

# Bayesian methods for meta-analysis of causal relationships estimated using genetic instrumental variables

CRP CHD Genetics Collaboration

November 20, 2009

## 1 Abstract

Genetic markers can be used as instrumental variables, in an analogous way to randomization in a clinical trial, to estimate the causal relationship between a phenotype and an outcome variable. Our purpose is to extend existing methods for such Mendelian randomization studies to the context of multiple genetic markers measured in multiple studies, based on analysis of individual participant data. First, for a single genetic marker in one study, we show that the usual ratio of coefficients approach can be reformulated as a regression with heterogeneous error in the explanatory variable. This can be implemented using a Bayesian approach, which is next extended to include multiple genetic markers. We then propose a hierarchical model for undertaking a meta-analysis of multiple studies, in which it is not necessary that the same genetic markers are measured in each study. This provides an overall estimate of the causal relationship between the phenotype and the outcome, and an assessment of its heterogeneity across studies. As an example, we estimate the causal relationship of blood concentrations of C-reactive protein on fibrinogen levels using data from eleven studies. These methods provide a flexible framework for efficient estimation of causal relationships derived from multiple studies. Issues discussed include weak instrument bias, analysis of binary outcome data such as disease risk, missing genetic data, and the use of haplotypes.

**Keywords:** Mendelian randomization; instrumental variables; causal association; meta-analysis; Bayesian methods

## 2 Introduction

In traditional observational epidemiological studies, associations between a risk factor or phenotype (X) and outcome (Y) are often biased by unmeasured confounders or reverse causation. Mendelian randomization [1] is a technique for using genetic markers (G) as instrumental variables (IV) to assess the true causal association without direct

experiment. It uses the random allocation of genes at conception in an analogous way to treatment assignment in a randomized control trial [2]. By finding genetic markers associated with levels of the phenotype, the different genotypes give rise to groups which, under certain assumptions, are randomly assigned and so independent of measured and unmeasured confounders ( $U$ ) and the effects of reverse causation. The assumptions underlying an IV analysis are depicted in the directed acyclic graph (DAG) in Figure 1. An IV is a variable  $G$  which is [3]:

- a) independent of any possible confounders (ie.  $G \perp\!\!\!\perp U$ ),
- b) associated with the phenotype (ie.  $G \not\perp\!\!\!\perp X$ ), and
- c) independent of the outcome given the phenotype and confounders (ie.  $G \perp\!\!\!\perp Y \mid X, U$ ).

There are many ways in which these assumptions may be violated [4], some of which we return to in the discussion.

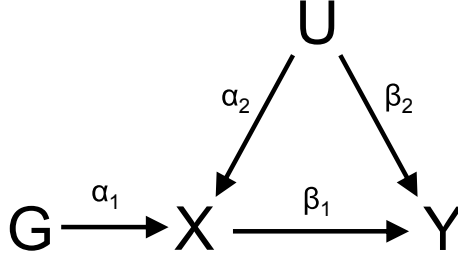


Figure 1: Directed acyclic graph (DAG) of Mendelian randomization assumptions

We consider the case where the outcome  $Y$  is continuous, and defer the issues relating to binary outcomes to the discussion. If the phenotype  $X$  and all confounders  $U$  are exactly measured, the causal association between  $X$  and  $Y$  ( $\beta_1$  in Figure 1) can be estimated using standard multiple regression of  $Y$  on  $X$  and  $U$ . In a study where  $G$ ,  $X$  and  $Y$  are measured, but some confounders  $U$  are not, several IV-based methods are available to estimate the causal association between  $X$  and  $Y$ . The most basic, the ratio of coefficients method [5, 3], can be used when there is one instrument, for example a single nucleotide polymorphism (SNP), providing genetic information: either the SNP is dichotomized (for example by combining the heterozygous group with one of the homozygous groups; recessive or dominant model), or linearity is assumed according to the number of variant alleles (additive model). In the latter case, the causal association is estimated as the ratio of the regression coefficient of the outcome  $Y$  on the number of variant alleles of the SNP to the regression coefficient of the phenotype  $X$  on the number of variant alleles of the SNP [4]. As the regression coefficients are asymptotically normal, approximate confidence intervals for the ratio can be calculated using Fieller's theorem [6] or the asymptotic variance can be estimated using a Taylor expansion [7]; the correlation between the two regression coefficients can be estimated by bootstrapping, but is often assumed to be zero since it is typically small [8].

The two-stage least squares (2SLS) method [4] can be used for multiple, polychotomous SNPs in one study. Least squares regression of the phenotype on the SNPs is used to obtain fitted values for the phenotype ( $\hat{X}|G$ ). Each SNP can be considered either as a continuous variable (per allele analysis) or as a factor with three levels (2df analysis). The effects of different SNPs can be combined additively; alternatively interactions can be included. The regression coefficient of the outcome  $Y$  on these fitted values for the phenotype is the estimate of the causal association. The point estimate from the 2SLS method performed per allele is equal to that from the ratio of coefficients method in the case of a linear effect of a single IV. If the uncertainty in the fitted values is ignored in the second stage regression, the standard error of the estimate of the causal association will be underestimated, and so a correction is needed [9]. To define confidence limits, the asymptotic normal distribution of the 2SLS estimator is used.

Methods based on genetic IVs are now being extensively used in practice. For example, they have been used for estimating the causal relationship between blood concentrations of C-reactive protein (CRP) and insulin resistance [10], CRP and carotid intima-media thickness [11], and folate levels and coronary heart disease [8]. A recurring problem is that the anticipated causal effects are only of moderate size, and the effects of genetic markers on the phenotype are typically small, so that IV techniques suffer from low power and poor precision. Typically sample sizes of tens of thousands are required [12, 13]. Meta-analysis of results from different studies is therefore often necessary, but current meta-analysis methods are restricted to studies all measuring the same single dichotomous or trichotomous SNP [8, 14].

We seek here to extend these established methods: firstly to gain power by using evidence from multiple studies, secondly to synthesize evidence across studies which use different SNPs as instrumental variables, thirdly to use multiple SNPs simultaneously [15], and finally to avoid the problems of “weak instruments” [16]. IV-based estimates using a weak instrument, where the association between phenotype and the IV is not statistically strong, suffer bias in the direction of the original observational association and deviation from an asymptotic normal to a more heavy-tailed distribution [17]. The F-statistic from the regression of phenotype on SNPs is commonly used as a measure of instrument strength [18].

We first describe a Bayesian approach to the estimation of causal effects using individual data on genetic characteristics. We present the simple case of a single genetic marker in one study (Section 3), and extend this to an analysis of multiple genetic markers in one study (Section 4). A hierarchical model for meta-analysis is then developed (Section 5) which efficiently deals with different genetic markers measured in different studies, and with heterogeneity between studies. The focus in this paper is on data from representative cross-sectional population studies; application to case-control studies is deferred to the discussion (Section 6).

### 3 A single genetic marker in one study

#### 3.1 Conventional methods

We first consider the case of a single SNP in one study, where confounding causes the observational estimate of the association of phenotype and outcome to be different from the causal relationship. Let individual  $i$  have phenotype level  $x_i$ , outcome  $y_i$ , genotype  $g_i$  taking a value in  $\{0,1,2\}$ , and unmeasured confounder  $u_i$ . We assume that all the confounders can be summarized by a single value  $u_i$ . Similarly to Palmer et al. [19], we consider the model represented in Figure 1:

$$\begin{aligned} x_i &= \alpha_0 + \alpha_1 g_i + \alpha_2 u_i + \epsilon_{xi} \\ y_i &= \beta_0 + \beta_1 x_i + \beta_2 u_i + \epsilon_{yi} \end{aligned} \tag{1}$$

with  $u_i \sim \mathcal{N}(0, \sigma_u^2)$ ,  $\epsilon_{xi} \sim \mathcal{N}(0, \sigma_1^2)$ ,  $\epsilon_{yi} \sim \mathcal{N}(0, \sigma_2^2)$  independently. As an example, we simulate data for a sample of size 300, containing 12 individuals with  $g_i = 2$ , 96 with  $g_i = 1$  and 192 with  $g_i = 0$ , corresponding to Hardy-Weinberg equilibrium for a minor allele frequency of 20%. We set the parameters  $(\alpha_0, \alpha_2, \beta_0, \beta_1, \beta_2, \sigma_u^2, \sigma_1^2, \sigma_2^2) = (0, 1, 0, 2, -3, 1, 0.25, 0.25)$ , and consider the cases of a weak instrument ( $\alpha_1 = 0.3$ , giving an expected F-value for the regression of X on G of 7), a moderate instrument ( $\alpha_1 = 0.5$ , F-value 20) and a strong instrument ( $\alpha_1 = 1$ , F-value 75). Figure 2 shows the simulated data grouped by genotype graphically.

The observational estimates obtained by regressing Y on X (Table 1) are far from the true causal association ( $\beta_1 = 2$ ) as expected because of the strong negative confounding (U is positively related to X but negatively to Y). The IV-based ratio method (assuming zero correlation between coefficients) gives estimates compatible with  $\beta_1 = 2$ , but with a wide confidence interval in the case of the weak or moderate instrument.

#### 3.2 A Bayesian method

Estimating the causal parameter by the ratio method is equivalent to determining the gradients in Figure 2 [4]. We can reformulate the problem as one of linear regression with heterogeneous error in X. For each genotype value  $j = 0, 1, 2$  we calculate the mean level of the phenotype  $\bar{x}_j$  with its variance  $\sigma_{xj}^2$  and mean outcome  $\bar{y}_j$  with its variance  $\sigma_{yj}^2$ . The model is:

$$\begin{aligned} \bar{x}_j &\sim \mathcal{N}(\xi_j, \sigma_{xj}^2) \\ \bar{y}_j &\sim \mathcal{N}(\eta_j, \sigma_{yj}^2) \\ \eta_j &= \beta_0 + \beta_1 \xi_j \end{aligned} \tag{2}$$

Thus we assume that each observed mean phenotype  $\bar{x}_j$  is from a normal distribution with unknown true mean  $\xi_j$  and known variance  $\sigma_{xj}^2$ , each observed mean outcome  $\bar{y}_j$  is from a normal distribution with unknown true mean  $\eta_j$  and known variance  $\sigma_{yj}^2$ , and

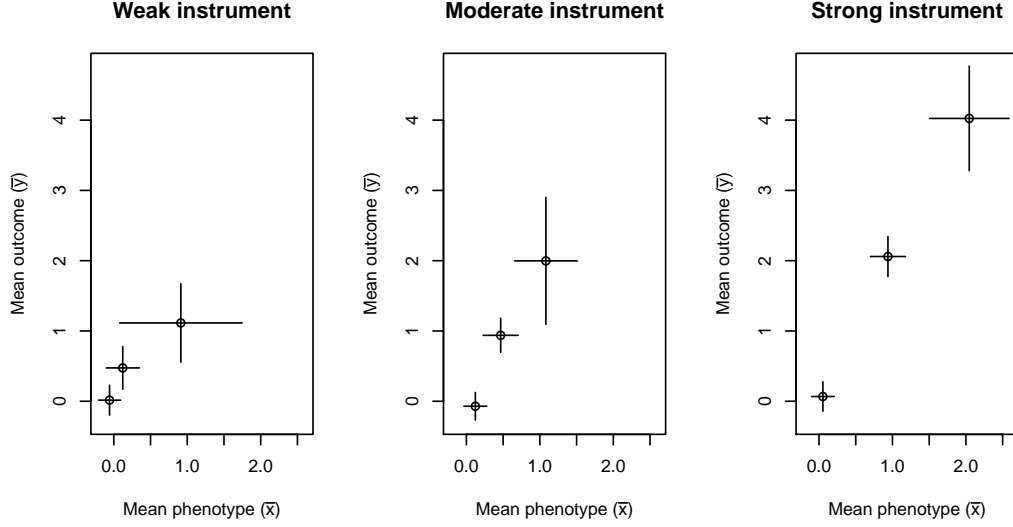


Figure 2: Graphs of mean outcome ( $\bar{y}$ ) against mean phenotype ( $\bar{x}$ ) in three genetic groups for the weak, moderate and strong instrument simulated examples of Section 3.1. Error bars are 95% CIs for the means.

there is a linear relationship between  $\eta$  and  $\xi$ .  $\beta_1$  represents the increase in outcome for unit increase in true phenotype and is the parameter of interest.

To implement this model, we employ Bayesian analysis and Markov Chain Monte Carlo (MCMC) methods with Gibbs sampling. This allows extension to more complicated situations, as in the next sections. We used vague priors (independent normals with zero mean and large variance of  $100^2$ ) for the regression parameters and each  $\xi_j$ . We performed this analysis in WinBUGS [20] using 150,000 iterations, discarding the first 1000 as “burn-in”, employing different starting values to assess convergence of the posterior distribution and sensitivity analyses to show lack of dependence on the prior distributions. The posterior distributions shown in Figure 3 are non-normal, with a heavier tail towards larger values especially for the weaker instruments. For this reason, the posterior median of the distribution of  $\beta_1$  is taken as the estimate of the causal association. Table 1 shows that the estimates and intervals from this Bayesian group-based method are similar to those from the ratio method. Other simulated examples (not shown) also demonstrated similar results. The 2SLS per allele method gives the same estimates as the ratio method, but the intervals are symmetric and so deviate from the ratio and Bayesian methods for the weaker instruments.

This Bayesian method assumes that the variances  $\sigma_{xj}^2$  and  $\sigma_{yj}^2$  are known, whereas in fact they need to be estimated from the data, an issue which is addressed in the next section.

Weak instrument - (F=7)	Estimate	95% CI/CrI
Observational estimate	-0.358	-0.506, -0.210
Ratio method	1.637	0.563, 6.582
Bayesian method	1.496	0.536, 7.190
2SLS method	1.637	-0.126, 3.400
Moderate instrument - (F=20)	Estimate	95% CI/CrI
Observational estimate	-0.251	-0.393, -0.109
Ratio method	2.555	1.481, 6.007
Bayesian method	2.417	1.473, 4.592
2SLS method	2.555	0.801, 4.309
Strong instrument - (F=75)	Estimate	95% CI/CrI
Observational estimate	0.108	-0.061, 0.276
Ratio method	2.136	1.632, 2.906
Bayesian method	2.107	1.633, 2.817
2SLS method	2.136	1.469, 2.804

Table 1: Causal parameter estimates and confidence/credible intervals using ratio, 2SLS and Bayesian methods compared with observational estimate for the weak, moderate and strong instrument simulated examples of Section 3.1

## 4 Multiple genetic markers in one study

### 4.1 Methods

If we have data in the study from more than one SNP then, provided they satisfy the IV assumptions above, all SNPs can be used simultaneously to divide the population into many subgroups. For each diallelic SNP, there are three genotypic categories, corresponding to 0, 1 or 2 variant alleles. For a dataset with  $n$  diallelic SNPs, we have a maximum  $3^n$  categories, for each of which we can measure the mean phenotype and outcome, and examine the regression as in (2) above to estimate  $\beta_1$ , the causal association. In practice, fewer than the maximum number of genotypic groups will be observed, due to correlation between SNPs caused by linkage disequilibrium (LD).

If the number of groups is large, and so their sizes  $n_j$  are small, then the assumption of exact knowledge of  $\sigma_{xj}^2$  and  $\sigma_{yj}^2$  for each group is not appropriate. Indeed if  $n_j = 1$ , the group-specific variance cannot even be calculated. It is then preferable to base the analysis on the standard deviation in the whole population for the phenotype ( $\sigma_x$ ) and the outcome ( $\sigma_y$ ), using an individual-based model for phenotype and outcome. For each individual  $i$  in category  $j$ , we have:

$$\begin{aligned}
x_{ij} &\sim \mathcal{N}(\xi_j, \sigma_x^2) \\
y_{ij} &\sim \mathcal{N}(\eta_j, \sigma_y^2) \\
\eta_j &= \beta_0 + \beta_1 \xi_j
\end{aligned} \tag{3}$$

The observed phenotype and outcome for each individual are here modelled using normal distributions, although other distributions might be more appropriate for some

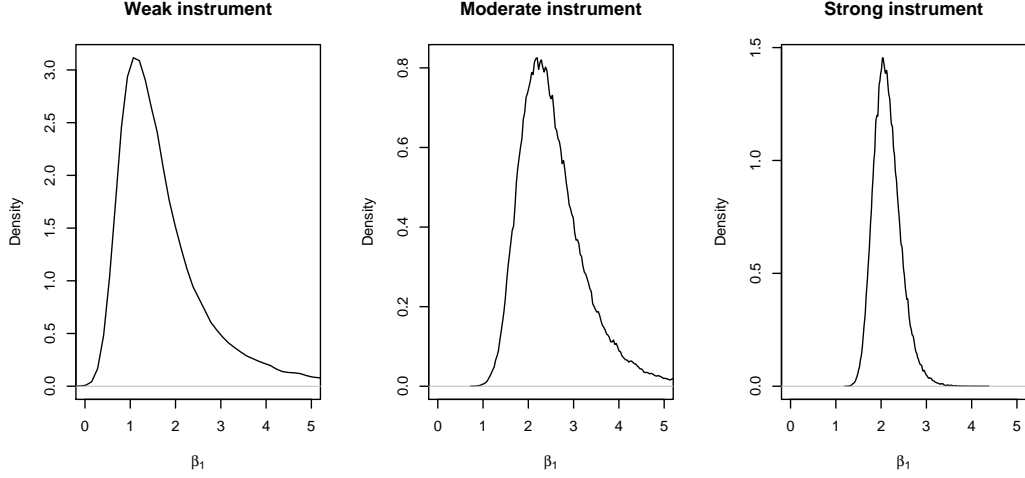


Figure 3: Kernel-smoothed density of posterior distribution of the causal parameter for the weak, moderate and strong instrument simulated examples of Section 3.1 using the Bayesian method of Section 3.2

applications. The information about  $\xi_j$  now depends on the population standard deviation for the phenotype as well as the size of the group. In the application below, vague Uniform[0,20] priors are used for  $\sigma_x$  and  $\sigma_y$ , while the other priors remain as before.

An alternative analysis is to assume a linear relationship between the phenotype and the number of variant alleles for each SNP which is also additive across SNPs. If this structure is appropriate, the analysis should be more efficient as the correlation between similar genotypes is accounted for and fewer parameters are estimated. Then we use these modelled values in the second stage regression. Writing  $G$  as the matrix of genotypes, so that  $G_{ik}$  is the number of variant alleles in SNP  $k$  for individual  $i$ , and  $\alpha_k$  as the first stage regression coefficients, then the model is:

$$\begin{aligned} \xi_i &= \alpha_0 + \sum_k \alpha_k G_{ik} \\ x_i &\sim \mathcal{N}(\xi_i, \sigma_x^2) \\ y_i &\sim \mathcal{N}(\eta_i, \sigma_y^2) \\ \eta_i &= \beta_0 + \beta_1 \xi_i \end{aligned} \tag{4}$$

Independent vague  $\mathcal{N}(0, 100^2)$  priors are now placed on the  $\alpha_k$  rather than the  $\xi_i$ . The values of  $\alpha_k$  depend, through feedback, on all the data including the outcome  $Y$ .

Models (3) and (4) are the equivalent of 2SLS in a Bayesian setting, except that there is feedback on the first stage coefficients from the second stage regression; the posterior distribution of the causal association parameter  $\beta_1$  naturally incorporates the uncertainty in the first stage regression, but with no assumption of asymptotic normality on its distribution.



## 4.2 Application to C-reactive protein and fibrinogen

C-reactive protein (CRP) is an acute-phase protein produced by the liver as part of the inflammation response pathway. Fibrinogen is a soluble blood plasma glycoprotein, which enables blood-clotting and is also associated with inflammation. The pathway of inflammation is not well understood, but is important as both CRP and fibrinogen are proposed as risk markers of coronary heart disease (CHD) [13]. Furthermore, although CRP is associated with CHD risk, this association reduces on adjustment for various risk factors, and attenuates to near null on adjustment for fibrinogen [21]. It is important therefore to assess whether CRP causally affects levels of fibrinogen, since if so adjusting for fibrinogen would represent an overadjustment. The CRP gene has several common variations which are associated with different blood concentrations of CRP. We use IV techniques to estimate the causal effect of CRP on fibrinogen. As CRP has a positively skewed distribution, we take its natural logarithm, and assume a linear relationship between fibrinogen and  $\log_e(\text{CRP})$ . All SNPs used here as IVs are in the CRP regulatory gene on chromosome 1.

The Cardiovascular Health Study [22] is an observational study of risk factors for cardiovascular disease in adults 65 years or older. We use cross-sectional baseline data for 4469 white subjects from this study, in which four diallelic SNPs relevant to CRP were measured: rs1205, rs1800947, rs1417938 and rs2808630. Each of these SNPs was found to be associated with CRP levels. We checked their associations with seven known CHD risk factors (age, body mass index, triglycerides, systolic blood pressure, total cholesterol, low and high density lipoproteins) for each SNP, and found no significant associations ( $P < 0.05$ ) out of the 28 examined. This suggests that the SNPs are valid instruments.

We used each of the techniques for estimating causal association mentioned above. The ratio method for each SNP separately is based on per allele regressions. For the 2SLS method, we use first a per allele model additive across SNPs and secondly a fully factorial version of the 2df model where each observed genotype is placed in a separate category. The 2SLS per allele model is equivalent to the structure-based Bayesian model (4) and the 2SLS factorial model is equivalent to the individual-based Bayesian model (3). When using the group-based regression (2), we excluded all genotypic groups with less than 5 subjects (14 subjects excluded, Figure 4). The individual-based (3), structure-based (4), ratio and 2SLS analyses include all subjects. A sensitivity analysis was performed excluding from the 2SLS factorial and Bayesian individual-based analyses all individuals from genotypic groups with less than 5 subjects. The observational increase in fibrinogen ( $\mu\text{mol/l}$ ) per unit increase in  $\log(\text{CRP})$  is 0.937 (s.e. 0.024) and correlation between fibrinogen and  $\log(\text{CRP})$  is 0.501. The  $F_{4,4464}$  statistic in the regression of  $\log(\text{CRP})$  on the SNPs additively per allele is 27.2, indicating that the instruments together are moderately strong, with a relative size bias less than 5% [16, 23]. As we have used more IVs than we have phenotypes, we can perform an overidentification test. The Sargan test [24] is a test of the validity of the IV and linearity assumptions in the model. The test statistic is 7.15, which compared to a  $\chi^2_3$  distribution gives a p-value of 0.067, meaning that the validity of the instruments is not rejected at the 5% level.



The ratio method gives a different point estimate for each SNP, all of which are compatible with zero association (Table 2). Using the 2SLS methods on all of the SNPs together, we obtain answers which synthesize all of the relevant data for each of the SNPs. The Bayesian methods give causal estimates consistent with the 2SLS estimates (Table 2). The Bayesian structural-based and 2SLS per allele models give lower estimates of causal association than the other models, with 95% CIs that include zero. The Bayesian credibility intervals are (appropriately) asymmetric, as no normal assumption has been made. The Bayesian structural-based and 2SLS per allele models give lower estimates of causal association than the other models. The Bayesian individual-based and the 2SLS factorial methods both give different results when individuals from small genotypic groups are excluded. This is due to weak instruments leading to a bias in the causal estimate in the direction of the confounded (observational) association [16]. We return to bias from weak instruments in the discussion.

Method	Estimate	95% CI
Ratio using rs1205	0.234	-0.169 to 0.660
Ratio using rs1417938	-0.608	-1.581 to 0.137
Ratio using rs1800947	0.203	-0.478 to 0.940
Ratio using rs2808630	2.722	$-\infty$ to $\infty$
2SLS factorial using all SNPs	0.376	0.088 to 0.665
2SLS factorial (excluding small groups)	0.280	-0.041 to 0.601
2SLS per allele using all SNPs	0.200	-0.138 to 0.538
Bayesian methods	Estimate	95% CrI
Group-based (excluding small groups)	0.342	0.004 to 0.698
Individual-based	0.389	0.049 to 0.728
Individual (excluding small groups)	0.300	-0.045 to 0.666
Structural-based	0.212	-0.157 to 0.586

Table 2: Comparison of the causal estimates of increase in fibrinogen ( $\mu\text{mol/l}$ ) per unit increase in  $\log_e(\text{CRP})$  in the Cardiovascular Health Study. 95% confidence/credible interval (CI/CrI) are shown. Small groups are genotypic groups with less than 5 subjects.

## 5 Multiple genetic markers in multiple studies

### 5.1 Methods

The above framework leads naturally to a model for meta-analysis across multiple studies. Assumption c) in Section 2 for IVs ensures that, in principle, the same parameter  $\beta_1$  is being estimated regardless of how many and which SNPs are available in each study. This is because the outcome is independent of the IV given the phenotype (which is measured) and the confounders (which are averaged over). We thus propose a hierarchical model for  $\beta_1$  estimated across multiple studies as follows. For

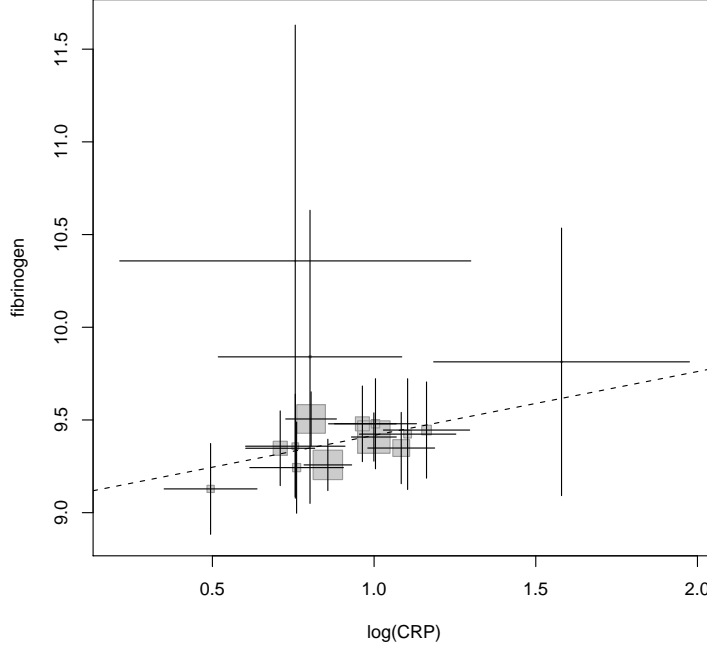


Figure 4: Plot of mean fibrinogen against mean  $\log(\text{CRP})$  in the Cardiovascular Health Study stratified by genotypic group. Error bars are 95% CIs. Groups with less than 5 subjects omitted. The size of the shaded squares is proportional to the number of subjects in each group. The dashed line is the estimate of causal association from the group-based method without random effects.

a fixed-effect meta-analysis, we assume the same value of  $\beta_1$  for each study. For a random-effects meta-analysis, we allow  $\beta_{1m}$  from study  $m$  to come from a distribution with mean  $\beta_1$  and variance  $\psi^2$ . This acknowledges the possibility that the causal parameters are somewhat different across studies, as is plausible due to the influences of different population characteristics, but that they are expected to have generally similar values.

For the group-based regression (2), for group  $j$  in study  $m$ , a fixed-effect meta-analysis is:

$$\begin{aligned}\bar{x}_{jm} &\sim \mathcal{N}(\xi_{jm}, \sigma_{xjm}^2) \\ \bar{y}_{jm} &\sim \mathcal{N}(\eta_{jm}, \sigma_{yjm}^2) \\ \eta_{jm} &= \beta_{0m} + \beta_1 \xi_{jm}\end{aligned}\tag{5}$$

Values for  $\beta_{0m}$ , the constant terms in the regression, will vary depending on the average level of outcome in the population in each study, and are thus given independent vague  $\mathcal{N}(0, 100^2)$  priors for each study.

For a random-effect meta-analysis, the last line of (5) is replaced by:

$$\begin{aligned}\eta_{jm} &= \beta_{0m} + \beta_{1m} \xi_{jm} \\ \beta_{1m} &\sim \mathcal{N}(\beta_1, \psi^2)\end{aligned}\tag{6}$$

We use a Uniform[0,20] prior for  $\psi$  in the example below.

These modifications to the simple group-based analysis (2) for a meta-analysis context can also be similarly made to the individual-based model (3), and to the structured model (4). For example, the full model using a structured model (4), assuming heterogeneity between studies, for individual  $i$  and SNP  $k = 1 \dots K_m$  in study  $m$  is:

$$\begin{aligned}\xi_{im} &= \alpha_{0m} + \sum_{k=1}^{K_m} \alpha_{km} G_{ikm} \\ x_{im} &\sim \mathcal{N}(\xi_{im}, \sigma_{xm}^2) \\ y_{im} &\sim \mathcal{N}(\eta_{im}, \sigma_{ym}^2) \\ \eta_{im} &= \beta_{0m} + \beta_{1m} \xi_{im} \\ \beta_{1m} &\sim \mathcal{N}(\beta_1, \psi^2)\end{aligned}\tag{7}$$

In this model, we assume that the first stage regression coefficients  $\alpha_{km}$  are unrelated in the different studies. An extra sophistication would be to assume that these coefficients are common or related when different studies involve the same set of SNPs. Example WinBUGS code is given in the appendix.

## 5.2 Application to C-reactive protein and fibrinogen

We give an example of meta-analysis of eleven studies [13] using the methods described. In addition to the Cardiovascular Health Study (CHS) used in Section 4.2, we incorporate data from a further eight general population cohort studies: British Women's Heart and Health Study (BWHHS), Copenhagen City Heart Study (CCHS), Copenhagen General Population Study (CGPS), English Longitudinal Study of Ageing (ELSA), Framingham Health Study (FRAM), Northwick Park Heart Study II (NPHS2), Rotterdam Study (ROTT), and Whitehall II Study (W2). In each of these the analyses presented here are cross-sectional, based on baseline measurements of CRP and fibrinogen. We also use data from two case-control studies, the Nurses' Health Study (NHS) and Stockholm Heart Epidemiology Program (SHEEP), again with CRP and fibrinogen measured at baseline. We use the data from controls alone since these better represent cross-sectional population studies. Details of these studies are summarized in Table 3.

To avoid problems with weak instruments, we want to choose genetic instruments which together are strongly related to  $\log(\text{CRP})$ . For this, the instrument was chosen to maintain the F statistic above 10 and to include sequentially, where available, each of SNPs rs1205, one of rs1130864 and rs1417938 (these SNPs are in complete

LD), rs3093077, rs1800947 and rs2808630. In the meta-analysis we use between 2 and 4 SNPs as instruments in each study; the Sargan overidentification tests were satisfied (Table 3). The choice of instruments here is not made a priori, as should ideally be the case, but pragmatically to exemplify the method. For comparison with the Bayesian methods, we use the study-specific 2SLS causal estimates and corresponding asymptotic standard errors in a standard two-step inverse variance weighted meta-analysis (using a moment estimator of the between-study variance in the case of random-effects meta-analysis). Mean log(CRP) and fibrinogen levels for the genotypic groups for six of the studies are shown in Figure 5.

Study	genes used <sup>a</sup>	Participants	Excluded	F value	df	Overid p-value
BWHHS	g1, g3, g5	3188	7	16.7	(3, 3184)	0.638
CCHS	g1, g2, g4	7998	5	29.6	(3, 7994)	0.358
CGPS	g1, g2, g4	35679	5	152.0	(3, 35675)	0.439
CHS	g1, g3, g5, g6	4469	15	27.2	(4, 4464)	0.067
ELSA	g1, g2, g4	4409	8	24.7	(3, 4405)	0.367
FRAM	g1, g2, g4	1575	4	10.0	(3, 1571)	0.447
NHS	g1, g6	414	0	13.2	(2, 411)	0.984
NPHS2	g1, g2, g4	2153	3	11.6	(3, 2149)	0.344
ROTT	g1, g2	2077	2	11.9	(2, 2074)	0.983
SHEEP	g1, g2, g4	1044	4	10.5	(3, 1040)	0.680
W2	g1, g2, g4	4354	5	21.5	(3, 4350)	0.469
Total		67361	58			

Table 3: Summary of studies in meta-analysis of Section 5.2: SNPs measured, number of participants with complete genetic data, number of participants in genotypic groups of size less than 5 excluded from some analyses, F value with degrees of freedom (df), p-value from Sargan test of overidentification from additive per allele regression of phenotype on SNPs used as IVs

<sup>a</sup>g1 = rs1205, g2 = rs1130864, g3 = rs1417938, g4 = rs3093077, g5 = rs1800947, g6 = rs2808630

Table 4 shows a causal association of log(CRP) on fibrinogen which does not significantly differ from the null, except for the structural-based fixed-effect meta-analysis, which suggests a weak negative causal association. Groups of size less than 5 have been omitted in the 2SLS factorial, group-based and individual-based analyses. There is no clear preference for the random-effects models from the Deviance Information Criterion (DIC) [25]. The DIC should only be used to compare between a fixed- or random-effect model, and not between models based on different data structures. Again, the structural-based models give lower estimates of causal association than the other methods.

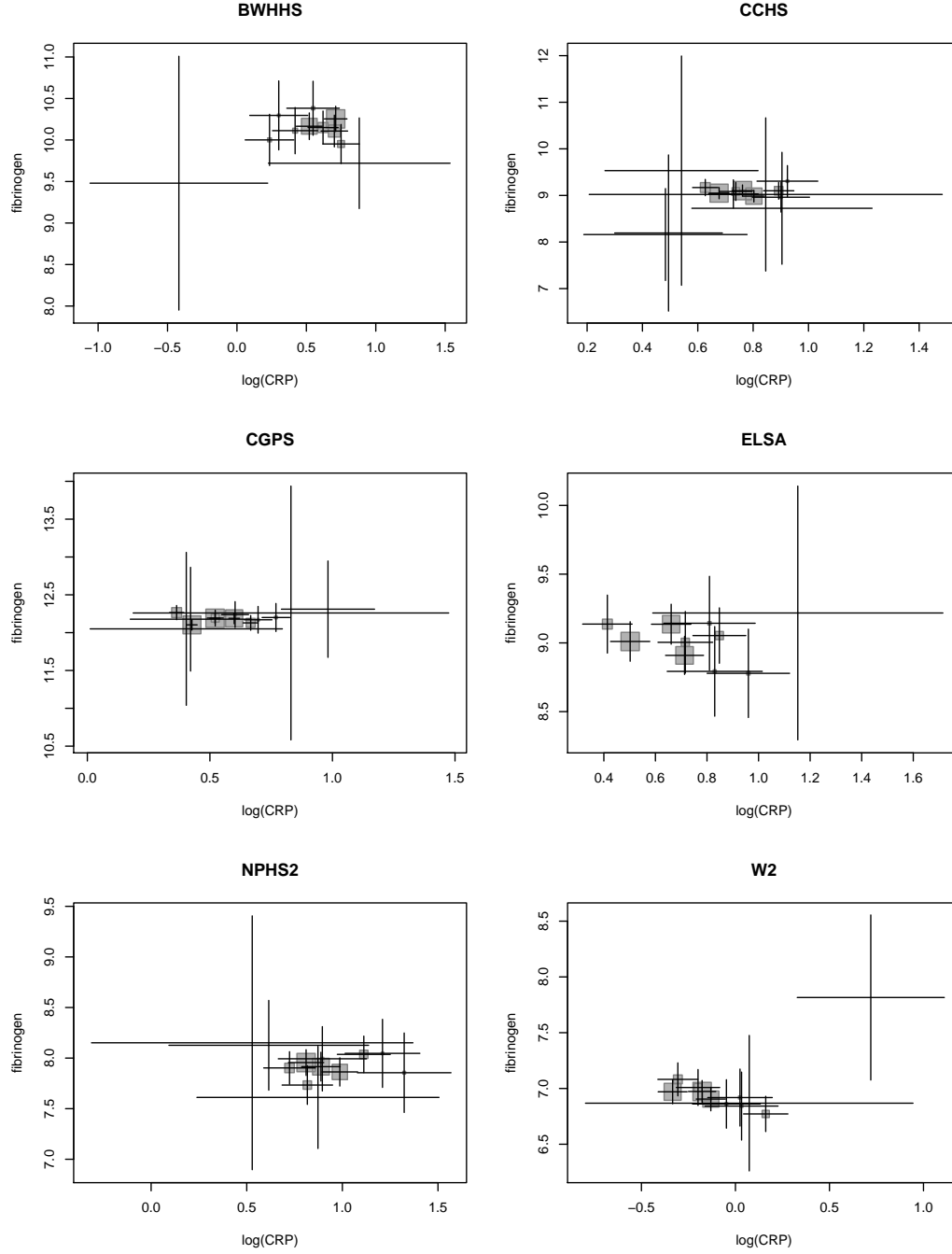


Figure 5: Plot of mean fibrinogen against mean  $\log(\text{CRP})$  for six studies from Section 5.2 stratified by genetic group. Error bars are 95% CIs. Groups with less than 5 subjects omitted. The size of the shaded squares is proportional to the number of subjects in each group.

Fixed-effect meta-analysis	Estimate	95% CI/CrI	DIC <sup>a</sup>	
2SLS factorial	-0.005	-0.139 to 0.130		
2SLS per allele	-0.086	-0.255 to 0.082		
Group-based	-0.008	-0.142 to 0.125	-242.1	
Individual-based	-0.036	-0.164 to 0.090	500692	
Structural-based	-0.136	-0.276 to -0.002	501037	
Random-effects meta-analysis	Estimate	95% CI/CrI	DIC	$\psi$
2SLS factorial	-0.007	-0.151 to 0.137		0.072
2SLS per allele	-0.086	-0.255 to 0.082		0.000
Group-based	-0.017	-0.234 to 0.177	-244.5	0.188
Individual-based	-0.039	-0.228 to 0.153	500692	0.155
Structural-based	-0.150	-0.365 to 0.048	501037	0.169

Table 4: Estimates of increase in fibrinogen ( $\mu\text{mol/l}$ ) per unit increase in  $\log_e(\text{CRP})$ , 95% confidence/credible interval (CI/CrI), deviance information criterion (DIC) and heterogeneity parameter ( $\psi$ ) in meta-analysis of eleven studies using 2SLS and Bayesian methods. Genotypic groups with less than 5 individuals excluded from the 2SLS factorial, group-based and individual-based analyses.

<sup>a</sup>Note: DIC should be used to compare between a fixed- or random-effect model and not between models.

## 6 Discussion

In this paper, we have described a Bayesian approach to analysis of Mendelian randomization studies. We introduced the approach in a simple example of a confounded association with one IV. We extended the method to use multiple IVs, to use individual participant data and to incorporate an explicit, here additive, genetic model. We then show how this leads naturally to a meta-analysis, which can be performed even with heterogeneous genetic data. These methods have been applied in the estimation of the causal association of CRP levels on fibrinogen.

### 6.1 Bayesian methods in IV analysis

The Bayesian approach has similarities to the 2SLS method. In both, fitted values of phenotype are estimated for each genotypic group, which are then used in a regression of outcome on phenotype. In 2SLS, these fitted values are assumed to be precisely known in the second stage regression, and a correction is made to the second stage standard error to account for this using sandwich variance estimators. In the Bayesian framework, the fitted values of phenotype and outcome are estimated simultaneously, and the standard error in the causal parameter is directly estimated from the MCMC sampling process. This means that no assumption is made on the distribution of the causal parameter, giving appropriately sized standard errors and skew CIs. The Bayesian approach allows us to be explicit about the assumptions made. This gives us flexibility to determine the model according to what we believe is plausible without

being limited to linear or normal assumptions.

Additionally, the Bayesian approach provides a framework to perform analyses that are not possible using 2SLS. These include meta-analysis in a single hierarchical model, imputation of missing data, use of haplotypes with uncertainty, and analysis of binary outcomes, each of which is discussed below. It also allows for more accurate inference when using instruments which are weak.

Bayesian methods have not been widely proposed for IV analyses or applied in Mendelian randomization studies. Although Bayesian methods for IV analysis have been suggested in the econometrics literature [26, 27], their use is not common and differences between the fields mean that methods cannot easily be translated into an epidemiological setting [28]. McKeigue et al. [29] have performed a Bayesian analysis in the single SNP and single study situation, but regarding the parameter of interest as the “ratio of the causal effect to crude [observational] effect”. We prefer to regard  $\beta_1$ , the causal association, as the parameter of interest.

## 6.2 Meta-analysis

Methods for meta-analysis of Mendelian randomization studies have not been extensively discussed, and have been restricted to studies measuring one identical SNP [8, 30, 14]. In applications, meta-analyses of studies have concentrated on testing for a causal effect, without accounting for the uncertainty in the estimated mean difference in phenotype values between genotypic groups [31, 32]. Where this uncertainty has been accounted for, confidence intervals for the causal association have been too wide to exclude a moderate causal association [33, 34]. Our proposed analysis thus extends this previous work in a number of ways: first by using a flexible Bayesian framework that eliminates the problems caused by non-normal causal estimates, second by presenting a coherent framework for estimation of the causal association using data from multiple studies, and third by allowing the use of different genetic markers in different studies.

An advantage of the Bayesian setting for meta-analysis is that the whole analysis can be performed in one step. This keeps each study distinct within the hierarchical model, only combining studies at the top level. This is more effective at dealing with heterogeneity, both statistical and in study design, than performing separate meta-analyses on each of the genotype-phenotype and genotype-outcome associations [8]. An alternative approach where the causal association estimate and its precision are estimated in each study, and these estimates combined in a meta-analysis in a second stage, is not recommended for two reasons. First, the distribution of each causal estimate is not normal (especially if the instrument is not strong), and so the uncertainty is not well represented by its standard error, and secondly, some causal estimates from individual studies may have infinite variance. Examples of these problems are apparent in Figure 3 and Table 2.



### 6.3 Weak instruments

A cause for concern in IV analysis is the bias created by using “weak instruments”, that is instruments not strongly associated with the phenotype. When several instruments are used, an instrument is strong if it explains sufficient variance in the phenotype given the other included instruments. Due to the correlation between SNPs, an instrument which is strong on its own may not be strong when considered in addition to other instruments. As the number of instruments increases, if no more of the variation in the phenotype is explained, then the overall instrument becomes weaker and two problems occur. Firstly, the theoretical bias of the IV estimator increases, due to the random correlation between the IVs and unmeasured confounders [35]. If the variation in the phenotype explained by confounders is relatively large compared to the variation in the phenotype caused by the IVs, then the instruments may model the variation in the confounders rather than the systematic variation from the genetic differences. This will give rise to a correlation between the confounders and the fitted phenotype values, which will lead to bias of the IV estimate in the direction of the confounded estimate [17, 23]. This bias affects all the methods considered: the ratio of coefficients method, 2SLS and the proposed Bayesian method. Secondly, as the instruments weaken, the distribution of the causal estimate becomes heavy-tailed, leading to possible underestimation of the size of a test based on 2SLS [23], although recent work on modifying the 2SLS method has concentrated improving properties of test size and confidence intervals with weak instruments [36, 37]. The Bayesian method has an advantage here, in that it makes no assumption of normality. The shape of the posterior distribution for the causal parameter  $\beta_1$  reflects its true uncertainty.

An extreme case of weak instruments is where multiple instruments place each individual in their own separate genetic group. Then the IV estimate, derived from the regression as in Figure 2, is equal to the observational confounded estimate. The generally quoted advice that an instrument with  $F > 10$  is strong is an oversimplification [17]. A modified list of values for a version of the F statistic that limit bias and preserve test size due to Stock et al. (2005) [23] is quoted in the function *ivreg2* in Stata [9].

Instruments should be specified in Mendelian randomization studies prior to data collection. When this is not possible, instruments should be chosen so as not to use those which give little additional strength to the G-X association. Overidentification tests, such as the Sargan or Basman test [9] can be performed to test the validity of instruments. However, it should not be thought that an overidentification test is a cure-all: instruments that pass the test can still give estimates that suffer from bias. Sensitivity analyses, especially when the F statistic is below 10, should be performed using different instruments and models of genetic association to informally investigate heterogeneity of estimates indicating possible bias or violation of IV assumptions.

Although the individual-level model includes all of the participants, the inclusion of many small groups potentially weakens the instruments in the model and so increases bias. Small groups will not have enough participants for confounder levels to be assumed equal between groups. The structural model may then be preferred, and will

be less biased and more efficient if the additive assumptions for the effects of genotype on phenotype are valid. Fewer parameters are estimated and groups with similar genotypes will have correlated estimated true levels of phenotype, which provides extra information in the analysis. In each of our analyses, the individual and group-based models give more positive estimates of causal association than the structural-based model. This is due to greater bias in the direction of the observed association in the individual and group-based models from weaker instruments with more degrees of freedom in the G-X association.

## 6.4 Missing data and haplotype assignment

If there are missing SNP data when applying the individual-based or group-based Bayesian models, and if we can assume that this missingness is not associated with any variable except possibly conditional on the genotype, then the pattern of missing data satisfies the IV assumptions. For each SNP, missing values can be included as a separate category when defining the subgroups. If there are missing genetic data when applying the structural model, we can impute the missing data  $M$  times using the correlation between SNPs with software such as fastPHASE [38]. These multiple imputations can then be included in the Bayesian model, for example using the WinBUGS function  $m \sim dpick(1, M)$ . Alternatively, if we believe that the variation in the phenotype is better explained by haplotypes than by genotypes, then we can use standard software to infer haplotypes [38] and instead use these haplotype assignments in a multiple imputation, with the true phenotype level as the sum of contributions from each haplotype [4].

## 6.5 Binary outcomes

The Bayesian group-level method can be applied to binary outcomes, such as disease events, using a normal approximation to the distribution of log-odds as the outcome. In this case, for example,  $\bar{y}_j$  and  $\sigma_{y_j}^2$  for genotype group  $j$  in model (2) are replaced by  $\text{logit}(p_j)$  and  $\frac{1}{n_{1j}} + \frac{1}{n_{2j}}$  respectively where  $p_j$  is the observed probability of the event in group  $j$ ,  $n_{1j}$  the number of events and  $n_{2j}$  the number of non-events. Alternatively, we can model the probability of an event directly using logistic regression. We replace the normal distribution of the outcome in each group in model (2) with a binomial distribution as follows:

$$\begin{aligned} n_{1j} &\sim \text{Binomial}(\pi_j, n_{1j} + n_{2j}) \\ \text{logit}(\pi_j) &= \beta_0 + \beta_1 \xi_j \end{aligned} \tag{8}$$

For an individual-level analysis of binary data, as in model (3), the outcome  $y_{ij}$  for individual  $i$  in genotype group  $j$  takes the value of 0 or 1. Then the first line in model (8) is replaced by  $y_{ij} \sim \text{Binomial}(\pi_j, 1)$ . Similar adaptations for binary outcomes can be made to the meta-analysis models (5) and (7).

Such models for binary outcomes are valid for testing the causal hypothesis  $\beta_1 = 0$ . However, they do not provide unbiased estimators of a non-zero causal parameter [2], due to the non-collapsibility of the log-odds function over the distribution of the unknown confounders [39]. Alternative methods, including marginal structural models [40] and structural mean models [41], have been proposed for binary outcomes in a classical setting, but these do not provide consistent estimators either [42]. An alternative estimator, usually called the control variable approach, involves including the residuals from the first-stage regression of  $X$  on  $G$  into the second stage regression of  $Y$  on  $\hat{X}$  [43]. This has been shown usually to reduce bias for IV analyses involving binary outcomes in the Mendelian randomization setting [19]. An alternative approach would be to use a relative risk model. This does not suffer the same problems of non-collapsibility and the parameter estimated from such an approach is the relevant causal parameter [2]. Application of these ideas and methods within our Bayesian model formulation needs investigation.

## 6.6 Conclusion

The validity of IV analyses relies on the assumptions specified in the introduction. These assumptions can only be partially verified from data, and there are a number of ways in which they may be violated for Mendelian randomization studies [4]. Whereas the association between the genotypes and measured confounders can be assessed, and checked that they are compatible with chance, associations with unmeasured confounders clearly cannot be checked. Potential correlations caused by linkage disequilibrium between the SNPs of interest and other genetic variants, which act on the outcome through different biological pathways, would also violate the assumptions implicit in Figure 1. The argument against the existence of such pathways is usually biological rather than statistical. Finally it is difficult to rule out the possibility that a genetic mutation leads, through developmental compensation or canalization, to feedback regulation which affects the distribution of confounders. Thus causal estimates derived from all IV analyses should be subject to the caveat that they rely on assumptions. Nevertheless, our proposed Bayesian method for meta-analysis of Mendelian randomization studies is a useful methodological advance. It should also find application in the context of the increasing number of consortia that are now collating the relevant genetic, phenotype and outcome data from multiple studies [13].

---

**Authors/Writing committee:** S. Burgess, MRC Biostatistics Unit, Cambridge; and S.G. Thompson, MRC Biostatistics Unit, Cambridge. Both authors are supported by the UK Medical Research Council (grant U.1052.00.001)

**Authors/Steering group** Atherosclerosis Risk in Communities Study (ARIC): G. Andrews, University of North Carolina. BHF/MRC Family Heart and GRACE Studies: N.J. Samani, University of Leicester and A. Hall, University of Leeds. British Regional Heart Study (BRHS): P. Whincup and R. Morris, University College London. British Women’s Heart and Health Study (BWHHS): D.A. Lawlor, G. Davey Smith and N. Timpson, University of Bristol; S. Ebrahim, London School of Hygiene and Tropical Medicine. Caerphilly Study: Y. Ben-Shlomo, G. Davey Smith and N. Timpson, University of Bristol. Cambridge

Heart Antioxidant Study (CHAOS): M. Brown, S. Ricketts and M. Sandhu, University of Cambridge. Cardiovascular Health Study (CHS): A. Reiner and B. Psaty, University of Washington; L. Lange, University of North Carolina; M. Cushman, University of Vermont. Carotid Ultrasound Disease Assessment Study (CUDAS) and Carotid Ultrasound in Patients with Ischaemic Heart Disease (CUPID): J. Hung, P. Thompson, J. Beilby, and N. Warrington, University of Western Australia; L.J. Palmer, Western Australia Institute for Medical Research. Copenhagen City Heart Study (CCHS), Copenhagen General Population Study (CGPS) and Copenhagen Ischaemic Heart Disease Study (CIHDS): B.G. Nordestgaard, A. Tybjaerg-Hansen and J. Zacho, University of Copenhagen. CRP Gene Variants and Coronary Artery Disease in a Chinese Han Population (CRP-Han): C. Wu, Shanghai Institute of Cardiovascular Disease. Edinburgh Artery Study (EAS): G. Lowe, University of Glasgow; I. Tzoulaki, Imperial College London. English Longitudinal Study of Aging (ELSA): M. Kumari, University College London. European Prospective Investigation into Cancer and Nutrition (EPIC), Norfolk Centre: M. Sandhu, University of Cambridge. Framingham Offspring Study: J.F. Yamamoto, Boston University. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico (GISSI): B. Chiodini, King's College London and M. Franzosi, Mario Negri Institute for Pharmacological Research. Health in Men Study (HIMS): G.J. Hankey, University of Western Australia; K. Jamrozik, University of Queensland; L. Palmer, Western Australia Institute for Medical Research. Health Professionals Follow-up Study (HPFS): E. Rimm and J. Pai, Harvard School of Public Health. Heart and Vascular Health Study: B. Psaty, S. Heckbert and J. Bis, University of Washington. INTERHEART Study: S. Anand, McMaster University; J. Engert, McGill University. International Study of Infarct Survival (ISIS): R. Collins and R. Clarke, University of Oxford. Malmö Diet and Cancer Study: O. Melander, Malmö University Hospital; G. Berglund, University of Lund. Northern Swedish Cohorts Study: P. L腺envall, Östra Sjukhuset Göteborg; L. Johansson and J-H. Jansson, Skellefteå Hospital; G. Hallmans, Umeå University. Northwick Park Heart-II Study: A. Hingorani and S. Humphries, University College London. Nurses' Health Study (NHS): E. Rimm, J. Manson and J. Pai, Harvard School of Public Health and Harvard Medical School. Procorious Coronary Artery Disease Study (PROCARDIS): H. Watkins, R. Clarke and J. Hopewell, University of Oxford. Pakistan Risk of Myocardial Infarction Study (PROMIS): D. Saleheen and R. Frossard, Center for Non-Communicable Diseases; J. Danesh, University of Cambridge. Prospective Study of Pravastatin in the Elderly at Risk (PROSPER): N. Sattar, M. Robertson and J. Shepherd, University of Glasgow; E. Schaefer, Tufts University. Rotterdam Study: A. Hofman, J.C.M. Witteman and I. Kardys, Erasmus University Medical Centre. Speedwell Study: Y. Ben-Shlomo, G. Davey Smith and N. Timpson, University of Bristol. Stockholm Heart Epidemiology Program (SHEEP): U. de Faire and A. Bennet, Karolinska Institute. West of Scotland Coronary Prevention Study (WOSCOPS): N. Sattar, I. Ford and C. Packard, University of Glasgow. Whitehall II Study: M. Kumari, University College London. Women's Health Initiative (WHI): J. Manson, Harvard Medical School. The contribution of Debbie A Lawlor & George Davey Smith to this paper is supported by a UK Medical Research Council grant (G0601625).

**Authors/Operations Group** S. Anand, McMaster University; R. Collins, University of Oxford; J.P. Casas, London School of Hygiene and Tropical Medicine; J. Danesh, University of Cambridge; G. Davey Smith, University of Bristol; M. Franzosi, Mario Negri Institute for Pharmacological Research; A. Hingorani, University College London; D.A. Lawlor, University of Bristol; J. Manson, Harvard Medical School; B.G. Nordestgaard, University of Copenhagen; N.J. Samani, University of Leicester; M. Sandhu, University of Cambridge; L.

Smeeth, London School of Hygiene and Tropical Medicine.

**Authors/Coordinating Centre** F. Wensley (coordinator), University of Cambridge; S. Anand, McMaster University; J. Bowden, MRC Biostatistics Unit; S. Burgess, MRC Biostatistics Unit; J.P. Casas, London School of Hygiene and Tropical Medicine; E. Di Angelantonio, University of Cambridge; J. Engert, McGill University; P. Gao, University of Cambridge; T. Shah, University College London; L. Smeeth, London School of Hygiene and Tropical Medicine; S.G. Thompson, MRC Biostatistics Unit; C. Verzilli, London School of Hygiene and Tropical Medicine; M. Walker, University of Cambridge; J. Whittaker, London School of Hygiene and Tropical Medicine; A. Hingorani (co-principal investigator), University College London; J. Danesh (co-principal investigator), University of Cambridge.

## 7 Appendix:

### 7.1 WinBUGS code for random-effects meta-analysis of group-based model

```
model {
# prior for hierarchical causal estimate (parameter of interest)
  betatrue ~ dnorm(0, 0.000001)
# prior for standard deviation of individual study estimates
  betasd ~ dunif(0, 20)
  betatau <- pow(betasd, -2)
for(m in 1:S) {          # S = number of studies
# prior for regression intercept parameter
  beta0[m] ~ dnorm(0, 0.000001)
# distribution of study-specific causal estimates
  beta[m] ~ dnorm(betatrue, betatau)
  for (j in 1:G[m]) {    # G[m] = number of genetic subgroups in study m
# distribution of phenotype in subgroup j, study m
  x[j, m] ~ dnorm(xi[j, m], xtau[j, m])
# distribution of outcome in subgroup j, study m
  y[j, m] ~ dnorm(eta[j, m], ytau[j, m])
# prior for true value of phenotype in subgroup j, study m
  xi[j, m] ~ dnorm(0, 0.000001)
# linear model of true outcome on true phenotype
  eta[j, m] <- beta0[m] + beta[m] * xi[j, m]
  }
  }
}
```

### 7.2 WinBUGS code for fixed-effect meta-analysis of structure-based model

```
model {
```

```

# prior for fixed causal estimate (parameter of interest)
beta ~ dnorm(0, 0.000001)
for(m in 1:S) {
# prior for regression intercept parameter
beta0[m] ~ dnorm(0, 0.000001)
alpha0[m] ~ dnorm(0, 0.000001)
# prior for study phenotype standard deviation
xsd[m] ~ dunif(0, 20)
xtau[m] <- pow(xsd[m], -2)
# prior for study outcome standard deviation
ysd[m] ~ dunif(0, 100)
ytau[m] <- pow(ysd[m], -2)
for(k in 1:G[m]) { # G[m] = number of genes in study m
# prior for gene-phenotype regression parameters
alpha[k, m] ~ dnorm(0, 0.000001)
}
for (i in 1:N[m]) { # N[m] = number of individuals in study m
# linear model of true phenotype on genes
xi[i, m] <- inprod(alpha[1:G[m], m], gene[i, 1:G[m], m]) + alpha0[m]
# distribution of phenotype in individual i, study m
x[i, m] ~ dnorm(xi[i, m], xtau[m])
# distribution of outcome in individual i, study m
y[i, m] ~ dnorm(eta[i, m], ytau[m])
eta[i, m] <- beta0[m] + beta * xi[i, m]
}
}
}

```

## References

- [1] Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 2003; **32**(1):1–22, doi:10.1093/ije/dyg070.
- [2] Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* 2007; **16**(4):309–330, doi:10.1177/0962280206077743.
- [3] Greenland S. An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.* 2000; **29**(4):722–729, doi:10.1093/ije/29.4.722.
- [4] Lawlor D, Harbord R, Sterne J, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* 2008; **27**(8):1133–1163, doi:10.1002/sim.3034.



- [5] Wald A. The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics* 1940; **11**(3):284–300.
- [6] Fieller E. Some problems in interval estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* 1954; **16**:175–185.
- [7] Thomas D, Lawlor D, Thompson J. Re: Estimation of Bias in Nongenetic Observational Studies Using “Mendelian Triangulation” by Bautista et al. *Annals of epidemiology* 2007; **17**(7):511–513.
- [8] Thompson J, Minelli C, Abrams K, Tobin M, Riley R. Meta-analysis of genetic studies using mendelian randomization—a multivariate approach. *Statistics in medicine* July 2005; **24**(14):2241–2254, doi:10.1002/sim.2100.
- [9] Baum C, Schaffer M, Stillman S. Instrumental variables and gmm: Estimation and testing. *Stata Journal* 2003; **3**(1):1–31.
- [10] Timpson N, Lawlor D, Harbord R, Gaunt T, Day I, Palmer L, Hattersley A, Ebrahim S, Lowe G, Rumley A, *et al.* C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *The Lancet* 2005; **366**(9501):1954–1959, doi:10.1016/S0140-6736(05)67786-0.
- [11] Kivimäki M, Lawlor D, Davey Smith G, Kumari M, Donald A, Britton A, Casas J, Shah T, Brunner E, Timpson N, *et al.* Does high C-reactive protein concentration increase atherosclerosis? The Whitehall II Study. *PLoS One* 2008; **3**(8):e3013, doi:10.1371/journal.pone.0003013.
- [12] Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int. J. Epidemiol.* 2004; **33**(1):30–42, doi:10.1093/ije/dyh132.
- [13] CRP CHD Genetics Collaboration. Collaborative pooled analysis of data on c-reactive protein gene variants and coronary disease: judging causality by mendelian randomisation. *European Journal of Epidemiology* August 2008; **23**(8):531–540, doi:10.1007/s10654-008-9249-z.
- [14] Palmer T, Thompson J, Tobin M. Meta-analysis of Mendelian randomization studies incorporating all three genotypes. *Statistics in Medicine* 2008; **27**(30):6570–6582, doi:10.1002/sim.3423.
- [15] Pai J, Mukamal K, Rexrode K, Rimm E. C-reactive protein (CRP) gene polymorphisms, CRP levels, and risk of incident coronary heart disease in two nested case-control studies. *PLoS One* 2008; **3**(1):e1395, doi:10.1371/journal.pone.0001395.
- [16] Stock J, Wright J, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* October 2002; **20**(4):518–529, doi:10.1198/073500102288618658.



- [17] Nichols A. Weak instruments: An overview and new techniques. *Technical Report* July 2006. Accessed on 04/03/09 at <http://www.stata.com/meeting/5nasug/wiv.pdf>.
- [18] Cragg J, Donald S. Testing identifiability and specification in instrumental variable models. *Econometric Theory* 1993; **9**(02):222–240, doi:10.1017/S0266466600007519.
- [19] Palmer T, Thompson J, Tobin M, Sheehan N, Burton P. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. *Int. J. Epidemiol.* 2008; **37**(5):1161–1168, doi:10.1093/ije/dyn080.
- [20] Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**(4):325–337, doi:10.1023/A:1008929526011.
- [21] Emerging Risk Factors Collaboration. C-reactive protein concentration and the risk of major cardiovascular diseases and nonvascular mortality 2009. In press.
- [22] Fried L, Borhani N, Enright P, Furberg C, Gardin J, Kronmal R, Kuller L, Manolio T, Mittelmark M, Newman A. The Cardiovascular Health Study: Design and rationale. *Annals of Epidemiology* 1991; **1**(3):263.
- [23] Stock J, Yogo M. Testing for Weak Instruments in Linear IV Regression. *SSRN eLibrary* 2002; **11**:T0284.
- [24] Sargan JD. The estimation of economic relationships using instrumental variables. *Econometrica* 1958; **26**(3):393–415.
- [25] Spiegelhalter D, Best N, Carlin B, Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(4):583–639.
- [26] Kleibergen F, Zivot E. Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics* 2003; **114**(1):29–72, doi:10.1016/S0304-4076(02)00219-1.
- [27] Conley T, Hansen C, McCulloch R, Rossi P. A semi-parametric Bayesian approach to the instrumental variable problem. *Journal of Econometrics* 2008; **144**(1):276–305.
- [28] Lawlor D, Windmeijer F, Davey Smith G. Is Mendelian randomization “lost in translation”: Comments on “Mendelian randomization equals instrumental variable analysis with genetic instruments” by Wehby et al. *Statistics in Medicine* 2008; **27**(15):2750–2755, doi:10.1002/sim.3308.
- [29] McKeigue P, Campbell H, Wild S, Vitart V, Hayward C, Rudan I, Wright A, Wilson J. Bayesian methods for instrumental variable

analysis with genetic instruments (“Mendelian randomization”): example with urate transporter SLC2A9 as instrumental variable for effect of urate levels on metabolic syndrome. Accessed on 04/03/09 at <http://homepages.ed.ac.uk/pmckeigu/mendelrand/instrumuric.pdf>.

- [30] Minelli C, Thompson J, Tobin M, Abrams K. An Integrated Approach to the Meta-Analysis of Genetic Association Studies using Mendelian Randomization. *Am. J. Epidemiol.* 2004; **160**(5):445–452, doi:10.1093/aje/kwh228.
- [31] Lewis S, Ebrahim S, Davey Smith G. Meta-analysis of MTHFR 677C - T polymorphism and coronary heart disease: does totality of evidence support causal role for homocysteine and preventive potential of folate? *British Medical Journal* 2005; **331**(7524):1053, doi:10.1136/bmj.38611.658947.55.
- [32] Keavney B, Danesh J, Parish S, Palmer A, Clark S, Youngman L, Delepine M, Lathrop M, Peto R, Collins R, *et al.*. Fibrinogen and coronary heart disease: test of causality by ‘Mendelian randomization’. *Int. J. Epidemiol.* 2006; **35**(4):935–943, doi:10.1093/ije/dyl114.
- [33] Davey Smith G, Lawlor D, Harbord R, Timpson N, Rumley A, Lowe G, Day I, Ebrahim S. Association of C-Reactive Protein With Blood Pressure and Hypertension: Life Course Confounding and Mendelian Randomization Tests of Causality. *Arterioscler Thromb Vasc Biol* 2005; **25**(5):1051–1056, doi: 10.1161/01.ATV.0000160351.95181.d0.
- [34] Lawlor D, Harbord R, Timpson N, Lowe G, Rumley A, Gaunt T, Baker I, Yarnell J, Kivimäki M, Kumari M, *et al.*. The Association of C-Reactive Protein and CRP Genotype with Coronary Heart Disease: Findings from Five Studies with 4,610 Cases amongst 18,637 Participants. *PLoS ONE* 08 2008; **3**(8):e3011, doi: 10.1371/journal.pone.0003011.
- [35] Nelson C, Startz R. The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *Journal of Business* 1990; **63**(1):125–140.
- [36] Mikusheva A, Poi B. Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal* 2006; **6**(3):335–347.
- [37] Imbens G, Rosenbaum P. Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society Series A* 2005; **168**(1):109–126, doi:10.1111/j.1467-985X.2004.00339.x.
- [38] Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics* 2006; **78**(4):629–644, doi:10.1086/502802.

- [39] Greenland S, Robins J, Pearl J. Confounding and collapsibility in causal inference. *Statistical Science* 1999; **14**(1):29–46, doi:10.2307/2676645.
- [40] Ten Have T, Joffe M, Cary M. Causal logistic models for non-compliance under randomized treatment with univariate binary response. *Statistics in medicine* 2003; **22**(8):1255–1283, doi:10.1002/sim.1401.
- [41] Vansteelandt S, Goetghebeur E. Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 2003; **65**(4):817–835.
- [42] Clarke P, Windmeijer F. Instrumental variable estimators for binary outcomes. *The Centre for Market and Public Organisation 09/209*, Department of Economics, University of Bristol, UK Jan 2009.
- [43] Rivers D, Vuong Q. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 1988; **39**(3):347–66.