STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:1133–1163
Published online 20 September 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3034


Paper Celebrating the 25th Anniversary of *Statistics in Medicine*

Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology

Debbie A. Lawlor[1,2,*,†], Roger M. Harbord[1,3], Jonathan A. C. Sterne[1,3], Nic Timpson[1,2] and George Davey Smith[1,2]

[1]*Department of Social Medicine, University of Bristol, U.K.*
[2]*MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, U.K.*
[3]*MRC Health Services Research Collaboration, University of Bristol, U.K.*

SUMMARY

Observational epidemiological studies suffer from many potential biases, from confounding and from reverse causation, and this limits their ability to robustly identify causal associations. Several high-profile situations exist in which randomized controlled trials of precisely the same intervention that has been examined in observational studies have produced markedly different findings. In other observational sciences, the use of instrumental variable (IV) approaches has been one approach to strengthening causal inferences in non-experimental situations. The use of germline genetic variants that proxy for environmentally modifiable exposures as instruments for these exposures is one form of IV analysis that can be implemented within observational epidemiological studies. The method has been referred to as 'Mendelian randomization', and can be considered as analogous to randomized controlled trials. This paper outlines Mendelian randomization, draws parallels with IV methods, provides examples of implementation of the approach and discusses limitations of the approach and some methods for dealing with these. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: Mendelian randomization; instrumental variables; genetics; causal models; econometrics; epidemiology; confounding

1. INTRODUCTION

The aim of this paper is to describe how Mendelian randomization—the random assignment of an individual's genotype from his or her parental genotypes that occurs before conception—can be exploited to make causal inferences in aetiological epidemiology. The principle of Mendelian

*Correspondence to: D. A. Lawlor, MRC Centre for Causal Analyses in Translational Epidemiology, Canynge Hall, Whiteladies Rd, Bristol BS8 2PR, U.K.
†E-mail: d.a.lawlor@bristol.ac.uk

Contract/grant sponsor: MRC Health Services Research Collaboration
Contract/grant sponsor: MRC
Contract/grant sponsor: U.K. Department of Health Career Scientist Award

Received 10 November 2006
Accepted 6 July 2007

randomization dates back to at least 20 years and has been attributed [1] to Katan [2], although within the context of observational epidemiology the term itself was not used until the turn of the millennium [3]. Thomas and Conti [4] pointed out that Mendelian randomization is an application of the method of instrumental variables (IVs), which are commonly used in econometrics. In Mendelian randomization studies genetic information is used as an IV. The use of Mendelian randomization is growing rapidly in popularity, with increasing recognition of its statistical challenges [5–9].

In this introductory section, we explain the limitations of conventional means of making causal inferences and introduce the concept of Mendelian randomization with its associated terminology. Section 2 provides a description of IV methods and the assumptions that underlie them: these provide a formal framework for causal inferences from Mendelian randomization studies. We illustrate these ideas with a specific example in Section 3. In Section 4 we outline potential problems with the application of IV methods in Mendelian randomization studies and describe potential solutions to these problems. Section 5 describes power calculations for Mendelian randomization studies, and Section 6 contains concluding remarks.

1.1. The limitations of observational epidemiology for making causal inference

Conventional observational epidemiology has made important contributions to understanding disease aetiology. Notable examples include work pioneered by Sir Richard Doll that identified the link between cigarette smoking and lung cancer, heart disease and other adverse health outcomes [10, 11], and work establishing the causal associations of both high blood pressure and dyslipidaemia with cardiovascular disease [12, 13]. These findings have been important in the development of interventions (lifestyle and pharmacological) that have resulted in population-level declines in these risk factors and the treatment of those at highest risk, which together have led to decreases in smoking-related cancers and in cardiovascular disease morbidity and mortality.

Despite the successes of observational epidemiology, the discipline has also produced its share of high-profile failures, in that it has identified many exposures apparently increasing or decreasing disease risk that have later been revealed by randomized controlled trials (RCTs) to be non-causal. The most likely explanations are confounding by lifestyle and socioeconomic factors, or by baseline health status and prescription policies, together with reverse causation and selection bias [14–17]. Many potentially health-modifying factors (such as use of vitamin E supplements or hormone replacement) are strongly related to a range of socioeconomic and lifestyle factors that will confound their associations with health outcomes. It is often not fully appreciated that the ability to deal with this issue statistically is limited [18]. These recent examples have important consequences for public health. Many post-menopausal women have taken hormone replacement for the prevention of cardiovascular disease, but RCT evidence shows that for most women hormone replacement will not prevent cardiovascular disease and may increase overall mortality and other adverse health outcomes [17, 19]. Similarly, large numbers of individuals take vitamin E supplements (falsely) believing that they prevent cancer and cardiovascular disease.

1.2. Limitations of RCTs for making causal inferences

Well-conducted RCTs are appropriately considered to be the gold standard for making causal inferences in health sciences. However, for a number of important modifiable exposures (e.g. smoking, alcohol) involved in the aetiology and prognosis of major diseases (cardiovascular disease, diabetes, cancer, mental ill-health), it would be unethical or impractical to randomly allocate

DOI: 10.1002/sim

individuals to these exposures. Because of the long lead time between exposure (e.g. to diet) and the development of diseases such as cancer, trials may take many decades to produce robust results. Participant compliance with dietary and other manipulations is often poor. One also needs to carefully examine the generalizability of an RCT to the whole population for which one would like to intervene to prevent disease or improve prognosis. RCTs tend to be conducted on very specific groups of people without co-morbidities, who are not using other interventions and who fall into a particular age group, whereas in real life clinicians and public health practitioners wish to know about lifestyle and therapeutic interventions to prevent and treat disease in all individuals, including those who do have co-morbidities. Those who give their informed consent to participate in RCTs agree to an equal chance of being allocated to any of the randomization groups, and such individuals may not be representative of the whole population for which we would like to consider the intervention. RCTs are expensive, resource intensive and should be conducted only when there is good evidence from basic science and observational epidemiology for a particular intervention.

1.3. The use of genetic epidemiology to make causal inference: Mendelian randomization

Mendelian randomization is the term that has been given to studies that use genetic variants in observational epidemiology to make causal inferences about modifiable (non-genetic) risk factors for disease and health-related outcomes [1, 3, 20]. Such studies exploit what is known as Mendel's second law or the law of independent assortment:

> that the behavior of each pair of differentiating characters in hybrid union is independent of the other differences between the two original plants, and, further, that the hybrid produces just so many kinds of egg and pollen cells as there are possible constant combination forms [21].

In simple terms this means that the inheritance of one trait is independent of (i.e. randomized with respect to) the inheritance of other traits. In contemporary times, Mendel's law is described in terms of genotype, alleles and phenotypes. Table I defines the genetic terms that will be used in this paper; further detailed descriptions of genetic terms can be found in two recently published text books on statistical analyses in genetic epidemiology [22, 23]. Mendel examined the inheritance of seven traits of pea-plants—shape of the ripe seeds; colour of the seed interiors; colour of the unripe pods; colour of the petals; inflated or pinched pods; long or short plant stems; and axial or terminal flowers—and noted that they were all inherited independently of each other.

Subsequent work showed that Mendel's second law was not always true. Punnett and Bateson published the first report of 'gene linkage' in peas, showing that when a broader set of traits were explored not all of them assorted independently. Specifically, they found that petal colour and pollen shape were not inherited independently—there was an excess of plants with purple petals and long-shaped pollen, an excess of plants with red petals and round pollen, but fewer plants that were either purple-round or red-long, than would be expected if these two traits were inherited independently [24]. Thomas Hunt Morgan and colleagues then went on to demonstrate that independent assortment is generally true for genes found on non-homologous chromosomes (see Table I for definitions), but is not always true for a set of genes located on a homologous chromosome, particularly if the genes are located close to each other [24]. The term linkage disequilibrium (LD) is used to describe this situation, which represents a departure from the hypothetical situation in which all alleles exhibit complete independence (linkage equilibrium, LE).

DOI: 10.1002/sim

Table I. A glossary of genetic terms.

- *Alleles* are the variant forms of a single-nucleotide polymorphism (SNP), a specific polymorphic site or a whole gene detectable at a locus
- *Canalization* is the process by which potentially disruptive influences on normal development from genetic (and environmental) variations are damped or buffered by compensatory developmental processes
- A *Chromosome* carries a collection of genes located on a long string of DNA. A *non-homologous chromosome* carries a unique collection of genes on a long string of DNA that is different from the gene collection of another non-homologue. Normal non-homologous chromosomes are not attached to each other during meiosis, and move independently of one another, each carrying its own gene collection. Two *homologous chromosomes* carry the same collection of genes, but each gene can be represented by a different allele on the two homologues (a heterozygous individual). A gamete will receive one of those homologues, but not both. Humans have 22 pairs of autosomal homologous chromosomes and 1 pair of sex chromosomes
- *DNA* deoxyribonucleic acid (DNA) is a molecule that contains the genetic instructions used in the development and functioning of all living organisms. The main role of DNA is the long-term storage of information. It contains the instructions needed to construct other components of cells, including proteins and ribonucleic acid (RNA) molecules. DNA has four nuleotide bases A, T, G and C. The two strands of DNA in the double-helix structure are complementary (sense and anti-sense strands) such that A binds with T and G binds with C
- A *gene* comprises a DNA sequence, including introns, exons and regulatory regions, related to transcription of a given RNA
- *Genotype* of an individual refers to the two alleles inherited at a specific locus—if the alleles are the same, the genotype is homozygous, if different, heterozygous
- *Haplotype* describes the particular combination of alleles from linked loci found on a single chromosome
- *Linkage disequilibrium* (LD) is the correlation between allelic states at different loci within the population. The term LD describes a state that represents a departure from the hypothetical situation in which all loci exhibit complete independence (linkage equilibrium, LE)
- A *locus* is the position in a DNA sequence and can be a single-nucleotide polymorphism (SNP), a large region of DNA sequence, or a whole gene
- *Pleiotropy* is the potential for polymorphisms to have more than one specific phenotypic effect
- *Polymorphism* is the existence of two or more variants (i.e. SNPs, specific polymorphic sites or whole genes) at a locus. Polymorphism is usually restricted to moderately common genetic variants—at least two alleles with frequencies of greater than 1 per cent in the population
- *Recombination* is any process that generates new gene or chromosomal combinations not found previously in that cell or its progenitors. During meiosis, recombination is the process that generates haploid cells that have non-parental combinations of genes
- *Single-nucleotide polymorphism* (SNP) are genetic variations in which one base in the DNA is altered, e.g. a T instead of an A

Many genetic association (and Mendelian randomization) studies exploit LD to their advantage by using genetic markers (a detectable genetic signal) that are in LD with probably functional variants that have a less readily identifiable signal. However, as discussed in more detail in Section 4.5, there are possible situations in which LD might result in violation of Mendelian randomization/IV assumptions.

The independent distribution of alleles (or blocks of alleles in LD) from parents to their offspring means that a study relating health outcomes in the offspring to genetic variation transmitted from the parents will not suffer from confounding. This holds true for full siblings who are not monozygotic twins. Despite the actual random allocation of groups of alleles being at the level of parents to their offspring dyads, at a population level—when relating genetic variants to disease outcome—alleles are generally unrelated to those confounding factors (in particular, socioeconomic position

DOI: 10.1002/sim

and lifestyle factors) that distort the interpretations of findings from observational epidemiology [25, 26]. Furthermore, disease processes do not alter germline genotype and therefore associations between genotype and disease outcomes cannot be affected by reverse causality. Finally, genetic variants that are related to a modifiable exposure will generally be related to it throughout life from birth to adulthood and therefore their use in causal inference can also avoid attenuation by errors (regression dilution bias) [27].

Mendelian randomization studies have being likened to a 'natural' RCT [28, 29]. Unlike conventional RCTs, Mendelian randomization studies can be conducted in a representative population sample without the need for exclusion criteria or for volunteers amenable to being randomly allocated to treatment. However, as noted above, random allocation of alleles occurs within family groups (from parents to offspring), whereas Mendelian randomization studies are most commonly carried out within general population samples, which potentially makes them susceptible to distorting factors such as population stratification, which is discussed in greater detail in Section 4.4. The fact that randomization occurs at conception also means that these studies may be biased by developmental compensation, which is also discussed in greater detail in Section 4.8.

Mendelian randomization studies are therefore defined as any study that uses genetic variation that serves as a robust proxy for an environmentally modifiable exposure in order to make causal inferences about the outcomes of the modifiable exposure. We have utilized the Mendelian randomization approach to determine the magnitude of causal relationships between C-reactive protein (CRP) and both metabolic syndrome traits [30, 31] and carotid intima media thickness [32], between fibrinogen and coronary heart disease [33], alcohol consumption and oesophageal cancer [34], folate and schizophrenia [35], vitamin D and tuberculosis [36], and folate and breast cancer [37]. Other groups have also examined some of the same pathways [38, 39] and have applied the approach to additional exposures in relation to several important disease outcomes including cardiovascular diseases, cancer, mental health and physical functioning in older age groups [40–46].

It is important to note that the aim of Mendelian randomization studies is *not* to identify functional genes for an outcome that might be used in genetic screening; indeed, the approach depends on knowing about the function of genes before undertaking the study. In these studies, genotype is an *instrument* for the modifiable exposure of interest. Thus, in the example that we describe in Section 3, below, our aim is to determine the magnitude of the causal association between circulating levels of CRP and insulin resistance. For this reason, and to ensure that readers remain focused on the aim of making causal inference for modifiable exposures, throughout this paper we refer to a modifiable exposure of interest, whereas in previous papers [1, 27, 29, 47] the term 'intermediate phenotype' has frequently been used in place of modifiable exposure.

2. MENDELIAN RANDOMIZATION AND IVs

The use of genotype to provide causal inference for the effect of a modifiable (non-genetic) exposure on disease outcome is an application of the general theory of *IV analysis* [4, 48–52]. An *IV* is a variable associated with the outcome only through its robust association with an intermediary variable—the exposure of interest in this case.

IV methods form the basis of much statistical analysis in econometrics. Specifically, econometricians use IV methods to deal with what they term 'endogeneity', which is a broad term covering what epidemiologists would refer to as regression dilution bias, confounding and reverse

DOI: 10.1002/sim

causality. The use of IV methods is not new to medical statistics, but its most common application has been to deal with non-compliance in clinical trials, and there are papers introducing IVs to epidemiologists and health service researchers that focus on this as a motivating example [48–51]. Its formal use in Mendelian randomization studies was introduced only recently [31] and is already being debated in the literature [4–8].

Although econometrics shares much methodology and terminology with statistics, some of its terminology will be unfamiliar to medical statisticians and epidemiologists. We shall seek to explain terms from econometrics as we introduce them, but readers who wish to investigate the IV literature may find the online glossary of research economics edited by Peter Meyer [53] useful, or the recent introductory econometrics textbook by Murray [54], which also includes a glossary. The most commonly cited graduate-level text is Wooldridge [55].

2.1. Assumptions of Mendelian randomization studies and IV analysis

Formally, an IV is defined as a variable that satisfies the following assumptions:

1. The IV Z is associated with the exposure of interest X;
2. Z is independent of the confounding factors U that confound the association of X and the outcome Y;
3. Z is independent of outcome Y given X and the confounding factors U.

These assumptions are depicted in the directed acyclic graph (DAG) shown in Figure 1.

The theoretical underpinnings of the Mendelian randomization approach are that: the genotype is robustly associated with the modifiable (non-genetic) exposure of interest (equivalent to assumption 1 above); the genotype is not associated with confounding factors that bias conventional epidemiological associations between modifiable risk factors and outcomes (assumption 2); and that the genotype is related to the outcome only *via* its association with the modifiable exposure (assumption 3). Thus, the relationship between the assumptions of IV analysis and the theoretical underpinning of Mendelian randomization can be seen by comparing the two DAGs shown in Figures 1 and 2.

The three assumptions listed above are sufficient for the simple case of statistical testing (i.e. using genotype as an IV to test the null hypothesis that the modifiable exposure X is not associated

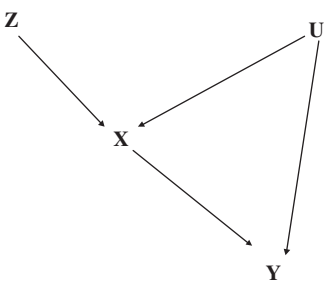

Figure 1. Directed acyclic graph (DAG) for the basic instrumental variables model. Z, instrumental variable; X, modifiable exposure of interest; Y, outcome of interest; and U, (unmeasured or measured with error) confounders.

DOI: 10.1002/sim

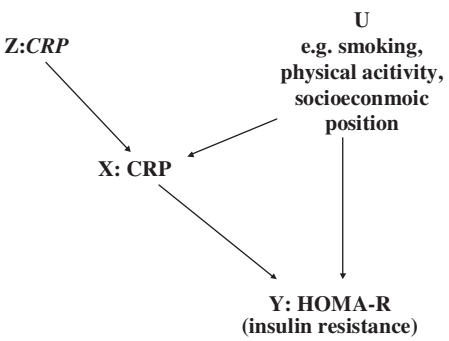

Figure 2. Directed acyclic graph (DAG) for the effect of circulating levels of C-reactive protein on insulin resistance (assessed using homeostasis model HOMA-R) determined using *CRP* as an instrumental variable. Z: *CRP*, c-Reactive protein genetic variant, the instrumental variable; X: CRP, circulating c-reactive protein levels, the modifiable exposure of interest; Y: HOMA-R, homeostasis model assessment of insulin resistance, the outcome of interest; U, (unmeasured or measured with error) confounders.

with outcome Y). However, in most epidemiological situations we are much more interested in obtaining a precise estimate of causal effects (i.e. answering the question 'what is the magnitude of the causal effect of X on Y?') than in simply testing a null hypothesis. Further, in Mendelian randomization studies where gene–outcome associations are expected to be weak, incorrect inference due to type 2 errors would be likely if we focused solely on statistical tests. Thus, the usual aim of Mendelian randomization studies is to provide an estimate of effect with reliable confidence intervals [7]. To estimate a causal effect with IV analysis, additional assumptions are needed; at a minimum the following is assumed:

4. All of the associations depicted in Figure 1 are linear and unaffected by statistical interactions [8, 9, 54].

This assumption is problematic for the situation with binary outcomes, where epidemiologists wish to present the effect estimate as an odds ratio or risk ratio (as opposed to an absolute difference in risk), since such associations are exponential (non-linear). Since this situation is frequent in epidemiology, we discuss the ways in which violations of this assumption can be dealt with in Mendelian randomization studies (see Section 4.9 for more details).

2.2. Estimating causal effects using IV methods

In the simplest case of linear associations and a continuous outcome variable Y (e.g. the log of fasting insulin), it can be shown that the IV estimate of the regression coefficient for the effect of exposure (X) on Y is [4, 8, 9, 54]

$$\hat{\beta}_{\mathrm{IV}} = \hat{\beta}_{ZY} / \hat{\beta}_{ZX} \tag{1}$$

where $\hat{\beta}_{ZY}$ is the coefficient for the regression of outcome (Y) on the IV (Z), and $\hat{\beta}_{ZX}$ is the coefficient for the regression of exposure (X) on the IV. The IV estimator $\hat{\beta}_{\mathrm{IV}}$ provides an estimate of the causal effect of exposure on outcome, even in the presence of unmeasured confounders of the exposure–outcome association. We may provide some intuitive motivation for equation (1) by noting that the change in outcome for a unit change in the instrument, $\hat{\beta}_{ZY}$, is the product of the

DOI: 10.1002/sim

change in the exposure for a unit change in the instrument, $\hat{\beta}_{ZX}$, and the change in the outcome per unit change in exposure, $\hat{\beta}_{\mathrm{IV}}$. If we use this to estimate the causal effect $\hat{\beta}_{\mathrm{IV}}$, we make use of only that part of the variation in the outcome Y that is due to the instrument Z, and (by assumption (2) above) avoid the contaminating effect of the variation in Y due to the confounding factors that affect the regression of Y on X. Figure 1 shows the DAG for the basic IV model.

The estimator $\hat{\beta}_{\mathrm{IV}}$ above is sometimes called the Wald estimator and applies only when there is a single IV. Usually there is more than one IV, for example for two contrasts between the three genotypes for a biallelic polymorphism. It is then necessary to combine the information available from each IV.

Several methods of IV estimation are available for the situation in which there is more than one IV and the outcome Y is a numerical variable and associations between variables are linear. The simplest and most widely available estimator for this situation is two-stage least squares (2SLS or TSLS). 2SLS is so named as the same estimate (though *not* its standard error) can be derived by the following procedure: First, perform least-squares regression of the exposure X on the IV(s) Z; second, perform least-squares regression of the outcome Y on the predicted values from the first regression. When the IVs are indicator variables for a categorical instrument, which is usually the case in Mendelian randomization using genetic information as the instrument, the predicted values are simply the means of X within each category.

Other estimators for IV analyses include limited information maximum likelihood (LIML) and the generalized method of moments (GMM), which are discussed in detail elsewhere [53, 55, 56]. There are methods for undertaking IV analyses in which the two estimates (X on Z and Y on Z) are obtained from different study samples. However, there are a number of advantages of using the same study sample for both estimates, including being in a better position to examine violation of IV method assumptions (see Section 4 for more details) and having greater precision, since the IV estimate will have smaller variance if estimated from the same study than if not [7]. Minelli *et al.* describe an integrated approach to meta-analysis of genetic association studies that provides an IV estimate in Mendelian randomization studies when the X–Z estimate and the Y–Z estimate are obtained from different studies [57]. Methods that deal with non-linear models are described in this paper in Section 4.9.

3. AN EXAMPLE OF THE USE OF IVs TO MAKE CAUSAL INFERENCES: THE EFFECT OF CRP ON INSULIN RESISTANCE

To demonstrate the use and methodological issues in application of this approach, we will present an example from one of our published studies: that of the effect of CRP on insulin resistance [31]. We refer the reader to the original publication of this study for full details of the study population, measurements of the modifiable exposure (CRP), outcome of interest (insulin resistance) and confounders and the derivation of *CRP* haplotypes [31]. Here we present a brief description of the background to this study and why we felt a Mendelian randomization approach was important, the statistical methods employed in our IVs analysis, the main results and how we interpret these.

3.1. Background

Findings from prospective cohort studies have suggested that CRP may have a causal effect on the development of insulin resistance, type 2 diabetes and cardiovascular disease, and it has been

DOI: 10.1002/sim

suggested that therapeutics aimed at reducing CRP levels should be developed to prevent and/or treat these conditions [58–63]. However, CRP is associated with a wide range of lifestyle and socioeconomic characteristics [64] and may be elevated by the presence of atherosclerosis or insulin resistance; hence, confounding or reverse causality may explain its association with insulin resistance and cardiovascular disease [65, 66]. Studies have indicated that genetic variation in *CRP* (C-reactive protein, pentraxin related, locus ID 1401) is associated with the concentration of circulating CRP, and thus this genetic variant provides an instrument for examining the association of CRP with insulin resistance. We applied Mendelian randomization using IV methods, with *CRP* haplotypes as instruments to determine the magnitude of the causal association between circulating CRP levels and insulin resistance. Figure 2 shows the DAG of the assumed associations between CRP, insulin resistance, *CRP* and potential confounding factors.

3.2. Study methods and results

We used data from the British Women's Heart and Health Study, a prospective cohort study of 4286 women who were aged 60–79 years at baseline examination (1999–2001). In total, 2809 women were included in the analyses [31]. These women had provided informed consent for genetic testing (less than 1 per cent of the eligible participants refused consent), were described by the research nurse as White (using a pre-specified list of options that included White, Afro-Caribbean, South Asian, Chinese, other; 99.6 per cent were described as White), were free of diabetes, had fasting glucose levels less than 7 mmol/l and had complete data on genotype, circulating CRP and homeostasis model assessment (HOMA-R), a measure of insulin resistance.

3.2.1. Statistical analyses. General analysis: Plasma CRP and HOMA-R scores were positively skewed and their logged values were used in all regression models. We used linear and logistic regression to assess the association of log plasma CRP concentration with insulin resistance (log HOMA-R) and potential confounding factors. We also used linear and logistic regression to demonstrate the association of *CRP* haplotypes with CRP levels and also to test the assumption that these genetic variations (haplotypes) would not be associated with lifestyle and socioeconomic confounding factors that tend to confound associations of CRP with insulin resistance. To aid interpretability, we used logarithms of CRP to base 2 and back-transformed regression coefficients to give the ratio of geometric means of HOMA-R (entered as natural logarithm into regression models) per doubling of CRP.

SIMHAP and haplotype analyses: Having collected genotype data, haplotypes were constructed using the genetic data analysis program SIMHAP (http://www.genepi.com.au). SIMHAP uses current estimation–maximization-based methods for the estimation of haplotypes from unphased genotype data [67]. The current implementation of SIMHAP utilizes the statistical computing package R (www.r-project.org) to resolve haplotypes and provide their posterior probabilities. All possible haplotype configurations are resolved for each individual within the program itself, and the posterior probability of each configuration calculated. Sensitivity analysis incorporating the uncertainty of haplotype assignment was not performed in this study as all probabilities for haplotypic inference were $\geqslant 0.99$ [31].

IV analyses: In these analyses, we used the most likely haplotype assignment for each woman. It is necessary to limit the number of variables used to model the effect of genotype on CRP in order to ensure that these variables are strongly correlated with CRP and thereby avoid problems of bias due to 'weak instruments' [68, 69]. We chose to use a model for the gene–CRP association

DOI: 10.1002/sim

which assumes that each of a woman's two haplotypes contributes additively to her value of log-CRP. We checked the fit of this model by examining the patterns of residuals for each possible pair of haplotypes, i.e. diplotypes. As a sensitivity analysis, we also fitted a model allowing for different log-CRP levels for each of the 10 possible diplotypes that an individual can possess. This model avoids the assumption that each woman's haplotypes have an additive effect on her CRP levels. However, this model has more parameters and so may be more prone to bias due to weak instruments since the amount of variation explained by each parameter will be lower.

We used 2SLS to fit the IVs models in the main analyses but checked results using LIML and the GMM [56]. We compared the IV estimates with those from ordinary least squares (OLS) regression using the Durbin form of the Durbin–Wu–Hausman statistic in the main analyses [56]. This tests the null hypothesis that the OLS estimator of the same equation would yield consistent estimates when compared with the IV analyses. In econometrics, rejection of the null hypothesis by this test is said to indicate problems of endogeneity (regression dilution bias, confounding, reverse causality) in the OLS estimate and, therefore, the IV analysis is required for an unbiased estimate of the effect of X on Y [56]. In addition to this main analysis, we also checked results comparing the OLS estimate of the effect of CRP on HOMA-score with the IV estimate (using *CRP* haplotypes as the IV) using the Hansen J statistics (LIML) and the Anderson–Rubin statistics (GMM). We examined F-statistics from the first-stage regressions to evaluate the strength of the instruments [68, 69]. Values greater than 10 are often assumed to indicate sufficient strength to ensure the validity of IV methods [68]. All of the IV analyses were performed in Stata (version 9.2) using the user-written `ivreg2` command [56] (usefully, Stata is one language that many epidemiologists and econometricians have in common).

3.2.2. Main results and interpretation. As anticipated (based on the principle of Mendelian randomization), there was no evidence of associations of *CRP* with cigarette smoking, alcohol consumption, physical activity, hormone replacement use or life course socioeconomic position [30, 31]. By contrast, there were strong associations between greater CRP levels and increased prevalence of smoking, lower levels of physical activity, more adverse socioeconomic position from across the participants' life course and decreased prevalence of daily alcohol consumption [30, 31].

In a conventional age-adjusted analysis without use of IVs, a doubling of CRP was associated with a ratio of geometric means of HOMA-R of 1.09 (95 per cent CI: 1.07, 1.10, $p<0.0001$), suggesting that a doubling of CRP resulted in a 9 per cent increase in HOMA-R (see Figure 3). With additional adjustment for 10 indicators of socioeconomic position from across the life course, smoking status, physical activity, alcohol consumption and hormone replacement use, this attenuated only slightly to 1.08 (95 per cent CI: 1. 07, 1.09, $p<0.0001$). We were concerned that this association might still be influenced by residual confounding and/or reverse causality and therefore went on to use *CRP* haplotype as an instrument to estimate the causal effect of CRP on HOMA-R.

The 2SLS approach can be illustrated graphically when the instrument is categorical, and this may help to build intuition. To do this, it is necessary to categorize women by the pair of haplotypes ('diplotype') they carry and use indicator variables for these categories as instruments (note that in our main analyses we used fewer parameters by categorizing each woman by her two haplotypes combined, assuming they had an additive effect—see above). We excluded the uncommon diplotypes carried by fewer than 10 women, leaving nine diplotypes and 2801 women. Figure 4 illustrates this approach.

DOI: 10.1002/sim

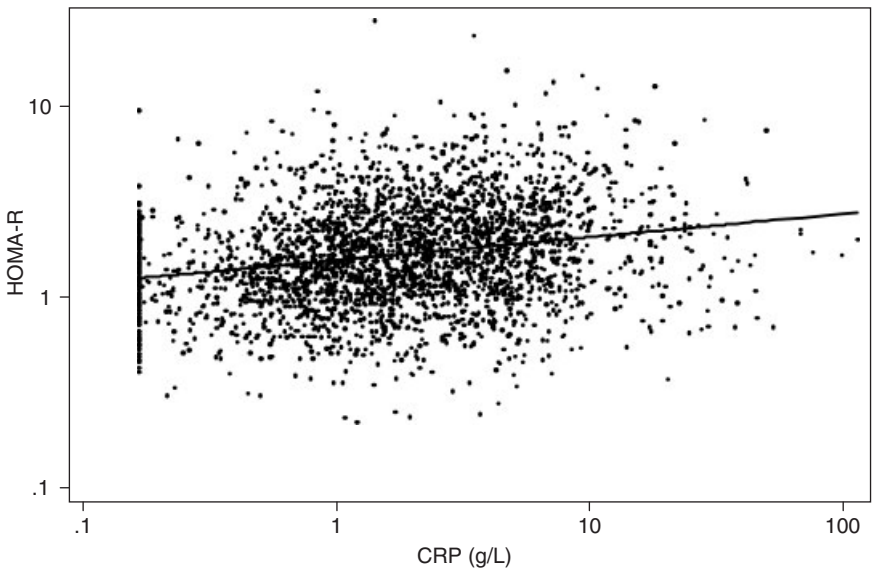

Figure 3. Observational association between HOMA-R score and CRP with fitted line from ordinary least squares using data from the British Women's Heart and Health Study as described in Section 3. Both axes are on a log scale. CRP values of 0.168 mg/L are recorded as 0.168 mg/L as this was the lower limit of detection of the laboratory assay used to measure CRP.

Note that both scales of Figure 4 needed to be greatly enlarged over those in Figure 3, as the variation between the means shown in Figure 4 is much smaller than the variation across the population shown in Figure 3. The small proportion of variation in the exposure of interest (log(CRP)) explained by the instruments (diplotype) leads directly to the much lower precision of the IV estimate compared with the observational estimate from OLS. The IV estimate is therefore less efficient than the OLS estimate if there is no confounding, reverse causality or measurement error, but unlike the OLS estimator the IV estimate should remain consistent when these are present [54, 55].

As discussed above, for the formal IV analysis we used an additive model for the effect of haplotype on log(CRP) that fitted the data well with a modest number of genetic variables (three indicator variables instead of the eight that would be needed for a saturated model of the nine diplotypes). It does not matter if this 'first-stage' does not represent a causal relationship [55] and its parameters are not of primary interest. Instead, it is often preferable to construct a parsimonious model to lessen the potential for problems of 'weak instruments' [68] (see Section 4.10).

The 2SLS IV analysis yielded an estimated causal effect for a doubling of CRP of 0.94 (95 per cent CI: 0.84, 1.05), suggesting that a doubling of CRP is associated with a 6 per cent reduction in HOMA-R, albeit with wide confidence intervals consistent with a marked inverse association or a positive association. Results were very similar using the other two IV estimators: maximum likelihood and GMM. The first-stage F-statistic was 13.7, suggesting that we should not have a problem with weak instruments. The important difference in magnitude between the confounder-adjusted association of HOMA-R on CRP from OLS regression analyses (1.08 (95 per cent CI: 1.07, 1.09)) and the IV analysis (0.94 (95 per cent CI: 0.84, 1.05)) confirms our concerns that

DOI: 10.1002/sim

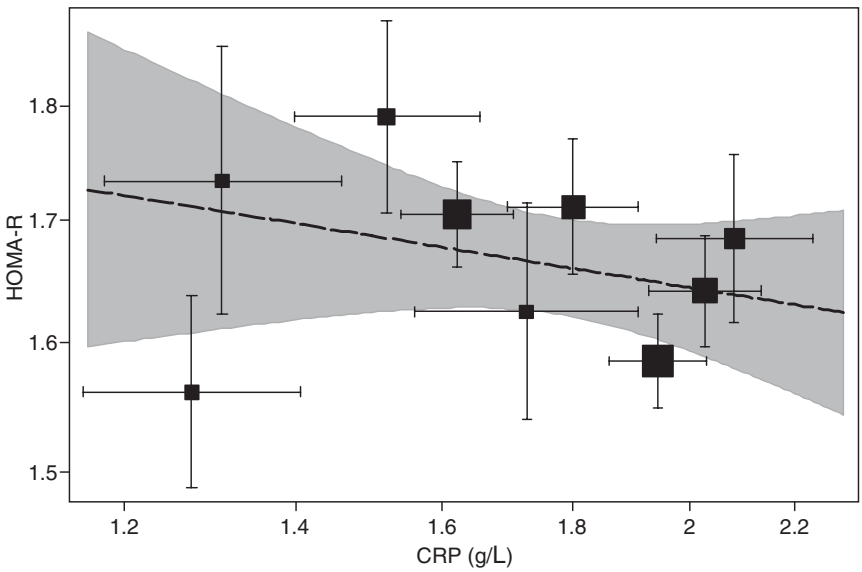

Figure 4. Illustration of IV analysis of the effect of CRP on HOMA-R score using diplotype (haplotype pair) as an instrument for CRP. Both axes are on a log scale but are greatly expanded compared with Figure 3. Squares show geometric mean CRP and HOMA-R by diplotype. The size of the square is scaled according to the number of women having each diplotype. Error bars show the standard errors of the geometric means. The thick dashed line gives the IV estimate (2SLS method) of the CRP—HOMA-R association, with the shading indicating the 95 per cent pointwise confidence interval around this estimate. The dashed line, but not the confidence interval, can also be obtained from weighted least-squares analysis of the logged geometric means.

the conventional observational association may be confounded or biased. The Durbin form of the Durbin–Wu–Hausman statistic provided statistical evidence that the IV estimate differed from that of the OLS regression estimate of effect ($p = 0.0139$). Corresponding tests for the LIML and GMM estimators gave similar results.

To summarize, this study suggests that circulating levels of CRP may not be causally associated with insulin resistance, and the difference in effect obtained from the conventional observational epidemiological approach (OLS regression of insulin resistance on circulating CRP) compared with that obtained from the Mendelian randomization IV approach suggests that the former was affected by residual confounding and/or reverse causality. This assertion is dependent upon the confidence that we had in *CRP* haplotypes fulfilling the assumptions, outlined in Section 2.1, that are required for it to be a valid IV for circulating CRP levels.

With respect to the first assumption—that *CRP* haplotypes are associated with CRP—this is supported by prior knowledge [70–78] and demonstrated within our own data set [31]. It is important to note that it is *not* a requirement for any IV analyses, or when genetic variation is used as an IV in Mendelian randomization studies, that the instrument (genetic variation in Mendelian randomization studies) is causally related to the risk factor of interest. Thus, in this specific example, the assumption is met because of both prior knowledge (well-replicated robust associations) and demonstration within our own data set of an association between *CRP* haplotypes and CRP levels.

DOI: 10.1002/sim

With respect to the second assumption—that *CRP* haplotype is unrelated to factors that are likely to confound the association between CRP levels and insulin resistance—the independent assortment of alleles from parents to offspring at conception provides the theoretical justification for this. Further, within our data set we have shown that *CRP* haplotypes and individual SNPs are not associated with established socioeconomic and lifestyle-confounding factors [30, 31].

The third assumption—that *CRP* haplotype is unrelated to insulin resistance, conditional upon circulating CRP levels and confounders of the CRP–insulin resistance association—is harder to prove. This requires the belief that there are no other pathways between *CRP* haplotype and insulin resistance than the one *via* CRP levels shown in Figure 2. Theoretical possibilities that could result in violation of this assumption through pleiotrophy, population stratification, LD and developmental canalization are discussed in more detail in Section 4. Further work in our study has confirmed that the variants used to generate the *CRP* haplotypes used here as IVs are in very close LD with variation within a transcription factor binding site located 5′ of the *CRP* gene that is associated with circulating concentrations of CRP and thought to be functional [79, 80]. It is unlikely that the variations in circulating CRP associated with this marker (or those in LD with it) are substantially involved in pleiotrophic events because of their role as a transcription factor binding site. Regarding population stratification, there is no published evidence that *CRP* has been importantly influenced by population selection. Further, the 23 centres used in this study were British towns that had a history of stable populations with little or no migration. All women included in our analyses were described by the examining nurse as 'white'. Thus, population stratification is unlikely to have resulted in a violation of assumption 3. Finally, we are currently extending this work by using a multiple IV approach to explore the extent to which our findings and conclusions are likely to be robust (see Section 4.7).

4. POTENTIAL PROBLEMS AND LIMITATIONS OF MENDELIAN RANDOMIZATION STUDIES

There are sound biological reasons for believing that genotype will not be associated with the socioeconomic and behavioural characteristics that commonly confound the effects of exposures on outcome, and this makes genotype a potentially valid IV where these confounders are unmeasured or measured with error. However, there are situations, summarized in Table II, in which Mendelian randomization (IV) assumptions (assumptions 1–3 above) may be violated. Table II also summarizes methods that can be employed to test whether violations are likely.

4.1. Finding a suitable genetic variant (instrument) for the modifiable exposure of interest

Mendelian randomization studies can be used only when there are established functional genetic variants that affect the modifiable exposure of interest. In the example given in Section 3 above, we genotyped women for variants in the *CRP* gene, which we knew from previous studies was related to CRP levels. If we wanted to use Mendelian randomization approaches to determine, for example, the magnitude of the causal association between vitamin D levels and insulin resistance, we would first need to find a genetic variant that we knew was associated with vitamin D levels and ensure that this association was robust (see Section 4.2). While it is likely that genotypes that

DOI: 10.1002/sim

Table II. Potential ways in which Mendelian randomization or instrumental variable assumptions may be violated.

Possible cause of violation	Approaches to assess and/or deal with possible violation
No suitable genetic variant (instrument)	Variant required. Progress in genomics means that in the next 5–10 years an increasing number of genetic variants (instruments) for modifiable exposures will become available
Unreliable gene associations	The Mendelian randomization approach requires variants that have had their association with the modifiable exposure of interest demonstrated and replicated in several independent data sets (i.e. needs this association to be reliable) The association should also be demonstrated within the Mendelian randomization study population
Social pressures on behaviours affected by genotype	Demonstrate association of genetic variant (instrument) with a modifiable behaviour (exposure) of interest in the population under study
Population stratification	Stratified analysis or analyses in homogeneous populations Use of genetic markers known to be related to stratification and/or ancestry. Statistical methods for detecting evidence of sub-populations within a sample Multiple instruments (genetic variants) Mendelian randomization studies within parent–offspring groups
Linkage disequilibrium	Knowledge of function of variant Multiple instruments (genetic variants)
Pleiotropy	Knowledge of function of variant Multiple instruments (genetic variants)
Developmental compensation	Extent of problem in Mendelian randomization studies unclear Requires understanding of developmental periods during which expression of gene and the age of effect are influential
Non-linear associations	Examine risk differences (rather than ratios) Latent variables approach Non-parametric (probit) models Bounds of effect Requires further research of use of these methods (widely used in econometrics) to the Mendelian randomization context
Weak instruments	F-statistic from first-stage regression (if less than 10, consider other approaches) Permutation method Other (e.g. likelihood based) methods Requires further research and exploration in Mendelian randomization studies

influence the levels or function of a majority of the modifiable exposures of interest in epidemiology will be discovered, only a small number of such genotypes have so far been described. The rapid expansion of knowledge in functional genomics makes it likely that in the next few years the potential of Mendelian randomization will be realized in an increasing number of areas.

DOI: 10.1002/sim

4.2. Ensuring gene associations are reliable

The Mendelian randomization approach requires reliable estimation of the associations between genotype and the modifiable exposure of interest and between genotype and the outcome of interest, but it is well recognized that genetic association studies often fail to replicate [81–83]. Lack of power related to the modest effects of genotype for most complex phenotypes, chance findings and publication bias contribute importantly to the noted failures to replicate initial gene association studies [82]. Strategies for improving this situation have been discussed in detail elsewhere and are beginning to be implemented [82]. In Mendelian randomization studies, one should be using genotypes with reliably established (replicated in several independent studies) associations with levels or function of the modifiable exposure of interest. This association should, where possible, be confirmed within the Mendelian randomization study. Thus, in the example given in Section 3 above, we knew that *CRP* had been shown to be associated with CRP levels in several independent data sets [70–78], and we demonstrated this established association within our own data set.

Unreliable estimates can also result from genotyping errors, when the observed genotype of an individual does not correspond to his or her true genotype. Such errors can result from a number of different reasons, several of which can operate together, including poor quality or quantity of DNA, biochemical artefacts, faulty equipment and human error in reading outputs and entering data [84]. Methods of assessing and maintaining high levels of quality control in genotyping, including typing of blind duplicate samples, comparison with standard DNA samples and data-entry checks, are described in detail elsewhere [84]. As discussed in Section 2.2, an IV estimate can be obtained by estimating $Z–X$ in a study sample different from that from which the $Z–Y$ estimate is obtained. However, an advantage of obtaining both estimates from the same study sample is that this provides the opportunity to demonstrate that genotyping errors are unlikely to explain null genotype–outcome associations. In the example presented in Section 3, the association between *CRP* haplotypes and circulating levels of CRP is in the same direction and of the same magnitude as of the same association found in previous studies. It is therefore not tenable to suggest that the lack of association between *CRP* haplotypes and insulin resistance is due to genotyping errors.

4.3. Social pressures on behaviours affected by genotype

For modifiable exposures of interest that are biological phenotypes, such as circulating CRP, individuals are unlikely to be aware of the effect of any genotype on phenotype and therefore will be unlikely to modify their lifestyles to change the phenotypic response to genotype. In the situation where genotypes are being used as an instrument for a behaviour, e.g. alcohol consumption, it is very important to test and demonstrate that the genotype does indeed result in the anticipated behaviour change in the population under study. For example, one could imagine that in populations where there are strong social pressures to consume moderate to large amounts of alcohol, any genotype that resulted in unpleasant symptoms after consuming alcohol may have limited effect on actual consumption (the social pressures to consume alcohol might out-weigh any unpleasant symptoms). Such effects are likely to be population specific. Thus, in the British Women's Heart and Health Study, a prospective cohort study of women born in Britain in the 1920s and 1930s [85], lactase-persistent genetic variants, which have been shown in other populations to be related to milk consumption [86–88], are not related to whether the women avoid milk (Davey Smith G, Lawlor DA *et al*, unpublished data, available from authors). While this lack

DOI: 10.1002/sim

of association might be related to measurement error in our assessment of milk consumption, it is also conceivable that in women of this generation, who experienced food rationing during the second world war and for some years afterwards, any gastrointestinal discomfort resulting from lactose intolerance is ignored because of the previous social pressures not to be a 'fussy eater'. Consequently in this population, and in any other population in which the variants do not relate to behaviour, lactase-persistent genetic variants will not be good instruments for determining the causal effect of milk consumption on outcomes such as osteoporosis, insulin resistance and diabetes [89]. The important point is that, with behavioural exposures in particular, it is essential to demonstrate the association of genotype with exposure in the population under study. Again, this highlights an advantage of obtaining both the Z–X and Z–Y estimates from the same study sample.

4.4. Population stratification

Population stratification occurs when there exist population subgroups that experience both different disease rates (or different distributions of traits) and have different frequencies of alleles of interest. This can result in spurious (confounded) associations between genotype and disease in the whole study population [90]. For example, in a study of nearly 5000 Native Americans of the Pima and Papago tribes, a strong inverse association between the HLA Gm Gm3;5,13,14 haplotype and type 2 diabetes was found (odds ratio 0.27; 95 per cent CI: 0.18, 0.40) [91]. This could have been interpreted as demonstrating that absence of this haplotype was causally associated with disease risk. However, the authors went on to demonstrate that the occurrence of both type 2 diabetes and the HLA haplotype was strongly related to ethnicity. In those of full American Indian heritage, haplotype frequency was 1 per cent and type 2 diabetes prevalence was 40 per cent, whereas in the Caucasian population in whom the haplotype frequency was 66 per cent, the prevalence of type 2 diabetes was 15 per cent. When analyses were undertaken within ethnic strata, there was no association between haplotype and diabetes [91].

The potential relevance of this to Mendelian randomization studies is that it could result in confounded results (see DAG in Figure 5), particularly in the case where the gene (IV)-modifiable risk factor association was undertaken in a different study population than the gene–outcome association. Consider the above example and imagine that the investigators had not gone on to demonstrate that the nature of the gene–diabetes association was in fact spurious, but had instead published this as a study demonstrating that this haplotype was associated with diabetes. If this haplotype were then used as an IV for type 2 diabetes in a separate study population, one that was ethnically homogeneous, and it were only possible to examine the association of the haplotype with cardiovascular disease outcomes (with an aim of determining whether diabetes is causally related to these outcomes), a null association would be wrongly interpreted as suggesting that diabetes was not causally related to cardiovascular disease. The null result would in fact occur because the haplotype is not really associated with diabetes and is therefore not suitable as an IV for its effect.

Concerns that population stratification may lead to false-positive inferences in population-based genetic association studies, and biased findings in Mendelian randomization studies, have led to calls for a greater use of family-based studies in genetic epidemiology in general [92] and also in Mendelian randomization studies [4]. However, while the example described above is a 'classic', which is frequently used to illustrate the potential of population stratification, other examples are hard to find in the literature and there is a general perception that population stratification is

DOI: 10.1002/sim

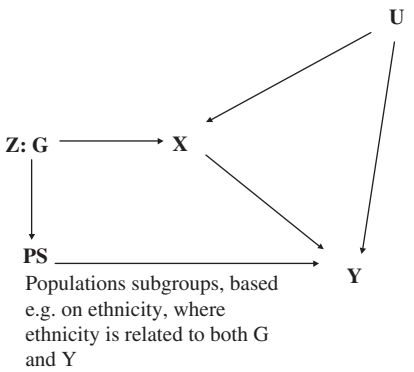

Figure 5. Illustration of Mendelian randomization study in which there is population stratification that results in violation of instrumental variable and Mendelian randomization assumptions. The assumptions are violated because there are sub-populations within the study sample (for example sub-groups of different ethnicities or ancestries) that have different allele frequencies for the variant being used as an instrumental variable and different levels of disease risk (unrelated to their genotype). Thus, these sub-groups 'confound' the association between IV and outcome and cause a violation of assumption 3. Z: G genetic variant that is being used as the instrumental variable; PS, population stratification, e.g. due to sub-groups in the sample that differ by ethnicity and where different ethnic groups have differing allele frequencies for G and differing levels of Y: X, modifiable exposure of interest; Y, outcome of interest; U, (unmeasured or measured with error), confounders.

unlikely to be a problem in most situations [93]. The authors of a recent review of population stratification concluded that

> A great deal of research effort seems to have been compromised to protect against a confounding factor that never realised its potential to bias allelic associations of complex traits [90].

Some caution may be warranted before applying these conclusions to Mendelian randomization, however, as a degree of population stratification too small to compromise conventional genetic association studies might still be enough to cause problems for Mendelian randomization, particularly when the effect of genotype on the modifiable exposure of interest is small (the instrument is weak). Furthermore, some of the genetic variants that are considered attractive candidates in Mendelian randomization studies—e.g. ALDH2 null variant, which is associated with alcohol consumption (individuals who are homozygous for the allele, resulting in inefficient breakdown of acetaldehyde, have severe adverse reactions to alcohol consumption and therefore tend to drink very little) [94]; lactase-persistence variants, which are related to lactose intolerance phenotype and would therefore be expected to be related to milk consumption [95]; and PTC-tasting variants, which are associated with bitter taste and might be expected to be related to dietary factors [96]—are ones that may be particularly prone to population stratification [97–100]. A salient example is that of Campbell *et al*. in which they found a positive association between procession of the T allele of *LCT*—13910C $\rightarrow$ T variant, which is associated with lactase persistence, and height in a European American population, all of whom had described themselves as 'white' or 'Caucasian' and all of whom had grandparents born in either the U.S. or Europe [100]. Use of conventional methods for detecting population stratification—including comparing observed with expected χ^2 association statistics for a panel of 111 random SNPs, tests of stratification using

DOI: 10.1002/sim

ancestry informative markers (AIM) and statistical packages designed to test for population structure—all suggested that there was no evidence of population stratification [100]. Since milk consumption is likely to be greater in those with the T allele of this variant, there is a biologically plausible reason for the association of this allele with height (better bone health/skeletal development might result in taller stature as a consequence of greater dietary calcium in those with genetically determined reduced risk of lactose intolerance). However, the investigators were concerned that because the allele frequency of this variant varies greatly within European populations (from 5–10 per cent in Southern Europe to 70–80 per cent in Northern Europe), along a similar pattern as variation in height within Europe, some or all of the association might be explained by population stratification. To explore this further, they stratified their sample on the basis of grandparental ancestry and found that the association of *LTC* with height diminished considerably. Further, when they examined this association in a sample from Poland and among family trios, all from Scandinavia, they found no association of the genetic variant with height. They concluded that the original association was largely or completely due to population stratification [100].

A number of methods are available in general genetic association studies for examining the possibility of population stratification. These include comparing observed with expected χ^2 association statistics for a number of random markers that have been used previously to examine stratification [101, 102] tests of stratification using AIM, which will have more power at detecting population stratification than random markers [103–105], and statistical packages, that use the genetic data in a study to search for evidence of sub-groups (sub-populations) within the study sample [106]. These methods can be used in Mendelian randomization studies, but, as the example above demonstrates [100], they are not fool proof. More recently Epstein has demonstrated a simple and improved method for dealing with population stratification, which uses a stratification score (similar to a propensity score) that works on the *LCT* and height data originally explored by Campbell *et al*. [107]. Studies performed within family parental–offspring triads would avoid problems of population stratification in Mendelian randomization studies, and as knowledge increases about the human genome we are likely to be in a better position to know which variants are most likely to have been affected by selection in the short-to medium-term future.

4.5. Linkage disequilibrium

As discussed in Section 1, LD is the correlation between allelic states at different loci within the population. The term LD describes a state that represents a departure from the hypothetical situation in which all loci exhibit complete independence (LE). Traditionally, LD has been used specifically to describe the correlation of genetic loci resulting from their close proximity to each other on non-homologous chromosomes. More recently, the term has been applied to any situation in which there is correlation between genetic variants, irrespective of whether this is due to physical genetic linkage. For example, many authorities use LD for association generated by population stratification [108]. In this paper we have used the definition in its more narrow and traditional expression to mean physically close alleles that are co-inherited and discuss issues of population stratification separately (Section 4.4).

When the genetic variant under study is correlated (shows high values of departure from LE and is described as being 'in LD') with another polymorphic locus and this locus influences the outcome of interest in the genetic association study, this may result in confounding. In Mendelian randomization studies, if the measured genotype (the one being used as an instrument) is correlated with a second polymorphism that is related to the modifiable exposure of interest (X), and X is

DOI: 10.1002/sim

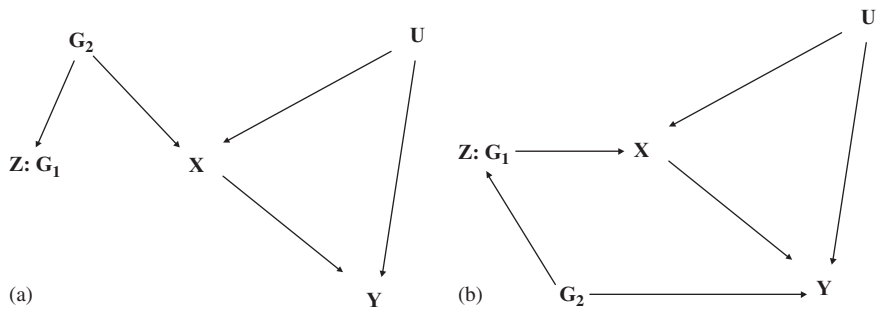

Figure 6. (a) Illustration of Mendelian randomization study in which there is linkage disequilibrium but where assumptions of instrumental variables analysis are *not* violated. The assumptions are not violated here because the genetic variant being used as an instrument is in linkage disequilibrium with a functional variant related to the exposure of interest and therefore is still a valid instrument. Assumptions are violated if the variant being used as an instrument is in linkage disequilibrium with a variant that is related to the outcome (see (b)). Z: G_1 genetic variant that is being used as the instrumental variable; G_2, genetic variant in linkage disequilibrium with G_1 and related to X, but not related to Y; X, modifiable exposure of interest; Y, outcome of interest; U, (unmeasured or measured with error) confounders. (b) Illustration of Mendelian randomization study in which there is linkage disequilibrium that leads to violation of the instrumental variables analysis. The assumptions are violated here because the genetic variant being used as an instrument is in linkage disequilibrium with a functional variant related to the outcome of interest (compare this with (a)). Z: G_1 genetic variant that is being used as the instrumental variable; G_2, genetic variant in linkage disequilibrium with G_1 and related to Y; X, modifiable exposure of interest; Y, outcome of interest; U, (unmeasured or measured with error) confounders.

observed so that the association Z–X can be estimated, none of the core IV assumptions (1)–(3) nor the underlying principles of Mendelian randomization will be violated. This is illustrated in the DAG (Figure 6(a)) and stems from the fact that an IV Z need not be *causally* associated with the levels or function of the exposure X of interest [9, 48, 109]. Indeed, in many Mendelian randomization studies this form of LD is exploited, in that it is LD of a genetic marker (a detectable genetic signal) with a functional variant that is driving the signal being used to characterize the IV.

However, if there is LD between the genotype being used as an instrument and a polymorphism that is associated with the *outcome* (Y), then assumption (3) will be violated and the estimate of the causal association between X and Y obtained from the IV analysis will be confounded. This is illustrated in the DAG in Figure 6(b). This latter situation is an extended example of pleiotropy, which is described in more detail in Section 4.6.

4.6. Pleiotropy

As seen in the case of LD, pleiotropy refers to a genetic variant having multiple functions. While pleiotropy is common, in the context of a Mendelian randomization study this will *not* result in violation of the core assumptions if the variant is associated with pleiotropic effects that do not (other than *via* the modifiable exposure of interest) influence the outcome (see DAG in Figure 7(a)); genetic variants with pleiotropic effects that do influence the outcome will, however, invalidate the Mendelian randomization approach (see DAG in Figure 7(b)). In the example presented in Section 2, for pleiotropy to explain the lack of evidence of a causal effect from the IV estimate,

DOI: 10.1002/sim

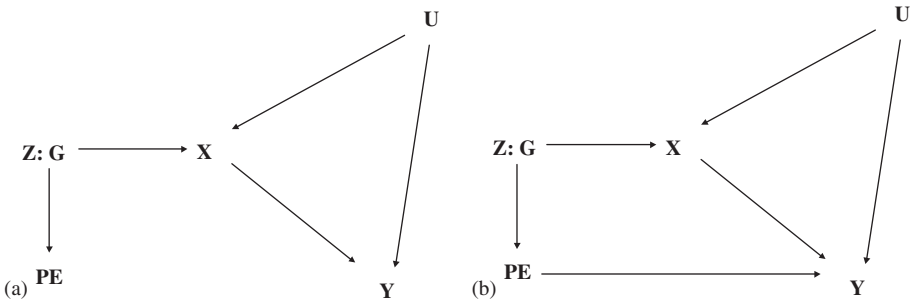

Figure 7. (a) Illustration of Mendelian randomization study in which there is pleiotropy but where assumptions of instrumental variables analysis are *not* violated. The assumptions are not violated here because the other functions that the genetic variant is related to are not in any way associated with the outcome (compare this with (b)). (b) Illustration of Mendelian randomization study in which there is pleiotropy and the assumptions of instrumental variables analysis are violated. The assumptions are violated here because the other functions that the genetic variant is related to are associated with the outcome (compare this with (a)). Z, G genetic variant that is being used as the instrumental variable; PE, pleiotropic effects of G that are not related to Y; X, modifiable exposure of interest; Y, outcome of interest; U, (unmeasured or measured with error) confounders.

there would have to be other functions of *CRP* haplotype (other than *via* CRP levels) that resulted in a decrease in insulin resistance to a magnitude that counterbalanced a putative positive effect of CRP on insulin resistance.

In some cases pleiotropy may distort Mendelian randomization studies. Consider the use of *APOE* genotype as an IV for examining the causal role of high-density lipoprotein cholesterol (HDLc) and low-density lipoprotein cholesterol (LDLc) on myocardial infarction risk [1]. The causal effect of these lipids on myocardial infarction risk is established from RCTs of the effects of statins, and the association of *APOE* with the apolipoprotein carriers of HDLc and LDLc (Apo A-1 and Apo B, respectively) is known. It was surprising then that the association of *APOE* with myocardial infarction was weaker than would be predicted from the RCT evidence of the causal effects of these on apolipoprotein and lipids [41]. One plausible explanation for this is pleiotropy [1]. The ε2 variant of *APOE*, as well as its association with lower LDLc and higher HDLc (both of which would result in reduced myocardial infarction risk), is also related to less efficient transfer of very low-density lipoproteins and chylomicrons from the blood to the liver, greater postprandial lipaemia, and an increased risk of type III hyperlipoproteinaemia (all of which would result in increased myocardial infarction risk) [1]. Thus, the multiple effects of *APOE*, each affecting myocardial infarction risk, make it unsuitable as an IV to examine the causal effect of any one of these pathways on myocardial infarction (equivalent to DAG in Figure 7(b)).

4.7. Using multiple genes to deal with population stratification, LD and pleiotropy

The specific problems of population stratification, LD and pleiotropy described above can to some extent be dealt with by using ethnically homogenous study populations, identifying and incorporating the population strata in the analysis and ensuring that the function of the genetic variant is fully understood (Table II). In addition, multiple instruments (genetic variants) can

DOI: 10.1002/sim

be used to examine the possibility that the Mendelian randomization/IV assumptions have been violated. Two or more genotypes known to functionally affect the modifiable exposure *via* different biological pathways can be used as multiple instruments. For example, in future studies we aim to undertake a multiple instrument approach to examine the causal effect of CRP on insulin resistance and other outcomes using both *CRP* and *IL6*.

Both these loci are known to be related to CRP levels, but *via* different pathways. If the IV estimate from both instruments is the same then it is implausible that this is explained by population stratification, LD or pleiotropy since this would imply the same magnitude of confounding/bias by one or more of these mechanisms occurring in both the *CRP* and *IL6* associations. This is analogous to comparing the results from two RCTs of different blood pressure-lowering drugs (i.e. completely different classes of drugs, each lowering the blood pressure by a different mechanism) on the risk of stroke. If each trial demonstrates the same magnitude of effect (e.g. a 4 mmHg reduction in blood pressure being associated with a 10 per cent decreased risk in stroke over 10 years), then this strengthens the argument that blood pressure is really causally related to stroke risk by this magnitude of effect. It argues against either of the trials being distorted by problems such as lack of concealment of allocation, that can result in non-random allocation and confounding in poorly conducted RCTs. Furthermore, such evidence suggests that it is blood pressure lowering itself—rather than any other effect of a particular antihypertensive agent—that influences stroke risk. As Imbens and Rosenbaum have said, ... *if each instrument is plausible but not certain* (with respect to complying with the basic assumptions of IV analysis), *there may be no reason why these different instruments should be biased in the same direction* [110].

The use of multiple instruments is common in the field of econometrics, where they are combined efficiently to obtain a single IV estimate of the magnitude of the causal effect, leading to a reduced variance of the IV estimator (i.e. their use is primarily to increase the precision of the causal estimator) [55]. When there are multiple instruments, it is also possible to construct a test of assumptions (2) and (3) by checking the estimates obtained using different instruments that are consistent (testing the so-called 'overidentifying restrictions' [54–56]). The use of multiple genes in Mendelian randomization studies to specifically test for violation of assumptions 2 and 3 due to population stratification, LD and pleiotropy is a developing area requiring further exploration.

4.8. Canalization (developmental compensation) and violation of Mendelian randomization/IV assumptions

Canalization (also known as developmental compensation) refers to the buffering of the effects of either environmental or genetic forces attempting to perturb development [111–114]. Several molecular mechanisms are likely to be involved in this buffering, including genetic redundancy, feedback regulation and cooperative biochemical interactions [113]. Whatever the mechanism, this buffering could invalidate findings from Mendelian randomization studies by altering the effect of a genotype on the outcome of interest in adulthood without any effect on the association between genotype and the modifiable exposure of interest. Hence, the estimate of association between Z and X (β_{ZX} in equation (1)) would be valid, whereas that between Z and Y (β_{ZY} in equation (1)) would not be valid; consequently, the IV estimate of causal effect would be biased. In effect, a polymorphism expressed during fetal development or post-natal growth may influence the expression of a wide range of other genes, leading to changes that may compensate for the effect of

DOI: 10.1002/sim

the polymorphism. For example, if an individual developed and grew (*in utero* or during the early post-natal period) within an environment in which one factor was disturbed (e.g. elevated CRP due to genotype), then he or she could be rendered resistant to the influence of lifelong elevated circulating CRP through permanent changes in tissue structure and function that counterbalance the effect of elevated CRP levels.

The most dramatic demonstrations of developmental compensation come from knockout studies in animals, in which a functional gene is rendered non-expressive but the animal shows less severe phenotypic effects than anticipated [1]. Unlike other potential violations of IV assumptions described above, a solution to the potential problem of developmental compensation is not easy to envisage. The use of multiple genotypes as multiple instruments is not an obvious solution, since the effects of all genotypes (e.g. 2 polymorphisms, *CRP* and *IL6* affecting CRP levels) resulting in perturbations during the developmental period are equally likely to be buffered. However, most examples of developmental compensation relate to dramatic genetic or environmental insults and it is unclear whether the generally smaller phenotypic differences induced by common functional polymorphisms will be sufficient to induce compensatory responses. Thus, it is unclear how important this theoretical problem will prove to be in practice. Nonetheless, as pointed out previously [1], when making analogies between Mendelian randomization studies and RCTs, it should be born in mind that in RCTs individuals are usually randomized to interventions in adulthood, after the developmental period during which compensation can occur, whereas in Mendelian randomization studies random allocation to alleles occurs at conception (i.e. at the very start of human development). By the same token, genetic variants that influence behaviours that first occur later in life, after developmental periods (for example, the effect of ALDH2 null variant on alcohol consumption), will not be affected by developmental compensation. It is, therefore, not a problem for these Mendelian randomization studies. Further, developmental decompensation will not be a problem when maternal genotype is being used as an IV for the effect of *in utero* exposures on later offspring outcomes. For example, the association of maternal *MTHFR* genotype with offspring neural tube defects illustrates the causal association of low levels of folic acid exposure during development on the offspring [1]. If there were some element of developmental decompensation for the effect of low folic acid levels on the risk of neural tube defects, this is appropriately (with respect to the need to know the causal effect) taken into account both in analyses using maternal genotype as an IV and in randomized controlled trials of the effect of periconceptual and antenatal folic acid supplementation on these neural tube defects.

4.9. Non-linear associations

Assumption (4) (linearity and absence of statistical interactions, see Section 2.2) can be particularly problematic for Mendelian randomization studies when the outcome is binary and we wish to express the causal effect as an odds ratio or risk ratio, as is often the case in epidemiology. Several solutions are possible. First, a linear model can be used to estimate the risk difference rather than the risk ratio [7, 9] (it might be argued that absolute causal effects should be presented more in epidemiology). Second, a full parametric model can be specified using a latent model formulation where the underlying outcome variable is assumed to be continuous and the observed binary outcome depends on whether a particular threshold has been exceeded (e.g. probit models) [55]. Such an approach requires additional assumptions (e.g. normality of unobservable error terms), and the sensitivity of Mendelian randomization estimates to violation of these assumptions is, as yet,

DOI: 10.1002/sim

unclear. While the natural regression parameter can be estimated with this latent variable approach, its causal interpretability may not be straightforward [9, 115, 116]. Third, non-parametric methods can be used to obtain bounds for the causal effect, at least when the instrument, exposure and disease are all categorical variables [9, 117], though this approach may prove less useful with the continuous exposures common in epidemiology [9, 117].

These IV methods for binary outcomes have been used in econometrics and for determining causal effects in RCTs when there are non-compliers [9, 48, 118, 119], but not yet, to our knowledge, in Mendelian randomization studies. Their utility in this context requires further exploration, and more methodological development may also be needed.

Even with continuously measured outcomes, the assumption of a linear effect of exposure on outcome may still be questioned. It will sometimes be useful to transform either or both of the exposure and outcome to make linearity assumptions more reasonable. However, it is often difficult to check linearity assumptions within the IV framework. As illustrated by Figures 3 and 4, the IV approach uses only the variation in the mean exposure *between* categories defined by the instrument, and such variation often covers a range of the exposure much smaller than the range in the population. A linear model will still often be a good way of testing the null hypothesis of no effect of exposure on outcome. But when the estimated size of the effect is of interest, the question arises as to what an IV estimator from a linear model is estimating when the true underlying exposure–outcome effect is non-linear, or when the effect differs between individuals. An interesting, if somewhat controversial [48], answer has been suggested in terms of a weighted average of the slope of the exposure–outcome effect over a range of the exposure determined by the instrument–exposure relationship [109, 120].

4.10. Weak instruments in Mendelian randomization studies

An instrument is 'weak' if it has only a slight effect on the exposure of interest (i.e. if the magnitude of β_{ZX} is small) [68, 69, 110]. If the instrument is extremely weak, it may provide little or no information. In principle, this should simply lead to very imprecise estimates of the causal effect. However, in recent years, econometricians have realized that with very weak instruments the usual asymptotic methods of IV estimation perform badly, giving biased estimates and confidence intervals that are too narrow, even in samples that would normally be thought large enough for asymptotic results to hold [68, 69, 110, 121, 122]. F-statistics from the first-stage regression of 2SLS can be used to check whether the combination of instrument strength and sample size will make asymptotic results reasonable. Values greater than 10 are often taken to indicate sufficient strength to ensure the validity of IV methods [68], although detailed tables are available showing how the bias of the IV estimate and the size of tests vary with instrument strength [54]. For a dichotomous instrument, for example, $F<10$ corresponds to $p<0.0015$, i.e. there needs to be strong evidence for an effect of gene on the modifiable exposure to be sure that weak instruments are not a problem. Imbens and Rosenbaum have recently suggested that permutation-based methods can be used to provide valid estimates and confidence intervals even with very weak instruments [110]. However, as yet, there is no software for performing such analyses. Other alternative approaches have also been suggested recently, and one based on the conditional likelihood ratio has been programmed in the statistical software package Stata [123, 124]. We would suggest that currently researchers using IV analyses in Mendelian randomization studies always report F-statistics from the first-stage regressions to examine the strength of the instrument and avoid making causal inferences, or seek advice from an IV expert where this is close to or less than 10. A useful

DOI: 10.1002/sim

overview of weak instruments was presented recently by Nichols [125], including new techniques for their diagnosis and for conducting valid inference in their presence.

4.11. Are Mendelian randomization studies invalidated if the exposure of interest is influenced by many genetic and environmental exposures?

The modifiable exposure of interest in Mendelian randomization studies is likely in most cases to be influenced by a large number of environmental and genetic factors. Some epidemiologists have suggested that this implies that Mendelian randomization cannot provide causal inferences. For example, Jousilahti and Salomaa have suggested that Mendelian randomization approaches cannot be used to examine the causal effect of fibrinogen on coronary heart disease because this approach does not take into account the complex genetic background of multifactorial disease, and fails to recognize that genetic factors and environmental factors other than the variant under study influence fibrinogen levels [126]. Note that the authors are *not* referring to any of the other problems discussed in Sections 4.1–4.10. The suggestion here is that because other factors (than the genetic variant) influence the exposure of interest (fibrinogen), the IV estimate cannot be used to provide a valid estimate of the causal effect of fibrinogen on CHD.

This is a misunderstanding of the theoretical underpinning of the Mendelian randomization/IV approach, which is used, in part, to 'control' for potential confounding and thus by definition takes into account the many factors that are known to influence the exposure of interest. If groups in the population are defined by a genetic variant that is known to affect levels of fibrinogen then these groups will consistently differ with respect to mean fibrinogen levels. This can be illustrated by a thought experiment, suggested previously by one of the authors [47]. Within a population the use of antihypertensive medicines (even if these are widely and appropriately used) will only make a small contribution to the variance in blood pressure. However, this does not imply that antihypertensive drugs will not have an influence on the clinical consequences (stroke, coronary heart disease) of raised blood pressure. The fact that many other environmental and genetic factors contribute to variation in blood pressure is irrelevant, as is the wide range of factors contributing to variance in fibrinogen levels in the example discussed by Jousilahti and Salomaa [126].

Groups with initially similar blood pressure that then differ with respect to antihypertensive drug use will differ with respect to final blood pressure, and this will result in differences in rates of stroke and coronary heart disease. Similarly, groups that differ with respect to a genetic polymorphism related to fibrinogen levels will differ with respect to fibrinogen level, and if fibrinogen level were a cause of coronary heart disease (in the way that blood pressure is indeed a cause of coronary heart disease) these groups should have different rates of disease [47]. To make this more concrete, consider an RCT in which half of a population are randomized to an antihypertensive treatment and the other half to a placebo. If the antihypertensive therapy reduced blood pressure by a quarter of a standard deviation, which is approximately the situation for such pharmacotherapy, then assignment with in the whole population being randomized to treatment or placebo group will explain 1.54 per cent of the variance in blood pressure. In the example we use of *CRP* haplotypes as an instrument for CRP levels, these haplotypes explain 1.66 per cent of the variance in CRP levels in the population. As can be clearly seen, the quantitative association of genetic variance as instruments can be similar to that of randomized treatments with respect to biological processes that are modified and influence disease risk. Both the logic and quantification fail to support Jousilahti and Salomaa's criticisms of the Mendelian randomization approach.

DOI: 10.1002/sim

5. STATISTICAL POWER IN MENDELIAN RANDOMIZATION STUDIES

Conventional power calculations can be modified for Mendelian randomization studies. If a reasonable estimate is available of the effect of the gene (instrument) on the exposure, equation (1) can be used to calculate size of the gene–outcome (Z–Y) association corresponding to the exposure–outcome (X–Y) effect it is desired to detect. Conventional power or sample size calculations can then be applied to this gene–outcome association. This approach involves approximations, but often approximating power or sample size is all that is required. The resulting sample sizes may well prove sobering, particularly when the gene–exposure effect is modest compared with the variability of the exposure in the population [127, 128]. Meta-analysis is therefore often valuable in order to combine information from several studies, but this approach may have its own problems [57]. We intend to return to these issues in more detail in a future paper.

6. CONCLUSIONS

Mendelian randomization studies aim to use genotypes that affect modifiable exposures to make causal inferences about the aetiological effects of these exposures. The statistical methods required for such inferences—IV methods—have been used widely in econometrics, but not in epidemiology. The use of genetic variants as instruments for modifiable exposures has the potential to avoid some of the limitations of observational epidemiology (confounding, reverse causality, regression dilution bias) and RCTs (problems with generalizability, unfeasible for some exposures, expensive) for making causal inferences. However, the approach has a number of potential limitations, including lack of suitable genetic variants, unreliable associations, population stratification, LD, pleiotropy, developmental canalization, the need for large sample sizes and some potential problems with binary outcomes. In this paper, we have discussed the likely extent of these potential limitations and offered some approaches to deal with them. We believe that there is considerable potential for Mendelian randomization studies to improve causal inference in epidemiology over the coming years. Realizing this potential will lead to utilization of the rapid expanse in genomic knowledge that will continue over the coming years, and the continuing adoption, adaptation and development of IV methods for use in Mendelian randomization studies will help realize this potential. We look forward to close collaboration between genetic and conventional epidemiologists, molecular biologists, statisticians and econometricians in pursuit of this goal.

ACKNOWLEDGEMENTS

This work was supported by the MRC Health Services Research Collaboration and the MRC. D.A.L. is funded by a U.K. Department of Health Career Scientist Award. The views expressed in this publication are those of the authors and not necessarily those of the MRC or any of the funding bodies. The funding bodies have had no influence over the scientific work or its publication. Frank Windmeijer (University of Bristol) provided useful comments on an earlier draft of this paper. The authors take full responsibility for the content of the paper.

REFERENCES

1. Davey Smith G, Ebrahim S. ‘Mendelian randomisation’: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 2003; **32**:1–22.

DOI: 10.1002/sim

2. Katan MB. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1986; **1**:507–508.
3. Youngman LD, Keavney BD, Palmer A. Plasma fibrinogen and fibrinogen genotypes in 4685 cases of myocardial infarction and in 6002 controls: test of causality by 'Mendelian randomization'. *Circulation* 2000; **102**(Suppl. II): 31–32.
4. Thomas DC, Conti DV. Commentary: the concept of 'Mendelian Randomization'. *International Journal of Epidemiology* 2004; **33**:21–25.
5. Thompson JR, Minelli C, Abrams KR, Tobin MD, Riley RD. Meta-analysis of genetic studies using Mendelian randomization—a multivariate approach. *Statistics in Medicine* 2005; **24**:2241–2254.
6. Bautista LE, Smeeth L, Hingorani AD, Casas JP. Estimation of bias in nongenetic observational studies using 'Mendelian triangulation'. *Annals of Epidemiology* 2006; **16**:675–680.
7. Thomas DC, Lawlor DA, Thompson JR. Re: estimation of bias in nongenetic observational studies using 'Mendelian triangulation' by Bautista *et al*. *Annals of Epidemiology* 2007; DOI: 10.1016/j.annepidem.2006.12.005.
8. Didelez V, Sheehan NA. *Mendelian Randomisation*: *Why Epidemiology Needs a Formal Language for Causality*. 2006. University College London, http://www.homepages.ucl.ac.uk/∼ucakvdi/kent.pdf.
9. Didelez V, Sheehan N. Mendelian randomisation and instrumental variables: what can and what can't be done. *Research Report 05-02*, Department of Health Sciences, University of Leicester, Leicester, 2005.
10. Doll R, Peto R, Boreham J, Sutherland I. Mortality from cancer in relation to smoking: 50 years observations on British doctors. *British Journal of Cancer* 2005; **92**:426–429.
11. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *British Medical Journal* 2004; **328**:1519.
12. MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J, Abbott R, Godwin J, Dyer A, Stamler J. Blood pressure, stroke, and coronary heart disease. Part 1. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; **335**:765–774.
13. Stamler J, Wentworth D, Neaton JD. Is the relationship between serum cholesterol and risk of premature death from coronary heart disease continuous or graded? *Journal of the American Medical Association* 1986; **256**:2823–2828.
14. Lawlor DA, Davey Smith G, Kundu D, Bruckdorfer KR, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *Lancet* 2004; **363**:1724–1727.
15. Vandenbroucke JP. Commentary: the HRT story: vindication of old epidemiological theory. *International Journal of Epidemiology* 2004; **33**:456–457.
16. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet* 2004; **363**:1728–1731.
17. Lawlor DA, Davey Smith G. Cardiovascular risk and hormone replacement therapy. *Current Opinion in Obstetrics and Gynaecology* 2006; **18**:658–665.
18. Phillips AN, Davey Smith G. Bias in relative odds estimation due to imprecise measurement of correlated exposures. *Statistics in Medicine* 1992; **11**:953–961.
19. Writing committee for the Women's Health Initiative randomized controlled trial. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association* 2002; **288**:321–333.
20. Davey Smith G. Capitalising on Mendelian randomisation to assess the effects of treatment. *James Lind Library* 2006. Available from: www.jameslindlibrary.org.
21. Mendel G. Experiments in plant hybridization (1865). Available from: http://www.mendelweb.org/archive/Mendel.Experiments.txt (accessed May 2007).
22. Thomas DC. *Statistical Methods in Genetic Epidemiology*. Oxford University Press: Oxford, 2004.
23. Ziegler A, Koenig IR. *A Statistical Approach to Genetic Epidemiology*: *Concepts and Applications*. Wiley-VCH: Weinheim, Germany, 2006.
24. Morgan TH, Sturtevant AH, Muller HJ, Bridges CB. *The Mechanism of Mendelian Heredity*. Henry Hold & Company: New York, 1915.
25. Bhatti P, Sigurdson AJ, Wang SS, Chen J, Rothman N, Hartge P, Bergen AW, Landi MT. Genetic variation and willingness to participate in epidemiologic research: data from three studies. *Cancer Epidemiology Biomarkers and Prevention* 2005; **14**:2449–2453.
26. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Medicine* 2007, in press.
27. Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology* 2004; **33**:30–42.

DOI: 10.1002/sim

28. Hingorani A, Humphries S. Nature's randomised trials. *Lancet* 2005; **366**:1906–1908.
29. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *British Medical Journal* 2005; **330**:1076–1079.
30. Davey Smith G, Lawlor DA, Harbord R, Rumley A, Lowe GDO, Day INM, Ebrahim S. Association of C-reactive protein with blood pressure and hypertension: lifecourse confounding and Mendelian randomisation tests of causality. *Arteriosclerosis Thrombosis and Vascular Biology* 2005; **25**:1051–1056.
31. Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day IN, Palmer LJ, Hattersley AT, Ebrahim S, Lowe GD, Rumley A, Davey Smith G. C-reactive protein and its role in metabolic syndrome: Mendelian randomisation study. *Lancet* 2005; **366**:1954–1959.
32. Kivimaki M, Lawlor DA, Eklund C, Davey Smith G, Hurme M, Lehtimaki T, Viikari JS, Raitakari OT. Mendelian randomization suggests no causal association between C-reactive protein and carotid intima-media thickness in the young Finns study. *Arteriosclerosis Thrombosis and Vascular Biology* 2007; **27**:978–979.
33. Davey Smith G, Harbord R, Milton J, Ebrahim S, Sterne JA. Does elevated plasma fibrinogen increase the risk of coronary heart disease? Evidence from a meta-analysis of genetic association studies. *Arteriosclerosis Thrombosis and Vascular Biology* 2005; **25**(10):2228–2233.
34. Lewis SJ, Davey Smith G. Alcohol, ALDH2, and esophageal cancer: a meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach. *Cancer Epidemiology Biomarkers and Prevention* 2005; **14**(8):1967–1971.
35. Lewis SJ, Zammit S, Gunnell D, Davey Smith G. A meta-analysis of the MTHFR C677T polymorphism and schizophrenia risk. *American Journal of Medical Genetics. Part B*: *Neuropsychiatric Genetics* 2005; **135**:2–4.
36. Lewis SJ, Baker I, Davey Smith G. Meta-analysis of vitamin D receptor polymorphisms and pulmonary tuberculosis risk. *International Journal of Tuberculosis and Lung Disease* 2005; **9**:1174–1177.
37. Lewis SJ, Harbord RM, Harris R, Davey Smith G. Meta-analyes of observational and genetic association studies of folate intakes or levels and breast cancer risk. *Journal of National Cancer Institutes* 2006; **98**:1607–1622.
38. Casas JP, Shah T, Cooper J, Hawe E, McMahon AD, Gaffney D, Packard CJ, O'Rielly DS, Juhan-Vague I, Judkin JS, Tremoli E, Margaglione M, Di Minno G, Hamsten A, Kooistra T, Stephens JW, Hurel SJ, Livingstone S, Colhoun HM, Miller GJ, Bautista LE, Meade T, Sattar N, Humphries SE, Hingorani AD. Insights into the nature of the CRP-coronary event association using Medelian randomization. *International Journal of Epidemiology* 2006; **35**:922–934.
39. Keavney B, Danesh J, Parish S, Palmer A, Clark S, Youngman L, Delepine M, Lathrop M, Peto R, Collins R. Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'. *International Journal of Epidemiology* 2006; **35**(4):935–943.
40. Staton J, Sayer M, Hankey GJ, Cole V, Thom J, Eikelboom JW. Protein Z gene polymorphisms, protein Z concentrations, and ischemic stroke. *Stroke* 2005; **36**:1123–1127.
41. Keavney B, Palmer A, Parish S, Clark S, Youngman L, Danesh J, McKenzie C, Delepine M, Lathrop M, Peto R, Collins R. Lipid-related genes and myocardial infarction in 4685 cases and 3460 controls: discrepancies between genotype, blood lipid concentrations, and coronary disease risk. *International Journal of Epidemiology* 2004; **33**:1002–1013.
42. Hagiwara T, Kono S, Yin G, Toyomura K, Nagano J, Mizoue T, Mibu R, Tanaka M, Kakeji Y, Maehara Y, Okamura T, Ikejiri K, Futami K, Yasunami Y, Maekawa T, Takenaka K, Ichimiya H, Imaizumi N. Genetic polymorphism in cytochrome P450 7A1 and risk of colorectal cancer: the Fukuoka Colorectal Cancer Study. *Cancer Research* 2005; **65**:2979–2982.
43. Al-Zahrani A, Sandhu MS, Luben RN, Thompson D, Baynes C, Pooley KA, Luccarini C, Munday H, Perkins B, Smith P, Pharoah PD, Wareham NJ, Easton DF, Ponder BA, Dunning AM. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. *Human Molecular Genetics* 2006; **15**:1–10.
44. Muntjewerff JW, Kahn RS, Blom HJ, Den HM. Homocysteine, methylenetetrahydrofolate reductase and risk of schizophrenia: a meta-analysis. *Molecular Psychiatry* 2006; **11**:143–149.
45. Frayling TM, Rafiq S, Murray A, Hurst AJ, Weedon MN, Henley W, Bandinelli S, Corsi AM, Ferrucci L, Guralnik JM, Wallace RB, Melzer D. An interleukin-18 polymorphism is associated with reduced serum concentrations and better physical functioning in older people. *Journal of Gerontology Series A—Biological Sciences and Medical Sciences* 2007; **62**:73–78.
46. Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD. Homocysteine and stroke: evidence on a causal link from Mendelian randomisation. *Lancet* 2005; **365**:224–232.
47. Davey Smith G. Randomised by (your) god: robust inference from an observational study design. *Journal of Epidemiology and Community Health* 2006; **60**:382–388.

DOI: 10.1002/sim

48. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association* 1996; **91**:444–472.
49. Moffitt RA. Identification of causal effects using instrumental variables: comment. *Journal of the American Statistical Association* 1996; **91**:462–465.
50. Robins JM, Greenland S. Identification of causal effects using instrumental variables: Comment. *Journal of the American Statistical Association* 1996; **91**:456–458.
51. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000; **29**:722–729.
52. Zohoori N, Savitz DA. Econometric approaches to epidemiologic data: relating endogeneity and unobserved heterogeneity to confounding. *Annals of Epidemiology* 1997; **7**:251–257.
53. Meyer PB. *Glossary of Research Economics*. Available from: http://econterms.com (accessed May 2007).
54. Murray MP. *Econometrics*: *A Modern Introduction*. Addison-Wesley: Boston, 2006.
55. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA, 2002.
56. Baum CF, Schaffer ME, Stillman S. Instrumental variables and GMM: estimation and testing. *Stata Journal* 2003; **3**:1–32.
57. Minelli C, Thompson JR, Tobin MD, Abrams KR. An integrated approach to the meta-analysis of genetic association studies using Mendelian randomization. *American Journal of Epidemiology* 2004; **160**:445–452.
58. Ridker PM, Hennekens CH, Buring JE, Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *New England Journal of Medicine* 2000; **342**:836–843.
59. Yudkin JS, Stehouwer CD, Emeis JJ, Coppack SW. C-reactive protein in healthy subjects: associations with obesity, insulin resistance, and endothelial dysfunction: a potential role for cytokines originating from adipose tissue? *Arteriosclerosis Thrombosis and Vascular Biology* 1999; **19**:972–978.
60. Festa A, D'Agostino Jr R, Howard G, Mykkanen L, Tracy R, Haffner SM. Chronic subclinical inflammation as part of the insulin resistance syndrome. The insulin resistance atherosclerosis study (IRAS). *Circulation* 2000; **102**:42–47.
61. Pickup JC, Crook MA. Is type II diabetes mellitus a disease of the innate immune system? *Diabetologia* 1998; **41**:1241–1248.
62. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. *Journal of the American Medical Association* 2001; **286**:327–334.
63. Barzilay JI, Abraham L, Heckbert SR, Cushman M, Kuller LH, Resnick HE, Tracy RP. The relation of markers of inflammation to the development of glucose disorders in the elderly: the Cardiovascular Health Study. *Diabetes* 2001; **50**:2384–2389.
64. Lawlor DA, Davey Smith G, Rumley A, Lowe GDO, Ebrahim S. Associations of fibrinogen and C-reactive protein with prevalent and incident coronary heart disease are attenuated by adjustment for confounding factors: British Women's Heart and Health study. *Thrombosis and Haemostasis* 2005; **93**:955–963.
65. Lowe GD, Pepys MB. C-reactive protein and cardiovascular disease: weighing the evidence. *Current Atherosclerosis Reports* 2006; **8**:421–428.
66. Davey Smith G, Timpson N, Lawlor DA. C-reactive protein and cardiovascular disease risk: still an unknown quantity? *Annals of Internal Medicine* 2006; **145**:70–72.
67. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Molecular Biology of Evolution* 1995; **12**:921–927.
68. Staiger D, Stock JH. Instrumental variables with weak instruments. *Econometrica* 1997; **65**:557–586.
69. Stock JH, Wright JH, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 2002; **20**:518–529.
70. Zee RY, Ridker PM. Polymorphism in the human C-reactive protein (CRP) gene, plasma concentrations of CRP, and the risk of future arterial thrombosis. *Atherosclerosis* 2002; **162**:217–219.
71. Brull DJ, Serrano N, Zito F, Jones L, Montgomery HE, Rumley A, Sharma P, Lowe GD, World MJ, Humphries SE, Hingorani AD. Human CRP gene polymorphism influences CRP levels: implications for the prediction and pathogenesis of coronary heart disease. *Arteriosclerosis Thrombosis Vascular Biology* 2003; **23**:2063–2069.
72. Russell AI, Cunninghame Graham DS, Shepherd C, Roberton CA, Whittaker J, Meeks J, Powell RJ, Isenberg DA, Walport MJ, Vyse TJ. Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. *Human Molecular Genetics* 2004; **13**:137–147.
73. Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M, Liu K, Williams OD, Iribarren C, Lewis EC, Fornage M, Boerwinkle E, Gross M, Jaquish C, Nickerson DA, Myers RM, Siscovick DS, Reiner AP. Polymorphisms within the C-reactive protein (CRP) promoter region are associated with plasma CRP levels. *American Journal of Human Genetics* 2005; **77**:64–77.

DOI: 10.1002/sim

74. Kovacs A, Green F, Hansson LO, Lundman P, Samnegard A, Boquist S, Ericsson CG, Watkins H, Hamsten A, Tornvall P. A novel common single nucleotide polymorphism in the promoter region of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. *Atherosclerosis* 2005; **178**:193–198.
75. Szalai AJ, Alarcon GS, Calvo-Alen J, Toloza SM, McCrory MA, Edberg JC, McGwin Jr G, Bastian HM, Fessler BJ, Vila LM, Kimberly RP, Reveille JD. Systemic lupus erythematosus in a multiethnic US Cohort (LUMINA): association between C-reactive protein (CRP) gene polymorphisms and vascular events. *Rheumatology* (*Oxford*) 2005; **44**:864–868.
76. Miller DT, Zee RY, Suk DJ, Kozlowski P, Chasman DI, Lazarus R, Cook NR, Ridker PM, Kwiatkowski DJ. Association of common CRP gene variants with CRP levels and cardiovascular events. *Annals of Human Genetics* 2005; **69**:623–638.
77. Suk HJ, Ridker PM, Cook NR, Zee RY. Relation of polymorphism within the C-reactive protein gene and plasma CRP levels. *Atherosclerosis* 2005; **178**:139–145.
78. Kathiresan S, Larson MG, Vasan RS, Guo CY, Gona P, Keaney Jr JF, Wilson PW, Newton-Cheh C, Musone SL, Camargo AL, Drake JA, Levy D, O'Donnell CJ, Hirschhorn JN, Benjamin EJ. Contribution of clinical correlates and 13 C-reactive protein gene polymorphisms to interindividual variability in serum C-reactive protein level. *Circulation* 2006; **113**:1415–1423.
79. Szalai AJ, Wu J, Lange EM, McCrory MA, Langefeld CD, Williams A, Zakharkin SO, George V, Allison DB, Cooper GS, Xie F, Fan Z, Edberg JC, Kimberly RP. Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. *Journal of Molecular Medicine* 2005; **83**:440–447.
80. Timpson NJ, Davey Smith G, Ebrahim S. Letter by Timpson *et al*. regarding article, 'Contribution of clinical correlates and 13 C-reactive protein gene polymorphisms to interindividual variability in serum C-reactive protein level'. *Circulation* 2006; **114**:e256.
81. Cardon LR, Bell JI. Association study designs for complex diseases. *Nature Reviews Genetics* 2001; **2**:91–99.
82. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes: can we avoid being swamped by spurious findings? *Lancet* 2003; **361**:865–872.
83. Tabor HK, Risch NJ, Myers RM. Opinion: candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics* 2002; **3**:391–397.
84. Pompanon F, Bonin A, Bellemain E, Taberlet P. Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* 2005; **6**:847–859.
85. Lawlor DA, Bedford C, Taylor M, Ebrahim S. Geographic variation in cardiovascular disease, risk factors and their control in older women: British Women's Heart and Health Study. *Journal of Epidemiology and Community Health* 2003; **57**:134–140.
86. Obermayer-Pietsch BM, Bonelli CM, Walter DE, Kuhn RJ, Fahrleitner-Pammer A, Berghold A, Goessler W, Stepan V, Dobnig H, Leb G, Renner W. Genetic predisposition for adult lactose intolerance and relation to diet, bone density, and bone fractures. *Journal of Bone and Mineral Research* 2004; **19**:42–47.
87. Rasinpera H, Savilahti E, Enattah NS, Kuokkanen M, Totterman N, Lindahl H, Jarvela I, Kolho KL. A genetic test which can be used to diagnose adult-type hypolactasia in children. *Gut* 2004; **53**:1571–1576.
88. Mainguet P, Faille I, Destrebecq L, Devogelaer JP, Nagant de DC. Lactose intolerance, calcium intake, and osteopenia. *Lancet* 1991; **338**:1156–1157.
89. Lawlor DA, Ebrahim S, Timpson N, Davey Smith G. Avoiding milk is associated with a reduced risk of insulin resistance and the metabolic syndrome: findings from the British Women's Heart and Health Study. *Diabetic Medicine* 2005; **22**:808–811.
90. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003; **361**:598–604.
91. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *American Journal of Human Genetics* 1988; **43**:520–526.
92. Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. *Journal of the National Cancer Institute Monographs* 1999; **26**:31–37.
93. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *Journal of the National Cancer Institute* 2000; **92**:1151–1158.
94. Takeshita T, Morimoto K. Self-reported alcohol-associated symptoms and drinking behavior in three ALDH2 genotypes among Japanese university students. *Alcoholism-Clinical and Experimental Research* 1999; **23**: 1065–1069.
95. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I. Identification of a variant associated with adult-type hypolactasia. *Nature Genetics* 2002; **30**:233–237.

DOI: 10.1002/sim

96. Kim U, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science* 2003; **299**:1221–1225.
97. Goedde HW, Agarwal DP, Fritze G, Meier-Tackmann D, Singh S, Beckmann G, Bhatia K, Chen LZ, Fang B, Lisker R. Distribution of ADH2 and ALDH2 genotypes in different populations. *Human Genetics* 1992; **88**:344–346.
98. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics* 2004; **74**:1111–1120.
99. Wooding S, Kim UK, Bamshad MJ, Larsen J, Jorde LB, Drayna D. Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. *American Journal of Human Genetics* 2004; **74**:637–646.
100. Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG, Hirschhorn JN. Demonstrating stratification in a European American population. *Nature Genetics* 2005; **37**:868–872.
101. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D. Assessing the impact of population stratification on genetic association studies. *Nature Genetics* 2004; **36**:388–393.
102. Reich DE, Goldstein DB. Detecting association in a case–control study while correcting for population stratification. *Genetic Epidemiology* 2001; **20**:4–16.
103. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. *American Journal of Human Genetics* 2003; **73**:1402–1422.
104. Pfaff CL, Kittles RA, Shriver MD. Adjusting for population structure in admixed populations. *Genetic Epidemiology* 2002; **22**:196–201.
105. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, Malasky MJ, Scafe C, Le E, De Jager PL, Mignault AA, Yi Z, De TG, Essex M, Sankale JL, Moore JH, Poku K, Phair JP, Goedert JJ, Vlahov D, Williams SM, Tishkoff SA, Winkler CA, De LV, Woodage T, Sninsky JJ, Hafler DA, Altshuler D, Gilbert DA, O'Brien SJ, Reich D. A high-density admixture map for disease gene discovery in African Americans. *American Journal of Human Genetics* 2004; **74**:1001–1013.
106. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000; **155**:945–959.
107. Epstein MP, Allen AS, Satten GA. A simple and improved correction for population stratification in case–control studies. *American Journal of Human Genetics* 2007; **80**:921–930.
108. Terwilliger JD, Goring HH. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Human Biology* 2000; **72**:63–132.
109. Angrist JD, Graddy K, Imbens GW. The interpretations of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies* 2000; **67**:499–527.
110. Imbens GR, Rosenbaum PR. Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society*, *Series A* (*Statistics in Society*) 2005; **168**:109–126.
111. Wilkins AS. Canalization: a molecular genetic perspective. *Bioessays* 1997; **19**:257–262.
112. Rutherford SL. From genotype to phenotype: buffering mechanisms and the storage of genetic information. *Bioessays* 2000; **22**:1095–1105.
113. Gibson G, Wagner G. Canalization in evolutionary genetics: a stabilizing theory? *Bioessays* 2000; **22**:372–380.
114. Debat V, David P. Mapping phenotypes: canalization, plasticity and developmental stability. *Trends in Ecology and Evolution* 2006; **16**:555–561.
115. Blundell R, Powell J. Endogeneity in semiparametric binary response models. *Review of Economic Studies* 2004; **71**:655–679.
116. Stanghelli E, Wermuth N. On identification of path analysis models with one hidden variable. *Biometrika* 2005; **92**:337–350.
117. Balke A, Pearl J. Counterfactual probabilities: computational methods, bounds and applications. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence,* Seattle, WA, 1994; 46–54.
118. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics—Theory and Methods* 1991; **20**:2609–2631.
119. Robins J, Rotnitzky A. Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika* 2004; **91**:763–783.
120. Angrist JD, Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 1995; **90**:431–442.

DOI: 10.1002/sim

121. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association* 1995; **90**:443–450.
122. Hahn JY, Hausman J. Weak instruments: diagnosis and cures in empirical econometrics. *American Economic Review* 2003; **93**:118–125.
123. Andrews DWK, Moreira MJ, Stock JH. Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 2006; **74**:715–752.
124. Mikusheva A, Poi BP. Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal* 2006; **6**:335–347.
125. Nichols A. *Weak Instruments*: *An Overview and New Techniques*. 5th North American Users Group, 24 July 2006. Available from: http://pped.org/austin/wiv.pdf.
126. Jousilahti P, Salomaa V. Fibrinogen, social position, and 'Mendelian randomisation'. *Journal of Epidemiology and Community Health* 2004; **58**:883.
127. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001; **358**:1356–1360.
128. Davey Smith G, Harbord R, Ebrahim S. Fibrinogen, C-reactive protein and coronary heart disease: does Mendelian randomization suggest the associations are non-causal? *Quarterly Journal of Medicine* 2004; **97**:163–166.

DOI: 10.1002/sim