

# Estimating the causal effect of liability to disease on healthcare costs using Mendelian Randomization

Padraig Dixon<sup>a,b,\*</sup>, Sean Harrison<sup>b,c</sup>, William Hollingworth<sup>c</sup>, Neil M. Davies<sup>b,c,d</sup>, George Davey Smith<sup>b,c,e</sup>

<sup>a</sup> Nuffield Department of Primary Care Health Sciences, University of Oxford, United Kingdom

<sup>b</sup> MRC Integrative Epidemiology Unit, University of Bristol, United Kingdom

<sup>c</sup> Population Health Sciences, University of Bristol, United Kingdom

<sup>d</sup> K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway

<sup>e</sup> NIHR Biomedical Research Centre, University of Bristol, United Kingdom

## ARTICLE INFO

### JEL Classification:

H51

I10

I11

I18

### Keywords:

Genetics

Disease liability

Instrumental variables

Healthcare costs

Mendelian randomization

## ABSTRACT

Accurate measurement of the effects of disease status on healthcare costs is important in the pragmatic evaluation of interventions but is complicated by endogeneity bias. Mendelian Randomization, the use of random perturbations in germline genetic variation as instrumental variables, can avoid these limitations. We used a novel Mendelian Randomization analysis to model the causal impact on inpatient hospital costs of liability to six prevalent diseases and health conditions: asthma, eczema, migraine, coronary heart disease, Type 2 diabetes, and depression. We identified genetic variants from replicated genome-wide associations studies and estimated their association with inpatient hospital costs on over 300,000 individuals. There was concordance of findings across varieties of sensitivity analyses, including stratification by sex and methods robust to violations of the exclusion restriction. Results overall were imprecise and we could not rule out large effects of liability to disease on healthcare costs. In particular, genetic liability to coronary heart disease had substantial impacts on costs.

## 1. Introduction

This paper proposes and evaluates a new way to estimate healthcare costs associated with liability to specific diseases. Liability to disease is influenced by a wide range of factors including environmental, social, economic and biological processes. This complicates attempts to attribute economic outcomes such as healthcare costs to specific diseases, since measurement error in diagnosis and omitted confounding variables of the association between disease status and healthcare can lead to spurious results. This pervasive measurement error and endogeneity is particularly problematic for the pragmatic evaluation of medical public health and clinical interventions, which generally require robust estimates of the long-term causal effect of disease status on healthcare costs (Briggs et al., 2006; Dixon et al., 2016; Franklin et al., 2019; Mihaylova et al., 2011).

For example, taxes on “sin goods” such as alcohol, cigarettes or sugary beverages are often motivated (amongst other policy rationales) by an anticipated beneficial impact on future disease incidence, mortality and healthcare costs (Cawley and Ruhm, 2011; Allcott et al., 2019;

O’Donoghue and Rabin, 2006). Even where robust randomized interventional study designs have been implemented, decision-analytic models are often needed to assess the long-term consequences of policies and interventions (Sheldon, 1996; NICE, 2013; Brennan and Akehurst, 2000; Sculpher et al., 2005, 2006; Buxton et al., 1997; Siebert, 2003; Dakin et al., 2020), and will typically assess future costs and other outcomes under different health states (Briggs et al., 2006). If the estimated costs of, for example, coronary heart disease, are inaccurately estimated then robust comparisons of different types of interventions aimed at modifying liability to this disease will be compromised.

### 1.1. Rationale and overview of methods

Accurately estimating the costs of ill health is therefore essential for cost-benefit and cost-effectiveness analysis in health economics. Here we develop and evaluate a new approach to overcome endogeneity and measurement error in most or all existing analyses of the long-term impact of disease on healthcare costs. Our methods may improve the evaluation of healthcare policies with long-term consequences for

\* Corresponding author at: Nuffield Department of Primary Care Health Sciences, University of Oxford, United Kingdom.

E-mail address: [Padraig.dixon@phc.ox.ac.uk](mailto:Padraig.dixon@phc.ox.ac.uk) (P. Dixon).

<https://doi.org/10.1016/j.ehb.2022.101154>

Received 29 October 2021; Received in revised form 15 June 2022; Accepted 28 June 2022

Available online 30 June 2022

1570-677X/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

healthcare costs. To this end, we report a novel analysis of the causal effect of liability to disease on healthcare costs using Mendelian Randomization (Davey Smith and Ebrahim, 2003; Sanderson et al., 2022) which can, in principle, overcome the limitations of existing research designs.

Mendelian Randomization can use genetic variation as instrumental variables to estimate causal effects. Some genetic variation is known to associate with diseases, behaviours and traits. Those variants known to associate with disease are promising candidates for instrumental variable analysis because their conditionally random assignment from parents to offspring at conception means that they may be independent of omitted variables that might otherwise confound associations between diseases and their outcomes, including health care costs. Furthermore, estimates using genetic variation are unlikely to be affected by reverse causation since it is generally not possible for the environment to affect germline genetic variation after conception.

We study the costs of genetic liability to six prevalent, chronic health conditions: asthma, eczema, migraine, coronary heart disease, type 2 diabetes, and depression. We refer interchangeably to disease, conditions and health conditions below. These six conditions were obtained following a review of health conditions that were associated with at least 100 disability-adjusted life years lost per 100,000 adults in the UK, which had a prevalence of at least 2% in the UK Biobank cohort (our primary data source in this paper), and which had at least three genome-wide significant genetic variants identified in genome wide association studies. The prevalence criterion was used to maximize statistical power, and the restriction to conditions with some degree of known genetic influence was necessary to implement Mendelian Randomization analysis (Harrison et al., 2020).

Our goals were to assess the feasibility of studying the costs of illness using Mendelian Randomization, and to develop and interpret new estimates of the cost of liability to disease using these methods. We note that these new estimates of the impact of liability to disease on healthcare costs are best used as inputs into evaluations of interventions and healthcare policies, rather than as estimates of the “cost of illness” that are sometimes used to justify interventions in a specific disease area that is associated with high costs. This has long (Shiell et al., 1987) been recognized as a fallacy based on circular reasoning – the fact that high costs are associated with a particular disease does not, in itself, necessarily merit further spending or relatively higher greater attention by healthcare funders or policy makers. Instead, estimates of how healthcare cost change in relation to liability to disease could be used in modelling the cost-effectiveness of interventions targeting disease incidence, in policy evaluation in relation to disease, the pricing of insurance, and in accounting for externalities related to disease. For example, if the costs of incident disease are higher than previously estimated, the cost-effectiveness of interventions that reduce incidence is likely to increase, other things being equal.

## 1.2. Outline of paper

Below, we describe in more detail each assumption necessary for valid causal inference in the specific context of genetic variants as instrumental variables. We apply these methods to cost data drawn from UK Biobank (Collins, 2012; Sudlow, 2015) a very large and richly phenotyped prospective cohort study. We compare the findings of the instrumental variable analysis to conventional multivariable adjusted analyses of the association between disease status and hospital costs. We conclude with a discussion of the policy relevance of these new Mendelian Randomization estimates.

## 2. Methods – Genetic variants as instrumental variables

### 2.1. Instrumental variable assumptions and Mendelian Randomization

Econometric identification in Mendelian Randomization relies on

random perturbations to germline genetic variation that occur at conception. Germline genetic variation refers to genetic variation that is passed from parents to their children. This excludes genetic variation that is not passed from parent to child. For example, somatic mutations, a common cause of disease including cancer, occur after conception in cells other than the germ cells (the sperm and egg cells) and therefore cannot be passed to children.

Points of the genome that vary across the population are known as genetic variants. A common form of genetic variation is a single nucleotide polymorphism (SNP) – a change in a single base pair of DNA. At each SNP, everyone has two alleles, one allele that was inherited from the father and one that was inherited from the mother. There is a 50:50 chance of each of the father's alleles being inherited by the offspring and a 50:50 chance of each of the mother's alleles being inherited. Therefore, the inheritance of genetic variation from parents to offspring is random, conditional on parental genotype.

Instrumental variables for disease liability are defined by the following assumptions: 1) they associate with liability to the disease of interest, 2) they are independent of all (known and unknown) confounding omitted variables, and 3) they influence the outcome only via disease liability. We unpack the assumptions for valid instrumental variable analysis requirements below for genetic instrumental variables and also illustrate the requirements for valid inference using directed acyclic graphs in supplementary material. Below, we refer to disease status as the treatment variable, and interpret “odds of disease” as a genetic liability (Davey Smith, 2019; Tudball et al., 2021) toward the specific disease phenotype modelled.

#### 2.1.1. Relevance

Valid instruments must be associated with liability to disease. This is sometimes known as the relevance requirement. This assumes that the expected value of the treatment variable (binary disease status) differs when an instrument has a higher rather than a lower value. On the liability scale, the degree of liability to disease differs across values of the instrument.

Instrument relevance is best assessed in this context by using genome wide association studies (GWASs). These studies examine associations between millions of genetic variants across the genome and phenotypes such as disease status. Humans are diploid, meaning that they have two copies of each chromosome. For a single variant located at a specific part or locus of the genome, an individual may possess no copies of the risk increasing allele, one copy of the risk increasing allele (only on one chromosome) or two copies of the risk increasing allele (on both chromosomes).

The specific type of variants that we study are single nucleotide polymorphisms (SNPs), which refer to single changes in a nucleotide base pair (adenine (A), thymine (T), cytosine (C), or guanine (G)) at a specific point in the genome across individuals in a population. We use the terms variants and SNPs interchangeably below. For a single variant, there may be two or more alleles (genotypes). Our analysis is restricted to genetic variants that only have two common alleles in the sample. The reference allele can be defined as the one that confers the lowest disease risk. People with one or two copies of the other risk increasing allele, have a higher risk of, or liability to, disease. We can summarize the genotype – the genetic architecture at a particular locus – by simply counting the number of risk increasing alleles (0,1, or 2) for a disease.

Our measure of disease status is binary. A GWAS on a binary outcome such as disease status involves using independent logistic regressions to estimate each of  $J$  SNPs association with disease. The p-values obtained for each measured association are corrected for multiple testing, reflecting the number of independent genetic variants across the genome. We treated p-values  $< 5 \times 10^{-8}$  as “genome-wide significant” for the purposes of our analysis.

In our main analysis, we use only genetic variants obtained from replicated external GWAS and meet this stringent multiple testing criterion. This reduces the risk of selecting variants that fail the relevance

criterion, a problem that affected early examples of Mendelian Randomization in health economics, which sometimes relied on variants drawn from unreplicated “candidate gene” studies (Fletcher, 2018, 2011).

### 2.1.2. Independence

The independence assumption requires the instrument to be independent of potential outcomes conditional on disease status. An intuitive interpretation of this assumption is that the instrument is “as good as randomly assigned”. Germline genetic variation is set at conception, which occurs before post-natal events and life circumstances. However, while the inheritance of genetic variants from parents to offspring is almost entirely random, the distribution of genetic variants across parents may not be independent of the environment. One means by which the independence assumption may be violated is due to differential ancestry within a population (Tian et al., 2008). Genetic variants in samples of individuals with different ancestral background are likely to have differences in allele frequencies. These differences will spuriously associate with any phenotypic differences between the ancestral groups. For example, allele frequencies and genetic differences in susceptibility to Type 2 diabetes vary by ethnicity (Keaton et al., 2014; Carulli et al., 2005; Elbein, 2009; Cook and Morris, 2016; Fuchsberger et al., 2016). This mixing of ancestries would lead to spurious associations due to differences between groups that differ in ancestry, rather than the effect only of the genetic variants under investigation.

We minimize this problem by restricting analysis to individuals of European ancestry groups (by means of self-reported ethnicity and genetic ancestry as indicated by the genetic principal components, in addition we control for the genetic principal components (Mathieson and McVean, 2012) (Price et al., 2006)) and by conditioning on study centre. Latent population structure may remain even after these steps (Haworth et al., 2019; Lawson et al., 2020) and we consider possible implications of this in the Discussion section below.

Assortative mating (Hartwig et al., 2018) and selection into genetic studies (Hughes et al., 2019) may induce associations between the genetic variants and the treatment variable and outcome that violate the independence assumption. Assortative mating refers to non-random mating; that is, the selection of partners based on particular traits. This may lead to the clustering of particular types of individuals in particular environments; for example, educated people are more likely to marry other educated people. In relation to selection bias, note that conditioning on a consequence (or a collider in the language of directed acyclic graphs) of both the instrument and the outcome can induce an association between the instrument and the outcome even if these are otherwise independent in the population (Gkatzionis and Burgess, 2018; Rohrer, 2018). This can lead to bias that can over- or under-state effect sizes (Munafò et al., 2017). Studies necessarily condition on available participants, and if this set of individuals is selected rather than drawn randomly from the population then instrumental variable effect estimates may be biased. Again, we consider the implications of these kinds of potential biases in the Discussion.

### 2.1.3. The exclusion restriction

The third assumption is that of exclusion: this requires that the instrument is independent of the outcome (and all potential outcomes) and does not affect the outcome other than via disease status. Genetic variants should therefore be independent of the outcome, conditional on treatment. Establishing this independence is generally impossible since some omitted variables will not be known or measurable, and disease status may not be measured perfectly.

The two principal means by which genetic variants can violate this assumption are linkage disequilibrium and pleiotropy (Davey Smith and Hemani, 2014). Linkage disequilibrium refers to a correlation between genetic variants that arises when variants located in close physical proximity to one another tend to be inherited together. Since variants in linkage disequilibrium are not independent, it is possible that correlated

variants may influence outcomes other than through the disease of interest in violation of the exclusion restriction.

Pleiotropy describes a situation where a variant affects multiple phenotypes (Davey Smith and Hemani, 2014; Paaby and Rockman, 2013; Stearns, 2010). There are two types of pleiotropy: vertical and horizontal (Hemani et al., 2018a). Vertical pleiotropy refers a SNP that influences one trait, which in turn influences another. Horizontal pleiotropy refers to SNPs that influence traits through independent pathways. Horizontal pleiotropy will violate the exclusion restriction if these other phenotypes affect the outcome of interest other than via the disease of interest.

Consider a variant associated with coronary heart disease that may also influence some other disease (e.g. depression). An analyst using this variant may unwittingly attribute all the effect of coronary heart disease on cost to that variant, when in truth the reported effect captures the influence of both heart disease and depression on the cost outcome. This amounts to a violation of the exclusion restriction. We consider various methods for modelling and overcoming these violations of the exclusion restriction in the next section.

### 2.1.4. Estimation

The two-stage least squares (2SLS) instrumental variable estimator predicts disease status (in this case) from a regression of disease status on genetic variants; the second stage involves a regression of the outcome on these predicted values. We implemented just-identified models where we have one instrument – an allele score or polygenic risk score – using 2SLS, as well as overidentified models which estimate the causal effect using each variant independently. The over-identified models allow for a more robust interrogation of the exclusion restriction and the potential role of pleiotropic genetic variants than is possible using only a 2SLS just-identified model. We discuss these issues below in relation to our various sensitivity analyses.

We point identify our estimates by assuming that the instrument has a monotonic effect on the treatment variable (Angrist and Pischke, 2009; Angrist et al., 1996). This assumes that, in a hypothetical experiment, replacing an allele with no effect on disease liability with a liability-increasing allele would either increase liability or leave it unchanged. This implies that, across our study population, the direction of effect from modulating the value of our instrumental variables will be in the same direction for everyone. Monotonicity may have biological plausibility for genetic instruments (Burgess and Thompson, 2015), but this is generally difficult or impossible to test. Under monotonicity, our estimates reflect local average treatment effects – these are the effects of increases to disease liability in people whose liability to disease is altered by the instrument. This group of individuals may comprise everyone in our sample if liability to disease is continuous.

We also examine whether our instruments are weak. A weak instrument explains only a small portion of the variance in disease status. 2SLS is a form of Wald estimator, i.e. a ratio of the costs-SNP and disease-SNP associations. Intuitively, when the denominator of this ratio estimator is close to zero, the ratio becomes unstable and its variance becomes large. We examined first stage F-statistics from 2SLS allele score models to assess whether weak instruments were likely to be problematic (Stock and Yogo, 2002).

## 2.2. Sensitivity analysis

Our main sensitivity analyses comprised various tests designed to assess possible consequences arising from violations of the exclusion restriction. Sensitivity analysis directed at potential violations of the exclusion restriction relies on techniques influenced by meta-analysis (Bowden and Holmes, 2019). The motivation for this stems both from the independence of variants as well as from interpreting conditional random allocation of genetic variation at conception as analogous to a within-family randomized controlled trial in which the individual is allocated to higher or lower susceptibility to disease.

When many such variants are available, we may proceed, as in a conventional meta-analysis, to summarize their effects on the outcome in a manner that is at least as informative as the scrutiny of any one variant. We can do this because the variants we use are independent across the genome (i.e. they are not in linkage disequilibrium). A variety of these techniques are applied, each with different assumptions. Similarity or concordance between the results despite different assumptions provides a degree of reassurance that the same causal effect is being identified, and that gross biases specific to one form of analysis are not influencing the results.

The most common estimator is the inverse variance weighted (IVW) random-effects meta-analysis. This weights the effect sizes (i.e. the regression betas) by the inverse of the variance of the estimated associations between the variants and the outcome. The effects are estimated by the ratio of the SNP-costs association and the SNP-disease association for each SNP. The exclusion restriction requires that there should be no effect on the outcome for any SNP that is not mediated via the disease status.

A random-effects meta-analysis allows some or all variants to have pleiotropic effects in violation of the exclusion restriction, but assumes that the effect of this pleiotropy “balances out” so that pleiotropy that increases the causal effect estimate is matched by pleiotropy that reduces the causal effect estimate. Thus, the effect size is estimated without bias, although with a greater standard error than would be the case without variation induced by pleiotropy.

A natural test of the assumptions of this model (embodying the assumption of “no directional pleiotropy”) is to compare this IVW effect estimate with the effect estimate for each individual variant. If this difference is large (in a sense defined below), then heterogeneity of effect may be present. Heterogeneity may indicate horizontal pleiotropy since one mechanism to generate heterogeneity of this type would be the influence of a variant on the outcome through multiple independent channels.

This can be tested formally using Cochran’s Q statistic, which measures heterogeneity by comparing the squared difference to the critical values of a chi-squared distribution:

$$Q = \sum_{j=1}^J \frac{1}{\sigma_{\hat{\beta}_j}^2} (\hat{\beta}_j - \hat{\beta}_{IVW})^2$$

Here, we assume up to  $J$  variants, and measure effect estimates for the  $j$ th variant as  $\hat{\beta}_j$  and the overall inverse variance weighted estimate as  $\hat{\beta}_{IVW}$ , and measure the inverse variance weighting of the estimate using  $\sigma_{\hat{\beta}_j}^2$  which is the variance from the SNP-outcome association.

The intuition for the Q statistic is as follows. If one instrument implies a very different effect of the exposure on the outcome than all others (i.e. there is heterogeneity in the implied effects across instruments), then one explanation is that this instrument may be mediated via different mechanisms than the others. For example, a particular SNP may affect both the disease of interest and also a separate disease outcome (through a channel that is independent of the disease of interest) and this has an extremely large (relative to other SNPs) effect on the cost outcome because it captures the effect of not one but two health conditions on healthcare costs. In this case, our empirical model would attribute all of the influence of that instrument to disease of interest, when in reality it is a consequence of both a direct effect of this disease as well as a pleiotropic effect via the other disease. This violates the exclusion restriction.

The Q statistic is equivalent to the Sargan test for overidentification (Sargan, 1958) when all relationships (SNP-treatment and SNP-outcome) are measured in the same sample (Greco M et al., 2015). The Q statistic does not identify whether any heterogeneity is due to one, some or all IVs being invalid. It is therefore a “first step” in identifying potential violations of the exclusion restriction that may be caused by pleiotropy or other violations of the instrumental variable assumptions.

We estimated the Q statistic and implemented pleiotropy-robust methods for all diseases. These robust methods were of two broad types – the first type of sensitivity analysis assumes that some or all variants violate the exclusion restriction but are nevertheless retained in the analysis. The second type of approach comprises a principled approach to variant outlier detection and removal, and analysis may be applied to a smaller, restricted set of variants. We consider each approach in turn; supplementary material provides an intuitive explanation using simple graphical examples.

We start with estimators of the first type. Median-based estimators (Bowden et al., 2016) assume that the median SNP estimate, is unbiased (i.e. has no horizontally pleiotropic effects). This estimator gives no weight to SNPs at the extremes of the distribution, and thus excludes potentially pleiotropic variants. Thus, if at least half of the weighted variants do not have horizontally pleiotropic effects, this method will be asymptotically unbiased. We implement a penalized weighted median estimator, in which variant-specific effects are weighted by the precision with which they are estimated and penalized (or “down-weighted”) for contributing heterogeneity to the Q statistic.

A related method makes an assumption that modal estimate from the individual variants does not violate the exclusion restriction (Hartwig et al., 2017). This permits some SNPs, possibly most SNPs, to be pleiotropic and therefore in violation of the exclusion restriction, provided that the modal variant is, or modal set of variants are, unbiased. Again, we implement a weighted version of this estimator to account for the precision with which variants are estimated, and we also define a bandwidth parameter to identify the modal group.

We also implement MR Egger regression (Bowden et al., 2015). The IVW estimator constrains the intercept of a regression of the SNP-costs associations and SNP-disease liability associations to be zero so that the regression line passes through the origin. MR Egger does not constrain this intercept to be zero. Instead, it is included and can be interpreted as the average pleiotropic effect from all variants. An intercept that is distinguishable from zero indicates the presence of pleiotropy, and the causal effect estimate from MR Egger regression can be unbiased depending on the following assumptions. MR-Egger assumes independence between the direct pleiotropic effect of variants on costs (other than through liability to disease) and association with the liability to disease – for further details see Bowden et al (Bowden et al., 2015). and Burgess and Thompson (Burgess and Thompson, 2017).

We also implement an approach, known as Radial Mendelian Randomization (Bowden et al., 2018), to automate the identification of outlying variants and analyze their impact on the estimators described above. Radial MR plots the SNP-outcome association (y-axis) against the SNP-disease status association (x-axis). The absolute vertical distance between a variant represented by a scatter point and the fitted IVW line is equivalent to the square root of that variant’s contribution to Cochran’s Q heterogeneity statistic. This distance can therefore be used to identify outlying variants that contribute relatively high amounts of heterogeneity by using a leave-one-out analysis of all variants. We implement all estimators described above on exclusion of variants identified by this process and compare results to models using the full set of variants.

Finally, to complement the two-sample approach described above, we also performed a one-sample analysis using data only from UK Biobank. One concern with two-sample estimators (where SNP-outcome and SNP-exposure associations are estimated in different samples) is that the populations in each sample should not overlap (that is, not have the same individuals) to avoid inducing a correlation between SNP-disease and SNP-cost associations and biasing estimates towards the disease-cost association. However, the two samples should be drawn from the same population, even if the set of individuals in each sample is disjoint. The rationale for this is to avoid bias caused by associations estimated on samples that differ in the types of individuals included. Similarity may be assessed in relation to distributions of age and sex, similarity of association between SNPs and the phenotype, and the mean



and variance of the phenotype distribution (Haycock et al., 2016).

We split UK Biobank into two random, non-overlapping samples for this sensitivity analysis. This approach eliminates any biases caused by sample heterogeneity between the UK Biobank cohort and external GWAS source data. We conducted a de novo GWAS for each of the six diseases on each split sample. The result of each GWAS was used to identify SNPs for the alternative sample. We estimated just-identified allele score models using each sample, and then used fixed effect meta-analysis on these separate results to give a single estimate.

### 2.3. Issues of interpretation

The definition of disease reflects a binary classification of individuals as cases (those identified as having the condition) and controls (those without). The interpretation of our results differs according to the type of instrumental variable models used – those that use a single allele score and those that use each SNP as individual instruments.

The interpretation of the allele score instrumental variable models reflect the genetic liability of changing from control to case across the population. Put differently, these models may be interpreted as the average per person change in hospital costs caused by a (genetically influenced) change in liability to case status in this population. These models are estimated using 2SLS. Denoting individuals  $i = 1 \dots N$ ,  $k = 1 \dots K$  instruments ( $G = g_1, \dots, g_k$ ) and disease  $X = x_1, \dots, x_6$  the first stage of a 2SLS model is:

$$x_i = \alpha_0 + \sum_k \alpha_k g_{ik} + \varepsilon_{xi}$$

This yields predicted values of disease status:  $\hat{x}_i = \hat{\alpha}_0 + \sum_k \hat{\alpha}_k g_{ik}$ , which enter the second stage regression for outcome  $Y = y_1, \dots, y_i$ :

$$y_i = \beta_0 + \beta_1 \hat{x}_i + \varepsilon_{yi}$$

The  $\beta_1$  coefficient represents the estimated effect of a change in disease status.

The interpretation of the models that use summary data is different. An intuitive way of understanding this difference is to consider a ratio (Wald) estimator for instrumental variables. The instrumental variable point estimate may be obtained as the change in the outcome (Y) for a unit increase in the instrument (G), scaled by the change in disease status (as the treatment variable) for a unit increase in G:

$$\frac{\hat{\beta}_{Y|G}}{\hat{\beta}_{X|G}}$$

The interpretation of Mendelian Randomization models using each SNP as an individual instrument therefore reflects the measurement scale used in the source GWAS, in contrast to the just-identified allele score models estimated using 2SLS. Since the GWASs we used analyzed outcomes using logistic regression, our over-identified models using each SNP as an instrument therefore measure the ratio of the effect of the change in costs per unit change in the log-odds of disease status.

This type of association can be difficult to interpret in terms of causal effects since a unit change in a binary disease status variable will be on the log-odds scale. A unit change is therefore  $\exp(1)$ , corresponding to a 2.72-fold (since the scale is multiplicative) change in the odds of the treatment variable. Following Burgess and Labrecque (Burgess and Labrecque, 2018), we re-express associations estimated from models relying on individual variant data to aid interpretation. We interpret “odds of the treatment variable” as a genetic liability (Davey Smith, 2019) toward the specific disease phenotype modelled. We calculate causal effects in terms of a doubling in genetic liability ( $\ln 2$ ) and a 10 % increase ( $\ln 1.1$ ). In the main text we report estimates according to a doubling in genetic liability; inverse variance weighted estimates of the effect of a 10 % relative increase are reported in the supplementary material.

Estimates on the relative scale may therefore be interpreted as, for

example, the causal effect on inpatient hospital costs of a doubling in genetic liability to coronary heart disease. Note that this is a relative rather than absolute increase, where the relative increase can be understood as an increase from possessing some risk-increasing alleles to possessing more risk-increasing alleles.

Finally, we note that the focus of our analysis is the mean conditional effect of disease status on healthcare costs in the conventional multi-variable models, and the mean conditional effect of liability to disease on healthcare costs in the instrumental variable models. Instrumental variable estimates can be interpreted as either an average treatment effect, assuming a constant effect of liability to disease on costs, or a local average treatment effect, assuming a monotonic effect of the genetic variants on disease liability (see for example Zhao et al (Qingyuan et al., 2020).), including where the outcome variable includes some proportion of zeroes, as is the case with the present analysis (see Results section below).

### 3. Data

Our principal data source is the UK Biobank, a large population-based prospective cohort that recruited over 500,000 adults aged between 39 and 71 years over the period 2006–2010 (13–15). Participants provided data on demographic features such as date of birth, sex as well as a wide variety of phenotypic data. Approximately 9 million individuals registered with the National Health Service were invited to participate (Sudlow et al., 2015; Allen et al., 2012). The participating cohort of over 500,000 individuals is generally “healthier and wealthier” (Sudlow et al., 2015; Hughes et al., 2019; Fry et al., 2017) than the general UK population. We discuss some of the consequences of this feature of the UK Biobank cohort below.

Patients were categorized as having a health condition if they self-reported the condition at the UK Biobank baseline recruitment visit, or if there had the corresponding ICD-9 or ICD-10 code reported in linked healthcare records, comprising both the hospital episode statistics database and the national cancer registry. This represents all healthcare costs associated with inpatient care in National Health Service (NHS) hospitals in England and Wales, including for private patients treated in NHS hospitals.

The diabetes cases were only those known to have Type 2 rather than Type 1 diabetes (Harrison et al., 2020). The depression cases were defined (as in Tyrrell et al (Tyrrell et al., 2018). and in (Harrison et al., 2020)) as participants who self-reported seeing a doctor for nerves, anxiety or depression for a duration of at least two weeks, or who had ICD-9 or ICD-10 codes for depression. Only ten recruitment centres asked questions related to depression, and this reduced the size of the available sample (see Results below) relative to the other conditions.

We studied non-cancer diseases with known genetic determinants, for which prevalence was at least 2% in the UK Biobank, and which accounted for at least 100 disability-adjusted life years lost per 100,000 adults in the UK (Murray et al., 2013). Diseases with less than three genome-wide significant SNPs were excluded from consideration, which resulted in the exclusion of osteoarthritis. The final set of diseases resulted in analysis sample comprising the following six diseases: asthma, eczema, migraine, coronary heart disease, type 2 diabetes and depression.

The cost measure we use as our outcome variable is inpatient hospital costs per person per year. Costs were calculated for episodes of care beginning on or after April 2006. These costs were followed up until March 2015. Participants could therefore be followed up for a maximum of 9 years (if recruited in April 2006). Median follow-up, accounting for staggered recruitment, was 6.1 years. Further details on the creation of the cost variable are available in (Dixon et al., 2018, 2020). Most cohort participants (54%) reported some inpatient hospital care and therefore had non-zero costs. Those reporting zero costs were retained in the analysis.

We examined previous GWAS for data on variants associated with

disease at genome-wide significance ( $P \leq 5 \times 10^{-8}$ ). The source GWAS data for each condition was as follows: asthma (Moffatt et al., 2010), coronary heart disease (Schunkert et al., 2011), major depressive disorder (Wray et al., 2018), type 2 diabetes (Morris et al. 2012), eczema (Paternoster et al., 2015), migraine (Gormley et al., 2016)). The Wray et al. (2018). GWAS included individuals from the pilot sample of UK Biobank, and we excluded these individuals in the two-sample analysis to avoid overlap between the GWAS and analysis samples.

We cross-checked genome-wide significant SNPs against those measured in the UK Biobank cohort. In some cases we did not locate in UK Biobank the variant used in the external GWAS – the difference between the precise variant used typically arises because of the different technologies used to genotype different cohorts, and the areas of the genome searched. Where the genetic variant was not available in UK Biobank, we used a substitute variant with a high ( $R^2 > 0.6$ ) degree of correlation with the missing variant, or if no substitute could be found, we excluded the missing variant.

We also created allele scores for each condition, sometimes known as polygenic risk scores. The number of disease-increasing alleles was counted for each individual, then multiplied by the effect size, determined in the respective GWAS, for each allele a weighted per-allele effect was summed. This gives a weighted allele score, where the weights reflect the effect size of each allele.

To avoid bias induced by differential ancestry, we condition on sex, age and the first forty genetic principal components. The analysis sample was restricted to individuals of white British or closely related ancestry (determined by self-report and by examination of genetic principal components) to avoid bias from differential ancestry. Analyses were conducted in R (including with the MR Base package (Hemani et al., 2018b)) and in Stata version 15.1 (StataCorp, College Station, Texas). Analysis code is available at [www.github.com/pdixon-econ](http://www.github.com/pdixon-econ).

## 4. Results

### 4.1. Introduction

Up to 307,032 individuals were available for analysis, of whom 54% were female ( $n = 164,985$ ). Mean age at recruitment in this sample was 57 years ( $SD = 8.0$  years). The actual numbers available for the analysis of each specific condition were slightly lower than the overall total given the completeness of data on self-report of disease history at baseline and information available to identify cases in linked hospital records.

Multivariable models, estimated using linear regression and adjusted for age, sex and study centre but without any genetic information, indicated substantial impacts on inpatient hospital costs for most diseases. This is summarized in Table 1. The coefficients represent the association of disease status with hospital costs.

With the exception of eczema, all disease phenotypes are associated with material impacts on annual per-patient inpatient hospital costs. We emphasize that these models are minimally adjusted (accounting only for age, sex and centre) but the scale of some of these effects

**Table 1**

Multivariable estimates of association between disease status and annual inpatient hospital costs.

	Effect estimate for change in disease status	95 % confidence interval
<b>Disease phenotype</b>		
Asthma	£ 137	£ 125 to £ 149
Eczema	£ 5	-£ 21 to £ 30
Migraine	£ 45	£ 21 to £ 68
Coronary heart disease	£ 496	£ 477 to £ 516
Type 2 diabetes	£ 327	£ 308 to £ 347
Depression	£ 142	£ 122 to £ 162

Note to table: These models adjust for age, sex and recruitment centre only.

(particularly for coronary heart disease) is large relative to median (£88) and mean (£479) annual per-patient inpatient hospital costs.

These estimates are likely to be confounded in various ways, particularly by measurement error and by omitted variables that influence both hospital costs and disease status. This may underestimate the costs. This possibility of confounding constitutes our motivation for considering more robust causal methods using genetic liability to disease, to which we now turn.

We begin by considering the proportion of variance in the disease phenotype explained by respective allele scores (measured by pseudo-R-squared statistics obtained from unconditional logistic regressions) summarized in Table 2.

Table 2 makes clear that the proportion of variance explained by available SNPs is modest, being less than 1% in all cases other than for type 2 diabetes. However, the associated instrumental variables are unlikely to be affected by weak instrument bias given that the p-values associated with the first stage F statistic were  $< 0.0001$  for all diseases. Allele scores for each condition were not strongly related to sex for any condition, and only the Type 2 diabetes score showed possible evidence of an association with age at recruitment (p-value=0.04).

### 4.2. Main results

Table 3 summarises the causal effect estimates and associated 95% confidence intervals for all six disease phenotypes as estimated using just-identified allele score 2SLS models. The coefficients in Table 3 reflect inpatient hospital costs per person per year of a (genetically influenced) change in liability to case status.

Estimates for asthma, eczema, Type 2 diabetes, and depression were consistent with the null. Migraine and coronary heart disease were consistent with a large positive effect on costs. It is notable that confidence intervals are wide in all cases, reflecting the modest proportion of variance in disease status that is explained by each respective score.

Stratification of the 2SLS allele score models by sex did not reveal any material differences to unstratified 2SLS models (Fig. 1), although it is notable that effects for migraine may be most influenced by female cases, and effects for coronary heart disease by male cases.

Cochran's Q statistic revealed some evidence of heterogeneity for all

**Table 2**

Cases, number of SNPs and strength of instruments.

	N	N of cases (%)	N of SNPs	% of variance explained by allele score in disease status	F-statistic from first stage of 2SLS allele score model
<b>Disease phenotype</b>					
Asthma	306,245	39,159 (12.8 %)	8	0.53 %	1242
Eczema	306,245	8017 (2.6 %)	12	0.37 %	270
Migraine	306,245	9782 (3.2 %)	33	0.54 %	484
Coronary heart disease	306,245	14,261 (4.7 %)	15	0.43 %	520
Type 2 diabetes	304,885	13,444 (4.4 %)	13	1.1 %	1274
Depression	92,830	18,760 (20.0 %)	36	0.10 %	93

Notes to table: Depression is reported for fewer individuals than other phenotypes because not all Biobank recruitment centres asked individuals about depression and also because the principal source for these data (Wray et al (Wray et al., 2018).) included the UK Biobank pilot sample, information from which we have excluded to ensure independence of GWAS and analysis samples. Note also that the “depression” phenotype is based on the “major depressive disorder” outcome in the GWAS of Wray et al (Wray et al., 2018). Diabetes cases were limited only to those with Type 2 rather than Type 1 cases.

**Table 3**  
2SLS allele score estimates Beta 95% confidence interval.

Disease phenotype		
Asthma	-£ 30	-£ 222 to £ 161
Eczema	£ 621	-£ 239 to £ 1482
Migraine	£ 786	£ 199 to £ 1373
Coronary heart disease	£ 712	£ 238 to £ 1186
Type 2 diabetes	£ 173	-£ 132 to £ 479
Depression	-£ 348	-£ 997 to £ 300

diseases, although the evidence for asthma is relatively weak. Table 4. Inspection of forest plots measuring the causal effect of each SNP for each condition reveals this heterogeneity. The forest plots show the mean point effect and associated confidence interval for each genetic variant or SNP included in the analysis. We may calculate an overall weighted average of the impact of all SNPs on cost by using techniques drawn from meta-analysis. In particular, the “diamond” figure “IVW-all” measures the aggregate mean effect, with each effect weighted by its precision. In Fig. 2, heterogeneity is indicated by asymmetry about the null hypothesis. Some differences in the magnitude of effect are expected due to sampling variation.

To avoid imposing a sharp cut-off for “significant” evidence of heterogeneity, we conservatively apply various forms of pleiotropy-robust sensitivity analysis to all phenotypes. Consistency among estimators, and between the types of sensitivity analysis, offers some reassurance that similar causal effects are identified, irrespective of the type of assumptions made regarding violations of the exclusion restriction. Supplementary material presents scatter plots for all variant and conditions. Fig. 3 summarises the results for all conditions using all pleiotropy-robust methods. These forest plots contain point estimates and 95% confidence intervals on a condition-by-condition basis for the six

**Table 4**  
Cochran’s Q statistic and heterogeneity by phenotype.

	N of SNPs	Q statistic	Q p-value
<b>Disease phenotype</b>			
Asthma	8	8.6	0.28
Eczema	12	16.9	0.11
Migraine	33	38.2	0.09
Coronary heart disease	15	26.7	0.02
Type 2 diabetes	13	19.3	0.02
Depression	36	48.7	0.02

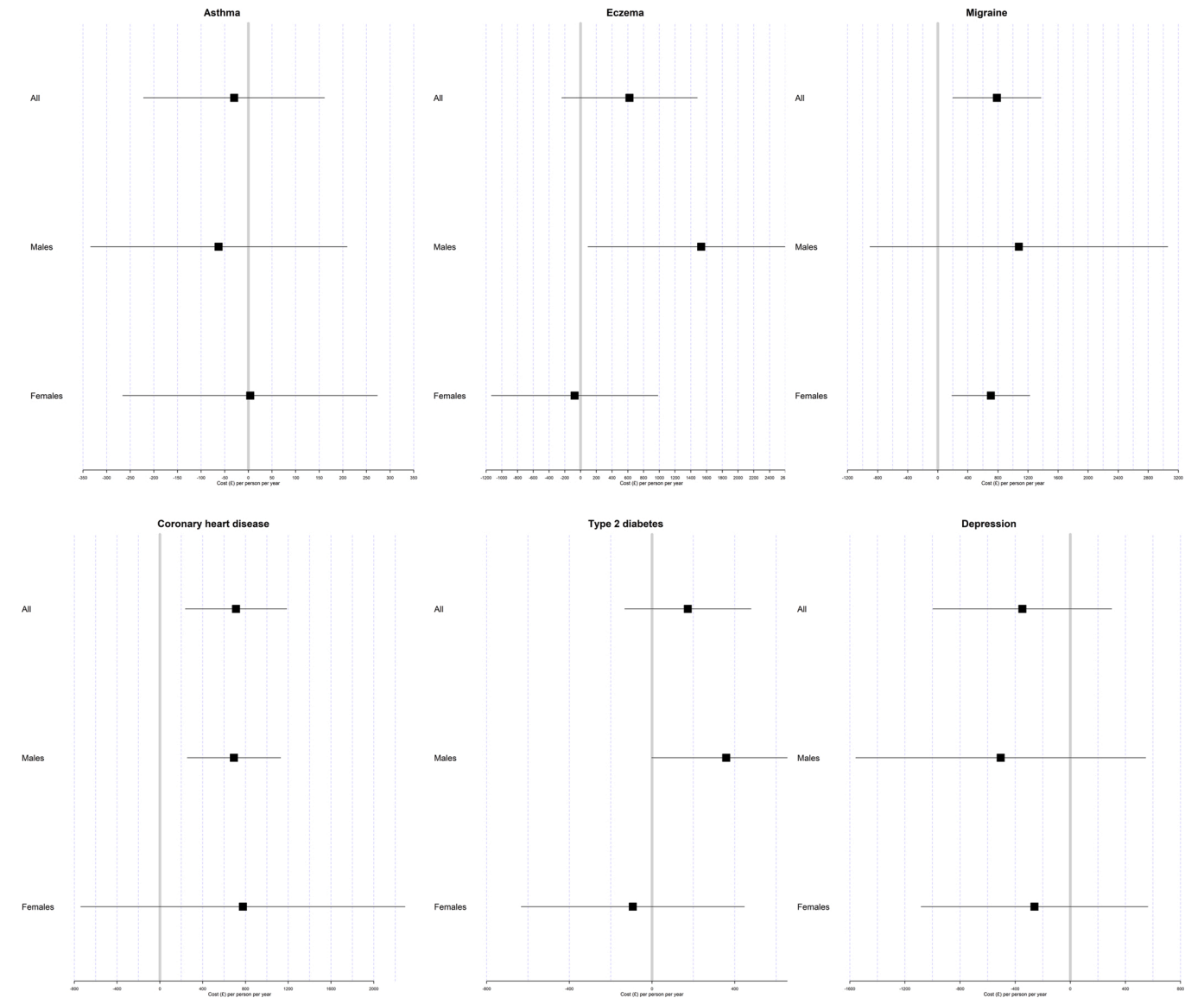


Fig. 1. 2SLS allele score estimates stratified by sex.

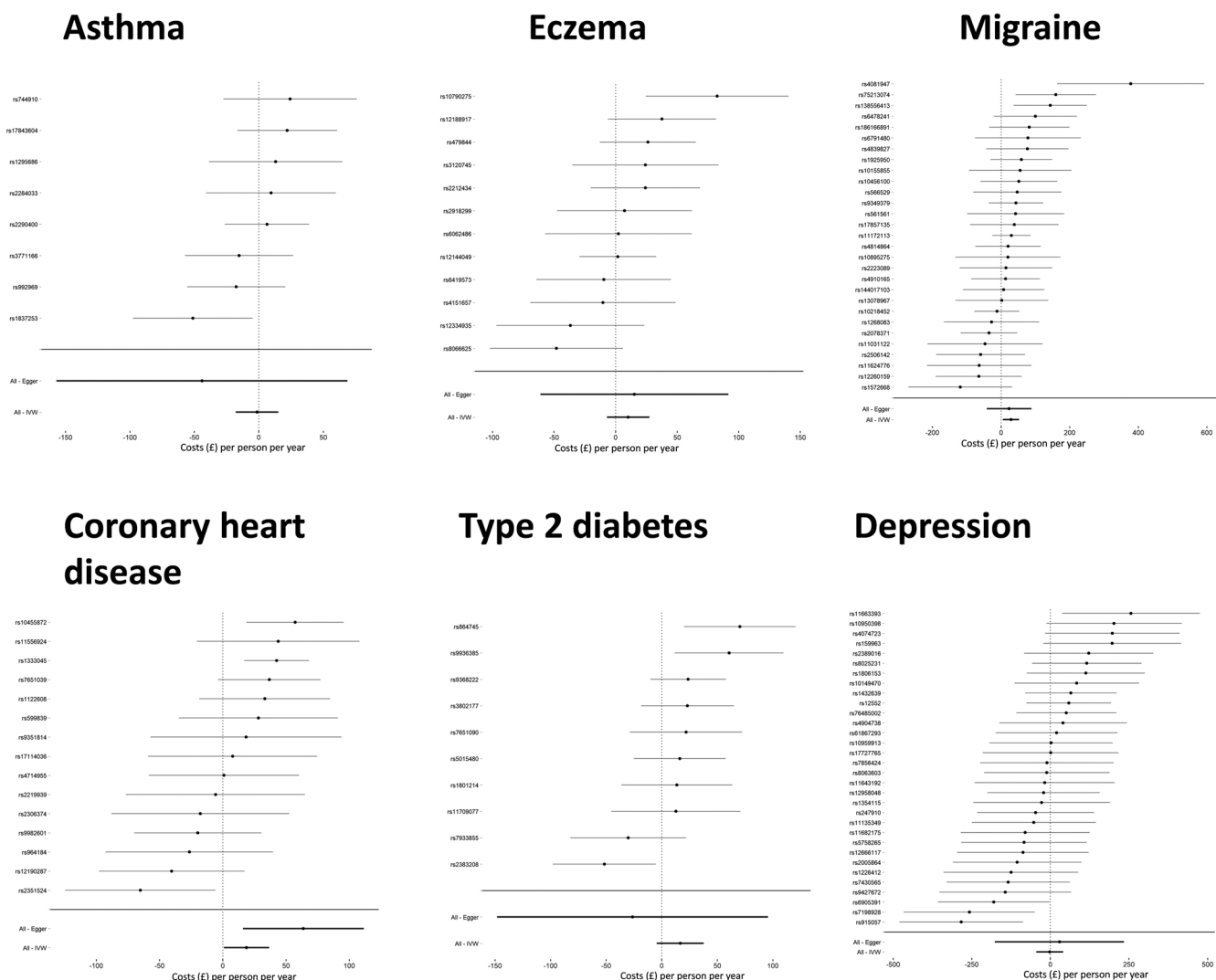


Fig. 2. Forest plot of genome-wide significant SNPs for all conditions (scaled to reflect a 100% relative increase in genetic liability).

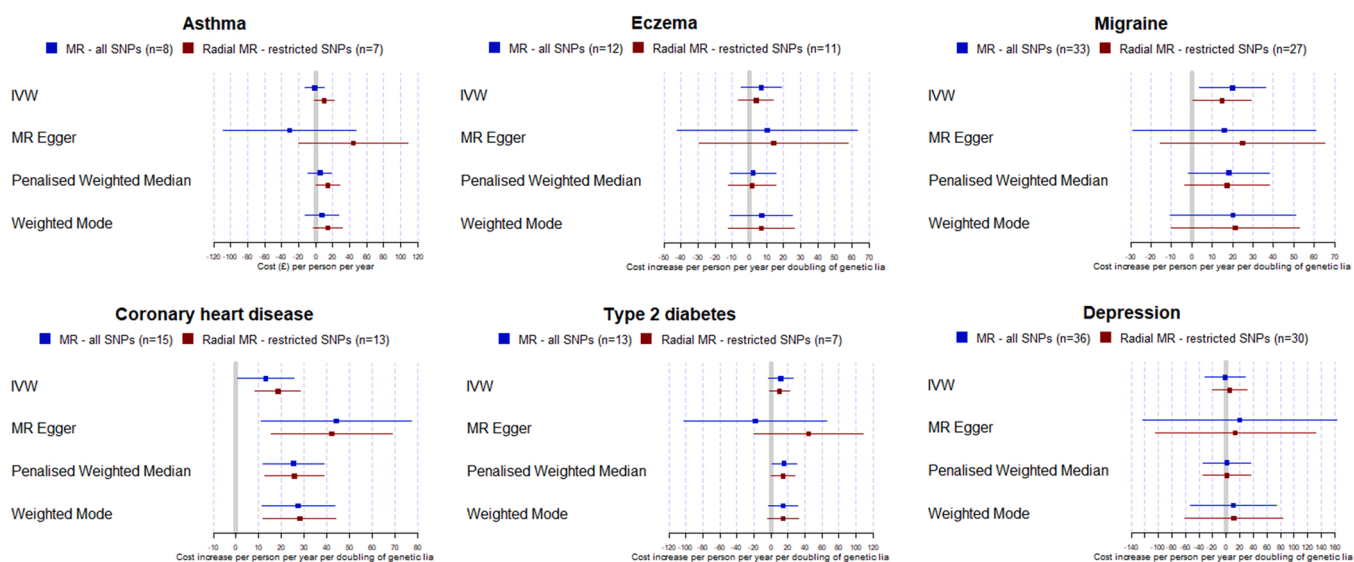


Fig. 3. Comparison of summary estimators Note to figure: IVW – inverse variance weighted.



diseases. These include both the full set of SNPs and restricted SNPs following removal of outliers identified by the radial methods described above.

It is notable that effect estimates are similar within each disease regardless of whether some or all SNPs are included in conventional or Radial Mendelian Randomization models. These estimates reflect per person, per year inpatient hospital costs on a relative scale: all estimators for all conditions imply modest effect sizes per doubling of relative genetic liability. The estimates of coronary heart disease were again more precise than for other disease.

The split sample results for the allele score models are presented in Table 5. We did not identify sufficient genome-wide significant SNPs to calculate allele scores for depression.

Comparisons between the results of the split sample with those of the main analysis are subject to several qualifications, particularly because these models comprise different sets of individuals. Confidence intervals for each phenotype overlap in both the main and split sample analyses. Asthma has the smallest effect size in both the main analysis and the split sample analysis, while coronary heart disease has the largest effect size in the split sample analysis but only the second largest in the main analysis. The impact of migraine on costs is consistent with the null in this analysis. We cannot calculate a Cochran Q value for the meta-analyzed split sample data given the differences in the SNPs identified as genome-wide significant in each sample.

## 5. Discussion

### 5.1. Summary and contribution to literature

There is a wealth of literature using more conventional study designs relating various disease-related exposures to healthcare cost (Larg and Moss, 2011; Van Den Akker-van Marle et al., 2005; Ellis et al., 2002; Berg and Stovner, 2005; Luengo-Fernández et al., 2006; Wang et al., 2003; Sobocki et al., 2006; Seuring et al., 2015). We have offered an alternative to these methods that may avoid some or all of the biases that affect conventional studies. We used the largest available genome-wide association studies to identify genome-wide significant SNPs related to six prevalent chronic diseases and health conditions that are each associated with substantial morbidity in the general population. We used these data in Mendelian Randomization analysis to relate liability to disease to healthcare costs. If the instrumental variable assumptions hold, our methods will produce unbiased estimates of the cost of increasing genetic liability to disease. Effect estimates represent a life-long exposure to genetic liability for each specific disease.

Mendelian Randomization effect sizes were roughly concordant across different types of sensitivity analysis. The similarity of results obtained from different Mendelian Randomization estimators offers some reassurance that the same causal effect was being identified even when assumptions regarding potential violations of the instrumental variable assumptions differed. There was evidence of a substantial and economically meaningful impact of coronary heart disease status on costs, but much less clear evidence for the other disease phenotypes given the uncertainty surrounding point estimates. No differences were apparent when 2SLS allele models were stratified by sex.

Our estimates should not be interpreted as indicating the need for

substantial spending on particular diseases, or the privileging of spending on one disease over another. Those types of arguments depend on circular reasoning (Cawley and Meyerhoefer, 2012) – the more “expensive” a disease appears to be, the more money should be dedicated to its prevention and treatment. Instead, the principal use of our Mendelian Randomization effect estimates is likely to be in decision-analytic models and other formal evaluations of interventions. These new effect estimates could, for example, be used as estimates as the marginal effect of disease status on healthcare cost.

The anticipated reduction in bias comes at the cost of greater imprecision that stems from the use of instrumental variables, since statistical power in instrumental variable models is a function of the proportion of variance in the that the particular exposure is explained by the instrument. Wide confidence intervals, which were consistent with both substantial positive and negative effects of disease status on health care cost, are attributable, at least in part, to the modest proportion of variation in disease status explained by available SNPs (Davey Smith and Munafò, 2019). The tradeoff between bias and precision will need be accounted for in specific applications, but the effect estimates and associated estimates of uncertainty that we present can readily be used in a variety of evaluative contexts.

Estimates were imprecise when considering either the just-identified allele score models, or the summary sensitivity analyses that rely on over-identified models using data on many individual genetic variants (including over-identified models that removed outliers), and in both the main analysis and the split-sample analysis. This is similar to Harrison et al (Harrison et al., 2020). in which causal diseases status associations with selected socioeconomic outcomes were also imprecise in some cases. This contrasts with other work using continuous exposures in relation to hospital costs, for which statistical power will be greater, such as body mass index (Dixon et al., 2020) and waist-hip ratio (Hazewinkel et al., 2022).

Further GWASs with larger sample sizes to increase the number of SNPs and proportion of variance explained in these disease phenotypes will increase the power of this analysis, as will even larger data resources that link to resource use and cost data. Future studies with more cases and improved precision will further clarify these associations.

A second mechanism that may have affected the size and precision of effect estimates is selection (Hughes et al., 2019) into UK Biobank. This cohort is more educated, wealthier and has a more favourable risk profile for adverse health outcomes than the general UK population from which the study population was drawn. We cannot quantify the extent of this bias without knowledge of the selection mechanism itself. Analyses of selection bias by (Hughes et al., 2019; Gkatzionis and Burgess, 2018) in Mendelian Randomization studies suggest that other sources of bias (such as pleiotropy) will be more consequential, although relative magnitudes of bias will depend to some extent on study context.

### 5.2. Interpreting Mendelian randomization

Our causal estimates are most appropriately interpreted as relating to liability to disease. It is possible in some cases that effect of liability on healthcare costs will be most relevant to a particular use of these findings; in other cases it may be less so. Nevertheless, there is an important direct link between disease liability and ultimate disease status. Interventions that reduce liability to the six diseases we study may reduce healthcare costs; for example, smoking cessation interventions may ultimately reduce the costs associated with liability to incident coronary heart disease, even in those individuals who do not develop that disease.

Our instrumental variable models estimated local average treatment effects of random allocation to lifelong liability to each of six diseases. Mendelian Randomization estimates may vary across populations, and the stable unit treatment value assumption (SUTVA) may not hold (Imbens and Rubin, 2015). Treatment, in the present context, refers to liability to disease. The assumption may not hold because we do not know if a (hypothetical) manipulation of in individual's genome to alter

**Table 5**  
2SLS allele score estimates from split sample analysis.

	Beta	95% confidence interval
<b>Disease phenotype</b>		
Asthma	£ 87	-£ 26 to £ 199
Eczema	£ 510	-£ 50 to £ 1069
Migraine	£ 116	-£ 610 to £ 841
Coronary heart disease	£ 1123	£ 672 to £ 1573
Type 2 diabetes	£ 222	-£ 25 to £ 470

liability has an identical effect to other means of managing disease liability.

For instance, various behavioral and socioeconomic factors are implicated alongside genetic influences in the etiology of coronary heart disease (Khera and Kathiresan, 2017; Marmot and Elliott, 2005). Our results estimating the causal effect from SNPs implicated in this disease do not necessarily (although they may) correspond to the results that might be obtained from a hypothetical manipulation of a specific risk factor through, for example, taking statins to manage the risk of heart disease. As another example, the point estimate obtained in 2SLS models for the effect of migraine is large. Although the uncertainty intervals around the point estimate are considerable, the large point estimate may indicate a difference in genetically influenced migraine and other types of migraine; it is the former type which is estimated in our Mendelian Randomization models. We also note that the results from the split sample analysis for migraine were consistent with the null and also associated with considerable uncertainty. Important alternative possibilities are that this large Mendelian Randomization point estimate meaningfully reflects the full impact of this condition on healthcare costs, or that it reflects the possibility that liability to migraine also reflects liability to other conditions.

### 5.3. Possible remaining sources of bias

Geographic stratification in allele distributions that is not removed by conditioning on genetic principal components may have induced bias in these estimates (Haworth et al., 2019; Abdellaoui et al., 2019). This can lead to confounding associated with environmental factors that differ by region. Again, the precise extent of this influence on our results is unknown, but merits consideration in their interpretation. Recent evidence suggests that bias is likely to be modest given an absence of geographic structure in allele scores for all diseases other than coronary heart disease (Abdellaoui et al., 2019).

The random allocation of genotypes that is the foundation of Mendelian Randomization is true conditional on parental genotype, meaning that the randomization occurs within the family unit (Davey Smith and Ebrahim, 2003; Davies et al., 2019a, 2019b; Brumpton et al., 2020; Koellinger and de Vlaming, 2019; Pingault et al., 2018). Dynastic effects (Fletcher, 2011) arise when the expression of the genetic variants in the parents affects disease status in the offspring independently of the child's genome. For example, it could be the case that genetic associations of asthma are influenced by parental smoking behaviour. Causal analysis relying on these SNPs would wrongly attribute causal effects to the asthma SNPs alone, when in truth the effects of the environment were also material to the outcome. However, health costs were measured later in life, when the influence of the early life familial environment may be modest.

Within-family studies would offer the best source of evidence on possible dynastic effects. However, there are far fewer available samples of siblings meaning that within-family estimates are unlikely to be informative because statistical power would be much reduced relative to the analysis of unrelated individuals that we report for our main analysis (Howe et al., 2019) (Davies et al., 2019a). Likewise, data on family trios (parents and offspring) would help assess whether assortative mating is likely to bias (Hartwig et al., 2018) our Mendelian Randomization analyses of unrelated individuals. However, there are no large samples of trios of offspring who have reached middle to old age. In practice, biases of this type may be less plausible for the types of diseases we study compared to traits such as BMI (Jacobson et al., 2007; Ajslev et al., 2012) and education (Kong et al., 2018), for which some evidence of these types of effect has been identified (Howe et al., 2022).

## 6. Conclusion

We report a novel Mendelian Randomization analysis of the causal effect of liability to disease on healthcare cost. Results were consistent

with both large and small effects of disease on inpatient hospital costs. There was concordance across methods bearing different assumptions. The modest precision of available data indicating genetic liability to prevalent long-term health conditions, and selection into the relatively healthy UK Biobank cohort, is likely to explain the imprecision, suggesting the absence of evidence rather than evidence of absence. This trade-off between bias and precision will challenge future work, but is likely to improve as further genetic variants are discovered and predictive power increases.

## Funding statement

PD, GDS, SH and NMD are members of the MRC Integrative Epidemiology Unit at the University of Bristol which is supported by the Medical Research Council and the University of Bristol (MC\_UU\_00011/1). PD acknowledges support from a Medical Research Council Skills Development Fellowship (MR/P014259/1). NMD is supported by a Norwegian Research Council Grant number 295989. SH was supported by Health Foundation grant Social and economic consequences of health: Causal inference methods and longitudinal, intergenerational data, part of the Health Foundation Social and Economic Value of Health Programme (Grant ID: 807293).

## CRediT authorship contribution statement

**Padraig Dixon:** Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft. Writing – review & editing. **Sean Harrison:** Methodology, Formal analysis, Data curation, Writing – review and editing. **William Hollingworth:** Methodology, Writing – review & editing. **Neil M Davies:** Methodology, Writing – review & editing. **George Davey Smith:** Methodology, Writing – review & editing.

## References

- Abdellaoui, A., et al., 2019. Genetic correlates of social stratification in Great Britain. *Nat. Hum. Behav.* 3 (12), 1332–1342.
- Ajslev, T.A., et al., 2012. Assortative marriages by body mass index have increased simultaneously with the obesity epidemic. *Front. Genet.* 3, 125.
- Allcott, H., Lockwood, B.B., Taubinsky, D., 2019. Should we tax sugar-sweetened beverages? An overview of theory and evidence. *J. Econ. Perspect.* 33 (3), 202–227.
- Allen, N., 2012. UK Biobank: current status and what it means for epidemiology. *Health Policy Technol.* 1 (3), 123–126.
- Angrist, J., Pischke, J.-S., 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Oxford.
- Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91 (434), 444–455.
- Berg, J., Stovner, L.J., 2005. Cost of migraine and other headaches in Europe. *Eur. J. Neurol.* 12, 59–62.
- Bowden, J., 2016. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* 40 (4), 304–314.
- Bowden, J., et al., 2018. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int. J. Epidemiol.* 47 (4), 1264–1278.
- Bowden, J., Holmes, M.V., 2019. Meta-analysis and Mendelian randomization: a review. *Res. Synth. Methods (n/a(n/a))*.
- Bowden, J., Davey Smith, G., Burgess, S., 2015. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* 44 (2), 512–525.
- Brennan, A., Akehurst, R., 2000. Modelling in health economic evaluation. *Pharmacoeconomics* 17 (5), 445–459.
- Briggs, A., Sculpher, M., Claxton, K., 2006. *Decision Modelling for Health Economic Evaluation*. Oxford University Press, p. 237.
- Brumpton, B., et al., 2020. Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. *Nat. Commun.*
- Burgess, S., Labrecque, J.A., 2018. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur. J. Epidemiol.* 33 (10), 947–952.
- Burgess, S., Thompson, S., 2015. *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation*. CRC Press, Boca Raton, Florida.
- Burgess, S., Thompson, S.G., 2017. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* 32 (5), 377–389.
- Buxton, M.J., et al., 1997. Modelling in economic evaluation: an unavoidable fact of life. *Health Econ.* 6 (3), 217–227.
- Carulli, L., et al., 2005. Review article: diabetes, genetics and ethnicity. *Aliment. Pharmacol. Ther.* 22 (s2), 16–19.

- Cawley, J., Meyerhoefer, C., 2012. The medical care costs of obesity: an instrumental variables approach. *J. Health Econ.* 31 (1), 219–230.
- Cawley, J. and C.J. Ruhm, Chapter Three - The Economics of Risky Health Behaviors, in *Handbook of Health Economics*, M.V. Pauly, T.G. McGuire, and P.P. Barros, Editors. 2011, Elsevier. p. 95–199.
- Collins, R., 2012. What makes UK Biobank special? *The Lancet* 379 (9822), 1173–1174.
- Cook, J.P., Morris, A.P., 2016. Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility. *Eur. J. Hum. Genet.* 24 (8), 1175–1180.
- Dakin, H.A., et al., 2020. Accurately reflecting uncertainty when using patient-level simulation models to extrapolate clinical trial data. *Med. Decis. Mak.* p. 0272989x20916442.
- Davey Smith, G., 2019. Post-modern epidemiology: when methods meet matter. *Am. J. Epidemiol.* 188 (8), 1410–1419.
- Davey Smith, G., Ebrahim, S., 2003. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32 (1), 1–22.
- Davey Smith, G., Hemani, G., 2014. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23 (R1), R89–R98.
- Davey Smith, G. and M.R. Munafò, in *Does schizophrenia influence cannabis use? How to report the influence of disease liability on outcomes in Mendelian randomization studies TARG Blog - The Tobacco and Alcohol Research Group blog*, Editor. 2019: University of Bristol.
- Davies, N.M., 2019a. Within family Mendelian randomization studies. *Hum. Mol. Genet.*
- Davies, N.M., 2019b. Within family Mendelian randomization studies. *Hum. Mol. Genet.* 28 (R2), R170–R179.
- Dixon, P., 2016. Estimating marginal healthcare costs using genetic variants as instrumental variables: Mendelian randomization in economic evaluation. *Pharmacoeconomics* 34 (11), 1075–1086.
- Dixon, P., 2020. Mendelian randomization analysis of the causal effect of adiposity on hospital costs. *J. Health Econ.* 70, 102300.
- Dixon, P., Davey Smith, G., Hollingworth, W., 2018. The association between adiposity and inpatient hospital costs in the UK biobank cohort. *Appl. Health Econ. Health Policy*.
- Elbein, S.C., 2009. Genetics factors contributing to type 2 diabetes across ethnicities. *J. Diabetes Sci. Technol.* 3 (4), 685–689.
- Ellis, C.N., 2002. Cost of atopic dermatitis and eczema in the United States. *J. Am. Acad. Dermatol.* 46 (3), 361–370.
- Fletcher, J.M., 2011. The promise and pitfalls of combining genetic and economic research. *Health Econ.* 20 (8), 889–892.
- Fletcher, J.M., 2018. *Economics and Genomics*. Oxford Research Encyclopedia of Economics and Finance. Oxford University Press.
- Franklin, M., 2019. An educational review about using cost data for the purpose of cost-effectiveness analysis. *Pharmacoeconomics* 37 (5), 631–643.
- Fry, A., 2017. Comparison of sociodemographic and health-related characteristics of UK biobank participants with the general population. *Am. J. Epidemiol.*
- Fuchsberger, C., 2016. The genetic architecture of type 2 diabetes. *Nature* 536, 41.
- Gkatzionis, A., Burgess, S., 2018. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int. J. Epidemiol.* 48 (3), 691–701.
- Gormley, P., 2016. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat. Genet.* 48 (8), 856–866.
- Greco M, F.D., 2015. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat. Med.* 34 (21), 2926–2940.
- Harrison, S., 2020. The causal effects of health conditions and risk factors on social and socioeconomic outcomes: Mendelian randomization in UK Biobank. *Int. J. Epidemiol.*
- Hartwig, F.P., Davey Smith, G., Bowden, J., 2017. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* 46 (6), 1985–1998.
- Hartwig, F.P., Davies, N.M., Davey Smith, G., 2018. *Bias in Mendelian randomization due to assortative mating*. *Genet. Epidemiol.* 42 (7), 608–620.
- Haworth, S., 2019. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat. Commun.* 10 (1), 333.
- Haycock, P.C., 2016. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am. J. Clin. Nutr.*
- Hazewinkel, A.-D., et al., 2022. Mendelian randomization analysis of the causal impact of body mass index and waist-hip ratio on rates of hospital admission. *Econ. Hum. Biol.* 44, 101088.
- Hemani, G., et al., 2018b. The MR-Base platform supports systematic causal inference across the human genome. *eLife* 7, e34408.
- Hemani, G., Bowden, J., Davey Smith, G., 2018a. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* 27 (R2), R195–R208.
- Howe, L.D., et al., 2019. Effects of body mass index on relationship status, social contact and socio-economic position: Mendelian randomization and within-sibling study in UK Biobank. *Int. J. Epidemiol.*
- Howe, L.J., 2022. Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nat. Genet.*
- Hughes, R.A., 2019. Selection bias when estimating average treatment effects using one-sample instrumental variable analysis. *Epidemiology* 30, 3.
- Imbens, G.W., Rubin, D.B., 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.
- Jacobson, P., et al., 2007. Spouse resemblance in body mass index: effects on adult obesity prevalence in the offspring generation. *Am. J. Epidemiol.* 165 (1), 101–108.
- Keaton, J.M., et al., 2014. A comparison of type 2 diabetes risk allele load between African Americans and European Americans. *Hum. Genet* 133 (12), 1487–1495.
- Khera, A.V., Kathiresan, S., 2017. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat. Rev. Genet.* 18 (6), 331–344.
- Koellinger, P.D., de Vlaming, R., 2019. Mendelian randomization: the challenge of unobserved environmental confounds. *Int. J. Epidemiol.* 48 (3), 665–671.
- Kong, A., et al., 2018. The nature of nurture: effects of parental genotypes. *Science* 359 (6374), 424.
- Larg, A., Moss, J.R., 2011. Cost-of-illness studies. *Pharmacoeconomics* 29 (8), 653–671.
- Lawson, D.J., 2020. Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Hum. Genet.* 139 (1), 23–41.
- Luengo-Fernández, R., 2006. Cost of cardiovascular diseases in the United Kingdom. *Heart* 92 (10), 1384.
- Marmot, M., Elliott, P., 2005. *Coronary Heart Disease Epidemiology: From Aetiology to Public Health*. Oxford Medical Publications.
- Mathieson, I., McVean, G., 2012. Differential confounding of rare and common variants in spatially structured populations. *Nat. Genet.* 44 (3), 243–246.
- Mihaylova, B., 2011. Review of statistical methods for analysing healthcare resources and costs. *Health Econ.* 20 (8), 897–916.
- Moffatt, M.F., 2010. A large-scale, consortium-based genomewide association study of asthma. *New Engl. J. Med.* 363 (13), 1211–1221.
- Morris, A.P., 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* 44 (9), 981–990.
- Munafò, M.R., 2017. Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.* 47 (1), 226–235.
- Murray, C.J.L., 2013. UK health performance: findings of the Global Burden of Disease Study 2010. *Lancet* 381 (9871), 997–1020.
- NICE, *Guide to the methods of technology appraisal*. 2013: Manchester.
- O'Donoghue, T., Rabin, M., 2006. Optimal sin taxes. *J. Public Econ.* 90 (10), 1825–1849.
- Paaby, A.B., Rockman, M.V., 2013. The many faces of pleiotropy. *Trends Genet.* 29 (2), 66–73.
- Paternoster, L., 2015. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat. Genet.* 47 (12), 1449–1456.
- Pingault, J.-B., 2018. Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* 19 (9), 566–580.
- Price, A.L., 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38 (8), 904–909.
- Qingyuan, Z., 2020. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann. Stat.* 48 (3), 1742–1769.
- Rohrer, J.M., 2018. Thinking clearly about correlations and causation: graphical causal models for observational data. *Adv. Methods Pract. Psychol. Sci.* 1 (1), 27–42.
- Sanderson, E., 2022. Mendelian randomization. *Nat. Rev. Methods Prim.* 2 (1), 6.
- Sargan, J.D., 1958. The estimation of economic relationships using instrumental variables. *Econometrica* 26 (3), 393–415.
- Schunkert, H., 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* 43 (4), 333–338.
- Sculpher, M., Claxton, K., Akehurst, R., 2005. It's just evaluation for decision-making: recent developments in, and challenges for, cost-effectiveness research. In: Sculpher, M. (Ed.), *Health Policy and Economics: Opportunities and Challenges*. Open University Press, Maidenhead, pp. 8–41.
- Sculpher, M.J., 2006. Whither trial-based economic evaluation for health care decision making? *Health Econ.* 15 (7), 677–687.
- Seuring, T., Archangelidi, O., Suhrcke, M., 2015. The economic costs of type 2 diabetes: a global systematic review. *Pharmacoeconomics* 33 (8), 811–831.
- Sheldon, T.A., 1996. Problems of using modelling in the economic evaluation of health care. *Health Econ.* 5 (1), 1–11.
- Shiell, A., Gerard, K., Donaldson, C., 1987. Cost of illness studies: an aid to decision-making? *Health Policy* 8 (3), 317–323.
- Siebert, U., 2003. When should decision-analytic modeling be used in the economic evaluation of health care? *Eur. J. Health Econ.*, Former: HEPAC 4 (3), 143–150.
- Sobocki, P., et al., *Cost of Depression in Europe. 2006, ICMPE: Italy*. p. 87–98.
- Stearns, F.W., 2010. One hundred years of pleiotropy: a retrospective. *Genetics* 186 (3), 767–773.
- Stock, J.H., Yogo, M., 2002. Testing for Weak Instruments in Linear IV Regression. National Bureau of Economic Research Cambridge, Mass., USA.
- Sudlow, C., 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12 (3), e1001779.
- Tian, C., Gregersen, P.K., Seldin, M.F., 2008. Accounting for ancestry: population substructure and genome-wide association studies. *Hum. Mol. Genet.* 17 (R2), R143–R150.
- Tudball, M.J., 2021. Mendelian randomisation with coarsened exposures. *Genet. Epidemiol.* 45 (3), 338–350.
- Tyrrrell, J., et al., *Using genetics to understand the causal influence of higher BMI on depression*. 2018.
- Van Den Akker-van Marle, M.E., Bruil, J., Detmar, S.B., 2005. Evaluation of cost of disease: assessing the burden to society of asthma in children in the European Union. *Allergy* 60 (2), 140–149.
- Wang, P.S., Simon, G., Kessler, R.C., 2003. The economic burden of depression and the cost-effectiveness of treatment. *Int. J. Methods Psychiatr. Res.* 12 (1), 22–33.
- Wray, N.R., et al., 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* 50 (5), 668–681.