



## Mendelian Randomization

# Using a two-sample Mendelian randomization design to investigate a possible causal effect of maternal lipid concentrations on offspring birth weight

Liang-Dar Hwang <sup>1</sup>, Deborah A Lawlor,<sup>2,3</sup> Rachel M Freathy,<sup>4</sup> David M Evans <sup>1,2,3†</sup> and Nicole M Warrington <sup>1,5\*†</sup>

<sup>1</sup>The University of Queensland Diamantina Institute, The University of Queensland, Brisbane, Australia, <sup>2</sup>Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, UK, <sup>3</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK, <sup>4</sup>Institute of Biomedical and Clinical Science, College of Medicine and Health, University of Exeter, Exeter, UK and <sup>5</sup>K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway

\*Corresponding author. University of Queensland Diamantina Institute, Level 7, 37 Kent St, Translational Research Institute, Woolloongabba, QLD 4102, Australia. E-mail: n.warrington@uq.edu.au

†Joint senior authors.

Editorial decision 21 June 2019; Accepted 10 July 2019

## Abstract

**Background:** The intrauterine environment is critical for fetal growth and development. However, observational associations between maternal gestational lipid concentrations and offspring birth weight (BW) have been inconsistent and ascertaining causality is challenging.

**Methods:** We used a novel two-sample Mendelian randomization (MR) approach to estimate the causal effect of maternal gestational high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and triglyceride concentrations on offspring BW. Single nucleotide polymorphisms (SNPs) associated with serum HDL-C, LDL-C and triglyceride concentrations identified in the Global Lipids Genetics Consortium genome-wide association study meta-analysis ( $n = 188\,577$  European-ancestry individuals; sample 1) were selected as instrumental variables. The effects of these SNPs on offspring BW were estimated using a structural equation model in the UK Biobank and Early Growth Genetics consortium ( $n = 230\,069$  European-ancestry individuals; sample 2) that enabled partitioning of the genetic associations into maternal- (intra-uterine) and fetal-specific effects.

**Results:** We found no evidence for a causal effect of maternal gestational HDL-C, LDL-C or triglyceride concentrations on offspring BW [standard deviation change in BW per standard deviation higher in HDL-C =  $-0.005$  (95% confidence interval:  $-0.039, 0.029$ ), LDL-C =  $0.014$  ( $-0.017, 0.045$ ), and triglycerides =  $0.014$  ( $-0.025, 0.052$ )].

**Conclusions:** Our findings suggest that maternal gestational HDL-C, LDL-C and triglyceride concentrations play a limited role in determining offspring BW. However, we cannot comment on the impact of these and other lipid fractions on fetal development more generally. Our study illustrates the power and flexibility of two-sample MR in assessing the causal effect of maternal environmental exposures on offspring outcomes.

**Key words:** Mendelian randomization, maternal effect, birth weight, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides

#### Key Messages

- We used a novel two-sample Mendelian randomization design to investigate the causal effect of maternal gestational lipid concentrations on offspring birth weight.
- We used summary statistics from genome-wide association studies that predict serum high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and triglyceride concentrations and birth weight, with the latter being partitioned into maternal- and fetal-specific effects.
- Our findings showed limited support for a causal effect of maternal gestational high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and triglyceride concentrations on offspring birth weight.
- The same method could be used to examine other maternal exposures putatively causally related to offspring outcomes assessed later in life.

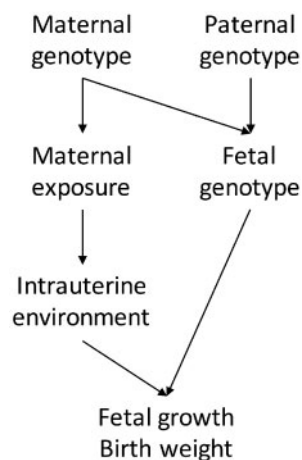
## Introduction

On average, babies born at term below the 25th percentile or above the 85th percentile of the population distribution for birth weight (BW) in high-income countries are at higher risk of adverse perinatal outcomes, such as still-birth, neonatal mortality, respiratory problems and the need for admission to neonatal intensive care.<sup>1</sup> Additionally, babies born of lower and higher BW are at increased risk of adverse future health. For example, inverse associations have been identified between BW and coronary heart disease,<sup>2</sup> positive associations with body mass index (BMI)<sup>3</sup> and most cancers<sup>4</sup> and U-shaped associations with future diabetes risk.<sup>5</sup> Understanding which maternal traits during pregnancy causally affect offspring BW will assist in understanding the mechanisms underlying these associations and defining a healthy BW range.

Maternal serum lipid concentrations are one such trait that could influence offspring BW. Observational studies have reported inconsistent results regarding the relationship between maternal lipid concentrations and offspring BW. For example, maternal serum high-density lipoprotein cholesterol (HDL-C) concentrations have been reported to have positive,<sup>6</sup> negative,<sup>7-9</sup> and no<sup>10-12</sup> association with offspring BW. In the case of maternal low-density lipoprotein cholesterol (LDL-C), most studies have reported no association with offspring BW<sup>7-10</sup> although there is some evidence for a negative association between smaller subclasses of LDL-C

and offspring BW.<sup>10</sup> Higher maternal triglyceride concentrations have been shown to be associated with higher offspring BW,<sup>6,11,12</sup> however, some associations were only observed when the study samples were stratified by maternal pre-pregnancy BMI (i.e. in mothers with pre-pregnancy BMI between 18.5 and 24.9 kg/m<sup>2</sup><sup>8,9</sup> or overweight/obese mothers with BMI >25 kg/m<sup>2</sup>).<sup>7</sup> Bivariate linkage disequilibrium (LD) score regression analyses<sup>13,14</sup> applied to a large-scale genome-wide association study (GWAS) meta-analysis of BW has shown negative genetic correlations of it with LDL-C and triglyceride concentrations, and a positive genetic correlation with HDL-C.<sup>15</sup> However, it is unclear whether these genetic correlations reflect pleiotropic effects of genetic loci influencing both traits or are a consequence of maternal blood lipids causally affecting offspring BW.

Mendelian randomization (MR) is a method commonly used in epidemiology to estimate the causal relationship between a modifiable environmental exposure of interest and a medically relevant trait or disease.<sup>16</sup> It uses genetic variants, most commonly single nucleotide polymorphisms (SNPs), robustly associated with the modifiable environmental exposure of interest as instrumental variables to estimate the causal effect. In the context of this manuscript, the environmental exposures of primary interest are maternal lipid concentrations during pregnancy (i.e. maternal HDL-C, maternal LDL-C and maternal triglycerides) and the medically relevant outcome is offspring BW.



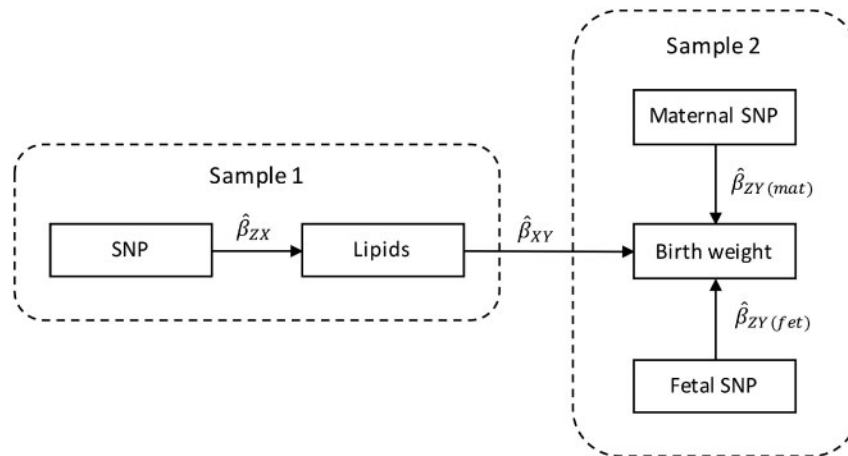
**Figure 1.** A directed acyclic graph (DAG) showing maternal and fetal genetic effects on offspring birth weight. Offspring birth weight can be influenced by both maternal and fetal genotypes, the former operating through the intrauterine environment and via fetal genotype. As fetal genotype is determined by maternal and paternal genotypes, to accurately estimate the association between maternal genotype and offspring birth weight we need to partition the overall genetic effect into maternal- and fetal-specific genetic components. This DAG shows the key pathways and assumptions of our analyses. We have not depicted the (unknown) paths through which fetal genotype would influence fetal growth and birth weight. Note, also that we are assuming that phenotypes relating to fetal genotype do not affect the intrauterine environment, whereas it is known that waste products from the fetus are transferred back across the placenta into the maternal circulation. If fetal genotype influences the production or transfer of waste products, this assumption will be violated.

Unlike MR studies investigating the causal relationship between two adult traits, estimating the causal effect of maternal exposures on offspring outcomes using MR is difficult due to the fact that maternal (and paternal) genotype determine offspring genotype. Specifically, any association between maternal genotype and offspring outcome could be mediated by the child's genotype rather than through the maternal genotype operating on the intrauterine environment (Figure 1). To ensure an unbiased causal estimate of the maternal environmental exposure on offspring outcome, one could perform MR analyses in genotyped mother-offspring duos and statistically correct for offspring's genotype by conditioning on it.<sup>6,17</sup> However, there are only a small number of cohorts worldwide that have genotyped mother-offspring duos, meaning that such analyses may be underpowered, especially in comparison to two-sample MR studies where different but maximized samples of individuals report SNP-exposure and SNP-outcome associations. Our previous MR study used genotyped mother-offspring duos in an attempt to examine the effect of maternal obesity related traits, including HDL-C and triglyceride concentrations, on offspring BW.<sup>6</sup> Results suggested a possible causal inverse effect of maternal

triglycerides on BW (i.e. opposite direction to most previous multivariable regression results). However, despite bringing together as many studies with genotyped maternal-offspring duos as possible in a sample size of 11 031 the effect was imprecisely estimated with wide confidence intervals: difference in mean BW per 1-standard deviation (SD) higher triglycerides = -33 g [95% confidence intervals (CIs): -86, 20].<sup>6</sup> For HDL-C the results were more clearly null [-1 g (-55, 54);  $n = 9176$ ].

An alternative solution, which we adopt, is to employ two-sample MR to estimate the causal effect of maternal lipid concentrations on offspring BW, but using unbiased estimates of the maternal genetic effect on BW from a very large GWAS meta-analysis of the UK Biobank and the Early Growth Genetics (EGG) Consortium.<sup>18</sup> Recently, we showed how structural equation modelling (SEM) could be used to partition genetic effects on BW into maternal-specific and fetal-specific components in the UK Biobank Study, even in the absence of genotyped mother-offspring duos.<sup>19</sup> Our design relies on the fact that genotyped individuals in the UK Biobank reported their own BW, and (in the case of women) the BW of their first child. The SEM can also include summary statistics from the EGG consortium GWAS of own or offspring BW, increasing the sample size and therefore the precision of the maternal- and fetal-specific genetic effect estimates.<sup>19</sup> Maternal genetic effects estimated using this study design represent the association between maternal genotype and offspring BW with the effect of child's genotype removed (i.e. similar to an analysis where child's genotype has been included as a covariate in a study of genotyped mother-offspring duos, but in vastly larger samples of individuals). Using a two-sample MR framework, these estimates of the maternal genetic effect on BW can subsequently be combined with SNP-exposure estimates for the maternal exposures we are interested in, allowing us to maximize sample size and statistical power (Figure 2).

In a recent study using this novel two-sample MR approach, we demonstrated a positive causal effect of maternal height and fasting glucose and inverse causal effect of blood pressure on offspring BW.<sup>18</sup> In the present study we used this two-sample MR design to investigate a possible causal effect of maternal lipid concentrations on offspring BW. We performed two-sample inverse variance weighted MR, and also a series of sensitivity analyses including MR Egger regression,<sup>20</sup> weighted median<sup>21</sup> and mode-based<sup>22</sup> analyses, with SNP-lipid associations from the Global Lipids Genetics Consortium<sup>23</sup> and the maternal SNP-BW associations from the UK Biobank and the EGG Consortium.<sup>18</sup>



**Figure 2.** Two-sample Mendelian randomization testing the causal effect of maternal and fetal serum lipid concentrations on offspring birth weight. Estimates of the SNP-lipid association ( $\hat{\beta}_{ZX}$ ) are calculated in sample 1. The association between these same SNPs and birth weight (BW) is then estimated in sample 2, except in this case the genetic association is partitioned into maternal ( $\hat{\beta}_{ZY(mat)}$ ) and fetal ( $\hat{\beta}_{ZY(fet)}$ ) effects on BW. These estimates are combined to yield estimates for each SNP of the maternal effect of serum lipid concentrations on offspring BW ( $\hat{\beta}_{XY(mat)} = \frac{\hat{\beta}_{ZY(mat)}}{\hat{\beta}_{ZX}}$ ), and the offspring effect of their own serum lipid concentrations on their own BW ( $\hat{\beta}_{XY(fet)} = \frac{\hat{\beta}_{ZY(fet)}}{\hat{\beta}_{ZX}}$ ). Finally, the estimates at each SNP are combined using the two-sample MR approaches, including inverse-variance weighted analysis ( $\hat{\beta}_{IVW}$ ) and the various sensitivity analyses, to produce an overall causal estimate of serum lipids on BW.

## Methods

### Two-sample inverse variance weighted Mendelian randomization analysis

To investigate the causal relationship between maternal lipid concentrations and offspring BW we initially performed two-sample inverse variance weighted (IVW) MR analyses<sup>24,25</sup> using GWAS summary results data from the Global Lipids Genetics Consortium ( $n = 188\,577$  European-ancestry individuals; sample 1)<sup>23</sup> and the summary results data from the GWAS meta-analysis of BW in the UK Biobank and the EGG Consortium ( $n = 230\,069$  European-ancestry individuals; sample 2).<sup>18</sup> The magnitude of the causal effect ( $\hat{\beta}_{IVW}$ ) was estimated as the average of the SNP-outcome effect ( $\hat{\beta}_{ZY(mat)}$ ) divided by the SNP-exposure effect ( $\hat{\beta}_{ZX}$ ) for all SNPs associated with an exposure, and weighted by the inverse variance of the SNP-outcome effect ( $\hat{\beta}_{ZY(mat)}$ ). The regression slope (ratio) in IVW analyses is forced through a zero intercept (i.e. assuming that there is no unbalanced horizontal pleiotropy and hence a zero difference in exposure results in zero difference in BW).

The exposure variables of interest were maternal concentrations of three serum lipid fractions: HDL-C, LDL-C and triglycerides. A total of 185 independent SNPs have previously been robustly associated with these three lipid fractions ( $P < 5.0 \times 10^{-8}$ ) and were selected as instrumental variables (IVs) for MR analyses (Supplementary Table 1, available as Supplementary data at *IJE* online). Of these SNPs, 96 were associated with HDL-C, 82 were associated with LDL-C, and 60 were associated with triglycerides, all at the genome-wide significance level of  $P < 5.0 \times 10^{-8}$ .

We refer to these SNPs as the ‘Unrestricted Set’ of SNPs for each of the three serum lipids. As many of these SNPs were associated with more than one lipid fraction, which potentially violates the exclusion restriction assumption of IVW MR analysis, we created a ‘Restricted Set’ of SNPs, which had strong evidence of association with one lipid fraction only (i.e.  $P < 5.0 \times 10^{-8}$  for one lipid and  $P > 0.05$  for the other two). This reduced the number of associated SNPs to 15 SNPs for HDL-C, 20 SNPs for LDL-C, and 3 SNPs for triglycerides (Supplementary Table 1, available as Supplementary data at *IJE* online). We did not perform MR analyses using the ‘Restricted Set’ for triglycerides because the 3 SNPs together only accounted for  $< 0.1\%$  of the phenotypic variance in serum triglyceride concentrations and calculations indicated low power to detect a causal relationship of maternal triglycerides on BW.

The outcome of interest was offspring BW. Because of concerns regarding confounding due to the effect of maternal genotypes on offspring genotypes, in our MR analyses we used maternal-specific genetic effects on offspring BW estimated using the SEM for each of the lipid-associated SNPs, as illustrated in Figure 2. The maternal SNP effects on BW estimated using the SEM are equivalent to the regression of offspring BW on maternal genotype conditional on offspring genotype. The maternal- (and fetal-) specific effects on offspring BW for each of the genetic instruments are presented in Supplementary Table 1, available as Supplementary data at *IJE* online. We note that the SNP effects on BW were obtained from a GWAS meta-analysis<sup>18</sup> and that BW was estimated in a variety of different ways including by retrospective self-report in the UK

Biobank cohort. Where possible, studies in the EGG consortium were adjusted for gestational age prior to GWAS analysis; as gestational age is not recorded in UK Biobank this adjustment was not possible.

To investigate the possibility of directional pleiotropy in our data, we used funnel plots<sup>20</sup> that plot instrument strength against the causal estimates for all IVs, and sensitivity analyses (see below) that relax some of the horizontal pleiotropy assumptions. Asymmetry in funnel plots suggests that causal estimates from weaker variants tend to be skewed in one direction and may indicate directional horizontal pleiotropy.

### Sensitivity analyses

As the IVW approach could yield biased estimates in the presence of horizontal pleiotropy (i.e. where some SNPs influence the outcome via additional paths other than via the exposure), we performed four sensitivity analyses using additional MR models, MR Egger regression,<sup>20</sup> the weighted median estimator,<sup>21</sup> and the simple and weighted mode-based estimator.<sup>22</sup>

MR Egger regression involves a weighted linear regression of  $\hat{\beta}_{ZY(Mat)}$  on  $\hat{\beta}_{ZX}$  where the intercept from the regression is free to vary. The slope of the regression provides an estimate of the causal effect of the exposure on the outcome that is free from directional horizontal pleiotropy.<sup>20</sup> The presence of an intercept that is different from zero indicates overall directional pleiotropy, which would bias the IVW estimate. As the assessment of directional pleiotropy using MR Egger can be affected by the orientation of the IVs,<sup>26</sup> we oriented the effect alleles of all IVs to have positive effects on the exposure prior to all MR Egger analyses. The validity of MR Egger regression relies on an assumption known as the 'InSIDE (INstrument Strength Independent of Direct Effect) assumption', which states that the pleiotropic effects of SNPs must be independent of their strength as instruments. In other words, the strength of association between the SNPs and the outcome via pathways other than through the exposure must be independent of the strength of the SNP-exposure associations. This is a weaker assumption than requiring the complete absence of horizontal pleiotropy to obtain valid causal estimates, which is the assumption required in traditional IVW MR analysis. However, although MR Egger regression provides a more robust estimate of the causal effect of the exposure on the outcome, its power to detect a causal effect is much lower than IVW analyses.<sup>20</sup> MR Egger regression also makes the assumption that the SNP-exposure associations are estimated without error. To check whether this 'NOME (NO Measurement Error) assumption' was violated in the two-sample MR Egger regression context, we

used the  $I_{GX}^2$  statistic<sup>27</sup> to estimate the potential relative (dilution) bias due to measurement error. The  $I_{GX}^2$  statistic is a value ranging between 0 and 1, with the value close to 1 indicating that the SNP-exposure associations are sufficiently heterogeneous and the uncertainty is relatively small compared with its variability so the violation of NOME is negligible.

The weighted median estimator approach uses the weighted median of the ratios (i.e. the ratio of  $\hat{\beta}_{ZY(Mat)}$  and  $\hat{\beta}_{ZX}$ ) of all the IVs, as opposed to the weighted mean that is used in the IVW approach. The approach provides an asymptotically consistent estimate of the causal effect and tolerates up to (but not including) 50% of the weight coming from invalid IVs (e.g. IVs that violate the exclusion restriction assumption).<sup>21</sup>

The simple mode-based estimator clusters the SNPs into groups based on similarity of causal effect estimates.<sup>22</sup> The causal effect is estimated based on the cluster that has the largest number of SNPs. The weighted mode-based estimator weights the number of SNPs within each cluster by the inverse variance of each SNP's effect on the outcome and returns the causal estimate based on the cluster that has the largest weighted number of SNPs. The mode-based methods return an unbiased estimate of the causal effect if the SNPs within the cluster with the largest number of SNPs (or the largest weighted number for the weighted model) are all valid instruments, which is also referred as the 'ZEMPA assumption (ZErO Model Pleiotropy Assumption)'.

If the five MR models (IVW, MR Egger regression, weighted median estimator, simple and weighted mode-based estimator), which make different assumptions regarding instrument validity, produce similar estimates of the causal effect, then we can be more confident in the robustness of our findings.

### Multivariable Mendelian randomization analysis

Two multivariable MR analyses, multivariable IVW regression<sup>28</sup> and multivariable MR Egger regression,<sup>29</sup> were performed as additional sensitivity analyses. These multivariable approaches take into account the correlation between the three lipid fractions and the fact that the SNPs used in the MR analyses using the 'unrestricted' SNP set are often associated with more than one lipid fraction (i.e. 147 and 44 SNPs in our analyses were associated with at least two lipid fractions with  $P < 0.05$  and  $5.0 \times 10^{-8}$ , respectively; [Supplementary Table 1](#), available as [Supplementary data](#) at *IJE* online).

The multivariable IVW regression was performed by regressing  $\hat{\beta}_{ZY(Mat)}$  on  $\hat{\beta}_{ZX}$  for all three lipid fractions (i.e. three sets of  $\hat{\beta}_{ZX}$  across all 185 SNPs) in a single regression model. Inverse variance weights of the  $\hat{\beta}_{ZY(Mat)}$  were used

in the regression model and the intercept was initially fixed at zero. The multivariable IVW regression analyses were performed using all 185 SNPs at once with the causal effect for each exposure estimated from the regression coefficients for the corresponding set of  $\hat{\beta}_{ZX}$ . A critical assumption, additional to the univariate assumptions, for multivariable MR is that the relationship between genetic instruments and the outcome is only mediated by the exposures considered in the analysis.<sup>28</sup>

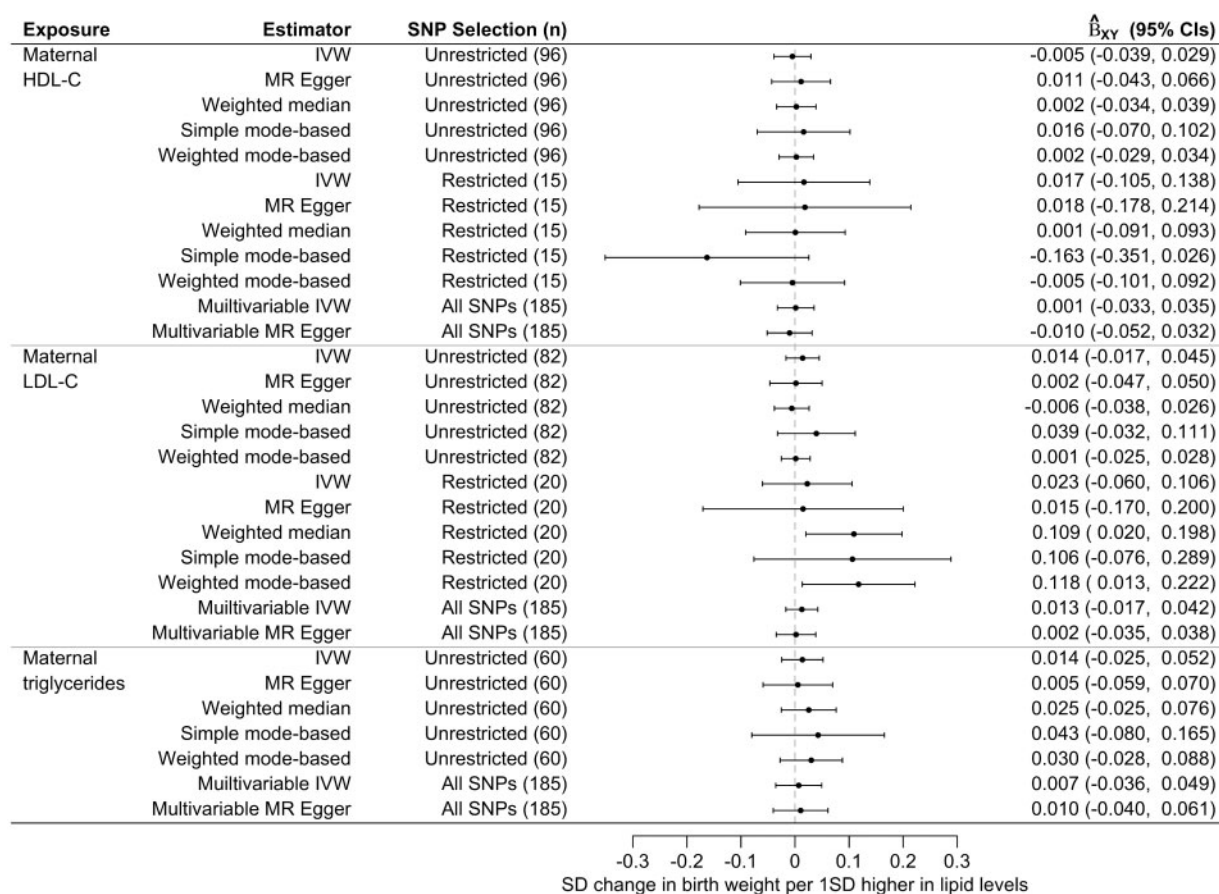
The multivariable MR Egger regression was performed also by including all 185 lipid-associated SNPs (i.e. three sets of  $\hat{\beta}_{ZX}$ ) in a single regression model. Similar to the univariate MR Egger regression model, the multivariable model allows the intercept to vary, with a non-zero estimate of the intercept suggestive of horizontal pleiotropy. We performed three separate analyses. In each of these analyses the SNP alleles were oriented to reflect the direction of association at a single lipid fraction (i.e. first HDL-C, then LDL-C and then triglycerides) to estimate its causal

effect. The advantage of the multivariable MR Egger regression model over the univariate one is that it reduces the bias when the amount of residual effect on the outcome (i.e. SNP-outcome associations not via the exposure) is decreased by including additional exposures.

We used the TwoSampleMR package<sup>30</sup> (version 0.4.11) in R (version 3.4.3) to perform IVW, MR Egger regression, weighted median, mode-based estimator and multivariable IVW MR analyses. We used the codes provided on <https://mrcieu.github.io/TwoSampleMR/> to run the analyses. We used the R scripts provided in a methods paper<sup>30</sup> to perform the multivariable Egger regression analysis.

## Results

There was little evidence for a causal effect of maternal gestational concentrations of HDL-C, LDL-C or triglycerides on offspring BW (Figure 3). Each estimate represents the estimated causal difference in mean SD of offspring



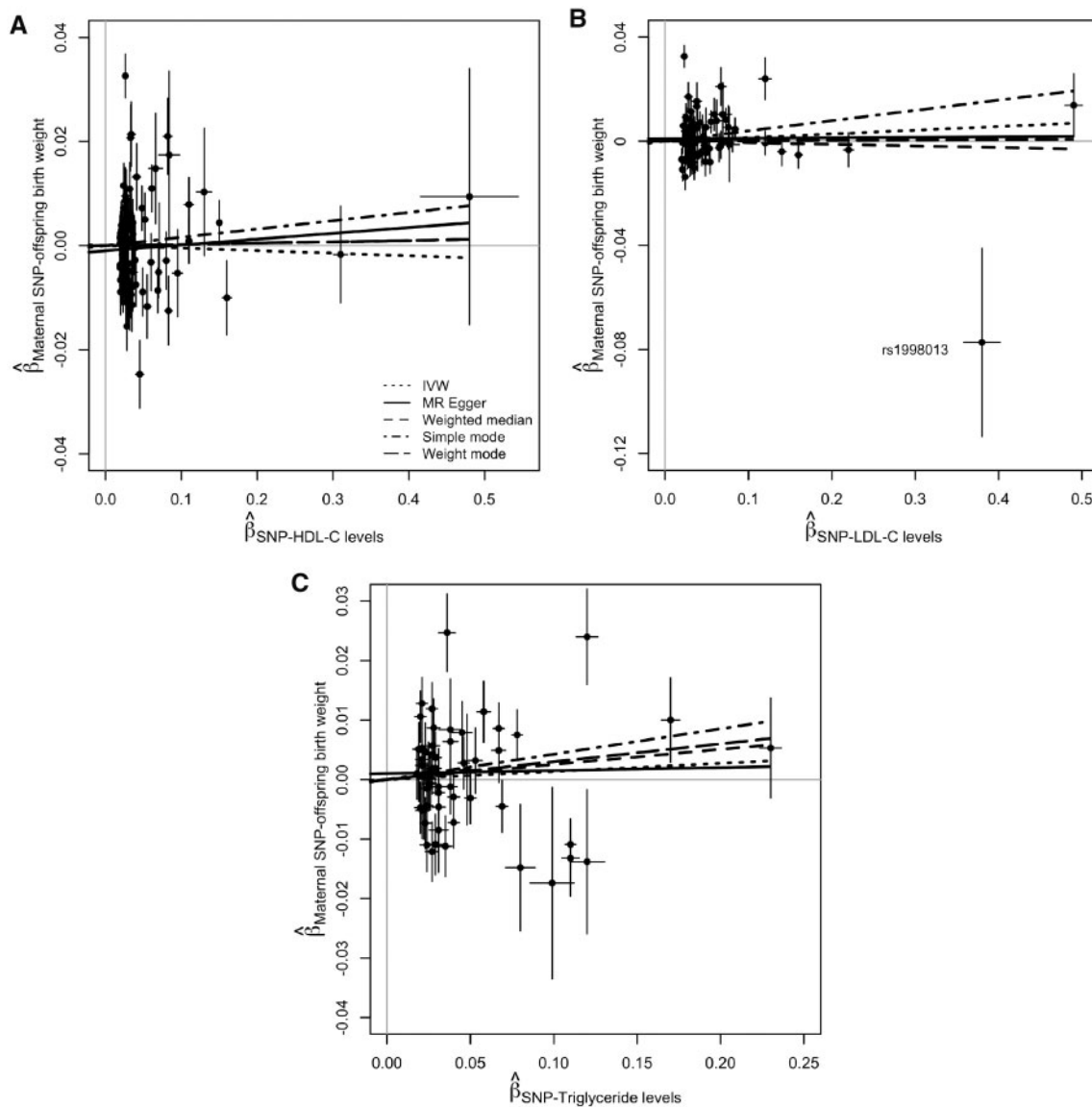
**Figure 3.** Mendelian randomization estimates of the causal effect of maternal serum high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein (LDL-C), and triglyceride concentrations on offspring birth weight. Causal effects are estimated using five univariate MR models [i.e. inverse variance weighted (IVW), MR Egger regression, weighted median, simple mode-based, and weighted mode-based] and two multivariable MR models (i.e. multivariable IVW regression and multivariable MR Egger regression). Each univariate MR model was fitted twice using an 'Unrestricted Set' of SNPs (i.e. all SNPs strongly associated with the exposure with  $P < 5.0 \times 10^{-8}$ ) and a 'Restricted Set' of SNPs (i.e. SNPs strongly associated with the exposure with  $P < 5.0 \times 10^{-8}$  but not associated with the other two lipids with  $P > 0.05$ ). Forest plots show the estimated causal change ( $\hat{\beta}_{xy}$ ) in standard deviations (SD) of offspring birth weight per 1SD higher in serum level of HDL-C, LDL-C or triglycerides with 95% confidence intervals (CIs).

BW per 1SD higher maternal concentration of HDL-C, LDL-C or triglycerides. We found no evidence of directional horizontal pleiotropy at any of the IVs for the three lipid fractions; the estimates of the intercept of both the univariate and multivariable MR Egger regressions were approximately zero in all analyses (Supplementary Table 2, available as Supplementary data at *IJE* online) and there was a lack of asymmetry of the funnel plots (Supplementary Figure 1, available as Supplementary data at *IJE* online). We also found little evidence for violation of the ‘NOME assumption’ for the MR Egger regression as the  $I^2_{GX}$  statistics were  $>97.5\%$  for all analyses (univariate and multivariable MR regressions for all three lipids;

Supplementary Table 2, available as Supplementary data at *IJE* online).

### Maternal serum concentrations of HDL-C and offspring birth weight

The two-sample MR IVW analysis showed no causal effect of maternal HDL-C level and offspring weight [SD change in BW per 1SD higher in HDL-C =  $-0.005$  ( $-0.039$ ,  $0.029$ )]. This was supported by all univariate and multivariable sensitivity analyses (Figures 3 and 4A), and regardless of whether SNPs from the ‘Unrestricted Set’ or the ‘Restricted Set’ were used in the analysis.



**Figure 4.** Scatter plots of the estimated SNP effects on offspring birth weight plotted against the estimated SNPs effects on the maternal (A) HDL-C, (B) LDL-C, and (C) triglyceride concentrations. The slopes of the lines are the estimated causal effects of the maternal lipid concentrations on offspring birth weight, estimated using different MR methods (i.e. inverse variance weighted, MR Egger regression, weighted median, simple mode-based, and weighted mode-based). An outlier SNP, rs1998013, is labelled in (B). It is distal from all other LDL-C associated SNPs and drives the causal effect toward the null.

### Maternal serum concentrations of LDL-C and offspring birth weight

While no causal effect was found in the univariate IVW analysis [SD change in BW per 1SD higher in LDL-C = 0.014 (−0.017, 0.045)], a higher genetically predicted level of maternal LDL-C tended to be associated with a higher offspring BW in the weighted median [SD change in BW per 1SD higher in LDL-C = 0.109 (0.020, 0.198)] and weighted mode-based [SD change in BW per 1SD higher in LDL-C = 0.118 (0.013, 0.222)] analyses using the ‘Restricted Set’ of SNPs (Figures 3 and Supplementary Figure 2, available as Supplementary data at *IJE* online). However, this estimated positive causal effect was not supported by all other univariate and multivariable analyses. The SNP rs1998013 appears to be an outlier in the scatterplot (Figure 4B). It is distal from all other LDL-C associated SNPs and it has the largest absolute effect on BW and the second largest effect on LDL-C that would drive the causal effect toward the null, which is against the overall effect of all other SNPs. We excluded it and re-ran the unrestricted and restricted analyses; the results remained unchanged (Supplementary Figure 3 and Supplementary Table 3, available as Supplementary data at *IJE* online).

### Maternal serum concentrations of triglycerides and offspring birth weight

Similar to the previous two lipid fractions, no causal effect of maternal triglycerides on offspring BW was detected in the univariate IVW analysis [SD change in BW per 1SD higher in triglycerides = 0.014 (−0.025, 0.052)], which was supported by the sensitivity analyses (Figures 3 and 4C).

## Discussion

We employed a novel methodological approach to assess the causal effect of maternal concentrations of LDL-C, HDL-C and triglycerides on offspring BW using a two-sample MR framework. The major difference between our approach and standard two-sample MR is that the SNP-outcome associations (i.e. SNP-BW associations) have been partitioned into maternal and fetal genetic components, which has allowed us to estimate the effect of maternal exposures on BW, independent of the fetal genetic effects. We applied this method to examine the causal effect of maternal gestational LDL-C, HDL-C and triglycerides on offspring BW, while controlling for fetal genotype, and found little support for a causal effect of these on offspring BW.

Results from this study are consistent with our previous MR study using a smaller sample of genotyped mother-offspring pairs for HDL-C although we do not find any

evidence of a suggested inverse effect of triglycerides on birthweight reported in that previous study.<sup>6</sup> In that study, polygenic scores of HDL-C and triglyceride concentrations were calculated using GWAS summary data on lipid concentrations<sup>31</sup> and the maternal genotypes, and no association was found between genetic scores and offspring BW after conditioning on offspring genotype, although a suggested effect of triglycerides was noted (albeit with very wide confidence intervals). The present study has extended this earlier work by using a much larger sample of individuals and using a range of sensitivity analyses to investigate possible bias through genetic pleiotropy. We were also able to explore effects of LDL-C which was not done in that previous study. As an indication of the power of the novel two-sample MR approach used here we can compare the results for the estimated effect of triglycerides on BW from our two studies. Assuming that the SD for BW is 454 g (the mean value in the EGG consortium) our result for an effect of maternal triglycerides on BW is 6.36 g (95%CI: −11.35, 23.61). Compared with our previous study result [−33 g (95%CI: −86, 20)] we now have a more precise estimate suggesting no evidence of an important effect.

We emphasize that the null associations we found do not necessarily mean that maternal LDL-C, HDL-C and triglyceride concentrations have no effect on the developing fetus. They suggest that these maternal circulating lipid fraction concentrations have no effect on the overall growth of the fetus that is reflected by BW. Maternal circulating LDL-C, HDL-C and triglycerides could still impact the development of a particular organ, growth during a particular period of pregnancy, or differential growth of fat, lean and skeletal tissue, but we could not detect these effects by looking only at BW. Several small detailed studies have documented how maternal lipid and fatty acid concentrations change across pregnancy, showing particularly marked increases in triglycerides and more modest increases in cholesterol and phospholipid concentrations together with an increase in the proportion of triglycerides carried in very low-density and low-density lipoprotein particles towards the end of pregnancy.<sup>32</sup> More recently larger epidemiological studies have shown marked changes in lipids and lipoproteins as women become pregnant, which then return to normal after pregnancy.<sup>33</sup> Just how different lipid fractions and fatty acids are carried across the placenta and are used by the fetus for healthy growth and development is unclear, but with the emergence of genome-wide analyses of a much wider range of lipids and fatty acids<sup>34,35</sup> it will be possible to extend the methods used in this study to determine the impact of more refined measures of maternal gestational lipid metabolism on BW. Additionally, BW was obtained from offspring born at full



term (>37 weeks of gestation), so we are unable to comment on whether these maternal lipid fraction concentrations have an effect on the BW of pre-term offspring.

A key assumption of our study design is that SNPs associated with lipid concentrations in the general population of men and (non-pregnant) women are appropriate instruments for lipid concentrations in mothers during pregnancy. Our concern arises from the fact that maternal lipid concentrations change dramatically during pregnancy,<sup>36,37</sup> there may be sex differences in the magnitude of association between SNPs and lipid concentrations, and that lipid concentrations measured at different gestational stages may have different observational associations with BW.<sup>12</sup> We previously showed that maternal gestational HDL-C and triglycerides were strongly correlated with polygenic risk scores constructed from HDL-C and triglyceride associated SNPs in the general population (although LDL-C was not examined), suggesting that the same genetic variants proxy lipid concentrations in pregnant and non-pregnant individuals (see eTable 5 of Tyrrell *et al.*).<sup>6</sup> Larger studies investigating the effect of the lipid-associated SNPs in pregnant women are required.

Another potential complication in interpreting the results of these sorts of MR analyses is the extent to which one can be sure that the causal estimates reflect the effect of the prenatal environment on the developing offspring. For example, it is possible that variation in circulating lipid concentrations before conception might affect oocyte quality and then have downstream effects on offspring BW.<sup>17,38</sup> It is also possible that the causal estimates could reflect the effect of postnatal influences on offspring phenotype. Although this is unlikely to be a concern for perinatal traits like BW where post-natal effects have not had time to exert appreciable effects on the offspring phenotype, it is a potential concern for later life phenotypes, particularly if the investigator is specifically trying to ascribe a causal role for the prenatal environment. One way to increase confidence that the MR results do indeed reflect causal effects of the prenatal environment would be to perform similar analyses using paternal genotype (conditional on offspring genotype). A significant causal relationship would suggest that the post-natal environment is also likely to be important in determining the offspring phenotype, though such a comparison, and indeed our results presented here, assume that the fetal genotype does not have an effect on the intrauterine environmental exposure. Fetal waste products cross the placenta into the maternal circulation and it is possible that fetal genotypes (that are determined by maternal genotypes for the intrauterine exposure of interest) might influence the secretion of fetal waste products into the maternal circulation, which in turn might influence the exposures of interest. Our method, like

two-sample MR in general, is statistically inefficient for exploring non-linear effects or stratified effects (e.g. by maternal BMI).

There are no obvious technical reasons why our method cannot be applied more generally to outcomes other than BW. The major practical limitation, however, is the availability of large-scale cohorts with maternal genotypes along with both maternal and offspring phenotypes so that unbiased estimates of maternal genetic effects on offspring outcomes can be made available to researchers for two-sample MR studies. One way to directly facilitate the availability of these kinds of data to the scientific community would be for researchers to publish summary GWAS results of maternal genotype and offspring phenotype, conditional on offspring genotype. However, performing conditional association analyses across the genome is computationally intensive, and may be difficult for many researchers to implement. A more practical alternative might be for investigators to deposit unconditional GWAS results of the association between maternal genotype and offspring phenotype and separate unconditional GWAS results between offspring genotype and offspring phenotype. These summary results could then be used in a structural equation model to generate unbiased estimates of maternal and fetal genetic effects on offspring phenotype, even when the degree of sample overlap is unknown.<sup>19</sup> Appropriate standard errors could be obtained in these models by estimating the degree of sample overlap using bivariate LD score regression<sup>13</sup> and weighting the likelihood appropriately. It is likely that similar results could also be obtained by using other methods for conditional GWAS.<sup>39–41</sup>

## Conclusion

Maternal gestational LDL-C, HDL-C and triglyceride concentrations do not appear to play a causal role in determining offspring BW. Our study illustrates how causal estimates of maternal environmental exposures on offspring BW can be generated in extremely large samples of individuals using a two-sample MR framework, obviating the requirement for mother-offspring duos with individual level genotypes. The implication is that the same method could be used to examine other maternal exposures putatively causally related to offspring BW, and potentially the relationship between maternal exposures and other offspring outcomes once large scale GWAS of maternal effects are publicly available.

## Supplementary Data

Supplementary data are available at *IJE* online.

## Funding

D.M.E. is supported by an NHMRC Senior Research Fellowship (GNT1137714) and this work was supported by a project grant (GNT1125200). D.A.L.'s contribution to this piece is supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013) / ERC Grant Agreement (Grant number 669545; DevelopObese), the European Union's Horizon 2020 research and innovation programme under grant agreement No 733206 (LifeCycle), the United States National Institutes of Health (NIH): National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK10324), and the British Heart Foundation (AA/18/7/34219). D.M.E and D.A.L work in a UK Medical Research Council (MRC) Unit that is supported by the University of Bristol and MRC (MC\_UU\_00011/6), and D.A.L. is a National Institute of Health Research Senior Investigator (NF-SI-0611–10196). N.M.W. is supported by an Australian National Health and Medical Research Council Early Career Fellowship (GNT1104818). R.M.F. is supported by Sir Henry Dale Fellowship (Wellcome Trust and Royal Society grant: WT104150). This research has been conducted using the UK Biobank Resource (Reference 12703). Access to the UKBB study data was funded by University of Queensland Early Career Researcher Grant (2014002959). L.-D.H is supported by a National Health and Medical Research Council project grant (GNT1125200). This research was carried out at the Translational Research Institute, Woolloongabba, QLD 4102, Australia. The Translational Research Institute is supported by a grant from the Australian Government. No funders influence the study design, data collection or interpretation of results. The views expressed in this paper are those of the authors and not necessarily any funder.

**Conflict of interest:** D.A.L has received research support from several national and international Government and charity funders, and from Medtronic Ltd and Roche Diagnostics in the last 10-years. All other authors report no conflicts of interest.

## References

- Iliodromiti S, Mackay DF, Smith GC *et al.* Customised and non-customised birth weight centiles and prediction of stillbirth and infant mortality and morbidity: a cohort study of 979, 912 term singleton pregnancies in Scotland. *PLoS Med* 2017;14:e1002228.
- Wang SF, Shu L, Sheng J *et al.* Birth weight and risk of coronary heart disease in adults: a meta-analysis of prospective cohort studies. *J Dev Orig Health Dis* 2014;5:408–19.
- Jelenkovic A, Yokoyama Y, Sund R *et al.* Association between birthweight and later body mass index: an individual-based pooled analysis of 27 twin cohorts participating in the CODATwins project. *Int J Epidemiol* 2017;46:1488–98.
- O'Neill KA, Murphy MF, Bunch KJ *et al.* Infant birthweight and risk of childhood cancer: international population-based case control studies of 40 000 cases. *Int J Epidemiol* 2015;44:153–68.
- Harder T, Rodekamp E, Schellong K, Dudenhausen JW, Plagemann A. Birth weight and subsequent risk of type 2 diabetes: a meta-analysis. *Am J Epidemiol* 2007;165:849–57.
- Tyrrell J, Richmond RC, Palmer TM *et al.* Genetic evidence for causal relationships between maternal obesity-related traits and birth weight. *JAMA* 2016;315:1129–40.
- Mudd LM, Holzman CB, Evans RW. Maternal mid-pregnancy lipids and birthweight. *Acta Obstet Gynecol Scand* 2015;94:852–60.
- Misra VK, Trudeau S, Perni U. Maternal serum lipids during pregnancy and infant birth weight: the influence of prepregnancy BMI. *Obesity (Silver Spring)* 2011;19:1476–81.
- Saleh J, Al-Riyami HD, Chaudhary TA, Cianflone K. Cord blood ASP is predicted by maternal lipids and correlates with fetal birth weight. *Obesity (Silver Spring)* 2008;16:1193–98.
- Zeljko A, Vekic J, Spasic S *et al.* Changes in LDL and HDL subclasses in normal pregnancy and associations with birth weight, birth length and head circumference. *Matern Child Health J* 2013;17:556–65.
- Kulkarni SR, Kumaran K, Rao SR *et al.* Maternal lipids are as important as glucose for fetal growth: findings from the Pune Maternal Nutrition Study. *Dia Care* 2013;36:2706–13.
- Geraghty AA, Alberdi G, O'Sullivan EJ *et al.* Maternal blood lipid profile during pregnancy and associations with child adiposity: findings from the ROLO study. *PLoS One* 2016;11:e0161206.
- Bulik-Sullivan BK, Loh PR, Finucane HK *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015;47:291–95.
- Zheng J, Erzurumluoglu AM, Elsworth BL *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2017;33:272–79.
- Horikoshi M, Beaumont RN, Day FR *et al.* Genome-wide associations for birth weight and correlations with adult disease. *Nature* 2016;538:248–52.
- Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32:1–22.
- Lawlor D, Richmond R, Warrington N *et al.* Using Mendelian randomization to determine causal effects of maternal pregnancy (intrauterine) exposures on offspring outcomes: Sources of bias and methods for assessing them. *Wellcome Open Res* 2017;2:11.
- Warrington NM, Beaumont RN, Horikoshi M *et al.* Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet* 2019;51:804–14.
- Warrington NM, Freathy RM, Neale MC, Evans DM. Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank. *Int J Epidemiol* 2018;47:1229–41.
- Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44:512–25.
- Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;40:304–14.
- Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017;46:1985–98.
- Willer CJ, Schmidt EM, Sengupta S *et al.* Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;45:1274–83.

24. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;**37**:658–65.
25. Lawlor DA. Commentary: two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol* 2016;**45**:908–15.
26. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* 2017;**32**:377–89.
27. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I<sup>2</sup> statistic. *Int J Epidemiol* 2016;**45**:1961–74.
28. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015;**181**:251–60.
29. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat Med* 2017;**36**:4705–18.
30. Hemani G, Zheng J, Elsworth B *et al*. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;**7**.
31. Teslovich TM, Musunuru K, Smith AV *et al*. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;**466**:707–13.
32. Sattar N, Greer IA, Galloway PJ *et al*. Lipid and lipoprotein concentrations in pregnancies complicated by intrauterine growth restriction. *J Clin Endocrinol Metab* 1999;**84**:128–30.
33. Wang Q, Wurtz P, Auro K *et al*. Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence. *BMC Med* 2016;**14**:205.
34. Shin SY, Fauman EB, Petersen AK *et al*. An atlas of genetic influences on human blood metabolites. *Nat Genet* 2014;**46**:543–50.
35. Kettunen J, Demirkan A, Wurtz P *et al*. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016;**7**:11122.
36. Alvarez JJ, Montelongo A, Iglesias A, Lasuncion MA, Herrera E. Longitudinal study on lipoprotein profile, high density lipoprotein subclass, and postheparin lipases during gestation in women. *J Lipid Res* 1996;**37**:299–308.
37. Montelongo A, Lasunción MA, Pallardo LF, Herrera E. Longitudinal study of plasma lipoproteins and hormones during pregnancy in normal and diabetic women. *Diabetes* 1992;**41**:1651.
38. Dunning KR, Russell DL, Robker RL. Lipids and oocyte developmental competence: the role of fatty acids and beta-oxidation. *Reproduction* 2014;**148**:R15–27.
39. Zhu Z, Zheng Z, Zhang F *et al*. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018;**9**:224.
40. Grotzinger AD, Rhemtulla M, de Vlaming R *et al*. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav* 2019;**3**:513–25.
41. Turley P, Walters RK, Maghzian O *et al*. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* 2018;**50**:229–37.