The Causal Impact of Health on Employment and Earnings: A Partial Identification Approach with Imperfect Instruments

Akbar Ullah*[1,2] Sara Jabeen[3] Doriane Mignon[1] and Luke Munford[1,2]

*Corresponding author address: Williamson Building, 176 Oxford Rd, Manchester M13 9QQ, email: akbar.ullah@manchester.ac.uk.
[1]Health Organisation, Policy and Economics (HOPE), The University of Manchester, Williamson Building, 176 Oxford Rd, Manchester M13 9QQ, United Kingdom. [2]National Institute for Health and Care Research Applied Research Collaboration for Greater Manchester (NIHR ARC-GM), Manchester, United Kingdom. [3]Department of Economics, The University of Manchester, Arthur Lewis Building, Oxford Street, Manchester M13 9PL, United Kingdom.

In observational studies, estimates of the impact of health on labour market outcomes often risk bias due to omitted variables, measurement errors, and reverse causality. Using data from the UK Longitudinal Household Survey, we apply a nonparametric partial identification strategy to bound the true causal effects of health on employment probability and labour earnings. We find that individuals with one health condition have a 0%–7% lower probability of employment, while two or three conditions reduce this by 2%–8% compared to individuals without health conditions. In terms of earnings, individuals with one condition earn £0–£1,490 less per month, two conditions result in £200–£1,520 less, and three conditions reduce earnings by £240–£1,520 compared to those without health conditions. Using biomarkers as an alternative measure of health, which is less prone to measurement error, reveals that an abnormal blood test is associated with a 0%–4% lower probability of employment and £0–£1,480 lower earnings, while two abnormal tests correspond to a 2%–5% reduction in employment probability and £1,500–£2,320 lower earnings. We also provide disease-specific estimates for several non-communicable diseases.

Keywords: Health Conditions, Non-communicable Diseases, Employment, Earnings, Causal Inference, Non-parametric Methods

JEL-Classification: C14, C21, C26, I10, I15, J22, J31

1 Introduction

Health is a fundamental component of human capital and plays a crucial role in shaping an individual's probability of being employed, their productivity, and their earning potential. Poor health can affect labour market participation and earnings through physical limitations, cognitive functioning, and personal beliefs and preferences (Pintor et al., 2024; Chirikos, 1993; Currie and Madrian, 1999; Germinario et al., 2022). A substantial body of empirical literature demonstrates a clear association between poor health and labour market outcomes, such as early retirement, lower employment rates, lower wages, and fewer working hours (Cai et al., 2014; García-Gómez et al., 2013; Gaulke, 2021; Blundell et al., 2023; Jones et al., 2020; Lenhart, 2019; García-Gómez et al., 2010; Cai et al., 2014; García-Gómez, 2011; Swaminathan and Lillard, 2000).

However, the causal interpretation of this association is often unclear, *i.e.*, whether poor health does in fact *cause* worse labour market outcomes. This is due to the potential endogeneity that arises from the omitted variable bias and reverse causality. Omitted variable bias occurs when unobservable time-variant and time-invariant factors (e.g., genetic traits, cognitive ability, childhood circumstances, personality, communication skills, financial or family issues) are correlated with both health and labour market outcomes. Reverse causality occurs when unemployment or reduced earnings lead to poorer health. The health and wellbeing impacts of lower personal or family income, as well as unemployment, are well documented (Reinhold and Jürges, 2012; Khanam et al., 2009; Propper et al., 2007; Currie et al., 2007; Currie and Stabile, 2003; Case et al., 2002; Adda et al., 2009; Eliason and Storrie, 2009; Sullivan and Von Wachter, 2009; Schmitz, 2011; Marcus, 2013; Mossakowski, 2009; Powdthavee, 2012; Kassenboehmer and Haisken-DeNew, 2009; Norström et al., 2017; Gedikli et al., 2023; Marmot, 2005; García-Gómez and López-Nicolás, 2006). This phenomenon, known as the "income gradient in health", underscores the importance of accounting for reverse causality when assessing the impact of health on labour market outcomes. A third issue in the survey data is measurement error, which can arise from misjudgement or justification bias, where respondents report poor health to rationalise their labour market performance (Blundell et al., 2023).

In this study, we estimate the impact of poor health on employment and earnings using data from the United Kingdom Household Longitudinal Survey (UKHLS). We contribute to the literature in four key ways. First, we employ a nonparametric partial identification (NPI) method to bound the true causal effect of health on employment and earnings (Nevo and Rosen, 2012; Manski and Pepper, 2000; Ban and Kédagni, 2022). Proposed by Ban and Kédagni (2022), this approach, based on imperfect instruments, addresses the endogeneity of health conditions arising from omitted variables or reverse causality, relying on more plausible assumptions that our data do not reject. This imperfect instrumental variable framework accommodates correlations between health and unobserved factors that may directly influence employment and earnings (Ban and Kédagni, 2022; Germinario et al., 2022).

Using the NPI approach, our second contribution is to provide bounds on the population average treatment effects (ATE), a parameter that is more policy relevant than the treatment effects for subpopulations (Germinario et al., 2022). In the presence of heterogeneous treatment effects, panel or IV regression analyses produce average treatment effects on the treated or local average treatment effects (Germinario et al., 2022; Ban and Kédagni, 2022). As a nonparametric method, NPI requires only the calculation of sample averages, making it robust to all forms of heterogeneity of the treatment effect (Manski and Pepper, 2000). Unlike traditional regression methods, it also eliminates the need to assume an error structure or specify a functional form (Manski and Pepper, 2000; Ban and Kédagni, 2022).

As a third contribution, we complement our estimates based on self-reported health data with a biomarker-based measure, which is less prone to measurement errors and captures the main illnesses in the UK and key physiological systems (Benzeval et al., 2014). Finally, our health measures provide a broader view, encompassing various disabilities and discomforts rather than focusing on a single condition, which may overlook comorbidities or lead to comparisons between individuals with different health issues (e.g., a study focusing solely on mental health could compare those with depression to individuals with physical health problems) (Blundell et al., 2023). We use data on 17 noncommunicable diseases (NCDs) and self-reported disabilities to compare outcomes between individuals with no reported health conditions (or abnormal blood tests) and those with one, two, or three conditions (or abnormal tests).

Our partial identification approach is based on three relatively weak and often testable assumptions (Christelis and Dobrescu, 2020; Germinario et al., 2022; Manski and Pepper, 2000; Ban and Kédagni, 2022). However, these advantages come at the cost of obtaining a range of possible values for the causal population ATE rather than a point estimate of it. We assume: 1) our imperfect instrumental variable (IIV) (parents' academic qualifications) and endogenous variable (individuals' health scores) have the same sign of correlation with unobserved latent variables (potential employment probability and earnings) (same direction of correlation assumption); 2) the IIV is less correlated with the latent variables than the endogenous variable (i.e., the instrument is less endogenous than the variable itself); 3) monotone treatment response (MTR), meaning that better health does not worsen labour market outcomes.

We estimate the impact of self-reported health measures from Wave 3 of UKHLS on employment and earnings in Wave 4. For the biomarker-test-based health measure, blood samples were collected in either Wave 2 or Wave 3 of UKHLS. Our estimated bounds suggest that individuals with one condition have a 0%–7% lower probability of employment (earning £0 to £1,490 less per month), two conditions result in a 2%–8% reduction in employment probability (£200 to £1,520 less in earnings), and three conditions lead to a 2%–8% decrease (£240 to £1,520 less in earnings), all relative to those without health conditions. The corresponding OLS estimates for employment and earnings losses also increase with the number of health conditions, indirectly supporting our MTR and same-direction-of-correlation assumptions. However, OLS estimates sometime lie outside the NPI bounds. For example, the OLS estimates for employment loss are 15% for one condition, 28% for two conditions, and 39% for three conditions— each falling below the corresponding lower NPI bounds.

Estimates from our biomarker-based health measure indicate that individuals with one abnormal blood test have a 0%–4% lower employment probability and earn £0 to £1,480 less than to those with no abnormal blood test. Similarly, individuals with two abnormal tests experience a 2%–5% lower employment probability and earn £1,500 to £2,320 less. Comparisons with returns to education show that the impact of health on labour outcomes is economically significant, comparable to the earnings and employment differences between individuals with A-levels and degree qualifications or A-levels and no qualifications.

The estimated bounds for specific conditions like asthma, arthritis, cancer, diabetes, and heart and liver conditions are generally wider, and thus cannot rule out zero-earning effects, as we lack information on the severity of these conditions and rely solely on binary disease indicators. Nevertheless, our NPI bounds suggest that the true causal employment effects of each NCD are significantly different from zero. For mental health, we can measure distress severity using the General Health Questionnaire (GHQ-12) caseness scores, ranging from 0 (least distressed) to 12 (most distressed). Our results show that mild depression (GHQ-12 score 3-6) reduces monthly earnings by £0 to £660, and moderate depression (GHQ-12 score 7-9) by £190 to £910.

The rest of the paper is organised as follows: Section 2 reviews the literature, Section 3 describes formally the empirical approach, Section 4 presents the data, Section 5 test the validity of the assumptions within our context, Section 6 contains the results, and Section 7 presents the conclusion of our study.

2 Related Literature

While Pintor et al. (2024) comprehensively review the effects of various health conditions and interventions on employment and earnings, and Shawa et al. (2024) focus on working hours, we provide an overview of the identification challenges these studies face in estimating the causal impact of health. We then briefly summarise the labour market impacts of different conditions. Our review, essentially partial, primarily examines studies on mental health (due to our ability to measure depression severity in our data) and those employing instrumental variables to estimate causal effects.

Most existing literature employs matching methods or panel fixed-effects regressions to identify the impact of health on labour market outcomes (see Pintor et al. (2024), Figure 3, for a summary of health measures, labour outcomes, and estimation methods). Studies combining matching methods with OLS regressions preprocess data by matching treatment and control groups based on observed characteristics such as gender, age, ethnicity, and pre-disease labour outcomes (Rumball-Smith et al., 2014; Heinesen and Kolodziejczyk, 2013; Lechner and Vazquez-Alvarez, 2011; García-Gómez et al., 2010; Aleksandrova et al., 2021; García-Gómez and López-Nicolás, 2006; Lenhart, 2019; Jones et al., 2020). These studies assume that matched groups share identical time-variant and time-invariant unobservable characteristics, and they cannot account for feedback effects from labour market status to health.

Studies using longitudinal data often employ panel fixed effects regressions or difference-in-differences methods, sometimes combined with matching, to address omitted variable bias (Vaalavuo, 2021; Tanaka, 2021; García-Gómez et al., 2013; Goryakin and Suhrcke, 2017; Mitra and Jones, 2017; Böckerman et al., 2017; Aleksandrova et al., 2021; García-Gómez and López-Nicolás, 2006; Lenhart, 2019). This approach controls only for time-invariant unobservable variables and is most effective when such bias is the primary concern. These studies typically assume that health conditions (treatment) are exogenous after accounting for observed variables and unit fixed effects. Furthermore, recent literature shows that when treatment timing varies (e.g., individuals experience health conditions at different times) and treatment effects differ between individuals or time, traditional regression-based difference-in-difference methods can produce severely biased results, even if parallel trends and no anticipation assumptions hold (De Chaisemartin and d'Haultfoeuille, 2020; Gibbons et al., 2019; Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021). In contrast, the NPI approach we use allows for arbitrary correlation with unobservables and accommodates any heterogeneity of the treatment effect, as the ATE is simply an average across the sample (Manski and Pepper, 2000; Ban and Kédagni, 2022; Christelis and Dobrescu, 2020).

Instrumental variables methods are a workhorse estimator in studies focused on causal relationships. These studies rely on the exclusion restriction—that instruments affect labour market outcomes only indirectly through their impact on health (i.e., no correlation between the instrumental variable and the regression error). However, it is difficult to validate that the instruments used in existing studies have no direct effect on employment or earnings (Germinario et al., 2022). Sometimes, instruments are used to address only measurement error; for instance, self-reported general health (SGH) is often instrumented with demographics, NCDs, and disability (Flores and Kalwij, 2019; García-Gómez et al., 2010; Cai, 2021; Swaminathan and Lillard, 2000), which mitigates SGH measurement errors. NCDs and disability can be measured more accurately than SGH but must have direct effects on earnings and employment (Blundell et al., 2023). Moreover, the number of health conditions measured in surveys is always limited and cannot completely eliminate the possibility of measurement error (Blundell et al., 2023).

Several studies on specific health conditions, such as mental health, obesity, and disability, have used genetic markers and early life circumstances (Norton and Han, 2008; Pitt et al., 2021; Böckerman et al., 2019; Gao and Smyth, 2010), family events (Frijters et al., 2014), healthcare prices (Mani et al., 2018), daylight length (Tefft, 2012), and health conditions of children or siblings (Cawley, 2004; Johar and Katayama, 2012) as instruments. Although this approach improves on naive regressions, it is challenging to fully exclude direct links between these factors and labour market outcomes. Genetic markers and childhood events can directly influence labour income through impacts on cognitive abilities, brain chemistry, preferences, and beliefs, and indirectly through academic performance and educational attainment (Pitt et al., 2021; Fletcher, 2008, 2013a; Scholder et al., 2012; Sabia, 2007; Gilleskie et al., 2017; Christelis and Dobrescu, 2020; Germinario et al., 2022).

As discussed in Section 2.3 of Von Hinke et al. (2016), many genetic variants have multiple functions or are correlated with other variants that directly affect economic outcomes. Family events can also directly affect worker productivity (Oswald et al., 2015). Similarly, siblings' data are not a valid instrument if siblings share unobserved endowment factors (such as innate abilities, personality traits, motivation, lifestyle, appearance, or inherited communication skills) that influence recruitment and earnings (Fletcher, 2013b). For example, the use of twin data as instrumental variables is questioned due to issues such as birth spacing, parental responses, and behaviour in utero (Bhalotra and Clarke, 2020; Rosenzweig and Zhang, 2009). Healthcare prices and annual pay rises are often affected by socio-political factors and unions not captured in labour or household surveys. The duration of daylight might also directly influence labour force participation and working hours, in addition to its impact on health (Neidell et al., 2021; Graff Zivin and Neidell, 2014). Rosenzweig and Wolpin (2000) notes that studies using natural instruments often lack the necessary behavioural, market, and technological assumptions needed to justify their instruments and interpret their estimates. In addition, given the challenge of finding the appropriate instruments, IV approaches often focus on specific health conditions rather than a comprehensive measure of health. The NPI approach we use relies on an imperfect instrument that does not need to satisfy the exclusion restriction required in traditional IV approaches.

A second major challenge in this research area, and where our study contributes, is addressing measurement error (Pintor et al., 2024; Blundell et al., 2023). Most evidence relies on health measures from survey data, where respondents self-report their health conditions (e.g., disability, acute conditions like stroke, diabetes, kidney disease, cancer, depression) or rate their general health (e.g., excellent, very good, good, fair, or poor) (Minor, 2013; Aleksandrova et al., 2021; García-Gómez and López-Nicolás, 2006; Riphahn, 1999; García-Gómez et al., 2013; Gaulke, 2021; Blundell et al., 2023; Jones et al., 2020; Lenhart, 2019; García-Gómez et al., 2010; Cai et al., 2014; García-Gómez, 2011; Swaminathan and Lillard, 2000). Measurement errors can arise from misjudgements (e.g., classifying health as very good, good, or fair) or from justification bias, where individuals may misreport their health to rationalise their labour market performance (Blundell et al., 2023). Self-reported general health (SGH) is considered a broader but less precise measure of health (Blundell et al., 2023; Pintor et al., 2024). To address concerns about measurement error in self-reported health conditions, we use alternative health measures derived from blood biomarkers that capture common illnesses in the UK.

The third key distinction of our study lies in the construction of our health measure, which is based on the count of health conditions. Most existing studies measure the effects of self-reported general health (SGH) (Lenhart, 2019; Aleksandrova et al., 2021; García-Gómez, 2011; García-Gómez and López-Nicolás, 2006; Riphahn, 1999) or specific health conditions on labour market outcomes (see Pintor et al. (2024) Table A5.1 for a comprehensive list of the conditions studied). The magnitude of the estimated association varies depending on the health measure used. The impact is often greater for SGH compared to objective health conditions such as cancer, stroke, disability, weight, lung disease, and diabetes (Heinesen and Kolodziejczyk, 2013; Böckerman et al., 2019; Blundell et al., 2023; García-Gómez et al., 2013; Gaulke, 2021; Jones et al., 2020; Lenhart, 2019; García-Gómez, 2011). This discrepancy is justified by the notion that SGH reflects a broad range of health conditions, as any health problem ultimately affects general health (Blundell et al., 2023).

Studies focusing on a single or limited number of objective health conditions (like cancer, stroke, diabetes, or arthritis) typically define treatment using a binary indicator that switches to one when the condition is first reported in the data. Such approaches can underestimate the impact of health on the labour market for two reasons: (1) the control group may include individuals with other health issues, and (2) the impact of comorbidities is only captured if different health conditions are strongly correlated. For example, Blundell et al. (2023) found that the difference between the impacts of SGH and the objective measures diminishes only when a sufficient number of objective measures are included in the analysis.

Our health measure is based on the count of health conditions, including disability and all seventeen NCDs reported in the data. We compare the outcomes of individuals with no health conditions with those who report one, two, or three health conditions. This approach captures the effect of co-morbidities and provides a comparator group of individuals without any health conditions or disabilities. Our biomarker-based measure also includes results from 14 blood tests covering the most common illnesses in the UK (Benzeval et al., 2014).

Our closest study on the labour market impact of health conditions is Germinario et al. (2022), who used a similar approach and NLSY79 data to estimate the employment and earnings effects of mental health. They estimated that depression reduced annual earnings in the US by \$0 to \$6,082 in 1993, with the negative impact depending on the severity of depression. They found that mild to severe depressive symptoms reduced employment by 3–18%. In comparison, our estimates suggest that moderate depression results in a reduction of 2 to 4% in the probability of employment.

Bryan et al. (2022) estimated the impact of mental health (defined by a binary indicator taking a value of one when the GHQ-12 caseness score is greater than 3) from the first nine waves of the UKHLS using fixed effects regression. They found that a transition to poor mental health reduces the probability of employment by 1.6%. Using Oster's (2019) bounding approach, they concluded that their fixed-effects regression is robust to omitted variable bias. This result is consistent with our NPI estimate for mild distress only, but much lower than our estimate for moderate or severe depression. As Blundell et al. (2023) show, regression results using a binary health categorisation, as in Bryan et al. (2022), may be biased toward zero due to their inability to capture the severity of the disease.

In other countries, Lundborg et al. (2014) found in a large-scale dataset covering almost the entire Swedish male population that adolescent mental health reduces the annual earnings of adulthood by 6.4% and the probability of employment by 5.3%. Using the 2012 Canadian Community Health Survey and family members' mental health problems as an instrument for poor mental health, Shen (2023) estimated significant negative employment effects of mental health. However, their estimated loss in employment probability varied widely with the mental health measure used: 3.4% with Kessler-6 (K6) scores, 2.0% with Kessler-10 (K10) scores and 26.5% with self-reported mental health. This large variation may be due to bias in self-reported mental health, or the K6 and K10 tools might miss the severity of depression. Using the Longitudinal Internet Studies for Social Sciences (LISS) panel (2008–2018) for the Netherlands and the Mental Health Inventory 5 tool (MHI-5), Ringdal and Rootjes (2022) found that a one-standard deviation increase in women's (men's) mental health increases the likelihood of having paid employment one year later by 1.1–1.6% (1.4–1.7%). They also found that the more severe the symptoms of depression, the more detrimental the labour market outcomes. Thus, the existing literature agrees that capturing the severity of the disease is crucial for accurately measuring its impact on labour market outcomes. Readers interested in related mental health literature can consult Section 3.2.8 of Pintor et al. (2024), and Sections 2 of Germinario et al. (2022) and Bryan et al. (2022).

For other NCDs, the literature is extensive, providing a range of estimates for the loss in earnings and employment due to each health condition, making comprehensive coverage challenging in this study. Briefly, in the UK, Jones et al. (2020) found that the incidence of cancer, stroke, or myocardial infarction reduces monthly earnings by £74 one year later, based on the first seven waves of the UKHLS and matching methods. Their estimate is closer to the lower bound of the loss in earnings from one health condition shown in Table 6 of our study, where the maximum loss can reach £14,90. For two health conditions, our estimate suggests a minimum loss of £200 in monthly earnings.

Lenhart (2019) used waves 10–18 (2000–2008) of the British Household Panel Survey, combining fixed effects regression and matching methods, to estimate the effects of poor general health and any of multiple health conditions (e.g., body pain, migraine, asthma, anxiety, heart or blood pressure issues) on annual earnings and employment. They found that these health problems could reduce annual earnings by £1,000 to £5,000 but concluded that they have no significant effect on the probability of employment. In contrast, our estimates suggest that the employment effects of one health condition range from 0-7%, and the effects of two or three health conditions range from 2-8%.

Lundborg et al. (2014) found that in Sweden, annual earnings decrease by 6.4% due to mental health issues, 1.8% due to respiratory conditions, 3.1% due to nervous system conditions, and 2.5% due to injuries. Earnings decrease by 16.2% with diabetes and by 2.4% with asthma. Employment effects of these conditions were at most 3%. Böckerman et al. (2019) combined various Finnish data sources and used genetic markers to measure the impact of body mass index (BMI) on employment and earnings from 2001 to 2011. With a narrower genetic risk score, a one-unit increase in BMI decreased earnings by 6.9% and employment by 1.8%. However, with a broader genetic risk score, they could not reject the null hypothesis of no effect. Stephens and Toohey (2022) found that the Multiple Risk Factor Intervention Trial (MRFIT) in the US, aimed at reducing coronary heart disease, increased annual earnings by 3%. Vaalavuo (2021) found that breast cancer survivors in Finland had overall annual earnings reduced by about 5% (around €1,678), with significant differences between earnings quantiles. The employment effects were less than one percent.

Using data from the Danish administrative register, Heinesen and Kolodziejczyk (2013) found that breast and cholesterol cancer could reduce employment by around 4%, but only breast cancer had significant earnings effects. Nwosu (2018) found that depression and diabetes were associated with a

4 to 6 percentage point decrease in the probability of employment in South Africa during 2008-2014. For additional studies, readers can consult Pintor et al. (2024).

3 Identification Strategy

The standard method for drawing causal inferences with endogenous variables is by instrumental variable (IV) techniques. However, as mentioned in the previous section, IV-based inference can be unreliable if the instruments are weakly correlated with the endogenous variables or are potentially endogenous themselves. Even if several instruments pass the over-identification test, this does not guarantee their validity, as this test assumes that at least one instrument (or a linear combination) is valid (Kiviet, 2020, 2013). This concern has led to a modern literature, introduced and popularised by Manski and Pepper (Manski and Pepper, 2009, 2000; Manski, 1997, 1990, 1989), which relaxes the validity assumption, replacing strict equalities with (weak) inequalities. This approach allows for restricted direct effects of instruments on outcomes or correlation with the regression error but only provides bounds that must contain the true causal parameter.

Before introducing our NPI approach, we briefly discuss linear regression-based methods that rely on the bounding approach. These methods allow for limited correlation between endogenous or instrumental variables and the regression error. For example, Conley et al. (2012) permit mild violations of exclusion restrictions by assuming a plausible range of direct effects for the instruments on the dependent variable. In essence, they suggest including the instrument as an explanatory variable but restrict its coefficient, say τ, to a permissible range, e.g., $\tau \in [\tau_{min},\ \tau_{max}]$. Estimation involves producing confidence intervals for the parameter of interest (e.g., coefficient of the endogenous variable) across this range. The final bounds are constructed by taking the union of the resulting intervals. However, it is often difficult to determine a plausible range for the instrument coefficient, and providing a broad speculative range often yields uninformative bounds that include zero.

Nevo and Rosen (2012) derive bounds for the size of the effect with imperfect instruments, assuming a known sign and maximum strength of the instrument correlation with the error term. As will become clear, we also use an imperfect instrument to infer the impact of health on labour outcomes. Unlike the above methods, Kiviet (2020, 2013) propose kinky least squares (KLS), an identification strategy that avoids exclusion restrictions, instead making assumptions about the degree of endogeneity. KLS achieves set identification by restricting the correlation between the endogenous variable and the error term within plausible bounds. KLS does not require instrument identification, but if a potential instrument is available, it can assess its validity in a just-identified model, something not possible with standard IV or 2SLS methods. We later used this approach to show that our instruments are indeed imperfect. Another instrument-free approach by Oster (2019) imposes assumptions about the magnitude of the correlation between the endogenous variable and the structural regression error term relative to its correlation with other control variables in the regression. In practice, bounding Oster's sensitivity parameter is challenging because of the lack of natural bounds.

To introduce our NPI strategy, let each individual i have a response function $y_i(\cdot) : D \to Y$, mapping treatments $d \in D$ to potential outcomes $y_i(d) \in Y$. Here, treatment d represents health scores (SHS and BHS), ranging from zero to four, with higher values indicating better health. The outcome Y refers to employment or earnings.

To illustrate how the bounds are derived, let $y_i(d_1)$ and $y_i(d_2)$ represent two potential outcomes for the individual i associated with health scores d_1 and d_2, where $d_2 > d_1$. We aim to estimate the ATE of health improvement on outcomes, that is,

$$ATE(d_2, d_1) = E[y(d_2)] - E[y(d_1)]. \tag{1}$$

If treatment were externally determined (e.g., randomised), the ATE estimation would be straightforward. However, with observational data, estimating the ATE in Equation (1) is challenging because the potential outcome $y(d_2)$ is unobserved for individuals with different treatment levels from d_2, and $y(d_1)$ is unobserved for those with different levels from d_1. This identification issue can be illustrated using the identity $\mathbb{1}\{D = d_2\} + \mathbb{1}\{D \neq d_2\} = 1$ to express the expected potential outcome $E[y(d_2)]$ as:

$$E[y(d_2)] = E[y(d_2)\mathbb{1}\{D = d_2\}] + E[y(d_2)\mathbb{1}\{D \neq d_2\}]. \tag{2}$$

A similar expression can be written for $E[y(d_1)]$. The data identify the sample analogues of all terms on the right-hand side of Equation (2) except for the counterfactual $E[y(d_2)|D \neq d_2]$. Therefore, we need to impose assumptions to identify this missing counterfactual. In randomised control trials, randomisation ensures that $E[y(d_2)|D = d_2] = E[y(d_2)|D \neq d_2]$, allowing estimation of the unobserved potential outcome and, subsequently, the ATE in Equation (1). In observational data, this equality is unlikely to hold because of nonrandom treatment assignment.

To address the issue of unobserved potential outcomes, Manski (1989) proposed bounding the counterfactual in Equation (2). Initially, Manski (1989) suggested a bounded support assumption (no-assumption bounds), using minimum (Y_{min}) and maximum (Y_{max}) values for the observed outcome in place of $E[y(d_2)|D \neq d_2]$ (or $E[y(d_1)|D \neq d_1]$ for $E[y(d_1)]$). For example, the employment probability ranges between zero and one, while the earnings cannot be negative. The maximum observed earnings can be used as Y_{max}, as in Germinario et al. (2022). Upon replacement of the counterfactuals with these bounds, the upper and lower bounds (UB and LB) for $E[y(d_2)]$ are:

$$E[y(d_2)\mathbb{1}\{D = d_2\}] + Y_{min}\mathbb{1}\{D{\neq}d_2\} \leq E[y(d_2)] \leq E[y(d_2)\mathbb{1}\{D = d_2\}] + Y_{max}\mathbb{1}\{D{\neq}d_2\}. \quad (3)$$

As shown by Manski (1990), the ATE can be bounded as follows:

$$LB.E[y(d_2)] - UB.E[y(d_1)] \leq ATE(d_2, d_1) \leq UB.E[y(d_2)] - LB.E[y(d_1)]. \quad (4)$$

The interval between the lower and upper bound on $ATE(d_2, d_1)$ is its identification region. The bounds for other ATEs, such as $ATE(d_4, d_1)$ are computed analogously.

The no-assumption bounds above can rule out extreme positive or negative values, but are typically wide and include zero by default. To narrow these bounds, we introduce additional assumptions as described below.

Assumption 1. Same direction of correlation (SDC)

Let Z denote the imperfect instrument (mother's qualifications). The SDC assumption implies:

$$Cov(y(d), D)Cov(y(d), Z) \geq 0 \quad \text{for all} \quad d \in D. \quad (5)$$

Before introducing our other assumptions, it is important to note that the SDC assumption is a weaker version of two well-known assumptions of Manski (1997) and Manski and Pepper (2000).

Assumption 1.1 Monotone treatment selection (MTS)

The MTS assumption posits that individuals selected into better health scores have either monotonically increasing or decreasing latent employment probabilities and earnings. In our context, we expect that, on average, individuals with better health scores have at least the same or better latent employment probabilities and earnings as those with poorer health, that is,

$$d_2 > d_1 \Rightarrow E[y(d)|D = d_2] \ \geq \ E[y(d)|D = d_1] \quad \text{for all} \quad d{\in}D. \quad (6)$$

The MTS assumption alone is not testable, but when combined with the MTR assumption, it leads to testable implications, as explained below.

Assumption 1.2 Monotone instrumental variable (MIV)

The MIV assumption states that the latent employment probabilities and earnings of individuals have a weakly monotonic relationship with the instrumental variable (Manski and Pepper, 2000). As noted, we expect individuals' employment probabilities and earnings to be nondecreasing, on average, in their parents' qualifications, that is:

$$z_2 > z_1 \Rightarrow E[y(d)|Z = z_2] \;\geq\; E[y(d)|Z = z_1] \quad \text{for all} \quad d{\in}D \text{ and } z{\in}Z. \tag{7}$$

Assumption 2. Monotone treatment response (MTR)

The MTR assumption states that the potential outcome (employment and earnings) weakly increases with health scores, i.e.:

$$y(d) \;\geq\; y(d') \quad \text{for all} \quad d > d'. \tag{8}$$

Combining MTR and SDC assumptions is less restrictive than the traditional combination of MTR, MTS, and MIV, as SDC can still hold even if MTS and MIV are violated (Ban and Kédagni, 2022). The MTR, MTS, and MIV combination requires MTR to hold within each MIV block, which is more restrictive. For example, in the empirical application by Ban and Kédagni (2022), the combination of MTS + MIV + MTR was often violated, whereas SDC + MTR + LEI held. The combination of MTS and MTR assumption implies that the observed employment probability and earning should not decrease with better health on average. We test this assumption later from our data.

Assumption 3. Less endogenous instrument (LEI)

Let $\rho_{x,y}$ denote the correlation coefficient between variables x and y. The LEI assumption states that:

$$|\rho_{y(d),D}| \geq |\rho_{y(d),Z}| \quad \text{for all} \quad d{\in}D. \tag{9}$$

The LEI assumption, used in Ban and Kédagni (2022) and Nevo and Rosen (2012) among others, asserts that the imperfect instrument Z is less correlated with the potential outcome than endogenous treatment D. In our context, it is reasonable to assume that parental education is less correlated with an individual's potential employment probability and earnings than their own health conditions. Although this assumption tightens our bounds, we demonstrate in the results section that our findings remain robust even without it.

Identification under SDC

Let $\tilde{D} = D - E(D)$ $\tilde{Z} = Z - E(Z)$. The SDC assumption can be written as $E[y(d)\tilde{D}]E[y(d)\tilde{Z}]\geq 0$. This is equivalent to either:

$$E[y(d)\tilde{D}] \geq 0 \quad \text{and} \quad E[y(d)\tilde{Z}] \geq 0, \tag{10}$$

or

$$E[y(d)\tilde{D}] \leq 0 \quad \text{and} \quad E[y(d)\tilde{Z}] \leq 0. \tag{11}$$

Let $(\eta, \upsilon) \in \mathbb{R}_+^2 \setminus \{(0,0)\}$ with $\alpha = \eta + \upsilon \in [0,1]$ and $\beta = \frac{\eta}{\eta+\upsilon}$. Then Inequality (10) implies that for any $(\alpha, \beta) \in [0;1]^2$, we have:

$$E[\alpha y(d)(\beta\tilde{D} + (1-\beta)\tilde{Z})] \geq 0$$

and

$$E[-\alpha y(d)(\beta\tilde{D} + (1-\beta)\tilde{Z})] \leq 0.$$

For a detailed derivation of these inequalities, see Ban and Kédagni (2022). Defining $\mu^+ = 1 + \alpha(\beta\tilde{D} + (1-\beta)\tilde{Z})$ and $\mu^- = 1 - \alpha(\beta\tilde{D} + (1-\beta)\tilde{Z})$, and using the above inequalities, we get $E[\mu^+ y(d)] \geq E[y(d)]$ and $E[\mu^- y(d)] \leq E[y(d)]$. Then, using the potential outcome framework, we write an expression similar to Inequality (3):

$$E[\mu^- y(d_2)\mathbb{1}\{D = d_2\} + \mu^- y(d_2)\mathbb{1}\{D \neq d_2\}] \leq E[y(d_2)] \leq E[\mu^+ y(d_2)\mathbb{1}\{D = d_2\} + \mu^+ y(d_2)\mathbb{1}\{D \neq d_2\}]. \tag{12}$$

Combining with the Manski (1990) no-bound assumption, Inequality (12) can be expressed as:

$$\begin{aligned} E[\mu^- y(d_2)\mathbb{1}\{D = d_2\} + \mathbb{1}\{D \neq d_2\} \min\{\mu^- Y_{min}, \mu^- Y_{max}\}] \leq E[y(d_2)] \leq \\ E[\mu^+ y(d_2)\mathbb{1}\{D = d_2\} + \mathbb{1}\{D \neq d_2\} \max\left\{\mu^+ Y_{min}, \mu^+ Y_{max}\right\}]. \end{aligned} \tag{13}$$

The final step involves taking the supremum and infimum of the lower and upper bounds in (13) over (α, β), respectively. Denote the lower and upper bounds in Inequality (13) by $\underline{L}(Y, D, \mu^-(\alpha, \beta))$ and $\overline{L}(Y, D, \mu^+(\alpha, \beta))$. Thus, we can derive the following bounds for $E[y(d_2)]$ under Inequality (10):

$$I^1_{SDC}(d_2) \equiv \left[\sup_{(\alpha,\beta)\in[0,1]^2} \underline{L}(Y, D, \mu^-(\alpha, \beta)) \ , \ \inf_{(\alpha,\beta)\in[0,1]^2} \overline{L}(Y, D, \mu^+(\alpha, \beta))\right].$$

Following similar steps, we can derive the corresponding bounds for $E[y(d_2)]$ under Inequality (11):

$$I^2_{SDC}(d_2) \equiv \left[\sup_{(\alpha,\beta)\in[0,1]^2} \underline{L}(Y, D, \mu^+(\alpha,\beta)) \ , \ \inf_{(\alpha,\beta)\in[0,1]^2} \overline{L}(Y, D, \mu^-(\alpha,\beta)) \right].$$

This results in a range of lower and upper bounds for $E[y(d_2)]$. The final bounds are obtained by taking the union of these results. Thus, the SDC assumption combined with the no-assumption bounds yields:

$$E[y(d_2)] \in I^1_{SDC}(d_2) \cup \in I^2_{SDC}(d_2) = I_{SDC}(d_2).$$

Once these bounds are available for the desired values of D, the ATE under the SDC assumption can be derived using the expression in (4).

The bounds derived under SDC are no wider than Manski's no-assumption bounds, as the latter are a special case of the former when $\alpha = 0$. Notably, *if the lower bounds of* $I^1_{SDC}(d_2)$ *and* $I^2_{SDC}(d_2)$ *turn out to exceed their respective upper bounds during estimation, the combination of no-assumption and SDC is rejected by the data.*

The bounds are implemented using the Stata package *clr3bound* (Chernozhukov et al., 2015), which implements the estimation and valid-inference intersection bounds framework of Chernozhukov et al. (2013) to construct confidence regions for $I^1_{SDC}(d)$ and $I^2_{SDC}(d)$. The union of these regions yields the desired $I_{SDC}(d)$, as outlined by Ban and Kédagni (2022).

Implementing the bounds in *clr3bound* requires one to construct and plug in the following sample analogues of the above inequalities for each value of our categorical health measures:

$$\mu^+_d = 1 + \alpha[\beta(D - E(D)) + (1-\beta)(Z - E(Z))]$$
$$\mu^-_d = 1 - \alpha[\beta(D - E(D)) + (1-\beta)(Z - E(Z))]$$

$$Lower^1_{sdc}(d) = (D = d)\mu^-_d Y + (D{\neq}d) \min\left\{\mu^- Y_{min}, \mu^- Y_{max}\right\}$$
$$Lower^2_{sdc}(d) = (D = d)\mu^+_d Y + (D{\neq}d) \min\left\{\mu^+ Y_{min}, \mu^+ Y_{max}\right\}$$
$$Upper^1_{sdc}(d) = (D = d)\mu^+_d Y + (D{\neq}d) \max\left\{\mu^+ Y_{min}, \mu^+ Y_{max}\right\}$$
$$Upper^2_{sdc}(d) = (D = d)\mu^-_d Y + (D{\neq}d) \max\left\{\mu^- Y_{min}, \mu^- Y_{max}\right\}$$

Conditional expectations are estimated using the parametric method, the default option in *clr3bound*. α and β are generated as continuous uniform distributions. Section S2 of the online appendix in Ban and Kédagni (2022) provides further details on the implementation of the NPI in *clr3bound*.

Y_{min} and Y_{max} are set to zero and one for the probability of employment and to minimum and maximum monthly earnings for the earnings analysis.

Identification under SDC+LEI+MTR

Before presenting the results, we briefly outline the identification of ATE when all three assumptions are combined. For detailed derivations under each assumption, see Section 3 of Ban and Kédagni (2022).

The LEI assumption implies that $|\frac{E[y(d)\tilde{D}]}{\theta_D}| \geq |\frac{E[y(d)\tilde{Z}]}{\theta_Z}|$, where θ represents the standard deviations of the respective variables. Combining this with the SDC assumption implies either:

$$E[y(d)(\theta_Z\tilde{D} - \theta_D\tilde{Z}])] \geq 0, \quad E[y(d)\tilde{D}] \geq 0 \quad \text{and} \quad E[y(d)\tilde{Z}] \geq 0, \tag{14}$$

or

$$E[y(d)(\theta_Z\tilde{D} - \theta_D\tilde{Z}])] \leq 0, \quad E[y(d)\tilde{D}] \leq 0 \quad \text{and} \quad E[y(d)\tilde{Z}] \leq 0. \tag{15}$$

Then, from Inequality (14), for any $\alpha, \beta, \upsilon \in [0,1]^3$, we have:

$$E[\alpha y(d)(\upsilon(\theta_Z\tilde{D} - \theta_D\tilde{Z}) + \beta\tilde{D} + (1-\beta-\upsilon)\tilde{Z})] \geq 0$$

and

$$E[-\alpha y(d)(\upsilon(\theta_Z\tilde{D} - \theta_D\tilde{Z}) + \beta\tilde{D} + (1-\beta-\upsilon)\tilde{Z})] \leq 0.$$

Finally, define $\hat{\mu}^+ = 1 + \alpha(\epsilon(1-\beta)(\theta_Z\tilde{D} - \theta_D\tilde{Z}) + \beta\tilde{D} + (1-\beta)(1-\epsilon)\tilde{Z})$ and $\hat{\mu}^- = 1 - \alpha(\epsilon(1-\beta)(\theta_Z\tilde{D} - \theta_D\tilde{Z}) + \beta\tilde{D} + (1-\beta)(1-\epsilon)\tilde{Z})$, where $\upsilon = (1-\beta)\epsilon$ and $\epsilon \in [0,1]$. Following similar steps as for the SDC assumption, bounds for ATE under SDC+LEI assumptions can be obtained. The difference in the bounds for $E[y(d)]$ under SDC versus SDC+LEI lies in the terms μ and $\hat{\mu}$. μ^+ (μ^-) is a special case of $\hat{\mu}^+$ ($\hat{\mu}^-$) when $\epsilon = 0$. Thus, the bounds under SDC+LEI are at least as tight as those under the SDC assumption alone.

The MTR assumption implies that $[y(d_2)|D = d']$ is no greater than $[y(d')|D = d']$ for all $d' > d_2$, identified in the data by $[Y|D = d']$. This tightens the upper bound, which would otherwise be Y_{max} with no assumption bounds. Similarly, for any $d'' < d_2$, MTR implies that $[y(d_2)|D = d'']$ is no less than $[y(d'')|D = d'']$, represented by $[Y|D = d'']$ in the data. This tightens the lower bound on $E[y(d_2)]$, which would have been Y_{min} under the no-assumption bounds. More generally, this can be expressed as:

$$Y_{min}\mathbb{1}\{D > d_2\} \leq y(d_2)\mathbb{1}\{D > d_2\} \leq Y\mathbb{1}\{D > d_2\}$$

For all $D > d_2$, as under the MTR assumption, for any $j > d_2$, $y(d_2) \leq y(j)$ and expressed in the potential outcome framework $y(j) = Y$ when $j = d_2$. Similarly:

$$Y\mathbb{1}\{D < d_2\} \leq y(d_2)\mathbb{1}\{D < d_2\} \leq Y_{max}\mathbb{1}\{D < d_2\}$$

For all $D < d_2$. These two expressions imply that the actual observed outcome Y corresponding to a higher treatment level $j > d_2$ is greater than or equal to the potential outcome under the lower treatment level d_2, and the actual outcome Y corresponding to a lower treatment level than d_2 is at most $y(d_2)$. For a detailed exposition of the MTR assumption, see Manski (1997), and for a graphical representation, see Germinario et al. (2022).

It is clear from the above expression that the more detailed the treatment variable, the more effective the MTR assumption is in tightening the bounds. Thus, knowing the severity of a condition allows for tighter bounds compared to those created for a binary treatment indicator. However, one must ensure that the data are compatible with the MTR assumption for each treatment variable value. In our context, our categorical health measures each take up to five possible values (from very poor to excellent health), and the joint validity of MTR+MTS is tested for each value of these health scores in Section 5.

Since our indicator functions represent mutually exclusive and exhaustive events, the sum of outcomes for $D > d_2$ and $D < d_2$ is equal to the outcome for $D \neq d_2$. Hence, we can write:

$$y(d_2)\mathbb{1}\{D < d_2\} + y(d_2)\mathbb{1}\{D > d_2\} = y(d_2)\mathbb{1}\{D \neq d_2\}.$$

Using this identity and combining the assumptions of SDC, LEI, and MTR, the bounds under Inequality (14) can be written as:

$$\begin{aligned} E[\hat{\mu}^- y(d_2)\mathbb{1}\{D = d_2\} + \mathbb{1}\{D{>}d_2\} \min\left\{\hat{\mu}^- Y_{min}, \hat{\mu}^- Y\right\} + \mathbb{1}\{D{<}d_2\} \min\left\{\hat{\mu}^- Y_{max}, \hat{\mu}^- Y\right\}] \leq E[y(d_2)], \\ E[\hat{\mu}^+ y(d_2)\mathbb{1}\{D = d_2\} + \mathbb{1}\{D{>}d_2\} \max\left\{\hat{\mu}^+ Y_{min}, \hat{\mu}^+ Y\right\} + \mathbb{1}\{D{<}d_2\} \max\left\{\hat{\mu}^+ Y_{max}, \hat{\mu}^+ Y\right\}] \geq E[y(d_2)]. \end{aligned} \tag{16}$$

Following similar steps as for SDC, the bounds under Inequality (15) can be derived. The final step involves taking the infimum and supremum over $(\alpha, \beta, \epsilon) \in [0, 1]^3$. As with the SDC assumption, the ultimate identification region for $E[y(d_2)]$ is estimated as:

$$E[y(d_2)] \in I^1_{MTR}(d_2) \cup \in I^2_{MTR}(d_2) = I_{MTR}(d_2).$$

To summarise, under the SDC+LEI+MTR assumptions, we need to plug-in the following sample analogues of the above inequalities for each value d in D:

$$\hat{\mu}_d^+ = 1 + \alpha \left[\epsilon(1-\beta)[\theta_Z(D-E(D)) - \theta_D(Z-E(Z))] + \beta(D-E(D)) + (1-\beta)(1-\epsilon)(Z-E(Z))\right]$$

$$\hat{\mu}_d^- = 1 - \alpha \left[\epsilon(1-\beta)[\theta_Z(D-E(D)) - \theta_D(Z-E(Z))] + \beta(D-E(D)) + (1-\beta)(1-\epsilon)(Z-E(Z))\right]$$

$$Lower^1_{MTR}(d) = (D=d)\hat{\mu}_d^- Y + (D>d)\min\left\{\hat{\mu}^- Y_{min}, \hat{\mu}^- Y\right\} + (D<d)\min\left\{\hat{\mu}^- Y_{max}, \hat{\mu}^- Y\right\}$$

$$Lower^2_{MTR}(d) = (D=d)\hat{\mu}_d^+ Y + (D>d)\min\left\{\hat{\mu}^+ Y_{min}, \hat{\mu}^+ Y\right\} + (D<d)\min\left\{\hat{\mu}^+ Y_{max}, \hat{\mu}^+ Y\right\}$$

$$Upper^1_{MTR}(d) = (D=d)\hat{\mu}_d^+ Y + (D>d)\max\left\{\hat{\mu}^+ Y_{min}, \hat{\mu}^+ Y\right\} + (D<d)\max\left\{\hat{\mu}^+ Y_{max}, \hat{\mu}^+ Y\right\}$$

$$Upper^2_{MTR}(d) = (D=d)\hat{\mu}_d^- Y + (D>d)\max\left\{\hat{\mu}^- Y_{min}, \hat{\mu}^- Y\right\} + (D<d)\max\left\{\hat{\mu}^- Y_{max}, \hat{\mu}^- Y\right\}$$

The conditional expectations are then estimated using the parametric method of *clr3bound*. $(\alpha, \beta, \epsilon) \in [0, 1]$ are generated as continuous uniform distributions.

4 Data

The analysis in this paper uses data from waves 2 to 5 of the United Kingdom Household Longitudinal Study (UKHLS), covering 2010 to 2014. The UKHLS, a nationally representative panel study launched in 2009, includes around 40,000 households (approximately 100,000 individuals) across the UK (University of Essex, 2020). The General Population Sample (GPS) is a stratified, clustered sample representative of the UK population in 2009. The survey builds on the British Household Panel Study (BHPS), which began in 1991, with 8,000 BHPS households incorporated into the UKHLS, and has been widely used in health and labour studies (Jones et al., 2020). The fieldwork for each wave spans two calendar years. In addition to a household questionnaire, all adults 16 years and older complete an individual questionnaire. These cover topics such as household composition, income, wealth, expenditures, demographics, education, health, disability, labour market history, job characteristics, and income sources.

In 2010-2012 (Waves 2 and 3), after the standard annual survey, respondents from the GPS and BHPS samples were invited to a nurse health assessment, which included various physical measures and blood samples. For the GPS sample, this assessment took place in Wave 2, and for the BHPS sample, in Wave 3. The blood samples were analysed to produce a set of biomarkers. These biomarkers measure key risk factors for major public health issues or reflect biological pathways

Table 1. Construction of health measures

$\sum_p NCD_p + Disab.$	SHS	$\sum_q Test_q$	BHS
0	Excellent self-reported health (4)	0	Excellent nurse-measured health (4)
1	Good self-reported health (3)	1	Good nurse-measured health (3)
2	fair self-reported health (2)	2	fair nurse-measured health (2)
3	Poor self-reported health (1)	3	Poor nurse-measured health (1)
>3	Very poor self-reported health (0)	>3	Very poor nurse-measured health (0)

Notes: NCD, Noncommunicable disease indicators; Disab. Disability indicator; SHS, Self-reported health score; BHS, biomarkers-based health score.
$NCD_p = 1$ if NCD_p is reported for $p \in (1, ..., 17)$, zero otherwise,
$Disab. = 1$ if disability reported, zero otherwise,
$Test_q = 1$ if $test_q$ is not normal for $q \in (1, ..., 14)$, zero otherwise.

linking social and environmental factors to health (Benzeval et al., 2014; National Institute of Health, 2001).

We use two health measures in our analysis: one based on biomarkers from 14 blood tests and another on self-reported health conditions and disabilities, as summarised in Table 1.

Our first health measure, Self-reported Health Scores (SHS), is based on counts of health conditions: including 17 non-communicable diseases (NCDs) and disability. In each wave, individuals report whether a doctor or health professional has diagnosed them with any of the following 17 conditions: asthma, arthritis, congestive heart failure, coronary heart disease, angina, heart attack (myocardial infarction), stroke, emphysema, chronic bronchitis, COPD, hypothyroidism, hyperthyroidism, liver conditions, cancer, diabetes, epilepsy, high blood pressure, or clinical depression. A positive response to any NCD is coded as one and zero otherwise. They are also asked if they have long-standing physical or mental impairment, illness, or disability (defined as lasting or expected to last 12 months or more). A yes response is coded as one and zero otherwise.

The 18 binary indicators (17 NCDs and disability) are summed so that the resulting variable ranges from zero (no conditions reported) to 18 (all 17 NCDs and disability). The Self-reported Health Score (SHS) is derived from total conditions, as detailed in Table 1. Thus, SHS can range from zero to four, with a higher value indicating better health.

Our second measure, the Biomarker-Based Health Score (BHS), is derived from 14 blood test results. These tests assess blood fat (cholesterol and triglycerides), diabetes (glycated haemoglobin HbA1c), inflammation and immune function, anaemia (haemoglobin and ferritin), and liver and kidney functions (Benzeval et al., 2014). These biomarkers are relevant for major illnesses in the UK and key physiological systems. Fourteen binary indicators are created from the blood test results, with abnormal results coded as one and normal results as zero. For details on measurement, clinical significance, and cut-off points, see Benzeval et al. (2014).

Briefly, normal scores for each blood test are as follows: total cholesterol $\leq$5 mmol/L, HDL-cholesterol > 1 mmol/L, triglycerides < 2 mmol/L, glycated haemoglobin (HbA1c) $\leq$48 mmol/mol, c-reactive protein (CRP) (a marker of inflammation/immune function) $\leq$3 mg/L, cytomegalovirus antibody measurement (CMV) was considered normal when both immunoglobulin G (IgG) and immunoglobulin M (IgM) tests turnout negative, haemoglobin (Hb) (a markers of anaemia) $\geq$13 g/dL for men and $\geq$12 g/dL for women, alanine transaminase (ALT) $\leq$40 U/L, aspartate transaminase (AST) $\leq$40 U/L, alkaline phosphatase (ALP) 30-130 U/L is treated normal, gamma glutamyl transferase (GGT) $\leq$70 U/L for men and $\leq$45 for women, and albumin of 35-50 g/L is considered normal. Creatinine is used to estimate glomerular filtration rate (eGFR), which is a standard measure of kidney function. A eGFR score of 60 plus is considered normal for the analyses. Finally, urea levels within 2.5-7.8 mmol/L are also considered normal.

The binary indicators for the 14 blood tests are summed, so the resulting variable ranges from zero (all tests normal) to 14 (all tests abnormal). The Biomarker-Based Health Score (BHS) is derived from these total blood test scores, as outlined in Table 1. Similarly to SHS, BHS ranges from zero to four, with a higher value indicating better health. As defined in Table 1, the number of health conditions (abnormal blood tests) and the health categorisation (ranging from excellent to very poor health) are used interchangeably throughout the article.

We estimate the effect of SHS (BHS), constructed from Wave 3 (Waves 2 and 3), on employment and gross monthly labour earnings in Wave 4 (Waves 4 and 5) to address concerns of reverse causality, as suggested by previous literature (Germinario et al., 2022). Since the nurse assessment in Waves 2 and 3 occurred after the main survey, using a later wave ensures that employment and earnings are not measured before the blood samples. For the biomarker-based measure (BHS), we combined Waves 4 and 5 due to the small sample size. The appendix reports results using only Wave 4 data.

Employment is indicated by a binary variable: one for paid or self-employment and zero for unemployment or sick/disability leave. Individuals not economically active or women on maternity leave are excluded. Waves 3 and 4 were chosen for SHS analyses based on nurse assessment data from Waves 2 and 3, used to construct BHS, ensuring consistency across health measures. Sensitivity analyses also used SHS scores from Wave 2 and outcomes from Wave 3.

The UKHLS records parents' educational qualifications when respondents were 14 years of age. We use the mother's qualification as our imperfect instrument in the main analysis. This choice aligns with our identification strategy, where the imperfect instrument must have the same directional

Table 2. Summary statistics

Full sample	Full	Men	Women	White	Non-white
Age	42.77	43.06	42.55	43.59	38.41
Qualification: Degree	0.38	0.37	0.38	0.37	0.44
Qualification: A-level/GCSE	0.53	0.54	0.52	0.54	0.45
Qualification: No qualification	0.09	0.09	0.10	0.09	0.12
Mother didn't go to school	0.03	0.03	0.03	0.01	0.19
Mother left school without qualifications	0.40	0.38	0.41	0.41	0.33
Mother left school with qualifications	0.32	0.33	0.31	0.33	0.25
Mother gained post school qualifications	0.18	0.17	0.18	0.19	0.13
Mother gained a uni. degree or higher	0.08	0.08	0.07	0.07	0.10
No. of non-communicable diseases	0.13	0.12	0.13	0.13	0.13
Disability	0.31	0.30	0.31	0.32	0.23
Monthly gross wage (000s, Wave 4)	1.94	2.41	1.58	1.97	1.81
Employed (Wave 4)	0.87	0.86	0.87	0.87	0.81
Weekly work hours (Wave 4)	32.44	36.83	28.98	32.56	31.67
Individuals	33356	14283	19073	28121	5164
Blood test sample					
Age	45.66	39.85	45.68	45.92	40.67
Qualification: Degree	0.39	0.48	0.39	0.39	0.46
Qualification: A-level/GCSE	0.51	0.52	0.51	0.52	0.46
Qualification: No qualification	0.10	0.00	0.10	0.10	0.07
Mother didn't go to school	0.01	0.00	0.01	0.01	0.12
Mother left school without qualifications	0.44	0.44	0.44	0.44	0.41
Mother left school with qualifications	0.31	0.41	0.31	0.31	0.24
Mother gained post school qualifications	0.18	0.15	0.18	0.18	0.12
Mother gained a uni. degree or higher	0.07	0.00	0.07	0.07	0.11
No. of positive tests	2.37	3.21	2.36	2.36	2.53
Monthly gross wage (000s, Wave 4 & 5)	1.68	2.49	1.66	1.67	1.81
Employed (Wave 4 & 5)	0.90	0.92	0.90	0.90	0.86
Weekly work hours (Wave 4 & 5)	29.37	37.23	29.18	29.30	30.66
Individuals	4523	17	4507	4294	229

relationship with potential outcomes (employment and earnings) as the health scores. We expect that people with educated parents have better employment and earnings potential on average, supported by substantial literature showing that children with more educated parents tend to have higher cognitive abilities, better education, and higher earnings (Mourifie et al., 2020; Heckman et al., 2018; Heckman and Mosso, 2014; Dubow et al., 2009; Chevalier et al., 2013; Haveman and Wolfe, 1995; Lemke and Rischall, 2003; Hoogerheide et al., 2012). Since parental qualifications are a strong predictor of their children's qualifications, they are commonly used as an instrumental variable in studies that estimate the impact of education on earnings (Mourifie et al., 2020). However, as with many existing studies, we show that parental qualifications also have direct effects on their children's employment and earnings potential. Thus, we use it as an imperfect instrument.

The summary statistics for our two samples are shown in Table 2. In the SHS (full) sample, 33,356 individuals have non-missing health condition data; earnings data are available for 19,513, mother's qualifications for 27,487, and employment data for 25,096. We have complete data for 15,539 individuals for the earnings analysis and 19,874 for the employment analysis. In the biomarker-based sample, 4,523 individuals have complete data on all 14 blood tests. Earnings data are available for 4,885 observations (Waves 4 and 5 combined), and employment data for 6,202 observations. Complete data is available for 4,275 observations in the earnings analysis and 5,422 in the employment analysis.

In the full sample (column 1), the average age is 42.77 years; 57% are female, and 15.4% are non-white. The mothers of 40% left school without qualifications, 38% obtained school qualifications, and 18% obtained post-school qualifications. Disabilities are more common than NCDs in the sample, with women reporting more of both than men. Although the employment rate is higher among women, their monthly earnings (£1,580) and weekly working hours (28.98) are substantially lower than those of men, £2,410 earnings and 36.83 hours, respectively. Whites also fare better than non-whites in terms of employment and earnings.

The sample with complete blood tests is predominantly women and largely white, with only 17 men and 5% non-white individuals. On average, the individuals in this sample have two abnormal blood test results. Due to the high proportion of women, monthly earnings and working hours are lower in this sub-sample compared to the full sample. The average age is 45.66 years.

5 Validation of Empirical Approach

As outlined in our identification strategy (Section 3), some of our assumptions are testable. Here, we assess the extent to which they are likely to hold.

Same Direction of Correlation (SDC) assumption SDC assumes that the correlation between parental qualifications (Z) and children's potential employment and earnings has the same sign as the correlation between children's health score (D) and their potential employment and earnings. In other words, individuals' potential employment and earnings are assumed to be non-decreasing in both their own health and their parents' qualifications, or vice versa. Parental education likely affects offspring's potential employment and earnings due to unobserved parental cognitive and non-cognitive investments, making it an imperfect instrument. However, such investments are unlikely to negatively affect the employment and earnings of their offspring (Mourifie et al., 2020; Heckman et al., 2018; Heckman and Mosso, 2014). Thus, it is commonly used as an imperfect instrument (Heckman et al., 2018; Mourifie et al., 2020; Nevo and Rosen, 2012; Ban and Kédagni, 2022). Similarly, as highlighted in Table 4 later, better health is unlikely to reduce employment

Table 3. Own and parental education by health scores

Health status	Very poor	Poor	Fair	Good	Excellent
Full sample	$SHS = 0$	$SHS = 1$	$SHS = 2$	$SHS = 3$	$SHS = 4$
Age	52.33	50.24	48.78	47.04	40.28
Male	0.46	0.39	0.39	0.42	0.44
Ethnicity: white	0.77	0.87	0.86	0.88	0.83
Ethnicity: mixed	0.01	0.01	0.02	0.02	0.02
Ethnicity: others	0.21	0.11	0.12	0.10	0.16
Qualification: Degree/Higher degree	0.17	0.22	0.28	0.34	0.41
Qualification: A-level/GCSE etc.	0.58	0.56	0.53	0.53	0.53
Qualification: No qualification	0.25	0.22	0.19	0.13	0.07
Mother didn't go to school	0.05	0.04	0.04	0.03	0.03
Mother left school without qualifications	0.69	0.55	0.55	0.46	0.35
Mother left school with qualifications	0.18	0.25	0.24	0.29	0.34
Mother gained post school qualifications	0.07	0.12	0.14	0.16	0.19
Mother gained a uni. degree or higher	0.01	0.04	0.03	0.06	0.09
Observations	142	336	1766	9294	21823
Blood sample	$BHS = 0$	$BHS = 1$	$BHS = 2$	$BHS = 3$	$BHS = 4$
Age	50.43	47.34	45.98	42.03	38.53
Male	0.01	0.01	0.00	0.00	0.00
Ethnicity: white	0.95	0.93	0.95	0.96	0.98
Ethnicity: mixed	0.01	0.00	0.01	0.01	0.01
Ethnicity: others	0.04	0.06	0.05	0.03	0.01
Qualification: Degree/Higher degree	0.29	0.34	0.42	0.47	0.42
Qualification: A-level/GCSE etc.	0.52	0.55	0.49	0.49	0.55
Qualification: No qualification	0.18	0.11	0.08	0.04	0.03
Mother didn't go to school	0.01	0.01	0.01	0.01	0.01
Mother left school without qualifications	0.57	0.48	0.45	0.32	0.28
Mother left school with qualifications	0.23	0.29	0.31	0.36	0.38
Mother gained post school qualifications	0.16	0.16	0.17	0.21	0.23
Mother gained a uni. degree or higher	0.03	0.06	0.06	0.10	0.10
Observations	1820	1831	2473	2026	730

Notes: SHS, Self-reported health score; BHS, Blood test based health score.

or earnings potential. Therefore, we expect both Z and D to have positive correlations with the outcomes on average, although SDC only requires them to be in the same direction. If the outcome variable has natural bounds (for example, employment probability between 0 and 1), the validity of SDC can be tested using the data.

We can assess the plausibility of the SDC assumption by examining average characteristics by health scores. Individuals with better health scores should intuitively have better education and more qualified parents, which are linked to better labour market outcomes. Table 3 shows the educational and demographic data for the full and blood samples, categorised by health scores. Parental education is lower for those with poorer health scores. For example, among individuals with more than three health conditions (compared to none), 69% (35%) of mothers left school without qualifications, 18% (34%) obtained school qualifications, 07% (19%) received post-school qualifications, and 01% (09%) achieved a university degree. In the blood sample, 57% of mothers lacked qualifications for individuals with more than three abnormal blood tests, versus 28% for those with all normal tests. In all other cells in Table 3, higher maternal qualifications are generally associated with better health.

For individual's own qualifications, the percentage with a degree or higher is positively associated with better health in both samples on average. Similarly, the percentage with no qualifications decreases with better health scores. Thus, the differences in qualifications (Table 3) and employment/earnings (Table 4 below) align with the SDC assumption.

Monotone Treatment Selection (MTS) and Monotone Treatment Response (MTR) Introduced by Manski (1997), and used in Manski and Pepper (2000), Manski and Pepper (2009), Christelis and Dobrescu (2020), Germinario et al. (2022), among many others, the MTR assumption must hold for all individuals, although it may not apply to each individual. Combined with the MTS assumption, it implies that the observed average employment probabilities and earnings should not decrease with better health, which is tested in Table 4.

Table 4 presents the average employment rates and earnings for the SHS and BHS values for different UKLHS waves. It shows that average employment rates and earnings increase with better health across all waves and both health measures. For example, in Wave 4, individuals with excellent self-reported health (no health conditions) have an employment rate (earnings) of 93% (£2,005), while those with very poor health (more than three health conditions) have 40% (£1,297). Individuals with poor health have a 53% employment rate (£1,597), and those with fair health have 64% (£1,845).

Similarly, for blood tests, individuals with very poor health (more than three abnormal tests) have an employment rate of 81% (£1,554 earnings) in Wave 4, while those with poor health (three abnormal tests) have 88% (£1,582), those with fair health have 93% (£1,681) and those with good health have 94% (£1,725). In all other waves reported in Table 4, the data support the assumptions of MTS and MTR under both health measures and aligns with the assumption of SDC as well (Table 3).

Theoretically, better health can improve employment and earnings through several channels. Health can directly improve productivity by increasing workers' ability to perform tasks, such as putting in greater physical effort or concentration. It also makes individuals more adept at using technology and adapting to changes in work, which can indirectly lead to higher wages. In addition, better health can positively impact cognitive functioning and education, reduce sick days and disruptions,

TABLE 4. Average labour outcomes by health scores, overall sample

	Very poor health	Poor health	Fair health	Good health	Excellent health
Full sample					
Wave 4	$SHS = 0$	$SHS = 1$	$SHS = 2$	$SHS = 3$	$SHS = 4$
Monthly wage	1296.78	1597.10	1844.89	1876.32	2005.19
Employment ratio	0.40	0.53	0.64	0.78	0.93
Observations	83	215	1141	6239	15795
Wave 3					
Monthly wage	1268.29	1517.43	1697.13	1853.37	1967.84
Employment ratio	0.34	0.48	0.65	0.76	0.93
Observations	103	250	1284	6286	16187
Waves 2-8					
Monthly wage	1387.54	1694.51	1838.25	1957.71	2096.61
Employment ratio	0.33	0.51	0.66	0.77	0.94
Observations	402	1083	6536	38683	102694
Blood sample					
Wave 4	$BHS = 0$	$BHS = 1$	$BHS = 2$	$BHS = 3$	$BHS = 4$
Monthly wage	1554.27	1582.63	1681.08	1725.60	1729.55
Employment ratio	0.81	0.88	0.93	0.94	0.96
Observations	599	627	894	808	306
Waves 4 & 5					
Monthly wage	1538.03	1587.63	1697.52	1732.30	1748.50
Employment ratio	0.81	0.88	0.92	0.94	0.95
Observations	1149	1239	1729	1577	611

Notes: SHS, Self-reported health scores; BHS, Blood test based health scores.

and increase available work time. Improved health can boost self-confidence and reduce uncertainty about the future, helping long-term planning and risk taking. Furthermore, better health provides more time for networking, which can create economic opportunities (Pintor et al., 2024; Germinario et al., 2022). It also reduces mobility barriers associated with disabilities or serious conditions, improving labour market matching. While better health might increase the utility of leisure and potentially reduce work hours, there is little evidence linking better health to worse labour market outcomes. In fact, numerous studies show that better health is associated with positive labour market outcomes (Shawa et al., 2024; Pintor et al., 2024).

To conclude this section, our data do not reject the validity of the SDC assumption or the combination of the MTR and MTS assumptions for both health measures across all analysed waves.

Before presenting the results, note that NPI methods are designed for cross-sectional data and often involve small samples (Germinario et al., 2022; Ban and Kédagni, 2022). The point estimates are sample averages, which do not require large samples. Confidence intervals are derived using the valid-inference intersection bounds framework of Chernozhukov et al. (2013), specifically designed to address small-sample bias.

However, a few key points are worth noting. The full sample includes only a small number of individuals with more than three health conditions (see Table 4). However, this does not affect our ATE estimates, as it is not included in their calculation. Rather, $SHS = 0$ serves to tighten the bounds for other ATE calculations. MTR ensures that the potential outcomes for individuals with $SHS = 1$ cannot be lower than those with $SHS = 0$, tightening the lower bound for $SHS = 1$. The lower bound for $SHS = 0$ remains unchanged, as no health states in our data are worse than very poor health (see the lower bounds for $SHS = 0$ in the results section).

In the blood sample, only 306 individuals have entirely normal blood tests (classified as excellent health). As this serves as the comparator group for the ATE calculations, we combine data from Waves 4 and 5 to ensure a sufficient sample size for estimation. Results using only Wave 4 are provided in the Appendix.

6 Results

Employment effects of health

Table 5 presents the 95% confidence intervals for the lower (LB) and upper (UB) bounds of the employment probability for each health state, along with the corresponding ATEs. ATEs are estimated by comparing the employment rate of individuals in excellent health (without health conditions) with the other groups, as in Equation (4). For the biomarker-based sample, the ATEs are derived by comparing individuals with all normal blood tests to the other groups. The ATEs in Table 5 present the impact of worsening health. For example, the effect on employment of moving from excellent health (no health condition or $SHS = 4$) to good health (one health condition or $SHS = 3$) is represented by $ATE(3, 4)$. The results can also be interpreted in reverse, moving from good health ($SHS = 3$) to excellent health ($SHS = 4$), represented by $ATE(4, 3)$. To do this, simply reverse the sign and positions of the LB and UB in Table 5. Similarly, other ATEs, such as $ATE(1, 3)$ or $ATE(2, 3)$, can be calculated using the formula in Equation (4).

The analysis covers both paid and self-employed individuals, resulting in a larger sample compared to the earnings analysis. The SDC bounds are ignored because they are similar to the SDC+LEI bounds. The SDC+LEI bounds in Table 5 are wide, as is typical for such estimates (Germinario et al., 2022; Ban and Kédagni, 2022).

The SDC+LEI bounds suggest that the true causal effect of moving from excellent health to good health ($ATE(3, 4)$) on employment ranges from -70% to 27%. The MTR assumption, which posits that poor health does not improve labour market outcomes, tightens these bounds. The SDC+LEI+MTR bounds indicate that moving from excellent to good health reduces the probability

TABLE 5. Estimated bounds of health scores effects on employment probability

	OLS	$SDC+LEI$		$SDC+LEI+MTR$		$SDC+MTR$	
Full sample		LB	UB	LB	UB	LB	UB
Very poor health ($SHS=0$)		0.02	0.98	0.02	0.87	0.02	0.88
Poor health ($SHS=1$)		0.03	0.98	0.87	0.87	0.88	0.88
Fair health ($SHS=2$)		0.06	0.97	0.87	0.88	0.88	0.88
Good health ($SHS=3$)		0.24	0.93	0.88	0.90	0.88	0.90
Excellent health ($SHS=4$)		0.66	0.94	0.90	0.95	0.90	0.95
$ATE(1,4)$	-0.39 (0.04)	-0.92	0.32	-0.08	-0.02	-0.08	-0.02
$ATE(2,4)$	-0.28 (0.02)	-0.88	0.31	-0.08	-0.02	-0.07	-0.02
$ATE(3,4)$	-0.15 (0.01)	-0.70	0.27	-0.07	0.00	-0.07	0.00
Blood sample	OLS	LB	UB	LB	UB	LB	UB
Very poor health ($BHS=0$)		0.22	0.99	0.22	0.93	0.22	0.93
Poor health ($BHS=1$)		0.22	0.98	0.93	0.95	0.93	0.95
Fair health ($BHS=2$)		0.27	0.99	0.95	0.97	0.95	0.97
Good health ($BHS=3$)		0.28	0.99	0.97	0.99	0.97	0.99
Excellent health ($BHS=4$)		0.13	1.00	0.99	1.01	0.99	1.01
$ATE(1,4)$	-0.07 (0.01)	-0.78	0.85	-0.08	-0.04	-0.08	-0.04
$ATE(2,4)$	-0.02 (0.01)	-0.73	0.85	-0.05	-0.02	-0.06	-0.02
$ATE(3,4)$	-0.01 (0.01)	-0.72	0.85	-0.04	0.00	-0.04	0.00
$ATE(1,3)$	-0.06 (0.01)	-0.77	0.71	-0.06	-0.02	-0.06	-0.02
$ATE(2,3)$	-0.01 (0.01)	-0.72	0.71	-0.04	0.00	-0.04	0.00

Notes: Robust standard errors for the OLS regressions in (.) in Column 1. OLS, Ordinary least square; SDC, Same direction of correlation; LEI, Less endogenous instrument; MTR, Monotone treatment response; LB, Lower bound (95%); UB, Upper bound (95%); SHS, Self-reported health score; BHS, Biomarker-based health score; ATE, Average treatment effects.
1. OLS coefficients come from linear regression of health scores (SHS or BHS) on the binary employment indicator (=1 if employed, zero otherwise).
2. ATE(1,4) provides the bounds on the average treatment effect of moving from excellent health (SHS=4) to poor health (SHS=1) (e.g., an effect ranging from 2% to 8% under $SDC+LEI+MTR$ in the full sample), while ATE(3,4) shows the bounds for moving from excellent to good health (SHS=3). All other ATEs can be interpreted similarly.

of employment by 0% to 7%, and from excellent to fair health (two conditions) ($ATE(2,4)$) by 2% to 8%, ruling out zero causal effects.

The true causal effect could lie within these bounds, with comparisons between excellent health and fair or poor health ruling out zero effects. The OLS estimates for $ATE(1,4)$, $ATE(2,4)$, and $ATE(3,4)$ all fall below the lower NPI bounds. In Germinario et al. (2022), consistent with the omitted variable bias, the OLS estimates of mental health effects on earnings and employment generally fall below the NPI bounds. This aligns with OLS exhibiting a statistically significant downward bias (Blundell et al., 2023).

In the blood sample, the estimated ATE bounds show that individuals with good health (one abnormal test) have a 0%–4% lower probability of employment, while those with fair health (two abnormal tests) have a 2%–5% lower probability of employment compared to individuals with excellent health (all normal tests). The estimated bounds for comparisons between none versus one, none versus two, none versus three, and one versus three abnormal tests rule out zero effects. The OLS estimates are within or near the boundary of the NPI bounds. In general, both self-reported and biomarker-based health scores reveal a similar pattern: declining health can reduce employment by 0 to 8%, depending on the severity or number of health conditions.

The ATE bounds of our estimates depend on the MTR assumption. Under this assumption, the lower bound of employment for a healthier group cannot be below the upper bound for a less healthy group (Manski, 1997; Germinario et al., 2022). In our context, this means that the upper bound of ATE cannot be greater than zero. In Table 5, the upper bounds of each preceding value of SHS are equal to the lower bounds of the subsequent values. Nonetheless, our sample agrees with both the MTR and SDC assumptions. Furthermore, OLS regression estimates increase consistently with the number of reported health conditions, indicating that the increase in lower bounds for successive SHS values is not solely due to the MTR assumption, but that it may indeed hold in our data.

Earning effects of health

Table 6 presents the 95% confidence intervals for the lower (LB) and upper (UB) bounds of earnings (in thousands pounds) for each value of SHS and BHS, along with the corresponding ATEs. The estimated ATE bounds for SHS suggest that moving from excellent to good health decreases earnings by up to £1,490, although null effects cannot be ruled out. The bounds comparing individuals with no health conditions with individuals with two or three conditions exclude zero effects. For example, moving from excellent to fair health (three conditions or $ATE(1,4)$) reduces earnings by £240 to £1,520. The $ATE(2,4)$ bounds exclude the OLS estimate in Column 1. However, if treatment effects are heterogeneous, OLS estimates the treatment effects for the treated, which differs from the ATE.

In our biomarker-based sample, the estimated ATE bounds for BHS suggest that an abnormal test reduces earnings by £0 to £1,480 per month, while two abnormal tests result in a monthly loss of £1,500 to £2,320. The bounds for none versus three ($ATE(1,4)$), none versus two ($ATE(2,4)$), and one versus three ($ATE(1,3)$) abnormal tests rule out zero effects and exclude the corresponding OLS estimate in Column 1.

The OLS estimates for the blood sample indicate a smaller impact than for self-reported health conditions, whereas the NPI bounds are generally larger. Overall, both self-reported and biomarker-based data show substantial health effects on earnings, which increase significantly with co-morbidities.

Comparison with return to qualifications

Table 6. Estimated bounds of health scores effects on monthly earnings (in thousands £)

	OLS	$SDC+LEI$		$SDC+LEI+MTR$		$SDC+MTR$	
Full sample		LB	UB	LB	UB	LB	UB
Very poor health ($SHS=0$)		0.35	2.96	0.35	2.03	0.36	2.04
Poor health ($SHS=1$)		0.21	4.00	2.03	2.03	2.04	2.04
Fair health ($SHS=2$)		0.21	7.60	2.03	2.07	2.04	2.08
Good health ($SHS=3$)		0.62	6.56	2.07	2.27	2.08	2.21
Excellent health ($SHS=4$)		1.48	3.56	2.27	3.55	2.21	3.48
$ATE(1,4)$	-0.40 (0.11)	-3.34	2.52	-1.52	-0.24	-1.44	-0.17
$ATE(2,4)$	-0.15 (0.06)	-3.35	6.12	-1.52	-0.20	-1.44	-0.13
$ATE(3,4)$	-0.10 (0.03)	-2.94	5.08	-1.49	0.00	-1.40	0.00
Blood sample	OLS	LB	UB	LB	UB	LB	UB
Very poor health ($BHS=0$)		0.53	5.44	0.56	1.79	0.55	1.82
Poor health ($BHS=1$)		0.59	6.05	1.79	2.77	1.82	2.70
Fair health ($BHS=2$)		0.84	5.61	2.77	3.61	2.70	3.39
Good health ($BHS=3$)		0.76	5.65	3.61	5.10	3.39	4.96
Excellent health ($BHS=4$)		0.49	5.08	5.10	5.08	4.96	5.15
$ATE(1,4)$	-0.12 (0.06)	-4.49	5.55	-3.29	-2.34	-3.33	-2.27
$ATE(2,4)$	-0.02 (0.06)	-4.24	5.11	-2.32	-1.50	-2.45	-1.58
$ATE(3,4)$	0.01 (0.06)	-4.32	5.16	-1.48	0.00	-1.76	0.00
$ATE(1,3)$	-0.14 (0.05)	-5.06	5.29	-3.31	-0.84	-3.14	-0.69
$ATE(2,3)$	-0.04 (0.05)	-4.82	4.85	-2.34	0.00	-2.27	0.00

Notes: Robust standard errors for the OLS regressions in (.) in Column 1. OLS, Ordinary least square; SDC, Same direction of correlation; LEI, Less endogenous instrument; MTR, Monotone treatment response; LB, Lower bound (95%); UB, Upper bound (95%); SHS, Self-reported health score; BHS, Biomarker-based health score; ATE, Average treatment effects.

1. OLS coefficients come from linear regression of health scores (SHS or BHS) on the monthly earnings.
2. ATE(1,4) provides the bounds on the average treatment effect of moving from excellent health (SHS=4) to poor health (SHS=1) (e.g., an effect ranging from £240 to £1520 under $SDC + LEI + MTR$ in the full sample), while ATE(3,4) shows the bounds for moving from excellent to good health (SHS=3). All other ATEs can be interpreted similarly.

Economically, how significant is the impact of job and earnings losses due to poor health? A common approach in economics is to compare the gains or losses of a policy or event to the returns on education, as qualifications are the key to the success of the labour market (Germinario et al., 2022). To assess the economic importance of health, we compare its effects on labour market outcomes with those of education. As in health analysis, maternal education is used as an imperfect instrument for individual qualifications. The results are presented in Table 7. We focus on the estimated differences between A-levels and no qualifications, and between degree/higher qualifications and GCSEs, as these have precise bounds. Our findings suggest that A-levels can increase the probability of employment by 3% to 86% and monthly earnings by £430 to £3,670. From Table 6, the earnings impact of having two or three fewer health conditions largely falls within these bounds. Similarly, the earnings effects of having no abnormal blood tests versus two (£1,500–£2,320) or three (£2,340–£3,290) align with the earnings gains associated with A-level qualifications. The bounds on the estimated

Table 7. Estimated bounds of education effects on employment and earnings (£000s)

	$SDC + LEI$		$SDC + LEI + MTR$		$SDC + MTR$	
Employment	LB	UB	LB	UB	LB	UB
No qualification (0)	0.08	0.97	0.08	0.89	0.09	0.89
GCSE (1)	0.27	0.96	0.89	0.92	0.89	0.92
A-level (2)	0.22	0.96	0.92	0.95	0.92	0.95
Degree (3)	0.46	0.97	0.95	0.97	0.95	0.97
$ATE(0,3)$	-0.88	0.51	-0.89	-0.06	-0.88	-0.06
$ATE(1,3)$	-0.70	0.50	-0.08	-0.03	-0.08	-0.03
$ATE(2,3)$	-0.75	0.51	-0.05	0.00	-0.05	0.00
$ATE(0,2)$	-0.88	0.75	-0.86	-0.03	-0.86	-0.03
$ATE(1,2)$	-0.70	0.74	-0.06	0.00	-0.06	0.00
$ATE(0,1)$	-0.87	0.70	-0.84	0.00	-0.83	0.00
Earning	LB	UB	LB	UB	LB	UB
No qualification (0)	0.20	4.85	0.20	2.05	0.20	2.05
GCSE (1)	0.59	6.37	2.05	2.48	2.05	2.47
A-level (2)	0.55	6.51	2.48	3.87	2.47	3.76
Degree (3)	1.28	5.46	3.87	5.46	3.76	5.35
$ATE(0,3)$	-5.26	3.58	-5.26	-1.82	-5.15	-1.71
$ATE(1,3)$	-4.87	5.09	-3.41	-1.39	-3.31	-1.29
$ATE(2,3)$	-4.91	5.24	-2.98	0.00	-2.89	0.00
$ATE(0,2)$	-6.31	4.30	-3.67	-0.43	-3.56	-0.42
$ATE(1,2)$	-5.93	5.82	-1.82	0.00	-1.71	0.00
$ATE(0,1)$	-6.16	4.27	-2.29	0.00	-2.27	0.00

Notes: SDC, Same direction of correlation; LEI, Less endogenous instrument; MTR, Monotone treatment response; LB, Lower bound (95%); UB, Upper bound (95%); ATE, Average treatment effects.
ATE(0,3) provides the bounds of the average employment and earning differences between no qualifications and degree level qualifications, while ATE(2,3) shows the bounds of the average employment and earning differences between A-level and degree level qualifications. All other ATEs can be interpreted similarly.

employment difference between A-levels and no qualifications are wider and less informative for comparison.

From Table 7, the difference in the probability of employment between degree and GCSE qualifications ranges from 3% to 8%, with earnings differences of £1,390–£3,410. The employment impact of having no versus two health conditions is similar (2%–8%) in Table 5. Likewise, having two abnormal blood tests reduces the probability of employment by 2%–5% and three abnormal tests by 4%-8%. These findings suggest that health capital may be as important as educational qualifications in shaping labour market outcomes.

Blundell et al. (2005) used several methods to estimate wage gains of around 24% for A-levels (compared to no qualifications), 30% for higher qualifications versus o-levels, and 48% for higher qualifications versus no qualifications, using data from the National Child Development Survey,

Table 8. Estimated bounds of health effects on employment and earnings (£000s)

	OLS	*SDC + LEI*		*SDC + LEI + MTR*		*SDC + MTR*	
$ATE(0,1)$ employment		LB	UB	LB	UB	LB	UB
Asthma	-0.15 (0.03)	-0.91	0.05	-0.91	-0.02	-0.91	-0.01
Arthritis	-0.23 (0.02)	-0.90	0.07	-0.91	-0.00	-0.90	-0.02
Heart diseases	-0.46 (0.05)	-0.91	0.05	-0.92	-0.01	-0.91	-0.00
Stroke	-0.62 (0.08)	-0.92	0.04	.	.	.	.
Liver condition	-0.43 (0.05)	-0.92	0.04	-0.92	-0.01	-0.92	-0.00
Cancer or malignancy	-0.18 (0.04)	-0.92	0.05	-0.92	-0.01	-0.91	-0.01
Diabetes	-0.31 (0.04)	-0.92	0.04	-0.92	-0.01	-0.92	-0.00
High blood pressure	-0.19 (0.02)	-0.88	0.07	-0.89	-0.01	-0.89	-0.01
Clinical depression	-0.35 (0.03)	-0.91	0.05	-0.91	-0.01	-0.90	-0.01
Severe depression	-0.29 (0.02)	-0.85	0.17	-0.85	-0.03	-0.85	-0.02
Moderate depression	-0.16 (0.01)	-0.82	0.17	-0.04	-0.02	-0.04	-0.01
Mild depression	-0.11 (0.01)	-0.80	0.17	-0.03	0.00	-0.03	0.00
$ATE(0,1)$ Earning		LB	UB	LB	UB	LB	UB
Asthma	-.33 (0.09)	-1.93	5.77	-1.93	0.00	-1.92	0.00
Arthritis	-.27 (0.06)	-1.97	4.02	-1.97	0.00	-1.96	0.00
Heart diseases	-0.13 (0.14)	-1.62	2.02	-1.62	0.00	-1.64	0.00
Stroke	0.06 (0.52)	-1.44	3.81	.	.	.	.
Liver condition	-0.24 (0.13)	-1.93	1.96	-1.93	0.00	-1.97	0.00
Cancer or malignancy	-0.13 (0.17)	-1.76	5.99	-1.76	0.00	-1.80	0.00
Diabetes	-.22 (0.12)	-1.94	4.98	-1.95	0.00	-1.96	0.00
High blood pressure	-0.09 (0.06)	-2.10	5.68	-2.10	0.00	-2.05	0.00
Clinical depression	-0.44 (0.07)	-2.03	4.67	-2.04	0.00	-1.99	0.00
Severe depression	-0.20 (0.05)	-2.66	5.05	-2.70	-0.43	-2.57	-0.31
Moderate depression	-0.21 (0.04)	-2.64	5.88	-0.91	-0.19	-0.76	-0.16
Mild depression	-0.12 (0.04)	-2.63	5.25	-0.66	0.00	-0.61	0.00

Notes: Robust standard errors for the OLS regressions in (.) in Column 1. OLS, Ordinary least square, SDC, Same direction of correlation; LEI, Less endogenous instrument; MTR, Monotone treatment response; LB, Lower bound (95%); UB, Upper bound (95%); ATE, Average treatment effects.
1. OLS coefficients come from linear regression of the binary disease indicator on the respective outcome.
2. ATE(0,1) provides the bounds on the average employment and earnings differences between having no health condition and having the specific health condition in the respective row.

which tracks individuals born in a single week in March 1958. At the mean wage for the healthy group, $SHS = 4$, our lower bound estimates suggest a 21% earnings difference between A-levels and no qualifications, and around 65% between degree or higher qualifications and GCSEs.

Disease-specific estimates

In this subsection, we estimate the effect on employment and earnings of various health conditions reported in the data. For the ATE estimates, the outcomes of individuals with each condition are compared with a group of healthy individuals, that is, those who do not report any NCD or disability. This differs from existing studies, which focus on specific health conditions (Pintor et al., 2024). Those studies typically compare individuals with a given condition, such as depression, with the rest of the sample, including individuals with other NCDs and disabilities in the control group.

This approach does not quantify the effect of mental health specifically, but instead measures how the outcomes of those with mental health conditions differ from the population average, which includes both healthy and unhealthy individuals.

The disease-specific impacts are shown in Table 8. Our data violates the MTR+MTS and SDC+LEI+MTR assumptions only for stroke. The results suggest that all health conditions could substantially reduce monthly earnings. However, NPI estimates cannot rule out statistically zero causal effects, as our assumptions are less useful when diseases are measured by binary indicators. In contrast, the NPI bounds for employment show that the true causal effects of each NCD are significantly different from zero. The earnings sample includes only those in paid employment, while some conditions, such as stroke, may be debilitating enough to completely prevent work, others, such as diabetes, may have less impact on work capacity. This difference in impact on earnings versus employment may stem from the earnings sample containing individuals with relatively milder conditions compared to the employment sample.

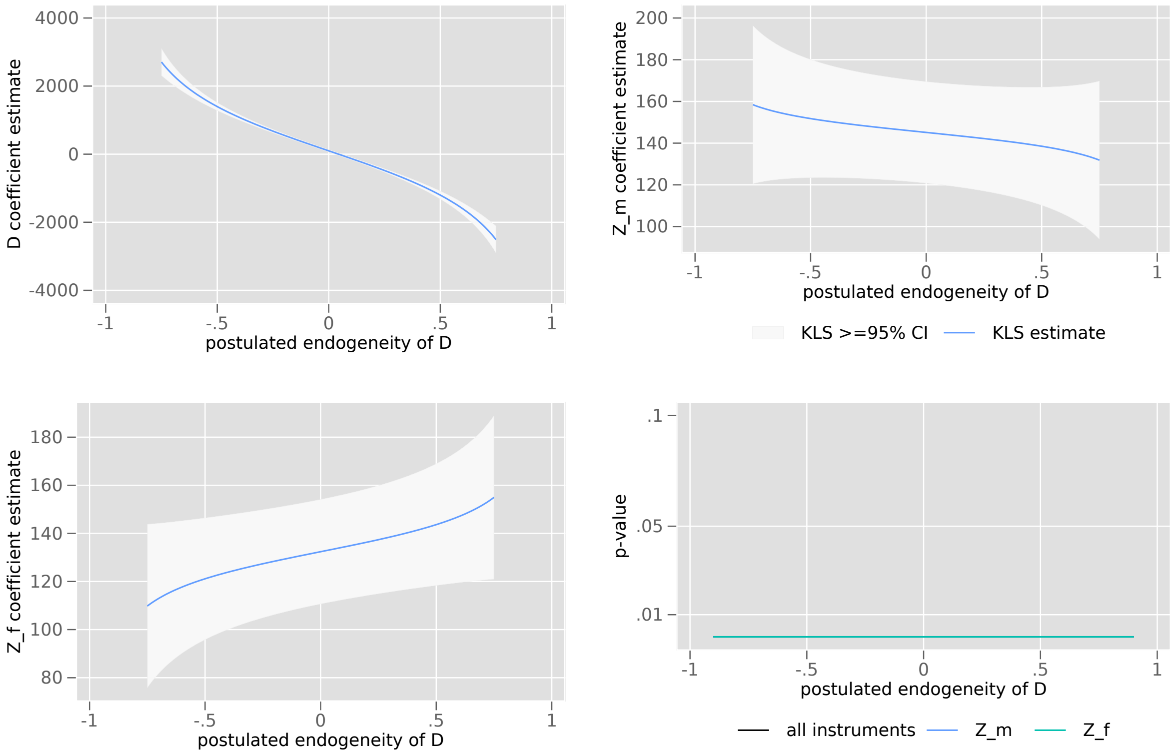
The negative earnings impacts are stronger for clinical depression and high blood pressure. Monthly earnings decrease by up to £2,040 for those with clinical depression and £2,100 for those with high blood pressure. These findings align with existing studies that show stronger negative effects of mental health compared to physical health, as noted in Section 3.2.8 of Pintor et al. (2024) and in Lundborg et al. (2014).

Our estimates using GHQ-12 caseness scores show that mild depression (scores 3-6) lowers the probability of employment by 0%-3% while moderate depression (scores 7-9) by 2%-4%. Mild depression reduces monthly earnings by £0 to £660, and moderate depression reduces earnings by £190 to £910.

Our estimate of mental health is consistent with those from a similar approach but significantly higher than estimates derived from fixed-effects regressions. Bryan et al. (2022) used the first nine waves of the UKHLS using fixed effects regression and suggested that a transition to poor mental health reduces the probability of employment by 1.6%. Their estimate aligns with our NPI estimate for mild distress, but is lower than our estimate for moderate depression.

Germinario et al. (2022) estimated that depression reduced annual earnings in the United States by \$0 to \$6,082 in 1993. Like our findings, their results show that the negative impact of mental health depends on the severity of depression. They found that mild to severe depressive symptoms reduce employment by 3% to 18%, while our estimate suggests a reduction of 2%-4% for mild depression and 3%-85% for severe depression. Thus, our results generally align with those of the US, although

Figure 1. KLS estimates of SHS effects on earnings and p-values for exclusion restriction tests



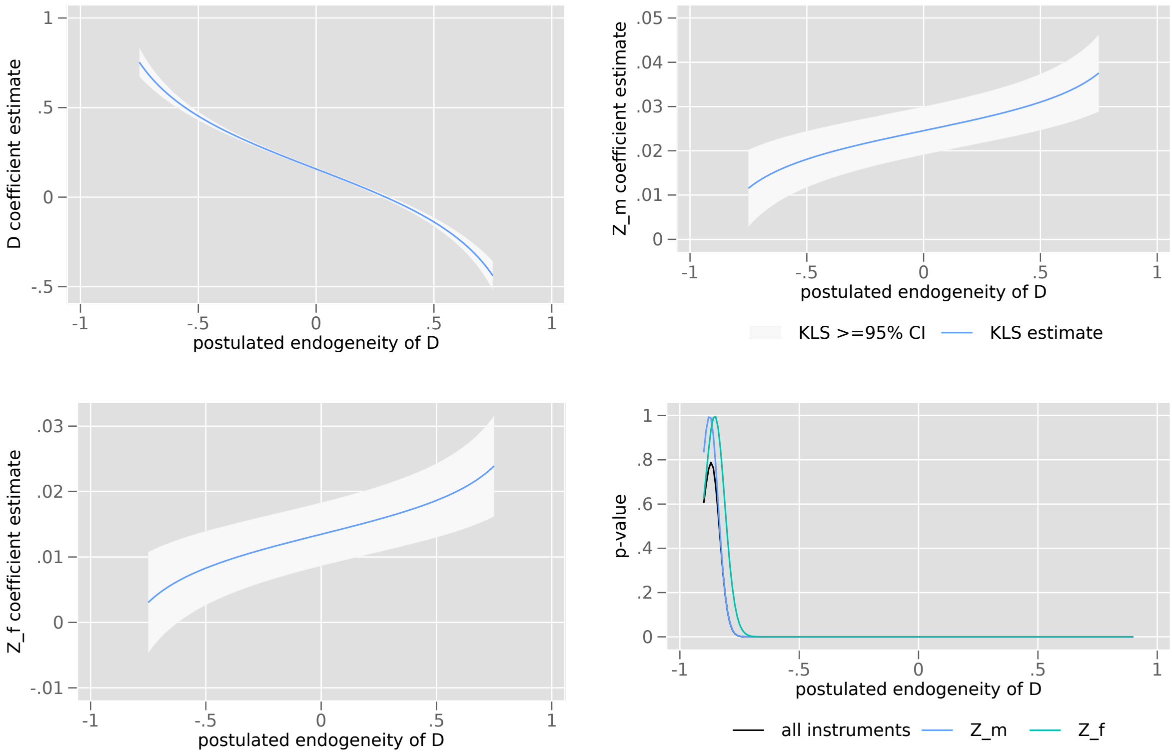
our lower bound for severe depression is wider due to reliance on relatively weaker assumptions compared to Germinario et al. (2022).

Parents' qualifications as instruments

The KLS approach described earlier does not require instrumental variables, but allows for sensitivity analysis in standard instrumental-variables inference to check instrument validity. It enables the assessment of instrument validity in a just-identified regression model, which is not possible in standard IV or 2SLS regressions. The KLS approach includes the instrument as an explanatory variable, while allowing for a range of correlations between the suspected endogenous variable and the error term. From this regression, the range of values consistent with the valid exclusion of the instrument can be determined. Note that in over-identified IV/2SLS, one can apply the over-identification test of instrument validity. However, that test is based on the assumption that at least one of the instrumental variables (or a linear combination of them) is a valid instrument (Kiviet, 2020, 2013).

To assess whether parental qualifications can serve as standard valid instruments, that is, have no direct effects on earnings and employment, we perform KLS regressions of D (SHS) on employment

Figure 2. KLS estimates of SHS effects on employment and p-values for exclusion restriction tests



and earnings. As the extent of correlation between SHS and the regression errors is uncertain, we allow it to vary between -0.9 and 0.9. The first panels in Figures 1 and 2 show the impact of SHS on earnings and employment for each assumed value of the correlation between SHS and regression errors. The second and third panels present the direct effects of the qualifications of mother (Z_m) and father (Z_f) on earnings and employment in the range of postulated correlations. The final panels in Figures 1 and 2 show the p-values of the Wald test for the null hypothesis that the instruments have direct effects on the outcomes, even after accounting for the correlation between SHS and the regression errors.

In the wage regression in Figure 1, the KLS exclusion restriction tests support our claim that parental qualifications are unlikely to be valid instruments for all postulated correlations between SHS and regression error. In the employment regression in Figure 2, only within a narrow range of highly negative endogeneity correlations do we fail to reject the null hypothesis that the instruments are validly excluded from the model. However, in this range, the corresponding effects of SHS on employment probability in panel one appear highly unrealistic, with each increase in SHS (or improvement in health) raising employment probability by approximately 60%. This puts serious doubts on the use of standard instrumental variable regressions in our context.

Sensitivity checks and gender-specific estimates

To test the robustness of our results concerning instrument selection and the UKHLS wave, we re-estimated our NPI bounds for self-reported health using father's qualifications as an imperfect instrument and Wave 3 instead of Wave 4. Biomarker-based estimates were repeated using only Wave 4 data. Finally, we conduct an estimation using a subsample of only women.

Table 9 in the Appendix presents father education alongside health scores. Similarly to mothers, fathers of individuals with poorer health scores generally have lower educational attainment. For example, among those with very poor health (more than three health conditions), 59% of fathers left school without qualifications, compared to 34% of those in excellent health. Furthermore, 16% (vs. 25%) obtained school qualifications, 18% (vs. 26%) achieved post-school qualifications, and 3% (vs. 13%) earned a university degree. In all other cells of Table 9, fathers' qualifications are non-decreasing in health scores, supporting the use of father qualifications as an imperfect instrument.

The estimated bounds on the ATE in Table 10 in the Appendix, using the father's qualifications as an instrument, suggest that transitioning from excellent to good health (one health condition), $ATE(3,4)$, reduces the probability of employment by up to 6%, while having two health conditions lowers it by 2% to 7% under the assumptions SDC+LEI+MTR. The bounds for ATEs comparing no health conditions with two or three conditions rule out zero effects.

For earnings, the estimated bounds indicate that moving from excellent to good health reduces earnings by up to £1,600 (compared to £1,490 when using mother's qualifications as an instrument). The bounds for comparisons between excellent health and fair or poor health also exclude zero effects. Specifically, transitioning from no to three health conditions, $ATE(1,4)$, reduces earnings by £230 to £1,630 (compared to the previous estimate of £240 to £1,520).

Table 11 in the Appendix, based on Wave 3 of UKHLS, indicates that moving from excellent to good health reduces the probability of employment by up to 1% to 7%, and all comparisons rule out zero effects. For example, moving from excellent to poor health, $ATE(1,4)$, reduces the probability of employment by 3% to 9%. The estimated bounds suggest that moving from excellent to poor health decreases earnings by £360 to 1,650 under the SDC+LEI+MTR assumptions. Comparisons between no health conditions and two and three also exclude zero effects. These results largely align with those obtained using mother qualifications as an instrumental variable or Wave 4 of the survey.

The final sensitivity analysis contains biomarker-based estimates with employment and earnings data from Wave 4 of the UKHLS. Table 12 in the Appendix shows that individuals in good health (one abnormal test) have a lower employment probability, 0%–2%, while those in fair health have a

1% to 4% lower probability than those in excellent health. The estimated ATE bounds for earnings suggest that an abnormal test reduces earnings by £0 to £1,270 per month, while two abnormal tests result in a £970–£2,490 monthly loss. Overall, the Wave 4 biomarker-based estimates align with the main results from Waves 4 and 5, except that employment effects are slightly lower when using Wave 4 alone.

Our final exercise estimates the population ATE for women based on self-reported health. Table 13 shows that transitioning from excellent to good health reduces women's employment probability by 0%–7% (as in the full sample) and from excellent to fair health by 2%–7% (compared to 2%–8% in the full sample). The estimated ATE bounds for SHS suggest that moving from excellent to good health decreases women's earnings by up to £1,800, while a change to fair health reduces earnings by £220–£1,840 (compared to £240–£1,520 in the full sample). Overall, the estimates for women are broadly consistent with those for the full sample.

7 Conclusion and Discussion

Identifying the causal effects of health on employment and earnings from observational data is challenging due to measurement errors, omitted variable bias, and reverse causality. To address omitted variable bias and reverse causality, some studies have used genetic markers, childhood events, and parental or sibling factors as instruments. However, these instruments are unlikely to meet the exclusion restrictions required for valid instrument selection. Most of the literature uses individual fixed effects, which account for unobserved time-invariant characteristics, but cannot control for time-varying confounders, such as family circumstances, new skills, workplace structure, or colleagues, all of which may influence both health and labour outcomes. Furthermore, when health effects vary across individuals and over time, IV and fixed-effect methods estimate treatment effects for specific subpopulations (local treatment effects and treatment effects for the treated, respectively) that may not align with population average treatment effects. Additionally, recent econometric advances have shown that standard fixed-effect methods are not suitable when individuals are exposed to diseases at different times and the effects vary between individuals, due to the negative weighting problem in these regressions (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021).

We contribute to the literature in three key ways, addressing measurement errors, omitted variable bias, and reverse causality in assessing the impact of health on earnings and employment. We apply a nonparametric partial identification method to provide bounds on the population average effect of health on these outcomes. This approach tackles the endogeneity of health conditions due to omitted variables or reverse causality, relying on more plausible assumptions that our data cannot reject. Being nonparametric, it requires only the calculation of sample averages, making it robust to all forms of treatment effect heterogeneity. We use two alternative health measures that offer a broader view of health, encompassing various disabilities and discomforts. One of these measures is derived from biomarkers in blood samples from respondents, eliminating the risk of misjudgement or justification bias, a common issue in self-reported health data (Blundell et al., 2023). We compare individuals who report one or more health conditions with those who report no NCD and disabilities, reinforcing the notion that "the grass is always greener on the other side" in our context.

Our results under the assumptions of SDC, MTR, and LEI show that individuals with one condition have a 0%-7% lower probability of employment, while two conditions reduce it by 2%-8% compared to those without health conditions. The corresponding OLS employment loss estimates also increase with the number of conditions but often exceed the lower bounds of NPI estimates, suggesting downward bias in OLS estimates of health effects on employment. Having one health condition affects earnings by £0 to £1,490 and two conditions reduce monthly earnings by £200 to £1,520 compared to those with no health conditions.

Estimates using our biomarker-based health measure indicate that individuals with an abnormal blood test have a 0%-4% lower probability of employment and earn £0 to £1,480 less compared to those without abnormal tests. Similarly, individuals with two abnormal tests have a 2%-5% lower employment probability and earn £1,500 to £2,320 less.

Estimated bounds for conditions such as asthma, arthritis, cancer, diabetes, clinical depression, and heart and liver diseases suggest that these can significantly reduce earnings and employment. However, the lack of information on the severity of the condition often widens these bounds, making it difficult to rule out zero-earning effects. However, our estimates indicate that the true causal effects of each NCD on employment are negative and significantly different from zero. For mental health, we measure the severity of the distress using the General Health Questionnaire (GHQ-12) caseness scores. Our results show that mild depression reduces monthly earnings by £0 to £660, and moderate depression by £190 to £910. This suggests that the reported zero-earnings effects of depression in some studies may be due to an inability to accurately capture the severity of health conditions (see the review of the literature in Germinario et al. 2022).

Our estimates of health impact bounds are somewhat comparable in magnitude to those for differences in employment and earnings between individuals with no qualifications and A-levels, or GCSEs and degrees. For example, the difference in the probability of employment between degree and GCSE qualifications ranges from 3% to 8%, with an earnings gap of £1,390 to £3,410. The employment effect of having no versus two health conditions is 2%-8%. Similarly, two fewer abnormal blood tests increase the probability of employment by 2%-5%. These findings suggest that health capital may be as significant as educational qualifications in improving labour market outcomes.

This study provides strong evidence for the significant link between good health and individual economic success. Comparison with education returns highlights substantial benefits of health on labour market outcomes. For policy makers, this emphasises the importance of integrating health into labour market strategies. Similarly, health interventions should be seen not only from a health and social care perspective, but also as a strategic economic investment that affects the wider economy and labour market.

The estimation method we employ provides bounds on the population Average Treatment Effect (ATE), which is often the relevant policy parameter, as opposed to a subpopulation treatment effect. It relies on relatively weak assumptions, which we have tested using the data. Our data do not reject the validity of these assumptions, lending further credibility to our estimate of the health impact on labour market outcomes. However, this approach yields a range of possible values for the population ATE rather than a single-point estimate. The range is more informative in the cases where we have data on the severity of a disease or the number of comorbidities that individuals may have. In studies focusing on a single condition, with only basic knowledge of its presence or absence, the resulting bounds may be less useful for policy decisions.

In addition to providing a range of possible values for the population ATE rather than a single-point estimate, this study has few other limitations. Health conditions may cause people to become economically inactive or retire early. Our analysis of employment effects is restricted to economically active populations (employed, unemployed, or on sick/disability leave). Second, the analysis of wage impact does not encompass self-employed individuals, as their income derives from a mix of labour and capital incomes. Third, the impact of health can vary by age or qualification, but this study does not explore these aspects due to space limitations. Finally, several criteria were used for eligibility to participate in nurse assessment and biomarker generations (see Benzeval et al. 2014), which could lead to selection bias in blood sample analyses.

Many surveys now include health-related quality of life measures (such as the EQ-5D or SF-12 questionnaires), various disability indicators (mobility, manual dexterity, continence, hearing, sight, communication, speech problems, lifting or moving objects, memory, concentration, learning, and understanding), as well as severity measures for conditions like mental health or diabetes. Similarly, routinely collected practice datasets often report the severity of conditions under study. Additionally, the NPI approach relies on imperfect instruments. Most surveys routinely report childhood factors, parental qualifications, occupations, household income, composition, and other characteristics, which can serve as imperfect instruments. Postcode or local area deprivation indices, often available to link to most datasets, can also act as imperfect instruments.

As such, the increasing availability of more detailed health and economic data offers greater opportunities to use these bounding approaches to complement standard regression-based approaches, enhancing our ability to reliably estimate the effects of health and other interventions on key outcomes.

Declarations

CRediT authorship contribution statement

Akbar Ullah: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Sara Jabeen**: Conceptualisation, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Doriane Mignon**: Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Luke Munford**: Formal analysis, Investigation, Methodology, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This paper benefited from helpful discussions with many individuals attending Health Economics Study Group (HESG) meetings (summer 2023 and winter 2025), and seminars attendees at the Manchester Centre for Health Economics (MCHE) and Health Organisation, Policy and Economics (HOPE).

Data Sharing

The data used in this study come from the UK Household Longitudinal Study (Understanding Society). We are not permitted to make the data publicly available. However, the data are available through the UK Data Service (https://ukdataservice.ac.uk/) under End User License conditions by registered users. We will provide all codes used for data cleaning and analysis to enable interested readers with access to the data to reproduce all tables presented in the main paper and online appendix.

Supplementary materials

The online version contains an additional file of supplementary appendix.

References

Adda, J., Banks, J., and Von Gaudecker, H.-M. (2009). The impact of income shocks on health: evidence from cohort data. *Journal of the European Economic Association*, 7(6):1361–1399.

Aleksandrova, E., Bagranova, V., and Gerry, C. J. (2021). The effect of health shocks on labour market outcomes in Russia. *Cambridge Journal of Economics*, 45(6):1319–1336.

Ban, K. and Kédagni, D. (2022). Nonparametric bounds on treatment effects with imperfect instruments. *The Econometrics Journal*, 25(2):477–493.

Benzeval, M., Davillas, A., Kumari, M., and Lynn, P. (2014). Understanding Society: The UK Household Longitudinal Study biomarker user guide and glossary. https://mpra.ub.uni-muenchen.de/114713/1/MPRA_paper_114713.pdf.

Bhalotra, S. and Clarke, D. (2020). The twin instrument: Fertility and human capital investment. *Journal of the European Economic Association*, 18(6):3090–3139.

Blundell, R., Britton, J., Dias, M. C., and French, E. (2023). The impact of health on labor supply near retirement. *Journal of Human Resources*, 58(1):282–334.

Blundell, R., Dearden, L., and Sianesi, B. (2005). Evaluating the effect of education on earnings: models, methods and results from the national child development survey. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 168(3):473–512.

Böckerman, P., Cawley, J., Viinikainen, J., Lehtimäki, T., Rovio, S., Seppälä, I., Pehkonen, J., and Raitakari, O. (2019). The effect of weight on labor market outcomes: An application of genetic instrumental variables. *Health economics*, 28(1):65–77.

Böckerman, P., Hyytinen, A., and Maczulskij, T. (2017). Alcohol consumption and long-term labor market outcomes. *Health economics*, 26(3):275–291.

Bryan, M. L., Rice, N., Roberts, J., and Sechel, C. (2022). Mental health and employment: a bounding approach using panel data. *Oxford Bulletin of Economics and Statistics*, 84(5):1018–1051.

Cai, L. (2021). The effects of health on the extensive and intensive margins of labour supply. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 184(1):87–117.

Cai, L., Mavromaras, K., and Oguzoglu, U. (2014). The effects of health status and health shocks on hours worked. *Health economics*, 23(5):516–528.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

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

Cawley, J. (2004). The impact of obesity on wages. *Journal of Human resources*, 39(2):451–474.

Chernozhukov, V., Kim, W., Lee, S., and Rosen, A. M. (2015). Implementing intersection bounds in stata. *The Stata Journal*, 15(1):21–44.

Chernozhukov, V., Lee, S., and Rosen, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica*, 81(2):667–737.

Chevalier, A., Harmon, C., O'Sullivan, V., and Walker, I. (2013). The impact of parental income and education on the schooling of their children. *IZA Journal of Labor Economics*, 2:1–22.

Chirikos, T. N. (1993). The relationship between health and labor market status. *Annual Review of Public Health*, 14(1):293–312.

Christelis, D. and Dobrescu, L. I. (2020). The causal effect of social activities on cognition: evidence from 20 European countries. *Social Science & Medicine*, 247:112783.

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

Currie, A., Shields, M. A., and Price, S. W. (2007). The child health/family income gradient: Evidence from England. *Journal of health economics*, 26(2):213–232.

Currie, J. and Madrian, B. C. (1999). Health, health insurance and the labor market. *Handbook of labor economics*, 3:3309–3416.

Currie, J. and Stabile, M. (2003). Socioeconomic status and child health: why is the relationship stronger for older children? *American Economic Review*, 93(5):1813–1823.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Dubow, E. F., Boxer, P., and Huesmann, L. R. (2009). Long-term effects of parents' education on children's educational and occupational success: Mediation by family interactions, child aggression, and teenage aspirations. *Merrill-Palmer quarterly (Wayne State University. Press)*, 55(3):224.

Eliason, M. and Storrie, D. (2009). Does job loss shorten life? *Journal of Human Resources*, 44(2):277–302.

Fletcher, J. (2013a). Adolescent depression and adult labor market outcomes. *Southern Economic Journal*, 80(1):26–49.

Fletcher, J. M. (2008). Adolescent depression: diagnosis, treatment, and educational attainment. *Health economics*, 17(11):1215–1235.

Fletcher, J. M. (2013b). The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior & Organization*, 89:122–135.

Flores, M. and Kalwij, A. (2019). What do wages add to the health-employment nexus? evidence from older European workers. *Oxford bulletin of economics and statistics*, 81(1):123–145.

Frijters, P., Johnston, D. W., and Shields, M. A. (2014). The effect of mental health on employment: evidence from Australian panel data. *Health economics*, 23(9):1058–1071.

Gao, W. and Smyth, R. (2010). Health human capital, height and wages in China. *The Journal of Development Studies*, 46(3):466–484.

García-Gómez, P. (2011). Institutions, health shocks and labour outcomes across Europe. *Journal of Health Economics*, 30(1):200–213.

García-Gómez, P., Jones, A. M., and Rice, N. (2010). Health effects on labour market exits and entries. *Labour Economics*, 17(1):62–76.

García-Gómez, P. and López-Nicolás, A. (2006). Health shocks, employment and income in the Spanish labour market. *Health economics*, 15(9):997–1009.

García-Gómez, P., Van Kippersluis, H., O'Donnell, O., and Van Doorslaer, E. (2013). Long-term and spillover effects of health shocks on employment and income. *Journal of Human Resources*, 48(4):873–909.

Gaulke, A. (2021). Individual and family labour market impacts of chronic diseases. *IZA World of Labour*, pages 482–482.

Gedikli, C., Miraglia, M., Connolly, S., Bryan, M., and Watson, D. (2023). The relationship between unemployment and wellbeing: an updated meta-analysis of longitudinal evidence. *European Journal of Work and Organizational Psychology*, 32(1):128–144.

Germinario, G., Amin, V., Flores, C. A., and Flores-Lagunes, A. (2022). What can we learn about the effect of mental health on labor market outcomes under weak assumptions? evidence from the nlsy79. *Labour Economics*, 79:102258.

Gibbons, C. E., Suárez Serrato, J. C., and Urbancic, M. B. (2019). Broken or fixed effects? *Journal of Econometric Methods*, 8(1):1–12.

Gilleskie, D. B., Han, E., and Norton, E. C. (2017). Disentangling the contemporaneous and dynamic effects of human and health capital on wages over the life cycle. *Review of economic dynamics*, 25:350–383.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Goryakin, Y. and Suhrcke, M. (2017). The impact of poor adult health on labor supply in the Russian Federation. *The European Journal of Health Economics*, 18:361–372.

Graff Zivin, J. and Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1):1–26.

Haveman, R. and Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of economic literature*, 33(4):1829–1878.

Heckman, J. J., Humphries, J. E., and Veramendi, G. (2018). Returns to education: The causal effects of education on earnings, health, and smoking. *Journal of Political Economy*, 126(S1):S197–S246.

Heckman, J. J. and Mosso, S. (2014). The economics of human development and social mobility. *Annu. Rev. Econ.*, 6(1):689–733.

Heinesen, E. and Kolodziejczyk, C. (2013). Effects of breast and colorectal cancer on labour market outcomes—average effects and educational gradients. *Journal of health economics*, 32(6):1028–1042.

Hoogerheide, L., Block, J. H., and Thurik, R. (2012). Family background variables as instruments for education in income regressions: A bayesian analysis. *Economics of Education Review*, 31(5):515–523.

Johar, M. and Katayama, H. (2012). Quantile regression analysis of body mass and wages. *Health economics*, 21(5):597–611.

Jones, A. M., Rice, N., and Zantomio, F. (2020). Acute health shocks and labour market outcomes: evidence from the post crash era. *Economics & Human Biology*, 36:100811.

Kassenboehmer, S. C. and Haisken-DeNew, J. P. (2009). You're fired! the causal negative effect of entry unemployment on life satisfaction. *The Economic Journal*, 119(536):448–462.

Khanam, R., Nghiem, H. S., and Connelly, L. B. (2009). Child health and the income gradient: evidence from Australia. *Journal of health economics*, 28(4):805–817.

Kiviet, J. F. (2013). Identification and inference in a simultaneous equation under alternative information sets and sampling schemes. *The Econometrics Journal*, 16(1):S24–S59.

Kiviet, J. F. (2020). Testing the impossible: Identifying exclusion restrictions. *Journal of Econometrics*, 218(2):294–316.

Lechner, M. and Vazquez-Alvarez, R. (2011). The effect of disability on labour market outcomes in Germany. *Applied Economics*, 43(4):389–412.

Lemke, R. J. and Rischall, I. C. (2003). Skill, parental income, and IV estimation of the returns to schooling. *Applied Economics Letters*, 10(5):281–286.

Lenhart, O. (2019). The effects of health shocks on labour market outcomes: evidence from UK panel data. *The European Journal of Health Economics*, 20(1):83–98.

Lundborg, P., Nilsson, A., and Rooth, D.-O. (2014). Adolescent health and adult labor market outcomes. *Journal of Health Economics*, 37:25–40.

Mani, S., Mitra, S., and Sambamoorthi, U. (2018). Dynamics in health and employment: Evidence from Indonesia. *World Development*, 104:297–309.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources*, pages 343–360.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323.

Manski, C. F. (1997). Monotone treatment response. *Econometrica: Journal of the Econometric Society*, pages 1311–1334.

Manski, C. F. and Pepper, J. V. (2000). Monotone instrumental variables: With an application to the returns to schooling. *Econometrica*, 68(4):997–1010.

Manski, C. F. and Pepper, J. V. (2009). More on monotone instrumental variables. *The Econometrics Journal*, 12(suppl_1):S200–S216.

Marcus, J. (2013). The effect of unemployment on the mental health of spouses–evidence from plant closures in Germany. *Journal of health economics*, 32(3):546–558.

Marmot, M. (2005). Social determinants of health inequalities. *The lancet*, 365(9464):1099–1104.

Minor, T. (2013). An investigation into the effect of type I and type II diabetes duration on employment and wages. *Economics & Human Biology*, 11(4):534–544.

Mitra, S. and Jones, K. (2017). The impact of recent mental health changes on employment: new evidence from longitudinal data. *Applied Economics*, 49(1):96–109.

Mossakowski, K. N. (2009). The influence of past unemployment duration on symptoms of depression among young women and men in the United States. *American journal of public health*, 99(10):1826–1832.

Mourifie, I., Henry, M., and Meango, R. (2020). Sharp bounds and testability of a roy model of stem major choices. *Journal of Political Economy*, 128(8):3220–3283.

National Institute of Health (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*, 69(3):89–95.

Neidell, M., Graff Zivin, J., Sheahan, M., Willwerth, J., Fant, C., Sarofim, M., and Martinich, J. (2021). Temperature and work: Time allocated to work under varying climate and labor market conditions. *PloS one*, 16(8):e0254224.

Nevo, A. and Rosen, A. M. (2012). Identification with imperfect instruments. *Review of Economics and Statistics*, 94(3):659–671.

Norström, F., Janlert, U., and Hammarström, A. (2017). Is unemployment in young adulthood related to self-rated health later in life? results from the Northern Swedish cohort. *BMC Public Health*, 17:1–12.

Norton, E. C. and Han, E. (2008). Genetic information, obesity, and labor market outcomes. *Health economics*, 17(9):1089–1104.

Nwosu, C. O. (2018). The relationship between employment and mental and physical health in South Africa. *Development Southern Africa*, 35(2):145–162.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Oswald, A. J., Proto, E., and Sgroi, D. (2015). Happiness and productivity. *Journal of labor economics*, 33(4):789–822.

Pintor, M. P., Fumagalli, E., and Suhrcke, M. (2024). The impact of health on labour market outcomes: a rapid systematic review. *Health Policy*, page 105057.

Pitt, M. M., Rosenzweig, M. R., and Hassan, M. N. (2021). Identifying the costs of a public health success: Arsenic well water contamination and productivity in Bangladesh. *The Review of Economic Studies*, 88(5):2479–2526.

Powdthavee, N. (2012). Jobless, friendless and broke: what happens to different areas of life before and after unemployment? *Economica*, 79(315):557–575.

Propper, C., Rigg, J., and Burgess, S. (2007). Child health: evidence on the roles of family income and maternal mental health from a UK birth cohort. *Health economics*, 16(11):1245–1269.

Reinhold, S. and Jürges, H. (2012). Parental income and child health in Germany. *Health economics*, 21(5):562–579.

Ringdal, C. and Rootjes, F. (2022). Depression and labor supply: Evidence from the Netherlands. *Economics & Human Biology*, 45:101103.

Riphahn, R. T. (1999). Income and employment effects of health shocks a test case for the German welfare state. *Journal of Population Economics*, 12:363–389.

Rosenzweig, M. R. and Wolpin, K. I. (2000). Natural "natural experiments" in economics. *Journal of Economic Literature*, 38(4):827–874.

Rosenzweig, M. R. and Zhang, J. (2009). Do population control policies induce more human capital investment? twins, birth weight and china's "one-child" policy. *The Review of Economic Studies*, 76(3):1149–1174.

Rumball-Smith, J., Barthold, D., Nandi, A., and Heymann, J. (2014). Diabetes associated with early labor-force exit: a comparison of sixteen high-income countries. *Health affairs*, 33(1):110–115.

Sabia, J. J. (2007). The effect of body weight on adolescent academic performance. *Southern economic journal*, 73(4):871–900.

Schmitz, H. (2011). Why are the unemployed in worse health? the causal effect of unemployment on health. *Labour economics*, 18(1):71–78.

Scholder, S. v. H. K., Smith, G. D., Lawlor, D. A., Propper, C., and Windmeijer, F. (2012). The effect of fat mass on educational attainment: examining the sensitivity to different identification strategies. *Economics & Human Biology*, 10(4):405–418.

Shawa, K. C., Hollingsworth, B., and Zucchelli, E. (2024). A systematic review and meta-analysis on the effects of ill health and health shocks on labour supply. *Systematic Reviews*, 13(1):52.

Shen, Y. (2023). Mental health and labor supply: Evidence from Canada. *SSM-Population Health*, 22:101414.

Stephens, M. J. and Toohey, D. (2022). The impact of health on labor market outcomes: evidence from a large-scale health experiment. *American Economic Journal: Applied Economics*, 14(3):367–399.

Sullivan, D. and Von Wachter, T. (2009). Job displacement and mortality: An analysis using administrative data. *The Quarterly Journal of Economics*, 124(3):1265–1306.

Swaminathan, S. and Lillard, L. (2000). Health and labour market outcomes: evidence from Indonesia. *paper written for Indonesia Family Life Survey, available at www. rand. org/Labour/FLS/IFLS*.

Tanaka, A. (2021). The effects of sudden health reductions on labor market outcomes: Evidence from incidence of stroke. *Health Economics*, 30(6):1480–1497.

Tefft, N. (2012). Mental health and employment: the sad story. *Economics & Human Biology*, 10(3):242–255.

University of Essex (2020). *Understanding Society: Waves 1-10, 2009-2019 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 13th Edition.* Institute for Social and Economic Research. UK Data Service.

Vaalavuo, M. (2021). The unequal impact of ill health: Earnings, employment, and mental health among breast cancer survivors in Finland. *Labour Economics*, 69:101967.

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

Appendix

Table 9. Own and father education by health scores

Full sample	$SHS = 0$	$SHS = 1$	$SHS = 2$	$SHS = 3$	$SHS = 4$
Age	52.33	50.24	48.78	47.04	40.28
Male	0.46	0.39	0.39	0.42	0.44
Qualification: Degree/Other higher degree	0.17	0.22	0.28	0.34	0.41
Qualification: A-level/GCSE etc.	0.58	0.56	0.53	0.53	0.53
Qualification: No qualification	0.25	0.22	0.19	0.13	0.07
Father didn't go to school	0.04	0.02	0.02	0.02	0.02
Father left school without qualifications	0.59	0.51	0.48	0.42	0.34
Father left school with qualifications	0.16	0.17	0.20	0.20	0.25
Father gained post school qualifications	0.18	0.23	0.24	0.26	0.26
Father gained a uni. degree or higher	0.03	0.07	0.07	0.09	0.13
Observations	142	336	1766	9294	21823

Notes: SHS, Self-reported health score.

Table 10. Estimated bounds for health scores on employment and earning (£000) (father qualifications as instrument)

	$SDC + LEI$		$SDC + LEI + MTR$		$SDC + MTR$	
Employment	LB	UB	LB	UB	LB	UB
$SHS = 0$	0.02	0.99	0.02	0.87	0.02	0.88
$SHS = 1$	0.03	0.98	0.87	0.88	0.88	0.88
$SHS = 2$	0.06	0.97	0.88	0.88	0.88	0.89
$SHS = 3$	0.25	0.93	0.88	0.90	0.89	0.90
$SHS = 4$	0.68	0.94	0.90	0.95	0.90	0.95
$ATE(1, 4)$	-0.92	0.31	-0.08	-0.02	-0.07	-0.02
$ATE(2, 4)$	-0.88	0.30	-0.07	-0.02	-0.07	-0.01
$ATE(3, 4)$	-0.70	0.26	-0.06	0.00	-0.06	0.00
Earning	LB	UB	LB	UB	LB	UB
$SHS = 0$	0.35	2.95	0.34	1.99	0.36	2.00
$SHS = 1$	0.20	3.98	1.99	1.99	2.00	2.00
$SHS = 2$	0.28	7.49	1.99	2.02	2.00	2.03
$SHS = 3$	0.64	6.42	2.02	2.23	2.03	2.17
$SHS = 4$	1.45	3.62	2.23	3.62	2.17	3.45
$ATE(1, 4)$	-3.42	2.53	-1.63	-0.23	-1.45	-0.16
$ATE(2, 4)$	-3.34	6.05	-1.63	-0.21	-1.45	-0.14
$ATE(3, 4)$	-2.98	4.97	-1.60	0.00	-1.42	0.00

Notes: SDC, Same direction of correlation; LEI, Less endogenous instrument; MTR, Monotone treatment response; LB, Lower bound (95%); UB, Upper bound (95%); ATE, Average treatment effects.

Table 11. Estimated bounds for health scores on employment and earning (£000) (Wave 3)

	$SDC + LEI$		$SDC + LEI + MTR$		$SDC + MTR$	
Employment	LB	UB	LB	UB	LB	UB
$SHS = 0$	0.02	0.97	0.02	0.86	0.02	0.87
$SHS = 1$	0.03	0.97	0.86	0.86	0.87	0.87
$SHS = 2$	0.07	0.96	0.86	0.87	0.87	0.87
$SHS = 3$	0.24	0.92	0.87	0.89	0.87	0.89
$SHS = 4$	0.67	0.94	0.89	0.95	0.89	0.95
$ATE(1, 4)$	-0.91	0.30	-0.09	-0.03	-0.08	-0.02
$ATE(2, 4)$	-0.87	0.29	-0.08	-0.02	-0.08	-0.02
$ATE(3, 4)$	-0.70	0.24	-0.07	-0.01	-0.07	0.00
Earning	LB	UB	LB	UB	LB	UB
$SHS = 0$	0.20	5.06	0.19	1.93	0.20	1.93
$SHS = 1$	0.29	7.88	1.93	1.94	1.93	1.93
$SHS = 2$	0.22	7.67	1.94	1.97	1.93	1.95
$SHS = 3$	0.57	6.54	1.97	2.30	1.95	2.20
$SHS = 4$	1.48	3.57	2.30	3.58	2.20	3.44
$ATE(1, 4)$	-3.28	6.40	-3.28	0.36	-1.65	0.27
$ATE(2, 4)$	-3.36	6.19	-3.36	0.33	-1.64	0.25
$ATE(3, 4)$	-3.00	5.06	-3.00	0.00	-1.61	0.00

Notes: SDC, Same direction of correlation; LEI, Less endogenous instrument; MTR, Monotone treatment response; LB, Lower bound (95%); UB, Upper bound (95%); ATE, Average treatment effects.

TABLE 12. Estimated bounds for health scores on employment and earning (£000) (blood sample, Wave 4 only)

	$SDC + LEI$		$SDC + LEI + MTR$		$SDC + MTR$	
Employment	LB	UB	LB	UB	LB	UB
$BHS = 0$	0.19	0.98	0.19	0.94	0.19	0.94
$BHS = 1$	0.20	0.97	0.94	0.96	0.94	0.97
$BHS = 2$	0.28	0.98	0.96	0.98	0.97	0.99
$BHS = 3$	0.27	0.98	0.98	0.99	0.99	1.00
$BHS = 4$	0.13	0.98	0.99	1.00	1.00	1.01
$ATE(1, 4)$	-0.78	0.84	-0.06	-0.02	-0.06	-0.02
$ATE(2, 4)$	-0.70	0.85	-0.04	-0.01	-0.04	-0.01
$ATE(3, 4)$	-0.72	0.85	-0.02	0.00	-0.02	0.00
$ATE(1, 3)$	-0.78	0.70	-0.05	-0.01	-0.05	-0.01
$ATE(2, 3)$	-0.70	0.71	-0.02	0.00	-0.02	0.00
Earnings	LB	UB	LB	UB	LB	UB
$BHS = 0$	0.54	5.47	0.53	1.84	0.55	1.85
$BHS = 1$	0.53	5.26	1.84	2.71	1.85	2.52
$BHS = 2$	0.78	5.49	2.71	3.92	2.52	3.53
$BHS = 3$	0.72	5.22	3.92	4.90	3.53	4.65
$BHS = 4$	0.51	5.19	4.90	5.20	4.65	5.21
$ATE(1, 4)$	-4.66	4.75	-3.36	-2.19	-3.35	-2.13
$ATE(2, 4)$	-4.41	4.98	-2.49	-0.97	-2.69	-1.12
$ATE(3, 4)$	-4.48	4.71	-1.27	0.00	-1.67	0.00
$ATE(1, 3)$	-4.69	4.54	-3.06	-1.22	-2.80	-1.01
$ATE(2, 3)$	-4.44	4.78	-2.19	0.00	-2.13	0.00

Notes: SDC, Same direction of correlation; LEI, Less endogenous instrument; MTR, Monotone treatment response; LB, Lower bound (95%); UB, Upper bound (95%); ATE, Average treatment effects.

Table 13. Estimated bounds for health scores on employment and earning (£000) (women only)

	$SDC + LEI$		$SDC + LEI + MTR$		$SDC + MTR$	
Employment	LB	UB	LB	UB	LB	UB
$SHS = 0$	0.02	0.98	0.02	0.88	0.02	0.89
$SHS = 1$	0.02	0.98	0.88	0.88	0.89	0.89
$SHS = 2$	0.06	0.97	0.88	0.89	0.89	0.90
$SHS = 3$	0.26	0.93	0.89	0.91	0.90	0.92
$SHS = 4$	0.65	0.95	0.91	0.95	0.92	0.96
$ATE(1, 4)$	-0.93	0.33	-0.08	-0.03	-0.08	-0.02
$ATE(2, 4)$	-0.89	0.32	-0.07	-0.02	-0.07	-0.02
$ATE(3, 4)$	-0.70	0.28	-0.07	0.00	-0.07	0.00
Earning	LB	UB	LB	UB	LB	UB
$SHS = 0$	0.35	1.98	0.35	1.65	0.35	1.72
$SHS = 1$	0.20	2.84	1.65	1.65	1.72	1.71
$SHS = 2$	0.26	7.38	1.65	1.69	1.71	1.75
$SHS = 3$	0.52	6.41	1.69	1.92	1.75	1.88
$SHS = 4$	1.21	3.49	1.92	3.49	1.88	3.31
$ATE(1, 4)$	-3.29	1.63	-1.84	-0.27	-1.60	-0.17
$ATE(2, 4)$	-3.23	6.17	-1.84	-0.22	-1.60	-0.14
$ATE(3, 4)$	-2.97	5.20	-1.80	0.00	-1.57	0.00

Notes: SDC, Same direction of correlation; LEI, Less endogenous instrument; MTR, Monotone treatment response; LB, Lower bound (95%); UB, Upper bound (95%); ATE, Average treatment effects.