit reduce precision in the self-reported health score and nearly double the coefficient size for BMI. As previously argued, we prefer the more conservative linear fit to avoid overfitting the ELSA data. Importantly, results for obesity are highly consistent across specifications, whereas results for heart problem diagnoses are mixed and noisy. This could be due to heart problems being a more extreme health condition that is less affected by early-life circumstances.

Second, the ELSA is a panel with several measurement periods for these health measurments. For the main results we have been using the maximal value observed for individauls accross the time series. Yet, different approaches could have been used. Table C.4 shows that using the first measurement ever observed or taking the minimal value would not change the patterns in these results.

Finally, we conduct a series of sanity checks to examine the effects of the NHS on outcomes that should not have been influenced by it. These placebo outcomes include household conditions during childhood, nationality, gender, and religion. Figure C.1 shows that, for most of these outcomes, the estimated effects are close to zero. There are two outcomes with larger effects, such as a nearly 20% increase in the likelihood of having a toilet in the household or in the number of rooms. Yet, the results are very imprecisely estimated, with confidence intervals largely overlapping zero. These results suggests that our analysis is not merely capturing a mechanical effect in the ELSA data, as outcomes that should not have been affected by the implementation of the NHS indeed show no meaningful impact.

7 Next Steps

This paper combines several databases that include restrictive information (genotypes and other sensitive data) and novel techniques that have not yet been established in the literature. We believe

that this novelty is one of the main contributions of our work. Nevertheless, this has also limited the pace and progress of this project because we depend on access to several resources, for which we are still waiting. For this reason, we have added a detailed 'Next Steps' section describing the analyses we will perform once we gain access to these resources.

To begin with, as already mentioned, a limitation of our method to detect survival bias is that it assumes the SNPs captured in certain PGIs do also predict resilience in-utero or early-life. This is a strong assumption, as many of these phenotypes seem to be far removed from relevant survival factors during early-life (for instance, mental health). Even with phenotypes that are more closely related (for instance, anthropometric measures), the biological mechanisms behind these PGIs remain uncertain, and it is unclear what they are actually capturing (Benjamin et al., 2024). In order to address this limitation, we will construct our own PGIs based on the GWAS of the Early Growth Genetics Consortium (EGG), which facilitates joint analysis of genetic data from multiple studies to identify human genomic loci that impact a variety of traits related to early growth. In particular, the summary statistics for birthweight obtained by (Warrington et al., 2019) seem particularly relevant, although there are many more that are worth exploring. A further advantage of this method is that it will allow us to compare the results from both ELSA and the UKB using the same summary statistics.

As a further step, we will conduct a GWAS analysis using exposure to the NHS as an outcome. This will identify SNPs directly associated with survival in early-life, allowing us to construct a relevant PGI. This seems particularly promising, as it will be the first time that an econometric technique such as RDD is used in GWAS analysis. For both approaches, we need access to the full genotypes for both ELSA and the UKB, for which access we are currently waiting.

Finally, we are currently in the process of gaining access to Understanding Society (USoc), a nationally representative survey of the British population that includes cohorts affected by the NHS implementation. It includes information about individuals' genotypes and health outcomes measured at older ages. Having a second nationally representative sample will shed light on the long-term results. Although this section will remain inconclusive, as the UKB results will not change, having evidence from a second nationally representative database will add valuable insights to that section.

8 Discussion

This section is forthcoming.

References

Ajnakina, O., & Steptoe, A. (2022). *The english longitudinal study of ageing (elsa) polygenic scores. 2020.*

Alberts, B., Bray, D., Hopkin, K., Johnson, A. D., Lewis, J., Raff, M., ... Walter, P. (2015). *Essential cell biology.* Garland Science.

Almond, D. (2006). Is the 1918 influenza pandemic over? long-term effects of in utero influenza exposure in the post-1940 us population. *Journal of political Economy*, *114*(4), 672–712.

Almond, D., & Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of economic perspectives*, *25*(3), 153–172.

Almond, D., Edlund, L., & Palme, M. (2009). Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in sweden. *The Quarterly journal of economics*, *124*(4), 1729–1772.

Amin, V., Fletcher, J. M., Sun, Z., & Lu, Q. (2022). Higher educational attainment is associated with longer telomeres in midlife: Evidence from sibling comparisons in the uk biobank. *SSM-Population Health*, *17*, 101018.

Arck, P. C., Rücke, M., Rose, M., Szekeres-Bartho, J., Douglas, A. J., Pritsch, M., ... others (2008). Early risk factors for miscarriage: a prospective cohort study in pregnant women. *Reproductive biomedicine online*, *17*(1), 101–113.

Backes, E. P., & Scrimshaw, S. C. (2020, February). Birth settings in america: Outcomes, quality, access, and choice. In (chap. 6, Maternal and Newborn Outcomes by Birth Setting). Washington, DC: National Academies Press (US). Retrieved from `https://www.ncbi.nlm.nih.gov/books/NBK555483/`

Bailey, L. B., & Berry, R. J. (2005). Folic acid supplementation and the occurrence of congenital heart defects, orofacial clefts, multiple births, and miscarriage. *The American journal of clinical nutrition*, *81*(5), 1213S–1217S.

Baker, S., Biroli, P., van Kippersluis, H., & von Hinke, S. (n.d.). Advantageous early-life environments cushion the genetic risk for ischemic heart disease. , *121*(27), e2314056121. Retrieved 2025-01-31, from `https://www.pnas.org/doi/10.1073/pnas.2314056121` doi: 10.1073/pnas.2314056121

Barker, D. J. (1990). The fetal and infant origins of adult disease. *BMJ: British Medical Journal*, *301*(6761), 1111.

Barker, D. J. (1995). Fetal origins of coronary heart disease. *Bmj*, *311*(6998), 171–174.

Becker, J., Burik, C. A., Goldman, G., Wang, N., Jayashankar, H., Bennett, M., ... others (2021). Resource profile and user guide of the polygenic index repository. *Nature human behaviour*, *5*(12), 1744–1758.

Benjamin, D. J., Cesarini, D., Turley, P., & Young, A. S. (2024). Social-science genomics: Progress, challenges, and future directions.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289–300.

Bhalotra, S. R., & Venkataramani, A. (2011). The captain of the men of death and his shadow: Long-run impacts of early life pneumonia exposure.

Biroli, P., Galama, T. J., von Hinke, S., Van Kippersluis, H., Rietveld, C. A., & Thom, K. (2022). The economics and econometrics of gene-environment interplay. *arXiv preprint arXiv:2203.00729*.

Black, S. E., Bütikofer, A., Devereux, P. J., & Salvanes, K. G. (2019). This is only a test? long-run and intergenerational impacts of prenatal exposure to radioactive fallout. *Review of Economics and Statistics*, *101*(3), 531–546.

Branson, Z., & Mealli, F. (2018). The local randomization framework for regression discontinuity designs: A review and some extensions. *arXiv preprint arXiv:1810.02761*.

Brown, D. W., Kowalski, A. E., & Lurie, I. Z. (2015). Medicaid as an investment in children: what is the long-term impact on tax receipts?

Bufe, S., Roll, S., Kondratjeva, O., Skees, S., & Grinstein-Weiss, M. (2021). Financial shocks and financial well-being: what builds resiliency in lower-income households? *Social Indicators Research*, 1–29.

Bütikofer, A., Løken, K. V., & Salvanes, K. G. (2019). Infant health care and long-term outcomes. *Review of Economics and Statistics*, *101*(2), 341–354.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., ... others (2018). The uk biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209.

Carrier, N. (1960). *The registrar-general's statistical review of england and wales for the year.* Wiley Online Library.

Case, A., Lubotsky, D., & Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, *92*(5), 1308–1334.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2023). A practical introduction to regression discontinuity designs: Extensions. *arXiv preprint arXiv:2301.08958*.

Cattaneo, M. D., & Titiunik, R. (2022). Regression discontinuity designs. *Annual Review of Economics*, *14*(1), 821–851.

Currie, J., Decker, S., & Lin, W. (2008). Has public health insurance for older children reduced disparities in access to care and health outcomes? *Journal of health Economics*, *27*(6), 1567–1581.

Currie, J., & Gruber, J. (2001). Public health insurance and medical treatment: the equalizing impact of the medicaid expansions. *Journal of Public Economics*, *82*(1), 63–89.

Currie, J., & Rossin-Slater, M. (2015). Early-life origins of life-cycle well-being: Research and policy implications. *Journal of policy Analysis and management*, *34*(1), 208–242.

Currie, J., & Schwandt, H. (2016). The 9/11 dust cloud and pregnancy outcomes: A reconsideration. *Journal of Human Resources*, *51*(4), 805–831.

Currie, J., & Thomas, D. (1995). Does head start make a difference? *The American Economic Review*, *85*(3), 341–364.

Currie, J., Zivin, J. G., Mullins, J., & Neidell, M. (2014). What do we know about short-and long-term effects of early-life exposure to pollution? *Annu. Rev. Resour. Econ.*, *6*(1), 217–247.

Diabelková, J., Rimárová, K., Urdzík, P., Dorko, E., Houžvičková, A., Andraščíková, Š., ... others (2022). Risk factors associated with low birth weight. *Cent Eur J Public Health*, *30*(88), S43–9.

Dominguez-Bello, M. G., De Jesus-Laboy, K. M., Shen, N., Cox, L. M., Amir, A., Gonzalez, A., ... others (2016). Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nature medicine*, *22*(3), 250–253.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American statistical association*, *56*(293), 52–64.

Erickson, J. (2019). *The road to stalingrad: Stalin's war with germany.* Routledge.

Furuya, S., Zheng, F., Lu, Q., & Fletcher, J. M. (2024). Separating scarring effect and selection of early-life exposures with genetic data. *Demography*, *61*(2), 363–392.

Gardiner, J. (2010). The blitz: the british under attack. *(No Title)*.

Gluckman, P. D. (2004). *The fetal matrix: evolution, development and disease*. Cambridge University Press.

Gluckman, P. D., & Hanson, M. A. (2006). The developmental origins of health and disease. *Early life origins of health and disease*, 1–7.

Godfrey, K. M., & Barker, D. J. (2001). Fetal programming and adult health. *Public health nutrition*, *4*(2b), 611–624.

Hardy, K., & Hardy, P. J. (2015). 1st trimester miscarriage: four decades of study. *Translational pediatrics*, *4*(2), 189.

Hjort, J., Sølvsten, M., & Wüst, M. (2017). Universal investment in infants and long-run health: Evidence from denmark's 1937 home visiting program. *American Economic Journal: Applied Economics*, *9*(4), 78–104.

Hoynes, H., Schanzenbach, D. W., & Almond, D. (2016). Long-run impacts of childhood access to the safety net. *American Economic Review*, *106*(4), 903–934.

James, W. H., & Grech, V. (2017). A review of the established and suspected causes of variations in human sex ratio at birth. *Early human development*, *109*, 50–56.

James, W. H., & Grech, V. (2018). Can sex ratios at birth be used in the assessment of public health, and in the identification of causes of selected pathologies? *Early human development*, *118*, 15–21.

Jayachandran, S. (2009). Air quality and early-life mortality: Evidence from indonesia's wildfires. *Journal of Human resources*, *44*(4), 916–954.

Jeffrey, G. (2023). *Essays in public provision of healthcare* (Unpublished doctoral dissertation). London School of Economics and Political Science.

Kawwass, J. F., Kulkarni, A. D., Hipp, H. S., Crawford, S., Kissin, D. M., & Jamieson, D. J. (2016). Extremities of body mass index and their association with pregnancy outcomes in women undergoing in vitro fertilization in the united states. *Fertility and sterility*, *106*(7), 1742–1750.

Kynaston, D. (2008). *Austerity britain, 1945-1951* (Vol. 1). A&C Black.

London, G., & Region, N. (1952). The registrar general's statistical review of england and wales.[1940–1945 text, vol. i, medical. hmso 6 s. 6 d. 1940–1945 text, vol. ii, civil. hmso 5 s. 0 d. 1946—1947 text, vol. i, medical. hmso 6 s. 6 d.] the war of 1939–45 interrupted the regular publication of the registrar. *JIA*, *78*, 0151–0167.

Marshall, L., Finch, D., Cairncross, L., & Bibby, J. (2019). *Mortality and life expectancy trends in the uk: stalling progress*. Health Foundation London.

Martin-Bassols, N., de New, S. C., Shields, M. A., & Johnston, D. W. (2024). Effect of in utero exposure to air pollution on adulthood hospitalizations. *Journal of Urban Health*, *101*(1), 92–108.

Menclova, A. K., & Stillman, S. (2020). Maternal stress and birth outcomes: Evidence from an unexpected earthquake swarm. *Health Economics*, *29*(12), 1705–1720.

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., Salichos, L., ... others (2016). The real cost of sequencing: scaling computation to keep pace with data generation. *Genome biology*, *17*, 1–9.

Murphy, M. (1993). The contraceptive pill and women's employment as factors in fertility change in britain 1963–1980: A challenge to the conventional view. *Population studies*, *47*(2), 221–243.

Nobles, J., & Hamoudi, A. (2019). Detecting the effects of early-life exposures: why fecundity matters. *Population research and policy review*, *38*, 783–809.

of Food, M. (1951). *The urban working-class household diet: 1940-1949*. HM Stationery Office.

Orzack, S. H., Stubblefield, J. W., Akmaev, V. R., Colls, P., Munné, S., Scholl, T., ... Zuckerman, J. E. (2015). The human sex ratio from conception to birth. *Proceedings of the National Academy of Sciences*, *112*(16), E2102–E2111.

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., ... Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics*, *118*(2), 511–521.

Rivett, G. (1998). *From cradle to grave: fifty years of the nhs*. King's Fund London.

Sanders, N. J., & Stoecker, C. (2015). Where have all the young men gone? using sex ratios to measure fetal death rates. *Journal of health economics*, *41*, 30–45.

Say, L., Chou, D., Gemmill, A., Tunçalp, Ö., Moller, A.-B., Daniels, J., ... Alkema, L. (2014). Global causes of maternal death: a who systematic analysis. *The Lancet global health*, *2*(6), e323–e333.

Scholte, R. S., van den Berg, G. J., & Lindeboom, M. (2015, January). Long-run effects of gestation during the Dutch Hunger Winter famine on labor market and hospitalization outcomes. *Journal of health economics*, *39*, 17–30. doi: 10.1016/j.jhealeco.2014.10.002

Song, S. (2015). Privation, stress, and human sex ratio at birth. *Early human development*, *91*(12), 823–827.

Sowers, J. R. (2003). Obesity as a cardiovascular risk factor. *The American journal of medicine*, *115*(8), 37–41.

Stein, A. D., Zybert, P. A., van de Bor, M., & Lumey, L. H. (2004, August). Intrauterine famine exposure and body proportions at birth: The Dutch Hunger Winter. *International Journal of Epidemiology*, *33*(4), 831–6. doi: 10.1093/ije/dyh083

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American statistical association*, *54*(285), 30–34.

Swanson, J. M. (2012). The uk biobank and selection bias. *The Lancet*, *380*(9837), 110.

Symonds, N. E., Vidler, M., Wiens, M. O., Omar, S., English, L. L., Ukah, U. V., ... others (2023). Risk factors for postpartum maternal mortality and hospital readmission in low-and middle-income countries: a systematic review. *BMC Pregnancy and Childbirth*, *23*(1), 303.

Torche, F., & Kleinhaus, K. (2012). Prenatal stress, gestational age and secondary sex ratio: the sex-specific effects of exposure to a natural disaster in early pregnancy. *Human reproduction*, *27*(2), 558–567.

van Alten, S., Domingue, B. W., Faul, J., Galama, T., & Marees, A. T. (2024). Reweighting uk biobank corrects for pervasive selection bias due to volunteering. *International Journal of Epidemiology*, *53*(3), dyae054.

von Hinke, S., & Sørensen, E. N. (2023). The long-term effects of early-life pollution exposure: Evidence from the london smog. *Journal of Health Economics*, *92*, 102827.

Warrington, N. M., Beaumont, R. N., Horikoshi, M., Day, F. R., Helgeland, Ø., Laurin, C., ... others (2019). Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nature genetics*, *51*(5), 804–814.

Webster, C. (2002). *The national health service: A political history*. Oxford University Press, USA.

Wherry, L. R., & Meyer, B. D. (2016). Saving teens: using a policy discontinuity to estimate the effects of medicaid eligibility. *Journal of Human Resources*, *51*(3), 556–588.

Zhang, T., Sun, Y., Chen, Z., & Li, T. (2018). Traditional and molecular chromosomal abnormality analysis of products of conception in spontaneous and recurrent miscarriage. *BJOG: An International Journal of Obstetrics & Gynaecology*, *125*(4), 414–420.

Zhou, J., Indik, C. E., Kuipers, T. B., Li, C., Nivard, M. G., Ryan, C. P., ... others (2024). Genetic analysis of selection bias in a natural experiment: Investigating in-utero famine effects on elevated body mass index in the dutch hunger winter families study. *American Journal of Epidemiology*, kwae376.

Zweiniger-Bargielowska, I. (1994). Rationing, austerity and the conservative party recovery after 1945. *The Historical Journal*, *37*(1), 173–197.

Zweiniger-Bargielowska, I. (2000). *Austerity in britain: rationing, controls, and consumption, 1939-1955*. OUP Oxford.

Appendices

A Extended Results: Short-term Effects

Table A.1: Short-term Effects of the NHS on Infant Mortality Rates: Quadratic Fit

	(1)	(2)
VARIABLES	IMR	IMR
RD_Estimate	-10.53***	-10.51***
	(1.947)	(1.947)
Observations	13,233	13,233
Baseline covariates	✓	✓
Mean	34.62	34.62

Note: Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation, standard errors in parentheses, *** p<0.001; ** p<0.01; * p<0.05. Linear fit and triangulate kernels. *Source*: Registrar General's Statistical Review.

Table A.2: Short-term Effects of the NHS on Total Mortality

VARIABLES	(1) Total Deaths	(2) Total Deaths Male	(3) Total Deaths Female
Panel A: Bandwidth 2 years			
T	-155.69***	-88.37***	-67.32***
	(14.665)	(7.799)	(9.372)
Observations	7,360	7,360	7,360
Baseline covariates	✓	✓	✓
Mean	1224	629.8	610.9
Panel B: Bandwidth 4 years			
T	-76.25***	-42.12***	-34.14***
	(9.305)	(4.958)	(5.849)
Observations	13,248	13,248	13,248
Baseline covariates	✓	✓	✓
Mean	1241	629.8	610.9

Note: Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation, standard errors in parentheses, *** p<0.001; ** p<0.01; * p<0.05. Linear fit and triangular kernels. *Source*: Registrar General's Statistical Review.

B Extended Results: Survival Bias

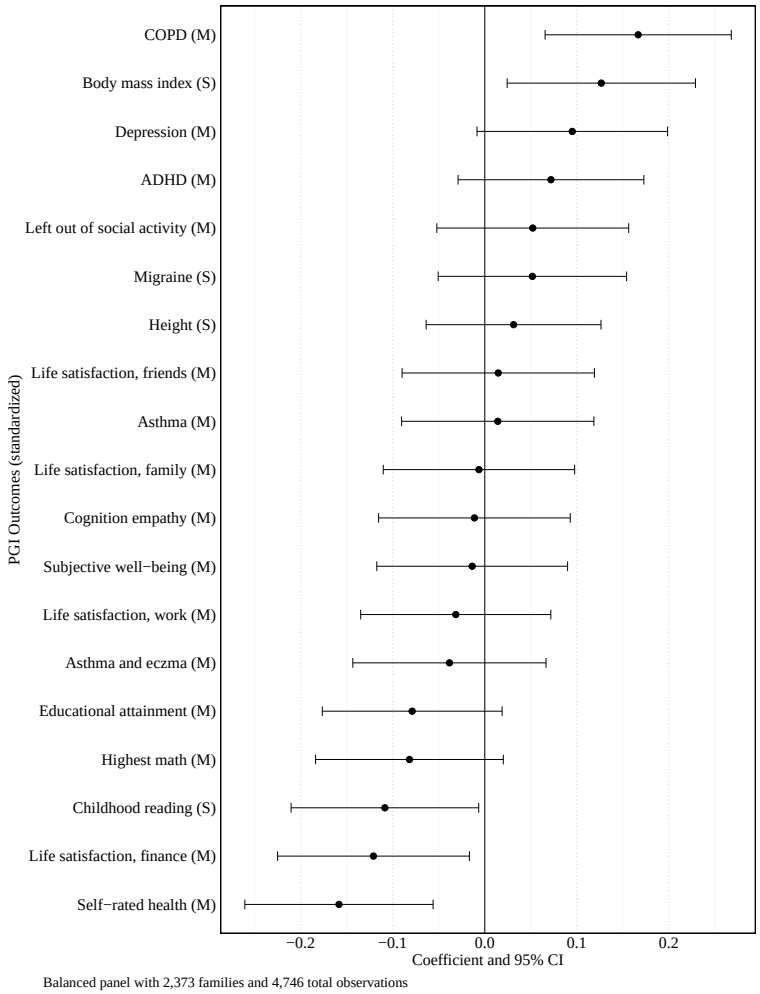


Figure B.1: Effect of the NHS on PGIs using UKB: Sample of Siblings and Family Fixed Effects Wider Bandwidth

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. The bandwidth is 36 months per- and post-NHS implementation. Controls include year-month of birth, interview date (age), gender, quarter of birth fixed effects, sibling characteristics including year-month of birth, gender, number of siblings (males, females, adopted males, adopted females), and number of older siblings, 10 ancestry-specific PCs, and family fixed effects. Source: UKB.

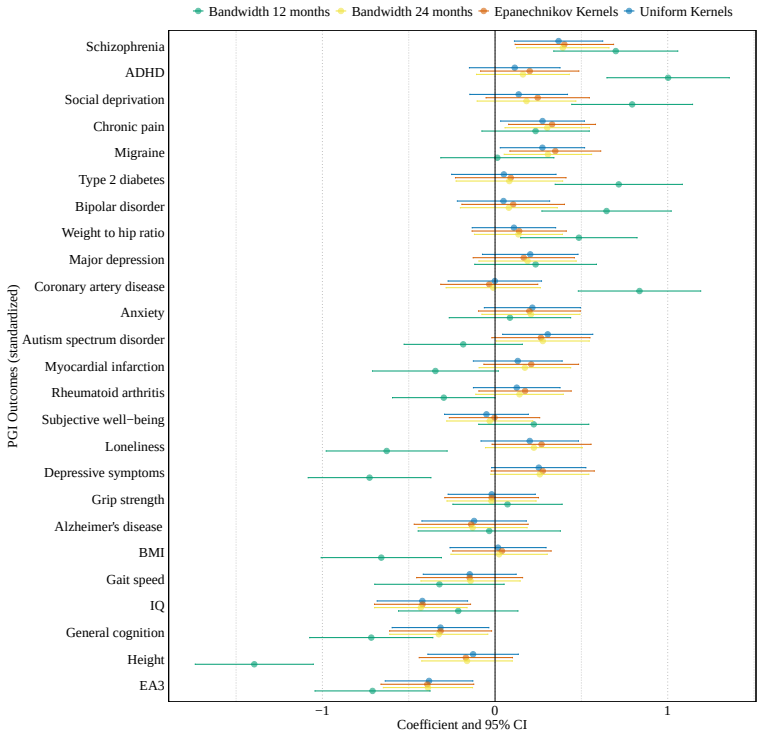


Figure B.2: Survival Effects of the NHS on PGIs using ELSA: Different Bandwidths and Kernels

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Standard errors clustered at the LAD level. Unless otherwise specified bandwidth is 18 months per- and post-NHS implementation. Linear fit. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals (Epachenikov and Uniform Kernls), x individuals (12 months bandwidth), and y individuals (24 months bandwidth). Source: ELSA.

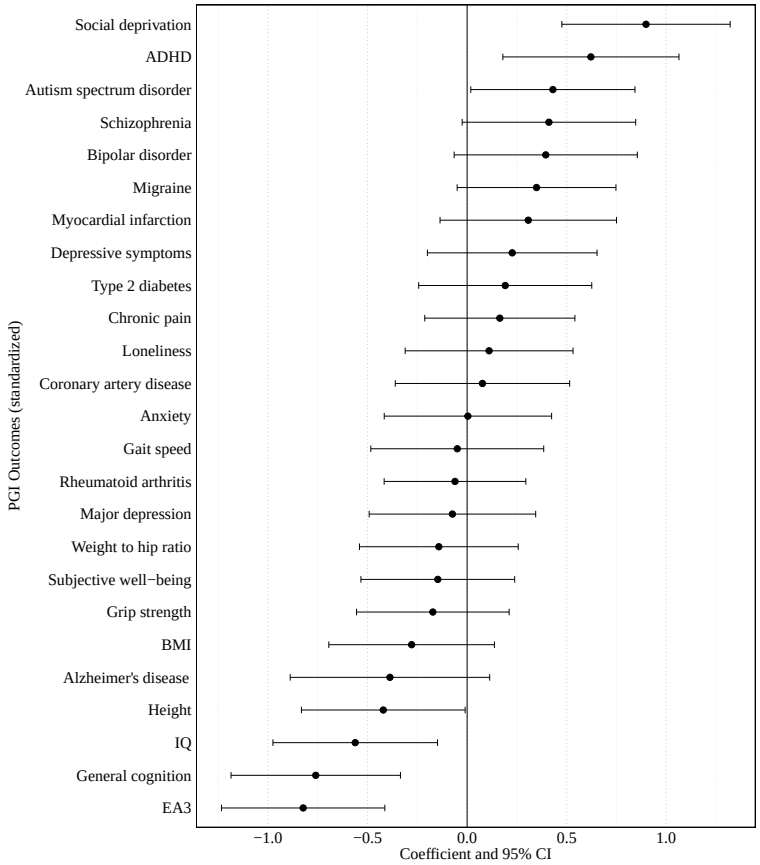


Figure B.3: Survival Effects of the NHS on PGIs using ELSA: Quadratic Fit

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Standard errors clustered at the LAD level. Bandwidth is 18 months per- and post-NHS implementation. Triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.

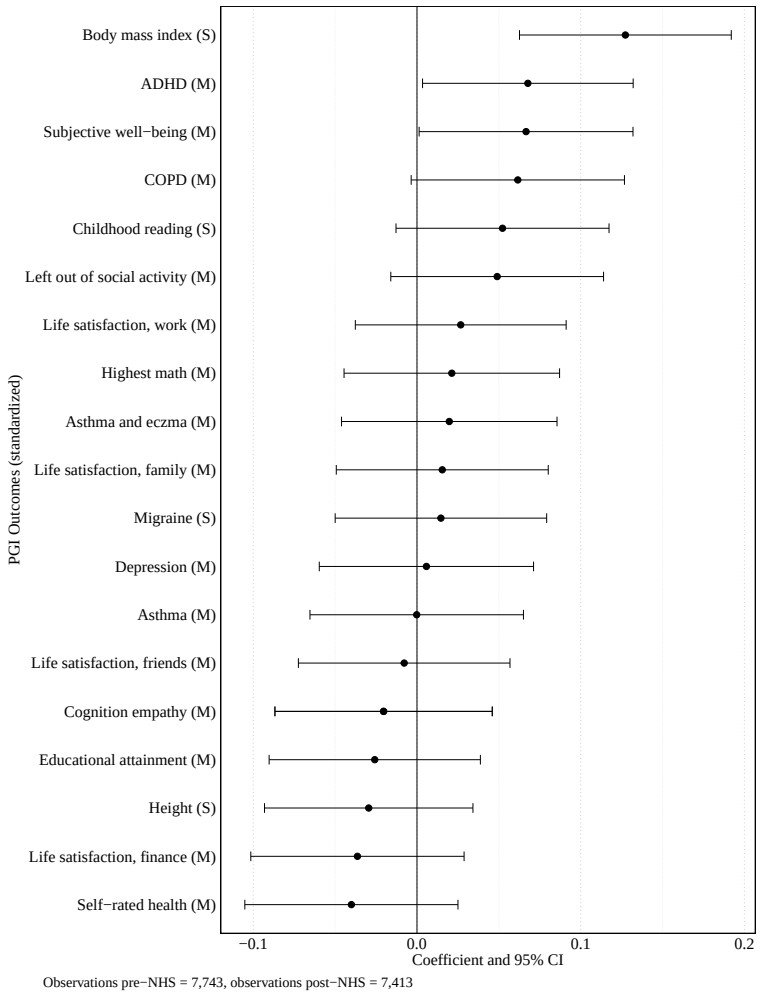


Figure B.4: Survival Effects of the NHS on PGIs using UKB: Bandwidth 12 Months

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

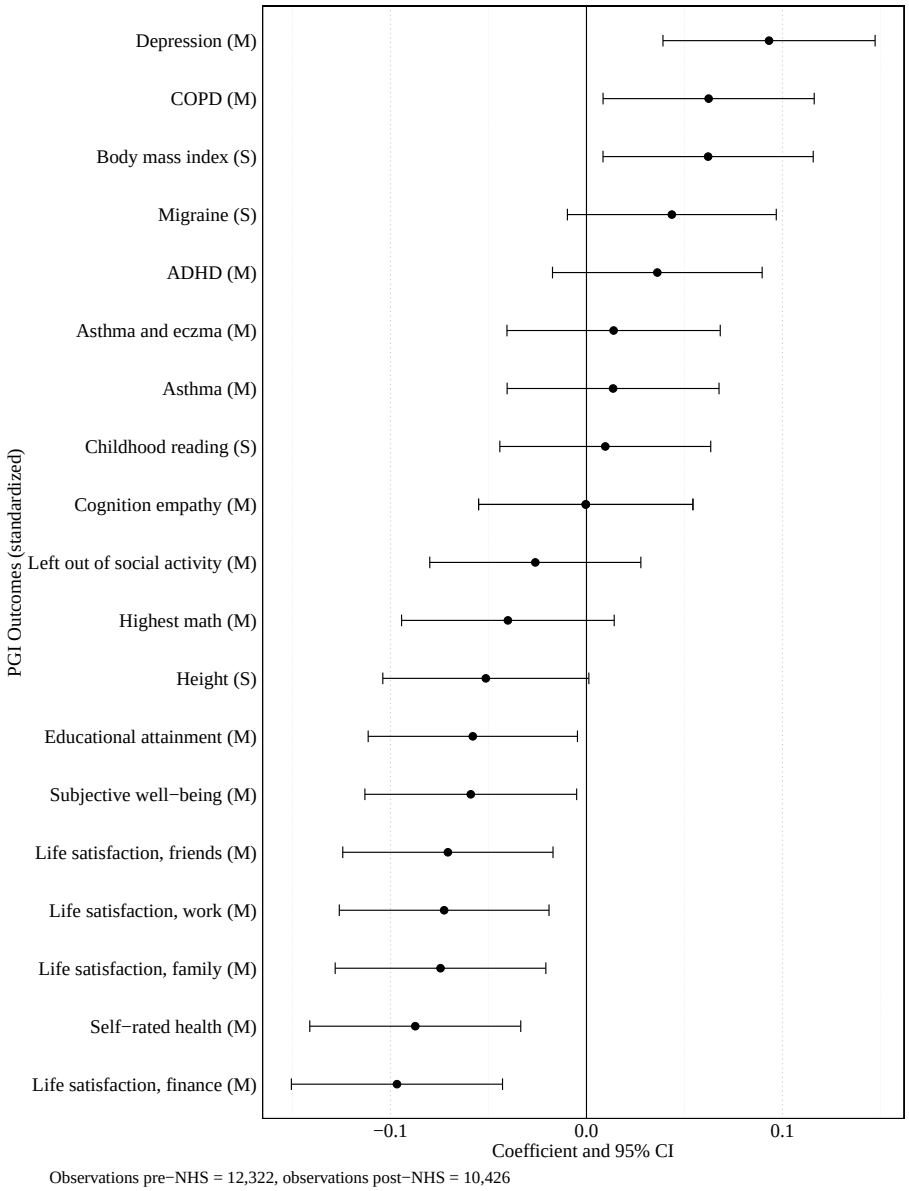


Figure B.5: Survival Effects of the NHS on PGIs using UKB: Bandwidth 24 Months

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

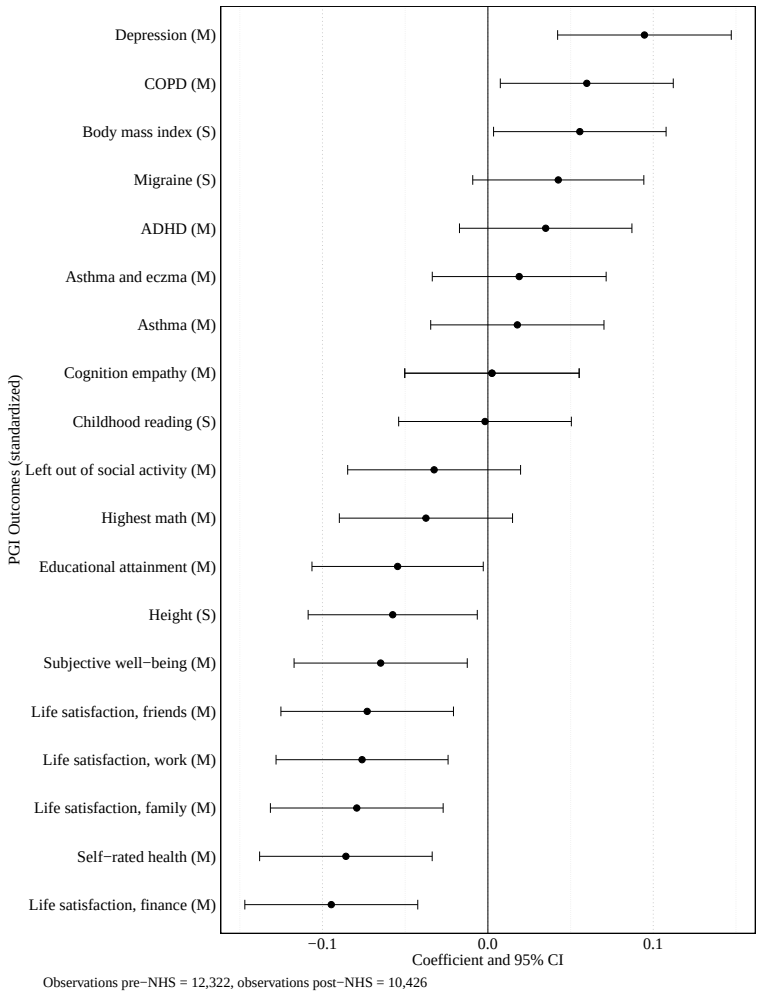


Figure B.6: Survival Effects of the NHS on PGIs using UKB: Uniform Kernels

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. Bandwidth is 18 months per- and post-NHS implementation. Linear fit. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

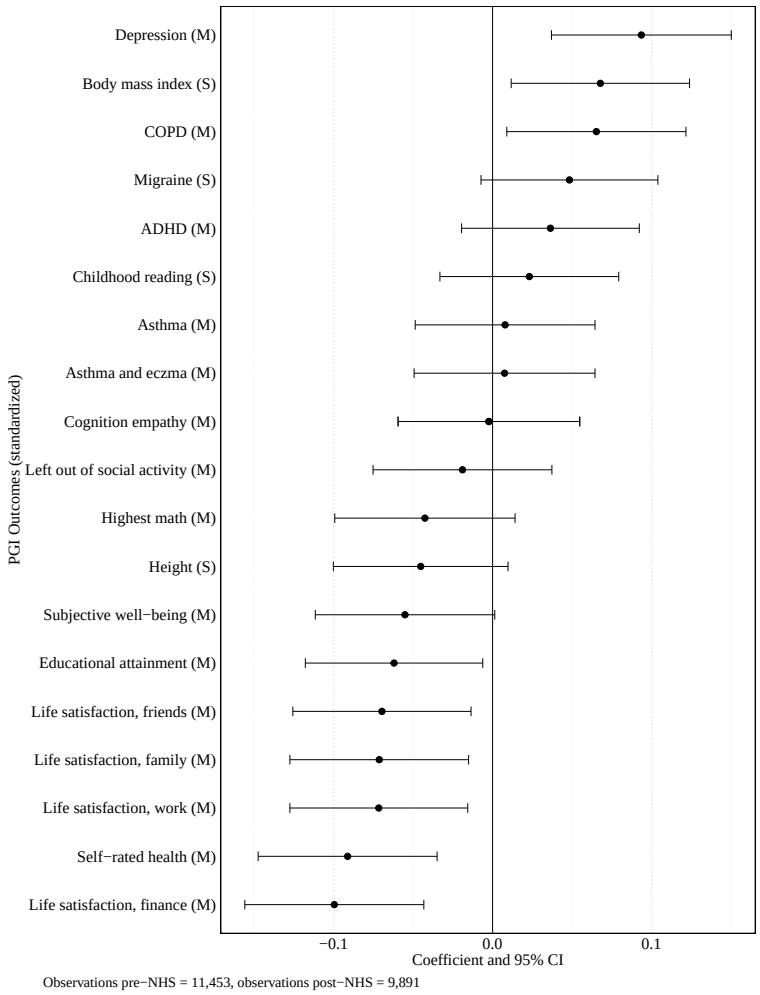


Figure B.7: Survival Effects of the NHS on PGIs using UKB: Epanechnikov Kernels

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. Bandwidth is 18 months per- and post-NHS implementation. Linear fit. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

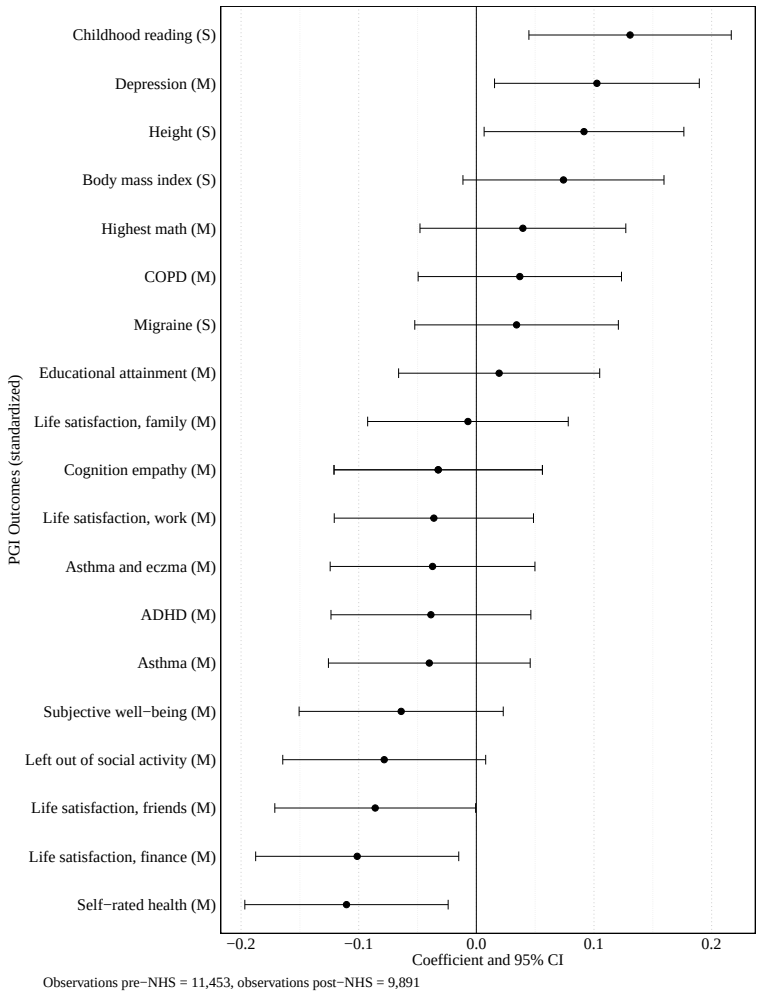


Figure B.8: Survival Effects of the NHS on PGIs using UKB: Quadratic Fit

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. Bandwidth is 18 months per- and post-NHS implementation. Triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

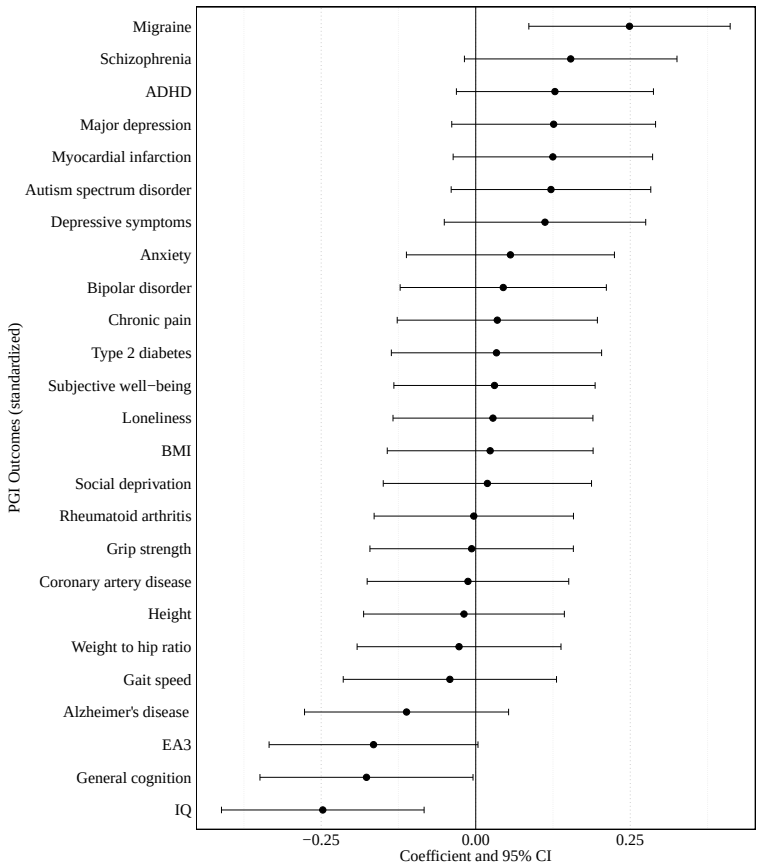


Figure B.9: Survival Effects of the NHS on PGIs using ELSA: RDD under Local Randomization

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Standard errors clustered at the LAD level. Bandwidth is 18 months per- and post-NHS implementation. Linear fit. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.

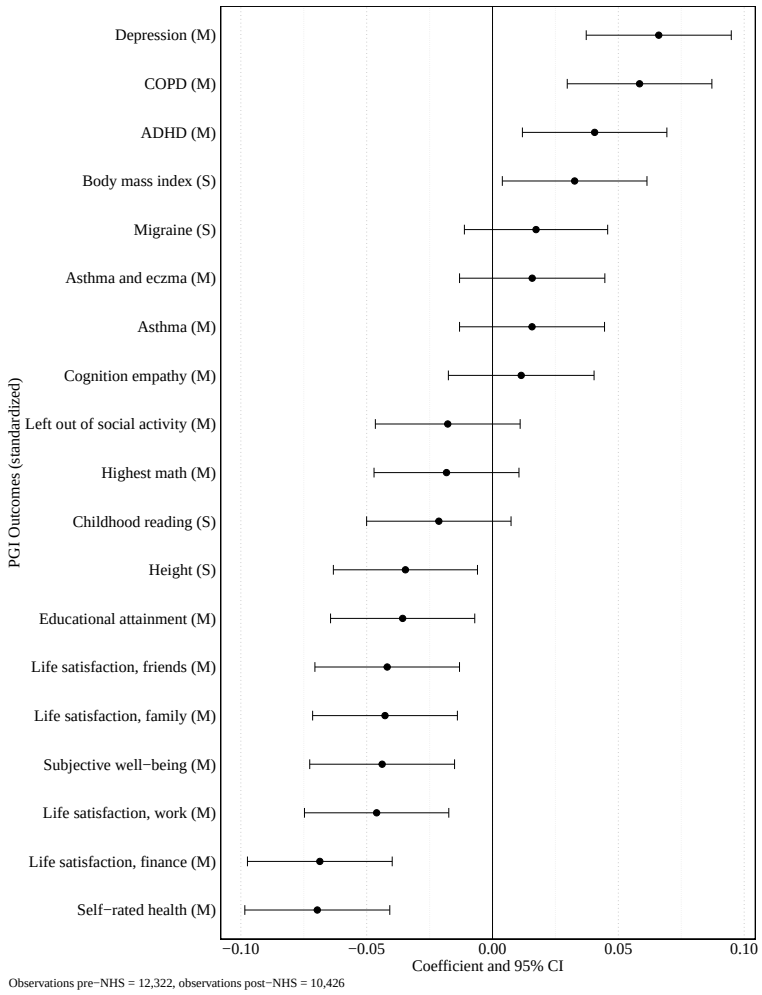


Figure B.10: Survival Effects of the NHS on PGIs using UKB: RDD under Local Randomization

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. Bandwidth is 18 months per- and post-NHS implementation. Triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

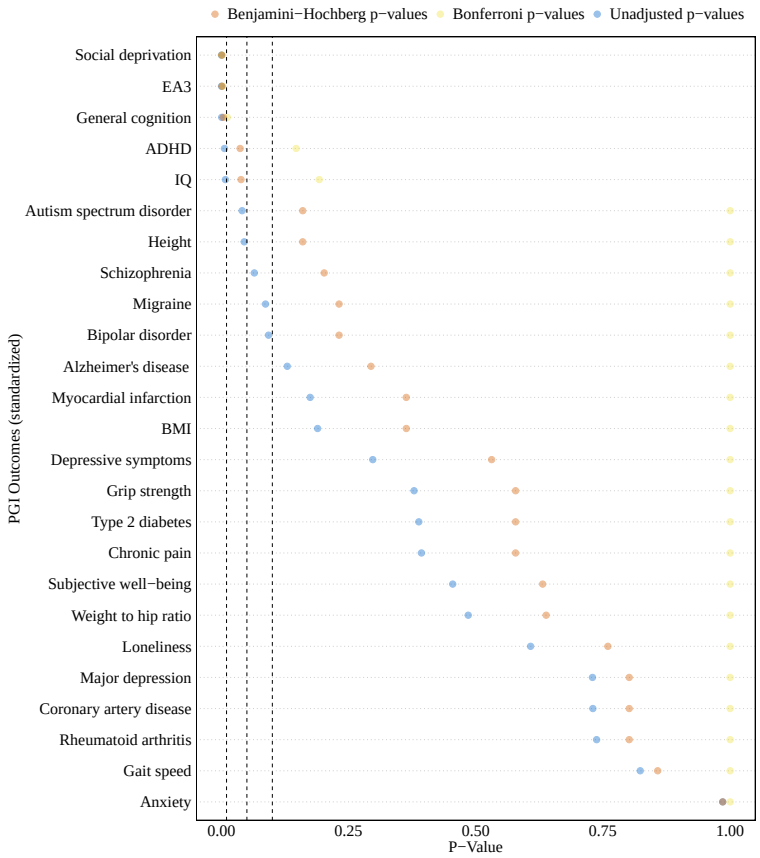


Figure B.11: Multiple Hypothesis Testing ELSA

Note: Standard errors clustered at the LAD level. Bandwidth is 18 months per- and post-NHS implementation. Linear fit. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.

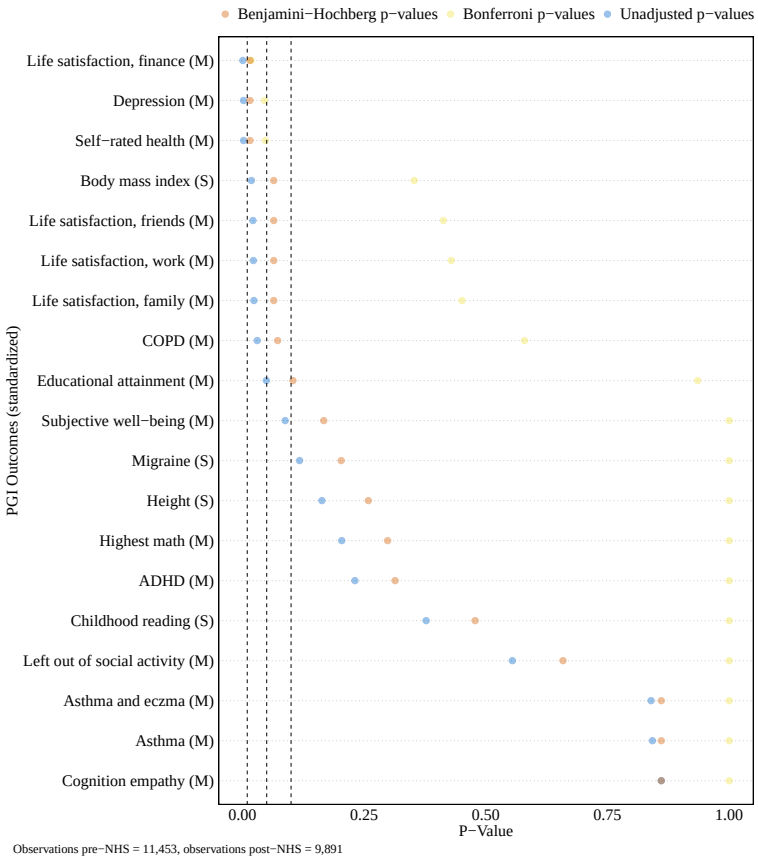


Figure B.12: Multiple Hypothesis Testing UKB

Note: Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. Unless otherwise specified bandwidth is 18 months per- and post-NHS implementation. Triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

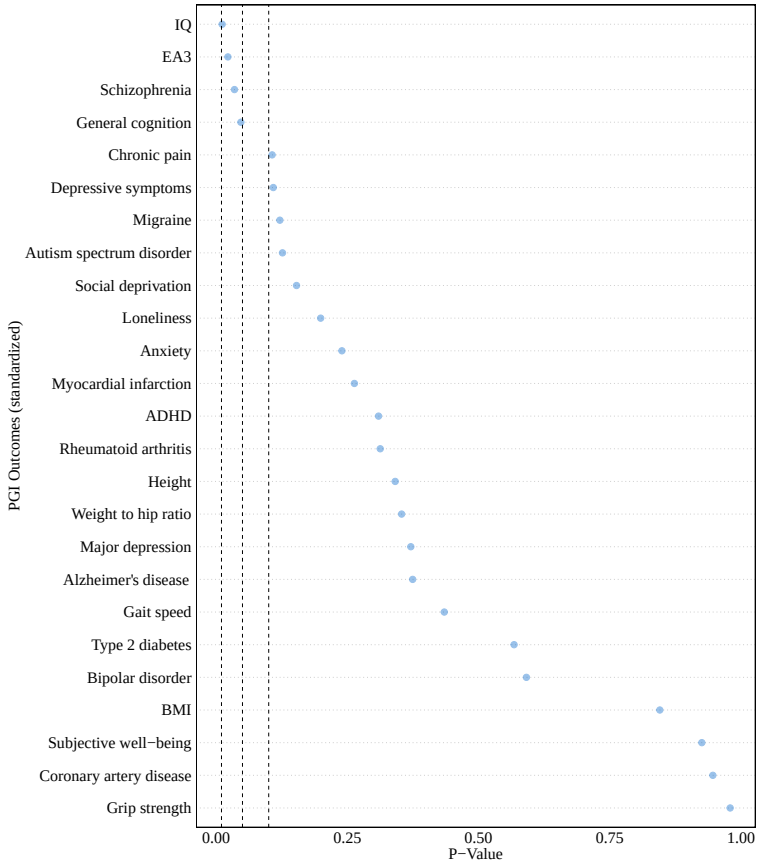


Figure B.13: Permutation Test ELSA

Note: Standard errors clustered at the LAD level. Bandwidth is 18 months per- and post-NHS implementation. Linear fit. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.

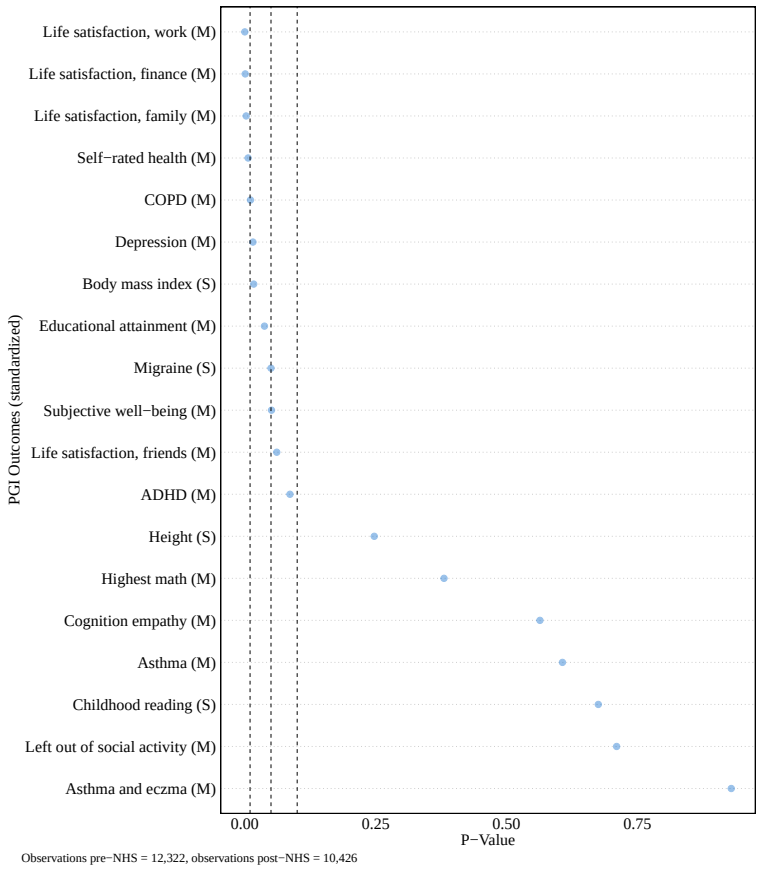


Figure B.14: Permutation Test UKB

Note: Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. Unless otherwise specified bandwidth is 18 months per- and post-NHS implementation. Triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

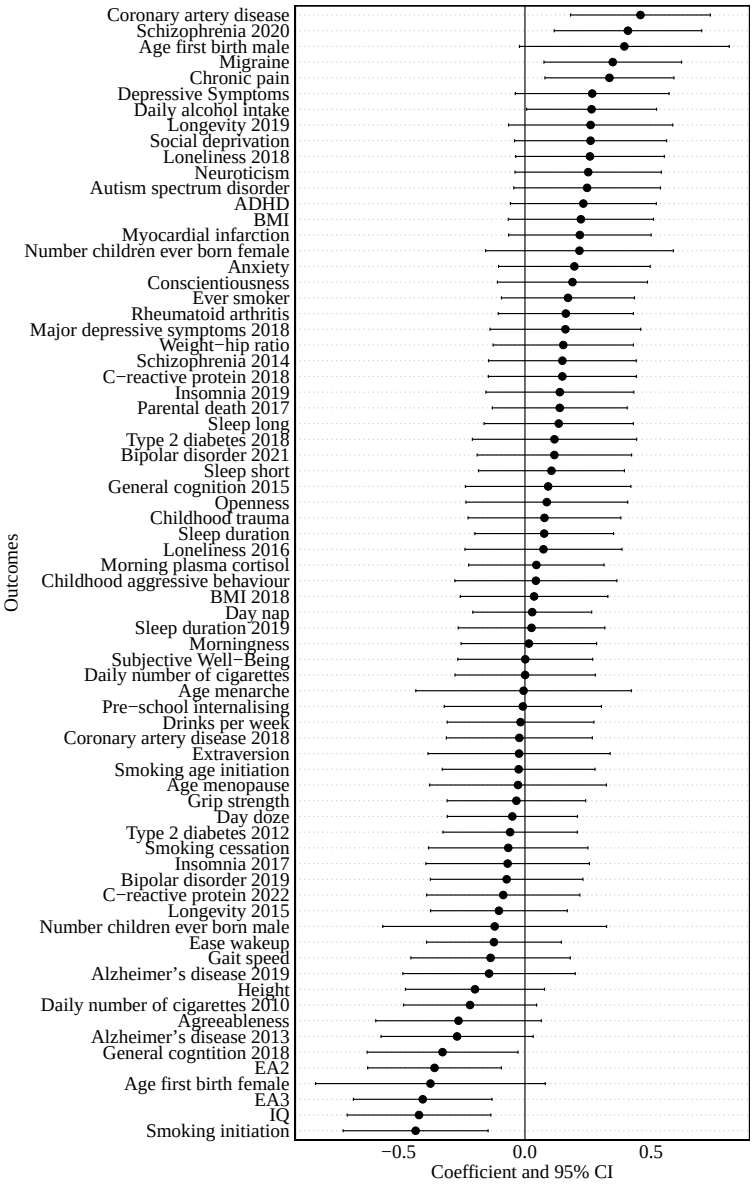


Figure B.15: Survival Effects of the NHS on PGIs using ELSA: All PGIs

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Standard errors clustered at the LAD level. The bandwidth is 18 months per- and post-NHS implementation. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.

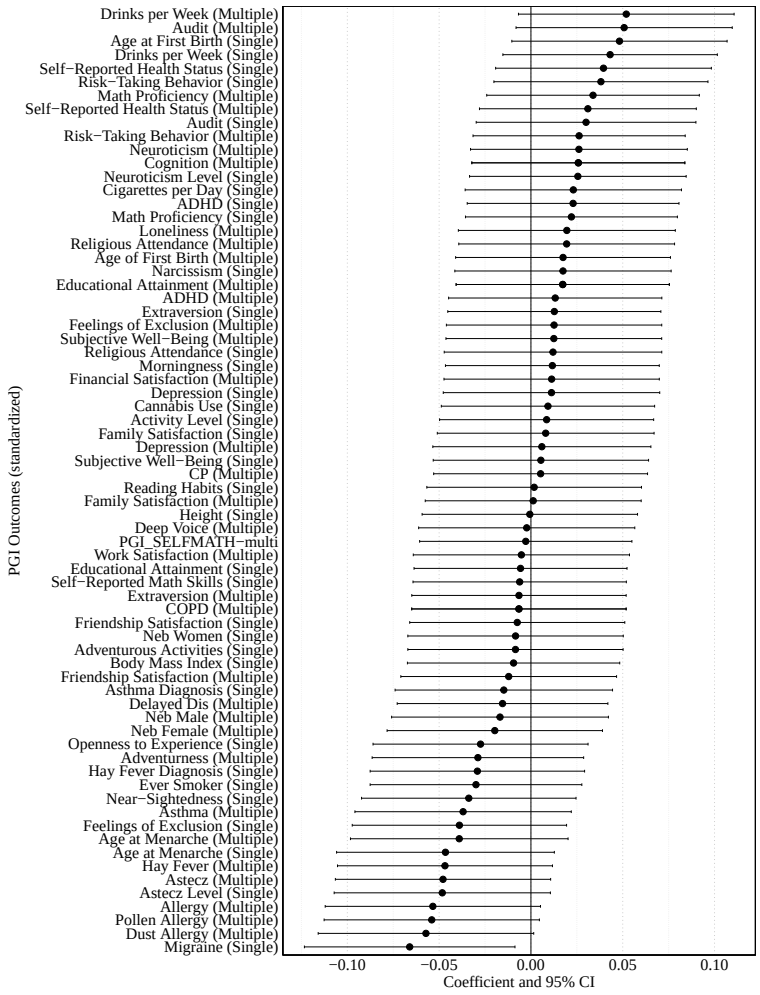


Figure B.16: Survival Effects of the NHS on PGIs using UKB: All PGIs

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. The bandwidth is 18 months per- and post-NHS implementation. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

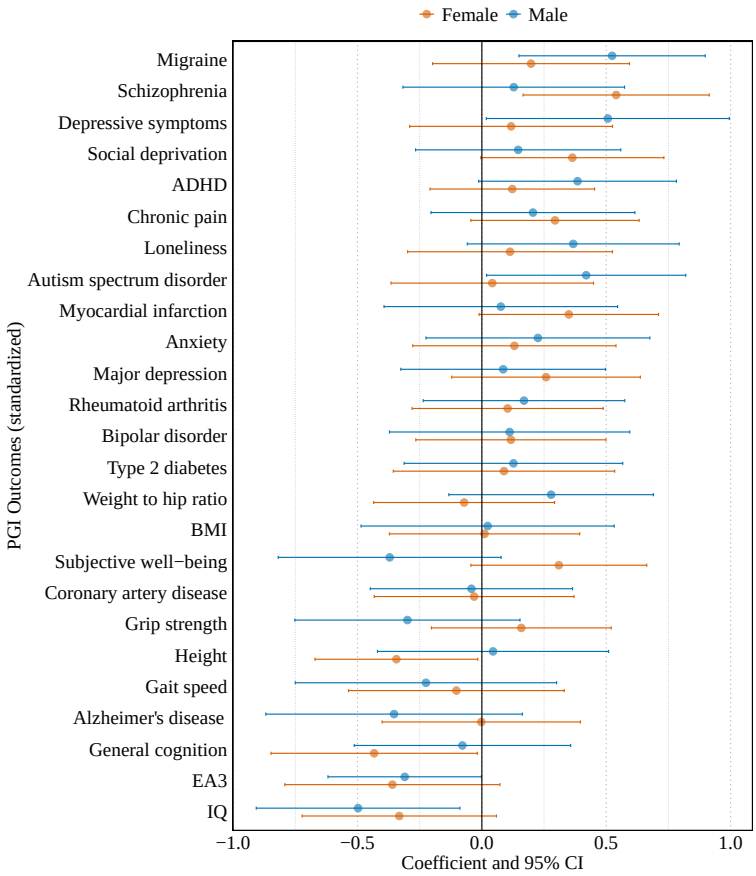


Figure B.17: Survival Effects of the NHS on PGIs using ELSA: Heterogeneity by Gender

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Standard errors clustered at the LAD level. The bandwidth is 18 months per- and post-NHS implementation. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.

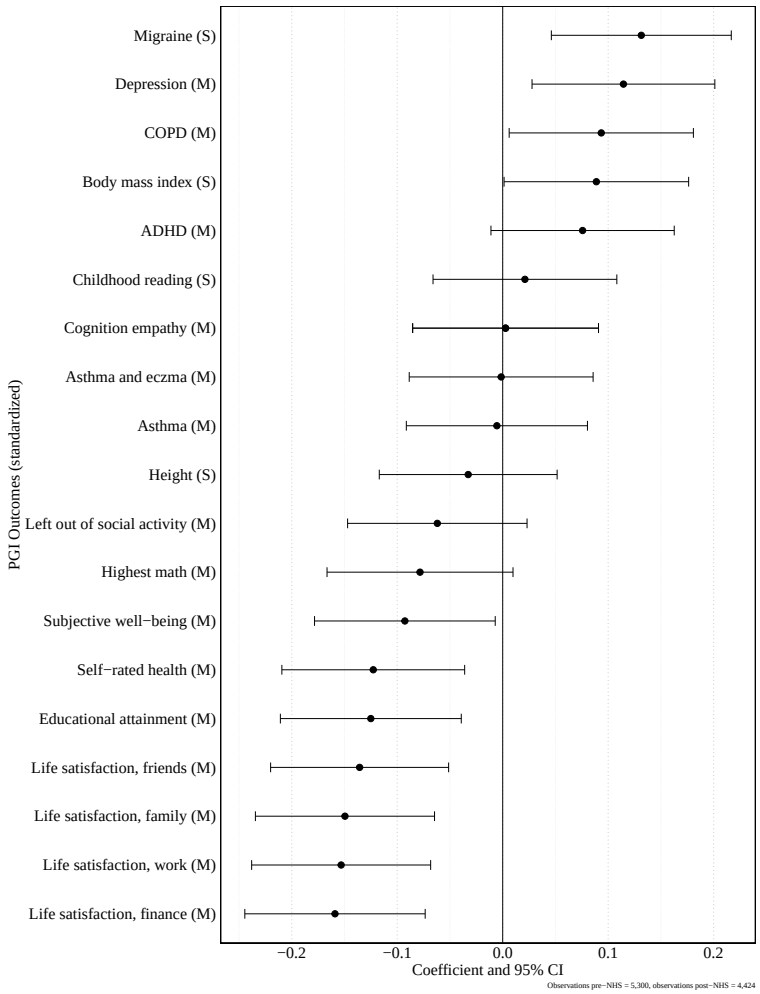


Figure B.18: Survival Effects of the NHS on PGIs using UKB: Heterogeneity by Males

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. The bandwidth is 18 months per- and post-NHS implementation. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

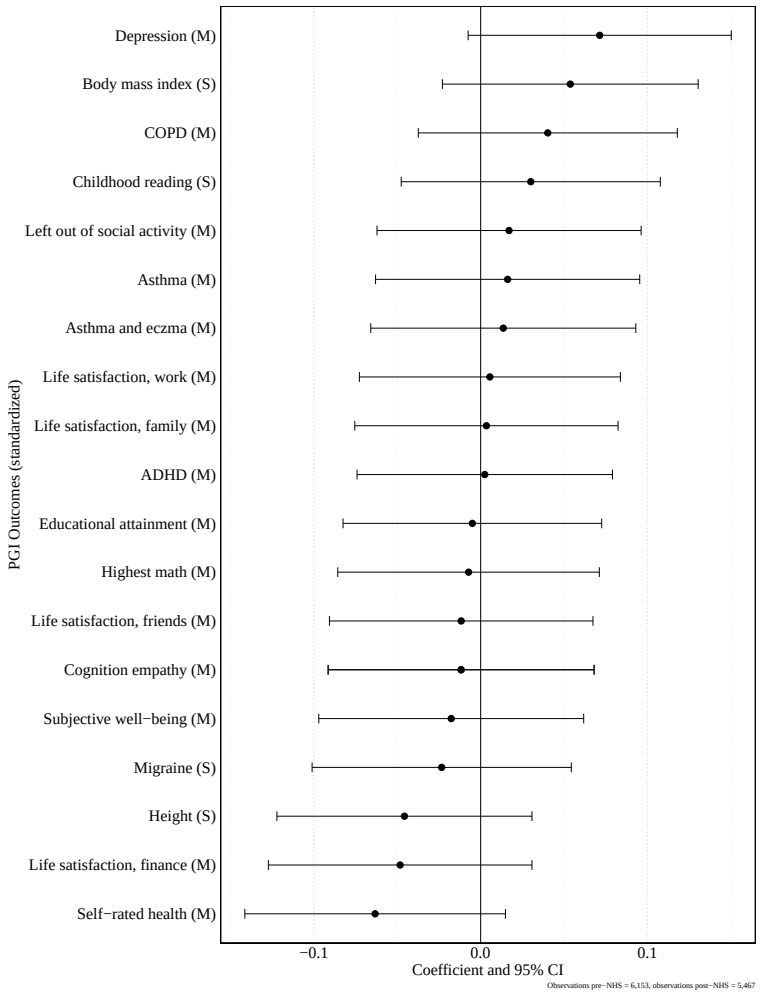


Figure B.19: Survival Effects of the NHS on PGIs using UKB: Heterogeneity by Females

Note: Each dot represents the causal effect of a separate regression on a different outcome. 95% confidence intervals. Cluster-robust variance estimation with degrees-of-freedom weights, combined with nearest neighbor variance estimation standard errors. The bandwidth is 18 months per- and post-NHS implementation. Linear fit with triangular kernels. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Source: UKB.

C Extended Results: Long-term Effects

Table C.1: Long-term Effects of the NHS on Self-reported Health, BMI, Obesity, Heart Problems using ELSA: Different Bandwidths

VARIABLES	(1) Bad Health Score (1-5)	(2) BMI (kg/m2)	(3) Obesity (0, 1)	(4) Ever heart (0, 1)
Panel A: Bandwith 12 months				
RDD Estimate	-0.85***	-1.58*	-0.04	0.00
	(0.178)	(0.929)	(0.073)	(0.051)
Observations	465	461	461	465
Panel B: Bandwith 24 months				
RDD Estimate	-0.29**	-1.40**	-0.14**	-0.07*
	(0.124)	(0.664)	(0.059)	(0.042)
Observations	948	942	942	948
Baseline covariates	✓	✓	✓	✓

Note: Standard errors clustered at the LAD level in parentheses, *** p<0.001; ** p<0.01; * p<0.05. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Linear fit and triangular kernels. Source: ELSA.

Table C.2: Long-term Effects of the NHS on Self-reported Health, BMI, Obesity, Heart Problems using ELSA: Different Kernels

VARIABLES	(1) Bad Health Score (1-5)	(2) BMI (kg/m2)	(3) Obesity (0, 1)	(4) Ever Heart (0, 1)
Panel A: Epanechnikov Kernels				
RDD Estimate	-0.32**	-1.55**	-0.14**	-0.07
	(0.140)	(0.734)	(0.066)	(0.046)
Panel B: Uniform Kernels				
RDD Estimate	-0.30**	-1.34*	-0.13**	-0.11**
	(0.130)	(0.697)	(0.062)	(0.046)
Observations	705	700	700	705
Baseline covariates	✓	✓	✓	✓
Mean	3.420	29.60	0.404	0.253
Observations left	379	377	377	379
Observations right	326	323	323	326
Bandwith right	18	18	18	18

Note: Standard errors clustered at the LAD level in parentheses, *** p<0.001; ** p<0.01; * p<0.05. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. The bandwidth is 18 months per- and post-NHS implementation. Linear fit. Source: ELSA.

Table C.3: Long-term Effects of the NHS on Self-reported Health, BMI, Obesity, Heart Problems using ELSA: Quadratic Fit

VARIABLES	(1) Bad Health Score (1-5)	(2) BMI (kg/m2)	(3) Obesity (0, 1)	(4) Ever heart (0, 1)
RD_Estimate	-0.31	-2.42**	-0.29***	0.06
	(0.221)	(1.147)	(0.088)	(0.063)
Observations	705	700	700	705
Baseline covariates	✓	✓	✓	✓
Mean	3.420	29.60	0.404	0.813
Observations left	359	357	357	359
Observations right	302	299	299	302
Bandwith right	18	18	18	18

Note: Standard errors clustered at the LAD level in parentheses, *** p<0.001; ** p<0.01; * p<0.05. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. The bandwidth is 18 months per- and post-NHS implementation. Triangular kernels. Source: ELSA.

Table C.4: Long-term Effects of the NHS on Self-reported Health, BMI, Obesity, Heart Problems using ELSA: Different Measurements ELSA

VARIABLES	(1) Bad Health Score (1-4)	(2) BMI (kg/m2)	(3) Obesity (0, 1)	(4) Ever heart (0, 1)
Panel A: First observed				
RDD Estimate	-0.20	-1.48**	-0.13**	0.01
	(0.149)	(0.705)	(0.062)	(0.068)
Mean	2.548	28.34	0.312	0.286
Panel B: Minimal value (better health)				
RDD Estimate	-0.31**	-1.47**	-0.07	0.10
	(0.137)	(0.693)	(0.058)	(0.072)
Observations	705	700	700	705
Baseline covariates	✓	✓	✓	✓
Observations left	359	357	357	359
Observations right	302	299	299	302
Bandwith right	18	18	18	18

Note: Standard errors clustered at the LAD level in parentheses, *** p<0.001; ** p<0.01; * p<0.05. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. The bandwidth is 18 months per- and post-NHS implementation. Triangular kernels. Source: ELSA.

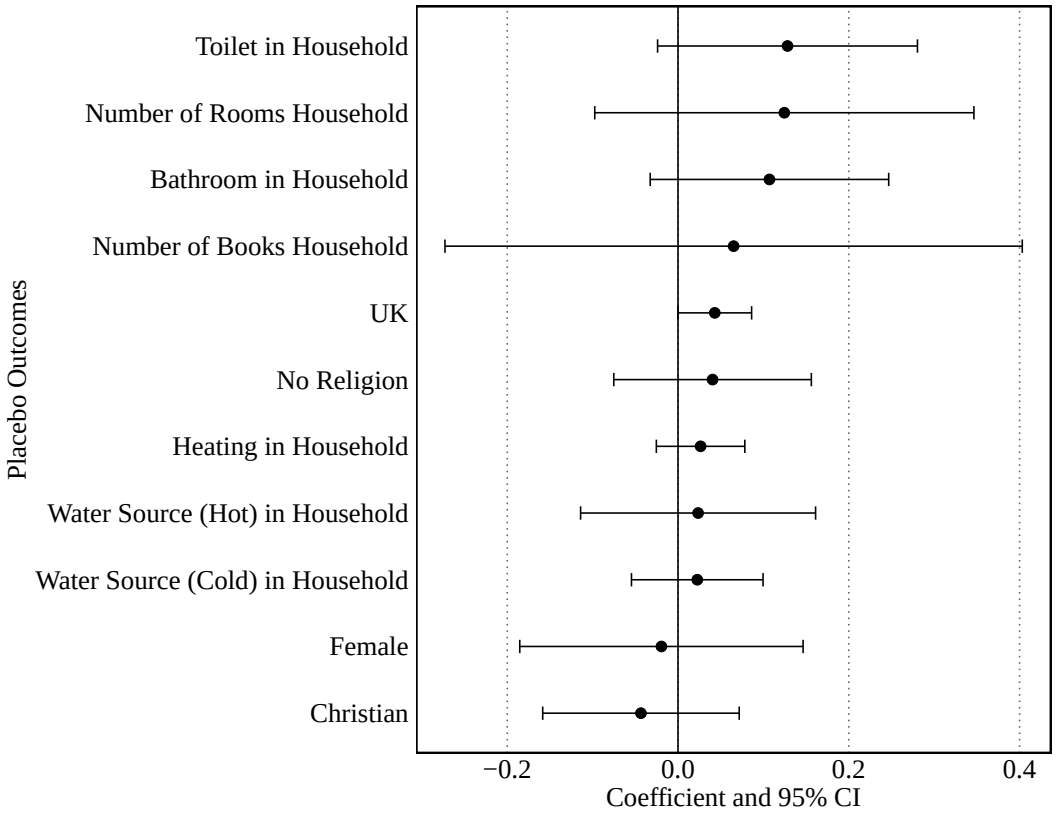


Figure C.1: Sanity Checks ELSA

Note: Standard errors clustered at the LAD level. Bandwidth is 18 months per- and post-NHS implementation. Linear fit. Controls include age, gender, quarter of birth fixed effects, and 10 ancestry specific PCs. Sample composed of 705 individuals. Source: ELSA.