





Original article

# Evidence of detrimental effects of prenatal alcohol exposure on offspring birthweight and neurodevelopment from a systematic review of quasi-experimental studies

Loubaba Mamluk <sup>1,2,3\*</sup> Timothy Jones,<sup>2,3</sup> Sharea Ijaz,<sup>2,3</sup>  
Hannah B Edwards,<sup>2,3</sup> Jelena Savović,<sup>2,3</sup> Verity Leach,<sup>2,3</sup>  
Theresa HM Moore,<sup>2,3</sup> Stephanie von Hinke,<sup>4</sup> Sarah J Lewis,<sup>2</sup>  
Jenny L Donovan,<sup>2,3</sup> Deborah A Lawlor,<sup>1,2,3,5</sup>  
George Davey Smith <sup>1,2,5</sup> Abigail Fraser<sup>1,2,5</sup> and Luisa Zuccolo<sup>1,2</sup>

<sup>1</sup>MRC Integrative Epidemiology Unit, Department of Population Health Sciences, University of Bristol, Bristol, UK, <sup>2</sup>Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK, <sup>3</sup>NIHR ARC West, University Hospitals Bristol NHS Foundation Trust, Bristol, UK, <sup>4</sup>Department of Economics, School of Economics, Finance and Management, University of Bristol, Bristol, UK and <sup>5</sup>NIHR Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust, University of Bristol, Bristol, UK

\*Corresponding author. NIHR Applied Research Collaboration (ARC) West, University Hospitals Bristol NHS Foundation Trust, 9th Floor, Whitefriars, Lewins Mead, Bristol BS1 2NT, UK. E-mail: l.mamluk@bristol.ac.uk

Editorial decision 25 November 2019; Accepted 8 January 2020

## Abstract

**Background:** Systematic reviews of prenatal alcohol exposure effects generally only include conventional observational studies. However, estimates from such studies are prone to confounding and other biases.

**Objectives:** To systematically review the evidence on the effects of prenatal alcohol exposure from randomized controlled trials (RCTs) and observational designs using alternative analytical approaches to improve causal inference.

**Search strategy:** Medline, Embase, Web of Science, PsychINFO from inception to 21 June 2018. Manual searches of reference lists of retrieved papers.

**Selection criteria:** RCTs of interventions to stop/reduce drinking in pregnancy and observational studies using alternative analytical methods (quasi-experimental studies e.g. Mendelian randomization and natural experiments, negative control comparisons) to determine the causal effects of prenatal alcohol exposure on pregnancy and longer-term offspring outcomes in human studies.

**Data collection and analysis:** One reviewer extracted data and another checked extracted data. Risk of bias was assessed using customized risk of bias tools. A narrative synthesis of findings was carried out and a meta-analysis for one outcome.

**Main results:** Twenty-three studies were included, representing five types of study design, including 1 RCT, 9 Mendelian randomization and 7 natural experiment studies, and reporting on over 30 outcomes. One study design–outcome combination included enough independent results to meta-analyse. Based on evidence from several studies, we found a likely causal detrimental role of prenatal alcohol exposure on cognitive outcomes, and weaker evidence for a role in low birthweight.

**Conclusion:** None of the included studies was judged to be at low risk of bias in all domains, results should therefore be interpreted with caution.

**Systematic review registration:** This study is registered with PROSPERO, registration number CRD42015015941

**Key words:** Alcohol, pregnancy, prenatal alcohol exposure, systematic review, quasi-experimental studies, negative control, Mendelian randomization, causal inference, neurodevelopment, FASD

### Key Messages

- Systematic reviews of prenatal alcohol exposure effects generally only include conventional observational studies. However, estimates from such studies are prone to confounding and other biases.
- We conducted a comprehensive systematic review of experimental human data and alternative analytical approaches to improve causal inference based on observational data.
- We also developed customized risk of bias tools for Mendelian randomization, natural experiments and parental and sibling comparison, and applied them to studies with these designs.
- Our results showed a likely causal detrimental role of prenatal alcohol exposure on cognitive outcomes, and weaker evidence for a decrease in birthweight, confirming results from conventional observational studies.
- Guidance should continue to advise abstinence from alcohol in pregnancy.

## Introduction

The effects of prenatal alcohol consumption have typically been studied using standard analytical approaches in observational studies.<sup>1</sup> Systematic reviews have used these types of studies to determine the effects of prenatal alcohol exposure on several outcomes with a wide range and varying definition of alcohol intake including low-moderate to binge drinking. Outcomes such as central auditory disorders in children,<sup>2</sup> orofacial clefts,<sup>3</sup> speech and language<sup>4</sup> and several birth outcomes including low birthweight, pre-term birth and small for gestational age<sup>1,5,6</sup> have been investigated. These have led to varying results from systematic reviews: an increased risk of detrimental outcomes at very heavy drinking levels,<sup>1,2</sup> inconsistent evidence regarding effects of moderate, heavy, or binge drinking (5+ drinks on any occasion),<sup>3</sup> inconsistent effects from low-moderate alcohol consumption (up to 83 g/week)<sup>5</sup> and some evidence that even light prenatal alcohol consumption is associated with harmful birth outcomes (up to 32 g/week).<sup>6</sup> However, estimates from such studies are prone to the effects of: (i) confounding by socio-demographic characteristics (age, ethnicity, education, socio-economic position) and behavioural factors

(smoking and substance use) and (ii) measurement error, namely under-reporting of alcohol intake and/or recall bias. Therefore, the direction and size of any potential causal relationships cannot be determined without bias.

In recent decades, novel analytical approaches have been increasingly applied to data from observational studies in order to improve causal inference when assessing potential effects of prenatal alcohol exposure. These approaches include Mendelian randomization (MR),<sup>7</sup> family-based designs such as paternal or sibling comparison studies<sup>8</sup> and natural experiments.<sup>9</sup> Their respective strengths and limitations are outlined in [Box 1](#).

We conducted a systematic review of human studies that used experimental data [randomized controlled trials (RCTs)] or alternative analytical methods to improve causal inference applied to observational data, in order to determine the causal effects of maternal alcohol consumption in pregnancy on offspring outcomes at birth and later in life. Additionally, as is being recognised elsewhere,<sup>11–13</sup> it is important in public health and in epidemiology to include work from other disciplines in order to avoid missing important contributions to the literature. We therefore present a co-citation analysis to evaluate whether studies of alcohol

**Box 1. Outline of alternative analytical approaches (adapted from<sup>10</sup>)**

Causal inference approach	Definition	Biases addressed	Strengths	Limitations
Randomized controlled trial (RCT)	Subjects are randomly allocated to either exposure or control groups with assumption that there is no difference between the two groups except for the intervention they are receiving	Confounding, reverse causality, selection bias, loss-to-follow up bias (using intention-to-treat analysis), measurement error	Gold standard for estimating causal effects. Any effect is very likely to be causal if study has large number and trial is reliably performed	Generalizability may be questionable; impossible or unethical to randomize to certain exposures; can be expensive
Mendelian randomization (MR)	MR is the use of a genetic variant robustly associated with an exposure/risk factor of interest as an instrumental variable to test and estimate the causal effect of that exposure/risk factor with a disease or health related outcome	Confounding by shared genetics and environment, reverse causality, selection bias, measurement errors	Genetic instruments are not subject to confounding from environmental or lifestyle factors, are not influenced by the outcome, do not change over time and are measured with high accuracy	Low power, lack of instrumentation/weak instruments, horizontal pleiotropy, population stratification, inability to assess non-linear associations/dose-response estimation, assortative mating, not specific to the intrauterine period, developmental canalization, subject to dynastic effects bias
Sibling comparison	Compares outcomes when siblings are discordant for an exposure. If causal, then there will be evidence of a difference in outcome in relation to discordant exposure levels within sibships	Confounding by genetics and environment, specificity of effect to intrauterine period	Improves causal inference of intrauterine exposures. Controls for familial background and related confounding factors	Assumes a stable family environment; potential for confounding by factors not perfectly shared by siblings; potential for measurement error of exposure; limited power
Natural experiment	Empirical study approach where a population is exposed to an external event or intervention at a specific time point. Associations are then compared with a similar cohort that was not exposed. The assumption is that exposure is caused by quasi-random assignment	Confounding by genetics and environment, reverse causality, specificity of effect to intrauterine period	Can include study settings which would be impractical or unethical to produce by researchers. Allows for long-term effects of the exposure of interest	Selection bias if treatment and control group are not sufficiently comparable, some unobserved confounding may remain, it may not be possible to study non-linear associations/dose-response estimation
Parental comparison	Maternal-child association is compared with paternal-child association for inferring causal effect of intrauterine exposure. If causal, maternal association is stronger than paternal association. Where associations are similar for both parents we assume that they are driven by genetic or postnatal environmental characteristics	Confounding by genetics and environment, specificity of effect to intrauterine period	Improves causal inference of intrauterine effect if exposures are measured in both parents at same time in pregnancy, and non-paternity is taken into account for phenotypic traits	Assumption that paternal exposures share same confounding structure as maternal exposures may not be correct; where parental associations are of similar magnitude this may be due to offsetting paternal pathways rather than shared confounding. Assumes no assortative mating

in pregnancy carried out in other disciplines, such as health economics, are currently being recognised in public health.

## Methods

### Selection criteria and search strategy

The protocol for this systematic review, carried out using PRISMA guidelines, is available from the PROSPERO systematic review register (registration number CRD42015015941); [http://www.crd.york.ac.uk/PROSPERO/display\\_record.asp?ID=CRD42015015941](http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015015941).

We reported results from prospective observational studies on low-moderate consumption, adopting standard analytical approaches, in a separate manuscript.<sup>6</sup> Here, we focus on RCTs and studies that used alternative analytical methods to improve causal inference (see Box 1). MR studies that only reported results of geneXenvironment analyses (i.e. stratified by levels of maternal alcohol consumption) were excluded, as these estimates may incur selection bias.<sup>14</sup>

We adopted study specific definitions for all outcomes. Outcomes included the following. (i) Pregnancy outcomes: still birth [pregnancy loss after week 24, miscarriage, gestational length and preterm delivery (<37 weeks gestation)]; hypertensive disorders of pregnancy; gestational diabetes; small for gestational age (SGA, <10th percentile in weight or <-2 standard deviation scores) and birth size [weight (including low birth weight defined as <2500 g), length and head circumference]; low amniotic fluid (oligohydramnios); placenta previa; placental abruption; assisted delivery (including vacuum extraction, forceps delivery, Caesarean section); Apgar score at birth; admission to neonatal unit; congenital malformations. (ii) Features of fetal alcohol spectrum disorder (FASD): childhood growth restriction; cranium size and head circumference; developmental delays; behaviour problems; cognitive impairment and intelligent quotient (IQ); facial malformations.

The databases that were searched included: MEDLINE, PsycINFO, EMBASE on Ovid; the Cochrane Library including CENTRAL (the Cochrane Central Database of Controlled Trials) on Wiley Interscience; and Science Citation Index, Social Science Citation Index, on Web of Science from inception to 21 June 2018 (Supplementary Table 1, available as Supplementary data at *IJE* online). The search was limited to papers in English and excluded letters, animal studies, editorials and conference proceedings without corresponding full-text papers. Investigators tailored searches to each database. The search did not include grey literature and was focused on published medical literature. Additionally, we performed manual searches of the reference lists of: (i) papers included in recent systematic reviews

of the effects of prenatal alcohol exposure on the outcomes of interest; and (ii) all recent papers citing those reviews.

Titles and abstracts, and full texts if necessary, were screened independently by two reviewers. Discrepancies were discussed between reviewers and resolved through consensus.

### Data extraction

A custom-built Microsoft Access database was used to extract data. The following information from each study was extracted: title, authors, publication year, country/region, population characteristics (sample size, methods of sampling, age distribution, and ethnicity), study design, measures of exposure, assessment methods for outcomes (including whether this was derived from medical records, obtained via a research interview and the person reporting the outcome e.g. parent, teacher, health professional, researcher or child), model adjustments, and study results. If a study reported more than one result for each outcome, we extracted all of them (e.g. relative to different timing of exposure, model adjustments, etc.). Information from each included paper was extracted by the lead reviewer (L.M.) and subsequently checked for accuracy and completeness by another reviewer (H.B.E.).<sup>15</sup> There were very few extraction errors and these were resolved through discussion between extractor and checker.

### Data analysis

Odds ratios (OR) and 95% confidence intervals (CI) were derived from count data from individual studies, if they were not reported. Studies were meta-analysed if they used the same analytical approach and estimated the same outcome (e.g. MR analyses of the same genotype–outcome association, discordant siblings' analyses looking at the same outcome, etc.). The  $I^2$  statistic was used to determine percentage of variation due to heterogeneity.<sup>16</sup> Where only two studies were available to meta-analyse, results were not pooled if they were very different from each other.<sup>17</sup> Alternatively, a narrative summary of the results was given.

### Risk of bias assessment

The Cochrane risk of bias tool was selected to explore risk of bias in eligible randomized control studies.<sup>18</sup>

There are currently no widely accepted risk of bias assessment tools for the alternative observational study designs included in this systematic review (MR, sibling comparison, paternal comparison and natural experiments). We therefore considered the previous work in this

area<sup>14,19,20</sup> and adopted key criteria presented in these studies to assess risk of bias. Separate checklists for each of the four study types were developed (Supplementary Tables 3–6, available as Supplementary data at *IJE* online). The checklists mainly focused on the assumptions required for causal inference in these methods (Box 1). Definitions for what would be considered high, medium or low risk of bias for each domain within each separate tool were given. The assessment of each study using the relevant checklist was carried out independently by two reviewers. Conflicts of interest were avoided by making sure any paper whose author was also a reviewer was allocated to another reviewer.

### Co-citation

Co-citation data were collected from Web of Science. These data were analysed using VOSviewer version 1.6.5. Weights/bubble size correspond to the strength of co-citation. The distant between bubbles corresponds to the number of times that journals are cited together in other journals. The colours correspond to ‘communities’ (clustering) identified by the software, and not pre-specified scientific disciplines.

### Results

A flowchart of the article review process is shown in Fig. 1. A total of 5424 citation records were identified from searching the four relevant databases. A manual search of recent systematic reviews identified 34 additional articles. After exclusions, 9 MR analyses, 6 negative control studies, 1 RCT and 7 papers based on natural experiments were included, giving a total of 23 studies.

### Risk of bias assessment

Table 1 shows the results of risk of bias assessments. No study was rated low risk of bias in all domains. The RCT was judged at low risk of bias in all except in the blinding domain as participants were not blinded and self-reported their alcohol use. For natural experiment studies the main concerns with regard to validity were the differential trends in outcome, instrument strength and selection bias. For paternal comparison studies potential for differential paternal and maternal confounding and non-paternity were the key threats to validity. In the 2 sibling-comparison studies differential assessment of exposure was the main concern in both studies. All MR studies were rated at moderate risk of having a weak instrument. Further concerns were non-genetic (two studies rated at high risk), genetic confounding and pleiotropy. Because none of the studies are at low

risk of bias in all domains for any of the study types, it is not possible to be fully confident in our findings or to predict the direction potential biases could move the results towards. Nevertheless, despite some concerns specific to these study designs, the included studies still provide more robust evidence that is less prone to the type of confounding typically affecting traditional observational epidemiological studies.

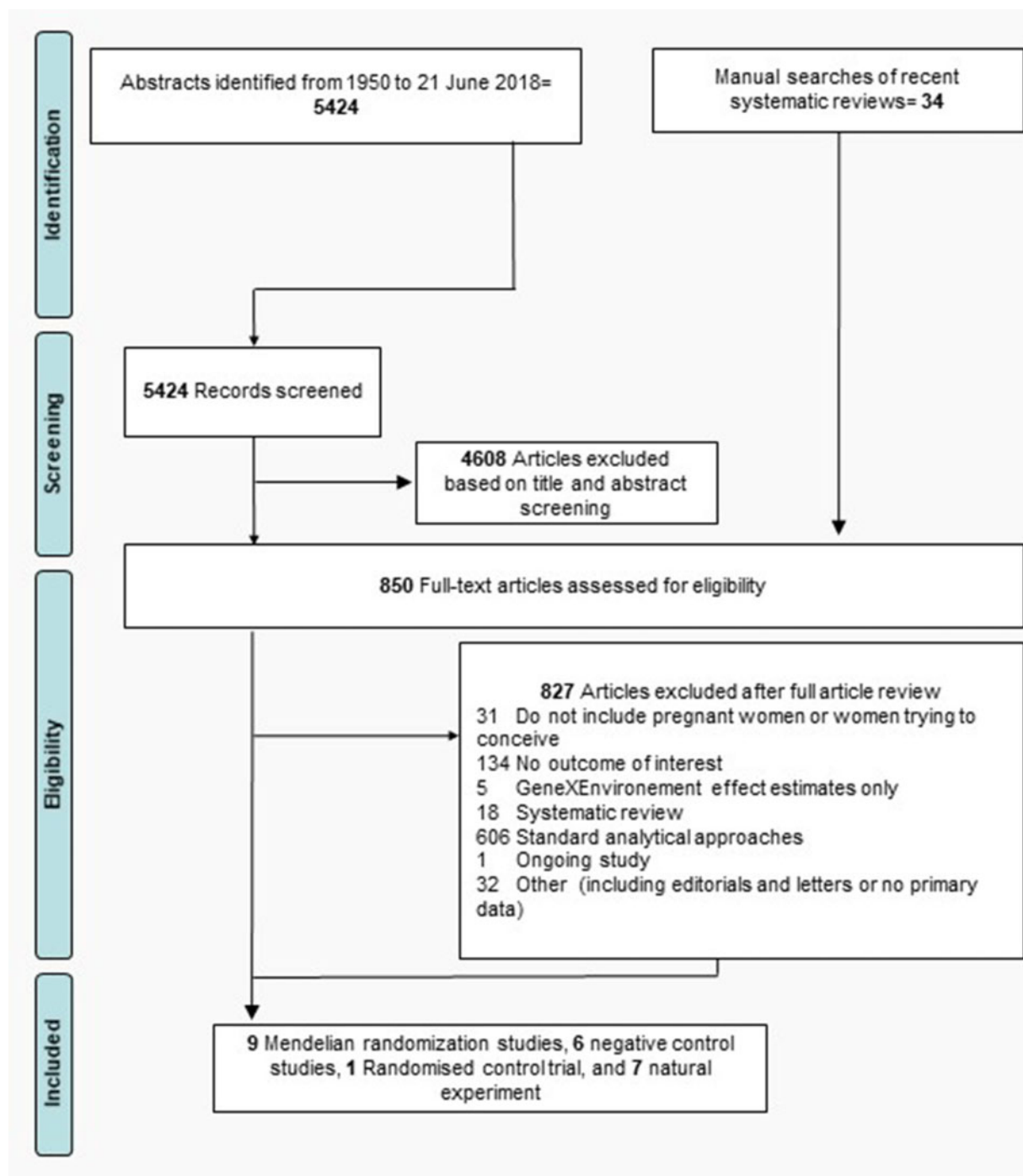
### Co-citation

Figure 2 illustrates patterns of journal co-citations. It shows four main journal clusters including (health) economics, clinical/alcohol research, genetics and epidemiology. The journal with the highest citation is ‘Alcoholism: Clinical and Experimental Research’. The two other journal disciplines with the highest tendency for co-citation are genetics and epidemiology. The (health) economics cluster has a weaker tendency for co-citation and is the most isolated. The weak cross-disciplinary citation between health economics and other public health/epidemiology/clinical journals could be due to several reasons including differences in the speed of publication as well as in the frequency of citations.

### Mendelian randomization studies

We identified 9 MR studies examining the effects of prenatal alcohol exposure on pregnancy or offspring outcomes (Table 2). All studies used known variants in alcohol dehydrogenase (ADH) genes in mothers and/or offspring as genetic proxies for the exposure: 5 employed a functional variant in ADH1B,<sup>21–23,27,29</sup> 2 a haplotype in ADH1C,<sup>24,25,28</sup> and 2 a number of ADH variants combined into an allele score<sup>26,29</sup> (Table 2). The ADH1B variant is known to alter alcohol metabolic rates<sup>44</sup> and has been shown to be robustly associated with alcohol consumption levels,<sup>45</sup> also in pregnant women.<sup>27</sup> There are two relevant ADH1B polymorphisms, rs1229984 and rs2066702, which define the ADH1B\*1, \*2 and \*3 alleles. The ADH1C haplotype affects alcohol metabolism to a lesser extent<sup>44</sup> and its effect on alcohol consumption is less clear.<sup>46</sup> Figure 3 shows a meta-analysis over 2 studies<sup>24,25</sup> exploring the impact of different maternal and fetal ADH1C alleles on development of infant oral cleft. For three allele comparisons (maternal \*2\*1 vs\*1\*1; fetal \*2\*1 vs\*1\*1 and fetal \*2\*2 vs\*1\*1) the  $I^2$  indicated results in the two studies were reasonably homogeneous, whereas for the maternal \*2\*2 vs\*1\*1 comparison, the  $I^2$  showed that the studies were not homogeneous, leading to a much larger overall confidence interval. The meta-analysis provided no evidence for an impact of any of the





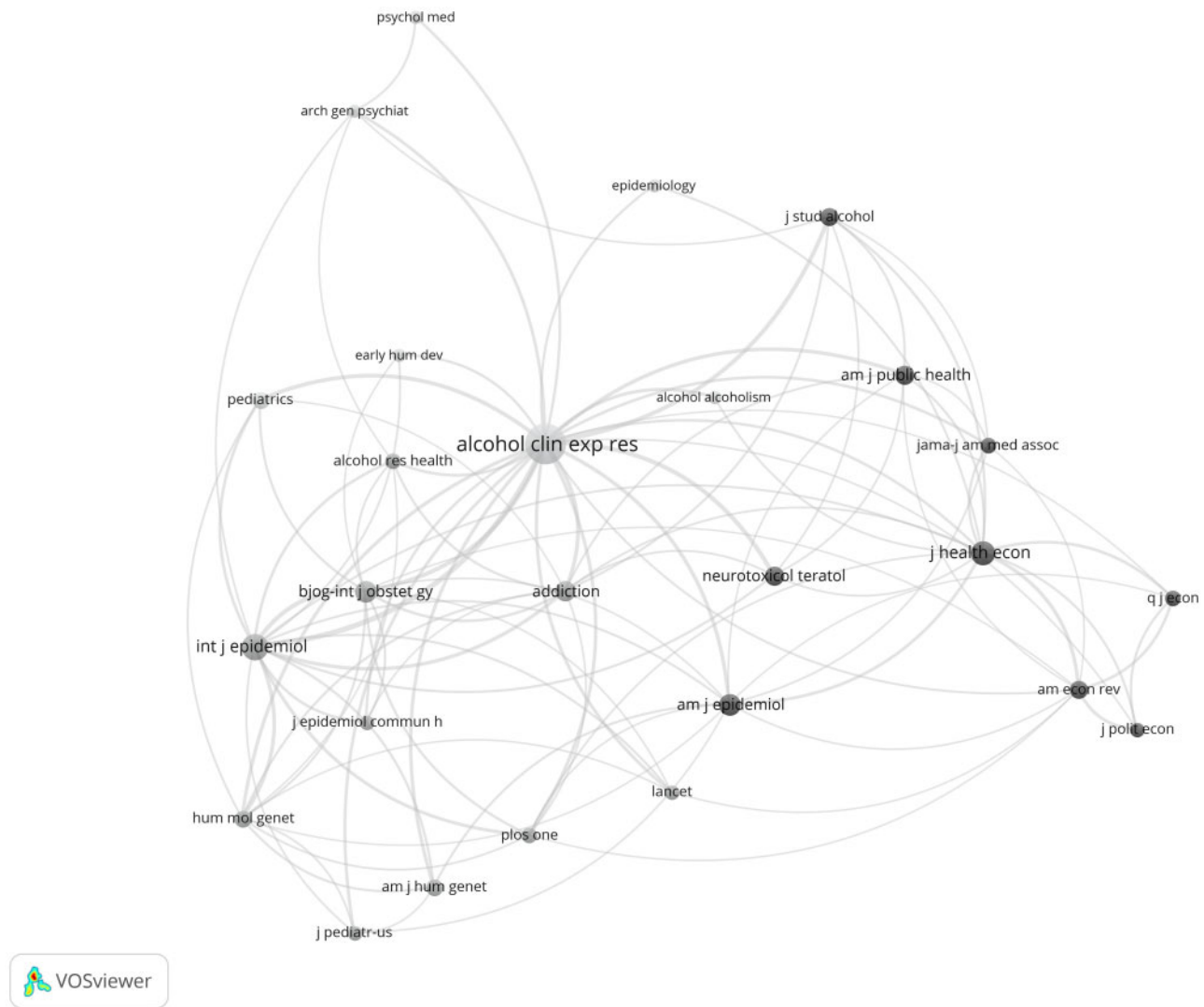
**Figure 1.** Flowchart of search strategy including primary reasons for article exclusion.

gene alleles on oral cleft. Two case-control studies examined the risk of oral cleft, comparing faster with slower metabolizers according to ADH1C maternal and fetal genotype. A French study found evidence of lower risk of non-syndromic cleft for ADH1C\*2\*2 compared with 1\*1 homozygotes, but did not report on whether genotype groups differed by alcohol consumption.<sup>24</sup> The study from Norway found no evidence of association with either

offspring cleft risk or maternal alcohol consumption (Table 2).<sup>25</sup>

#### Pregnancy outcomes

A study of African American infants found no strong evidence of association between infant ADH1B genotype and measures of birth size and gestational age, but did not report levels of maternal alcohol use by genotype<sup>23</sup> (Table 2).



**Figure 2.** Co-citation of journals. Bubble size corresponds to the magnitude of each journal's citation in the other journals (limit of minimum 8 citations per journal) with a total number of 26 journals. The distance between bubbles corresponds to the number of times with which journals are cited together in other journals. The colours correspond to communities identified by the software (VOS clustering). Produced in VOSviewer version 1.6.5.

### Features of FASD

The US-based study by Stoler *et al.*<sup>22</sup> found some evidence of higher odds of a FASD-like construct in offspring carrying the ADH1B\*3 allele compared with \*1\*1 homozygotes (Table 2). The latter metabolize alcohol more slowly and were also reported to have been exposed to lower levels of alcohol in pregnancy. The same direction of effect was observed comparing offspring of mothers carrying ADH1B\*3, and the evidence was stronger for those of black ethnicity.<sup>22</sup> Another study on fetal alcohol syndrome (FAS), from South Africa, found evidence of lower risk comparing carriers of maternal (or fetal) (fast metabolizing and lower alcohol intake) ADH1B\*2 with ADH1B\*1\*1 homozygotes (slower metabolizers and higher intake), and little evidence of an effect of ADH1B\*3 on FAS, in a mixed-ancestry South African

population (Table 2).<sup>21</sup> This study did not report on genotype–alcohol use association.

### Other outcomes

The four most recent (and by far the largest) MR studies reported on cognitive and behavioural childhood outcomes in the same UK-based cohort (Table 2).<sup>26–29</sup> Two used multiple offspring ADH variants known to be expressed in fetal life. One of these found evidence of association with IQ at 8 years old, but not when using the maternal allele score;<sup>26</sup> the effects were stronger for children of mothers reporting some alcohol consumption, but there was no evidence of association between the allele score and maternal alcohol use per se. The other study did not find an association between maternal genotype ADH1B\*2\*2 and an increased risk of children having early-onset-persistent

Table 1. Risk of bias assessment outcomes

Study ID	Random sequence generation	Allocation concealment	Blinding of participants/personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias
Tzilios 2011	Low risk	Low risk	Some concerns (Moderate risk)	Low risk	Low risk	Low risk	Low risk
Risk of bias assessment for Natural experiment studies <sup>a</sup>							
Study ID	Confounding 1—similar populations	Confounding 2—different trends in outcome	Confounding 3—population-level counterintervention	Instrument strength	Assessment bias—outcome	Selection bias	
Barraca 2015	Low risk	High risk	Low risk	Unclear risk	Low risk	Moderate risk	
Evans 2016	Moderate risk	Moderate risk	Low risk	Unclear risk	Low risk	High risk	
Fertig 2009	Low risk	Moderate risk	Low risk	Moderate risk	Low risk	High risk	
Nilsson 2017	Low risk	Low risk	Low risk	Moderate risk	Low risk	Low risk	
Zhang 2010	Moderate risk	Moderate risk	Moderate risk	Low risk	Low risk	Moderate risk	
Zhang 2011	Moderate risk	Moderate risk	Moderate risk	Low risk	Low risk	Moderate risk	
Cil 2017	Low risk	Moderate risk	Low risk	Low risk	Low risk	Moderate risk	
Risk of bias assessment for Parental Comparison studies <sup>b</sup>							
Study ID	Confounding 1—paternal and maternal confounding similar	Confounding 2—timing of both parents' exposure	Confounding 3—nonpaternity	Confounding 4—paternal effect	Assessment bias—exposure		
McCormack 2018	Moderate risk	Moderate risk	Moderate risk	Moderate risk	Low risk		
Alati 2008	Low risk	Low risk	Moderate risk	Low risk	Low risk		
Alati 2013	High risk	Low risk	Low risk	Low risk	Low risk		
Zuccolo 2016	Moderate risk	Low risk	High risk	Moderate risk	Low risk		
Risk of bias assessment for Sibling Comparison Studies <sup>c</sup>							
Study ID	Confounding 1—birth order	Confounding 2—individual level	Assessment bias—exposure	Assessment bias—outcome	Selection Bias		
D'Onofrio 2007	Moderate risk	Low risk	High risk	Low risk	Low		
Eliertsen 2017	Low risk	Low risk	Moderate risk	Moderate risk	Low risk		
Risk of bias assessment for Mendelian Randomization studies <sup>d</sup>							
Study ID	Weak instrument bias	Genetic confounding	*Non-genetic Confounding (e.g. lifestyle factors)	Pleiotropy—additional direct effects btw IV & outcome	Selection Bias		
Artfsten 2004	Moderate risk	Moderate risk	High risk	Moderate risk	Moderate risk		
Lewis 2012	Moderate risk	Low risk	Low risk	Low risk	Low risk		
Boyles 2010	Moderate risk	Moderate risk	Moderate risk	Moderate risk	Low risk		
Chevrier 2005	Moderate risk	Moderate risk	Moderate risk	Moderate risk	Low risk		
Stoler 2002	Moderate risk	Moderate risk	High risk	Moderate risk	Low risk		
Viljoen 2001	Moderate risk	Moderate risk	Low risk	Moderate risk	Moderate risk		
Zuccolo 2013	Moderate risk	Low risk	Moderate risk	Low risk	Low risk		

(Continued)



Von Hinke 2014	Moderate risk	Low risk	Low risk	Low risk	Low risk
Murray 2016	Moderate risk	Moderate risk	Moderate risk	Low risk	Low risk

<sup>a</sup>Natural experiment studies. Confounding 1—Similar population: were the populations compared similar with the exception of the naturally randomized exposure? Confounding 2—Differential trends: Has the outcome been changing over time differentially in the populations with and without the naturally randomized exposure? Confounding 3—Population level cointervention: have there been any other state-level changes coinciding with the natural experiment/intervention? Instrument strength: strength of adherence to law change (natural experiment) or other measures of compliance with instrument presented. Assessment bias—outcome: whether outcome assessment was valid and objective. Selection bias: whether the intervention/natural experiment caused a change in the distribution/characteristics of women getting pregnant, such as to introduce selection bias?

<sup>b</sup>Parental Comparison studies. Confounding 1—Paternal and maternal confounding similar: assumption that paternal exposures share same confounding structure as maternal exposure. Confounding 2—Timing of both parents' exposure: have exposures been measured in both parents at same time in pregnancy? Confounding 3—Nonpaternity: has nonpaternity been taken into account for phenotypic traits? Confounding 4—Paternal effect: is likelihood of a paternal effect also present?

<sup>c</sup>Sibling Comparison Studies. Confounding 1—Birth order: whether exposed siblings share the same birth order. Confounding 2—Individual level: whether additional intrauterine exposures and post-natal confounders were adjusted. Assessment bias—exposure: whether exposure assessment was free of recall bias. Assessment bias—outcome: whether outcome assessment was valid and objective. Selection bias: Whether there was likely loss to follow up bias for a cohort study.

<sup>d</sup>Mendelian Randomization studies. Weak instrument bias: Strength of association between instrument and exposure  $F$  statistic  $< 10$  in the same sample. Genetic confounding: whether the study reported test on association between confounders and instrumental variable (IV). Non genetic confounding: Whether the study reports on the distribution of genetic IVs and confounders other than ethnicity (e.g. lifestyle factors). Pleiotropy—additional direct effects btw IV & outcome: whether there is other known effect of genetic variants on outcome or its risk-factors, which is independent of alcohol. Selection bias: Whether the population was homogenous, stratified by ethnicity or adjusted for population stratification.

behavioural problems, however this may be due to lack of statistical power (Table 2).<sup>29</sup>

The other two studies both used the functional ADH1B variant, and found some evidence that the offspring of mothers genetically predisposed to consuming less alcohol had better academic performance at ages 7, 11, 14 and 16, but no association between offspring genotype and their educational outcomes,<sup>28</sup> nor was there evidence for an effect of genotype on IQ.<sup>27</sup> Both studies reported lower alcohol consumption in mothers carrying the rare ADH1B\*2 allele compared with the ADH1B\*1\*1 homozygotes.

### Sibling comparison studies

Two sibling-comparison studies compared behavioural outcomes in siblings differentially exposed to alcohol *in utero* (Table 2).

### Features of FASD

The study from the USA examined externalizing problems (measured through the Behaviour Problem Index) at ages 4–11 and found evidence that siblings exposed to moderate levels of prenatal alcohol had higher rates of conduct problems compared with their unexposed siblings, however there was no evidence of differences in attention or impulsivity problems.<sup>30</sup> The more recent study from Norway compared differentially exposed siblings in terms of their attention-deficit hyperactivity disorder (ADHD) at 5 years of age.<sup>31</sup> Results differed slightly depending on the ADHD scale used, with evidence of increased prenatal alcohol exposure being associated with higher ADHD levels according to the revised Conner's Parent Rating Scale, but less strong evidence for the Child Behaviour Checklist.<sup>31</sup>

### Parental comparison studies

Four maternal–paternal comparison studies met our inclusion criteria. These investigated the effects of prenatal alcohol exposure on neurocognitive domains in offspring: childhood educational achievement,<sup>33</sup> IQ,<sup>32</sup> cognitive development<sup>35</sup> and head circumference<sup>34</sup> (Table 2).

### Features of FASD

Two reports from the same UK-based study found no evidence of association between regular maternal alcohol use in pregnancy and either school results at 11<sup>33</sup> or IQ at 8 years of age.<sup>32</sup> One of the studies did find some evidence that increased levels of maternal binge drinking in pregnancy (consuming 32+g alcohol/occasion) were associated with decreased school results at age 11 years, whereas paternal exposure was associated with improved school results.<sup>33</sup> The other report did not find the same level of

**Table 2.** Quasi-experimental studies examining the effects of prenatal alcohol exposure (PAE) on pregnancy and childhood outcomes. \*3 Denotes rare allele found in sub-Saharan African populations; \*2 denotes rare allele found in European ancestry populations; \*1 denotes common allele found worldwide

Study (year)	Country	Sample size	Exposure definition (proxy for PAE)	Exposure categories	Age at outcome assessment (child)	Outcomes	Summary of results & conclusions as presented in the paper
<b>Mendelian randomization studies</b>							
Viljoen <i>et al.</i> (2001) <sup>21</sup>	South Africa	56 Mother-child pairs 178 Controls Khoisan-Caucasian mixed ancestry	Gene: ADH1B SNP-rs number: rs1229984	*1*1 Slow metabolizers (ref. category) *2* Fast metabolizers *3* Intermediate metabolizers	5-9 years	Fetal alcohol syndrome (FAS)	Fetal genotype: *3* vs *1*1 OR 0.85 (95% CI: 0.26, 2.72) *2* vs *1*1 OR 0.31 (95% CI: 0.10, 0.90) Maternal genotype: *3* vs *1*1 OR 0.85 (95% CI: 0.26, 2.72) *2* vs *1*1 OR 0.31 (95% CI: 0.10, 0.90) Maternal genotype (White women): OR *3*1 vs *1*1: 6.68 (95% CI: 0.89, 49.90) Maternal genotype (Black women): OR *3*1 vs *1*1: 2.62 (95% CI: 0.97, 7.04) Fetal genotype (White children): OR *3*1 vs *1*1: 4.75 (95% CI: 0.35, 64.74) Fetal genotype (Black children): OR *3*1 vs *1*1: 3.75 (95% CI: 1.00, 14.02) Fetal genotype: *3* vs *1*1 3211 vs 3196, P=0.83 *3* vs *1*1 39.6 vs 39.4, P=0.37 *3* vs *1*1 OR 0.85 (95% CI: 0.30, 2.39) *3* vs *1*1 OR 0.27 (95% CI: 0.06, 1.20) Maternal genotype: *2*1 vs *1*1 OR 0.93 (95% CI: 0.50, 1.90) *2*2 vs *1*1 OR 0.20 (95% CI: 0.10, 0.50) Fetal genotype: *2*1 vs *1*1 OR 1.37 (95% CI: 0.80, 2.40) *2*2 vs *1*1 OR 0.41 (95% CI: 0.20, 0.80) Maternal genotype: *2*1 vs *1*1 OR 0.93 (95% CI: 0.70, 1.24) *2*2 vs *1*1 OR 0.95 (95% CI: 0.67, 1.36) Fetal genotype: *2*1 vs *1*1 OR 1.05 (95% CI: 0.78, 1.41) *2*2 vs *1*1 OR 0.81 (95% CI: 0.56, 1.17) Fetal allele score <sup>a</sup> Beta -1.20 (95% CI: -1.89, -0.52)
Stoler <i>et al.</i> (2002) <sup>22</sup>	USA	173 White women 85 Black women	Gene: ADH1B SNP-rs number: rs1229984	*1*1 Slow metabolizers (ref. category) *3*1 Intermediate metabolizers	At birth	Blinded physician assessment of a composite trait including: growth restriction, microcephaly or 4 of 6 predefined facial features typical of FAS	
Arfsten <i>et al.</i> (2004) <sup>23</sup>	USA	305 African American infants	Gene: ADH1B SNP-rs number: rs1229984	*1*1 Slow metabolizers (ref. category) *3* Intermediate metabolizers	At birth	Birth weight (g) GA (weeks) LBW (<2500g) SGA	
Chevrier <i>et al.</i> (2005) <sup>24</sup>	France	205 Case children + parents (195 fathers and 211 mothers) 115 Control children and 54 mothers	Gene: ADH1C SNP-rs numbers: rs169342 rs698	Biallelic (2 haplotypes) *1*1 Faster metabolizers (ref. category) *2* Slower metabolizers	--	Non-syndromic oral clefts	
Boyles <i>et al.</i> (2010) <sup>25</sup>	Norway	Mother-child pairs (483 cases 503 controls)	Gene: ADH1C SNP-rs number: rs169342 rs698	Biallelic (2 haplotypes) *1*1 Faster metabolizers (ref. category) *2* Slower metabolizers	At birth	Oral cleft	
Lewis <i>et al.</i> (2012) <sup>26</sup>	UK	4 167 Mother-child pairs	Gene: ADH7 SNP-rs number: rs284779 Gene: ADH1B SNP-rs number: rs4147536 Gene: ADH1A SNP-rs number: rs975833 SNP-rs number: rs2866151	Allele score composed of 4 SNPs (unweighted) No information on how it affects metabolic rates	8 years	IQ: Wechsler Intelligence Scale for Children (WISC-III)	
Zuccolo <i>et al.</i> (2013) <sup>27</sup>	UK	6268 Mother-child pairs for KS2 analysis	Gene: ADH1B SNP-rs number: rs1229984	Biallelic SNP *1*1 Slow metabolizers (ref. category) *2* Fast metabolizers	8 years	Total IQ: Wechsler Intelligence Scale for Children (WISC)	Maternal genotype: *2* vs *1*1 MD -0.01 (95% CI: -2.8, 2.7) Maternal genotype: *2* vs *1*1 OR 0.85 (95% CI: 0.26, 2.72) *2* vs *1*1 OR 0.31 (95% CI: 0.10, 0.90) *3* vs *1*1 OR 0.85 (95% CI: 0.26, 2.72) *2* vs *1*1 OR 0.31 (95% CI: 0.10, 0.90) OR *3*1 vs *1*1: 6.68 (95% CI: 0.89, 49.90) OR *3*1 vs *1*1: 2.62 (95% CI: 0.97, 7.04) OR *3*1 vs *1*1: 4.75 (95% CI: 0.35, 64.74) OR *3*1 vs *1*1: 3.75 (95% CI: 1.00, 14.02) P=0.83 P=0.37 OR 0.85 (95% CI: 0.30, 2.39) OR 0.27 (95% CI: 0.06, 1.20) OR 0.93 (95% CI: 0.50, 1.90) OR 0.20 (95% CI: 0.10, 0.50) OR 1.37 (95% CI: 0.80, 2.40) OR 0.41 (95% CI: 0.20, 0.80) OR 0.93 (95% CI: 0.70, 1.24) OR 0.95 (95% CI: 0.67, 1.36) OR 1.05 (95% CI: 0.78, 1.41) OR 0.81 (95% CI: 0.56, 1.17) Beta -1.20 (95% CI: -1.89, -0.52)

(Continued)

**Table 2. Continued**

Study (year)	Country	Sample size	Exposure definition (proxy for PAE)	Exposure categories	Age at outcome assessment (child)	Outcomes	Summary of results & conclusions as presented in the paper
von Hinke Kessler Scholder <i>et al.</i> (2014) <sup>28</sup>	UK	4061 pairs for WISC analysis Mother-child pairs KS1: 3255 KS2: 3067 KS3: 2812 KS4: 3138	Gene: ADHDIB SNP-rs number: rs1229984	MR Biallelic SNP *1* Slow metabolizers (ref. category) *2* Fast metabolisers	7–16 years	Academic achievement: Key Stage 2 scores	*2* vs *1*1 MD 1.7 (95% CI: 0.4, 3.0) Fetal genotype: *2* vs *1*1 MD 0.68 (95% CI: -0.24, 1.60)
						Academic achievement: Key Stage 1 scores (age 7)	Maternal genotype: *2* vs *1*1 MD 0.198 (95% CI: 0.05, 0.35)
						Key Stage 2 scores (age 11)	Fetal genotype: *2* vs *1*1 MD -0.082 (95% CI: -0.24, 0.07)
						Key Stage 3 scores (age 14)	Maternal genotype: *2* vs *1*1 MD 0.239 (95% CI: 0.08, 0.40)
						Key Stage 4 scores (age 16)	Fetal genotype: *2* vs *1*1 MD -0.122 (95% CI: -0.30, 0.06) Maternal genotype: *2* vs *1*1 MD 0.192 (95% CI: 0.03, 0.36)
Murray <i>et al.</i> (2016) <sup>29</sup>	UK	Maternal genotype analysis: 3114 mother-child pairs	Gene: ADHDIB SNP-rs number: rs1229984	*1*1 Slow metabolizers (ref. category) *2* Fast metabolizers	4–13 years	Child's conduct problem trajectories aged 4–13 based on Strengths and Difficulties Questionnaire (SDQ), categorized as: low-risk (ref) early-onset-persistent	Maternal genotype: *2* vs *1*1 MD -0.075 (95% CI: -0.22, 0.07)
							Fetal genotype: *2* vs *1*1 MD 0.25 (95% CI: 0.10, 0.40)
Sibling control studies D'Onofrio <i>et al.</i> (2007) <sup>30</sup>	USA	8621	Siblings discordant for PAE	Moderate alcohol exposure vs none	4–11 years	Conduct problems	Moderate vs no alcohol intake: MD 0.05 SE: 0.02
						Attention/impulsivity problems	Moderate vs no alcohol intake: MD 0.03 SE: 0.02
Eilertsen <i>et al.</i> (2017) <sup>31</sup>	Norway	34 283	Siblings discordant for PAE	Not available	5 years	Attention-deficit hyperactivity disorder (ADHD) Scales: the revised Conner's Parent Rating Scale (CPRS-R) Child Behaviour Checklist (CBCL)	$\beta = 0.017$ , 95% CI: 0.005, 0.030 $\beta = 0.011$ , 95% CI: -0.002, 0.024
Parental control studies Alati <i>et al.</i> (2008) <sup>32</sup>	UK	4332	Mother and partner alcohol intake	First trimester (regular use): never <1 drinks/week 1–6 drinks/week 7+ drinks/week Binge drinking:	8 years	IQ: Wechsler Intelligence Scale for Children (WISC)	1st trimester regular alcohol use (pdiff = 0.43) MD per increase in maternal category: 0.03 (-0.58, 0.65) MD per increase in paternal category: 0.40 (-0.01, 0.82) Binge drinking: (Pdiff = 0.38)
						Maternal-paternal comparison	

(Continued)

Table 2. Continued

Study (year)	Country	Sample size	Exposure definition (proxy for PAE)	Exposure categories	Age at outcome assessment (child)	Outcomes	Summary of results & conclusions as presented in the paper
Alati (2013) <sup>33</sup>	UK	7062	Maternal-paternal comparison Mother and partner alcohol intake	1-4 drinks/occasion 5-10 drinks/occasion 10+ drinks/occasion First trimester (regular use): never <1 drinks/week 1-6 drinks/week 7+ drinks/week Binge drinking: 1-4 drinks/occasion 5-10 drinks /occasion 10+ drinks/occasion	11 years	Academic achievement: Key Stage 2 scores (standardized)	MD per increase in maternal category: -0.45 (-1.32, 0.43) MD per increase in paternal category: 0.10 (-0.36, 0.56) 1st trimester regular Alcohol use (Pdiff = 0.406) MD per increase in maternal alcohol category: Adjusted* MD: 0.10 (-0.17, 0.37) MD per increase in paternal alcohol category: 0.25 (0.07, 0.43), (pdiff 0.41) Binge drinking: (Pdiff <0.0001) MD per increase in maternal category: -0.68 (-1.03, -0.33) MD per increase in paternal category: 0.27 (0.07, 0.46)
Zuccolo <i>et al.</i> (2016) <sup>34</sup>	Norway	46 178	Maternal-paternal comparison Mother and partner alcohol intake	Non-drinker (ref) <1 unit/week 1-2 units/week 3-4 units/week 5+ units/week	At birth and at 3 months post-partum	Head circumference	At birth- Fully and mutually adjusted model <sup>c</sup> <1unit- Mother: mean difference (SD) 0.00 (95% CI: -0.02, 0.02) Father: mean difference (SD) -0.00 (95% CI: -0.05, 0.04) 1-2 units- Mother: mean difference (SD) -0.02 (95% CI: -0.05, 0.01) Father: mean difference (SD) 0.01 (95% CI: -0.03, 0.05) 3-4 units- Mother: mean difference (SD) 0.06 (95% CI: 0.02, 0.11) Father: mean difference (SD) 0.01 (95% CI: -0.03, 0.05) 5+ units- Mother: mean difference (SD) 0.01 (95% CI: -0.04, 0.06) Father: mean difference (SD) -0.01 (95% CI: -0.04, 0.03) At 3 months post-partum: fully and mutually adjusted model <sup>d</sup> <1unit- Mother: mean difference (SD) 0.02 (95% CI: -0.00, 0.05) Father: mean difference (SD) -0.02 (95% CI: -0.08, 0.04) 1-2 units- Mother: mean difference (SD) -0.02 (95% CI: -0.05, 0.02) Father: mean difference (SD) -0.03 (95% CI: -0.08, 0.02) 3-4 units- Mother: mean difference (SD) 0.04 (95% CI: -0.02, 0.10) Father: mean difference (SD) -0.04 (95% CI: -0.09, 0.01)

(Continued)

**Table 2. Continued**

Study (year)	Country	Sample size	Exposure definition (proxy for PAE)	Exposure categories	Age at outcome assessment (child)	Outcomes	Summary of results & conclusions as presented in the paper	
McCormack <i>et al.</i> (2018) <sup>35</sup>	Australia	2030	Maternal-paternal comparison	First 6 weeks:	12 months	Infant cognitive development (Bayley Scales for Infant Development, third edition)	5+ units- Mother: mean difference (SD) 0.05 (95% CI: -0.02, 0.12)	
				Abstinent			Trimester 1: first 6 weeks	Father: mean difference (SD) -0.05 (95% CI: -0.10, 0.00)
Natural experiments Fertig and Watson (2009) <sup>36</sup>	USA	All women 16 165 747 Native White women 11 426 203	Changes in minimum legal drinking age (MLDA)	Lower (18 years) vs higher (19-21 years)	At birth	Low birth weight (<2500g) Preterm birth (<37 weeks) Congenital anomalies	<1 unit- Mother: OR 0.68 (95% CI: 0.50, 0.94)	
				Moderate			Trimester 2:	At birth: fully and mutually adjusted model <sup>f</sup>
				Binge			Abstinence	<1 unit- Mother: OR 0.53, 1.88)
				Heavy			Low	Father: OR 1.00 (95% CI: 0.53, 1.88)
				Second 6 weeks			Abstinence	1-2 units- Mother: OR 1.13 (95% CI: 0.81, 1.59)
				Abstinence			Low	Father: OR 1.11 (95% CI: 0.66, 1.87)
				Low			Trimester 3	3-4 units- Mother: OR 0.97 (95% CI: 0.57, 1.68)
				Trimester 2:			Abstinence	Father: OR 1.21 (95% CI: 0.72, 2.06)
				Abstinence			Low	5+ units- Mother: OR 1.22 (95% CI: 0.68, 2.20)
				Trimester 3			Abstinence	Father: OR 1.36 (95% CI: 0.81, 2.28)
				Low			Low	At 3 months post-partum: fully and mutually adjusted model <sup>f</sup>
				Low			Low	<1 unit- Mother: OR 0.82 (95% CI: 0.66, 1.02)
Low	Low	Father: OR 1.25 (95% CI: 0.79, 1.95)						
Low	Low	1-2 units- Mother: OR 0.82 (95% CI: 0.62, 1.09)						
Low	Low	Father: OR 1.16 (95% CI: 0.79, 1.71)						
Low	Low	3-4 units- Mother: OR 1.08 (95% CI: 0.73, 1.59)						
Low	Low	Father: OR 1.38 (95% CI: 0.94, 2.03)						
Low	Low	5+ units- Mother: OR 0.82 (95% CI: 0.49, 1.39)						
Low	Low	Father: OR 1.33 (95% CI: 0.90, 1.95)						
Low	Low	Maternal alcohol use (compared with abstinence):						
Low	Low	Trimester 1: first 6 weeks						
Low	Low	Low: ( $\beta$ -0.45, SE 0.86)						
Low	Low	Moderate: ( $\beta$ 1.35, SE 1.61)						
Low	Low	Binge: ( $\beta$ -0.90, SE 0.96)						
Low	Low	Heavy: ( $\beta$ -0.13, SE 1.02)						
Low	Low	Trimester 1: second 6 weeks						
Low	Low	Low: ( $\beta$ 0.54, SE 0.86)						
Low	Low	Trimester 2						
Low	Low	Low: ( $\beta$ 2.11, SE 0.77)						
Low	Low	Trimester 3						
Low	Low	Low: ( $\beta$ 1.60, SE 0.77)						
Low	Low	Partners alcohol use (compared with abstinence):						
Low	Low	Low: ( $\beta$ 2.42, SE 1.55)						
Low	Low	Moderate: ( $\beta$ 0.67, S.E 1.81)						
Low	Low	Heavy: ( $\beta$ 2.19, SE 1.98)						
Low	Low	Binge: ( $\beta$ 2.00, SE 1.58)						
Low	Low	MLDA of 18 main effect; -0.17% (SE 0.07) <sup>#</sup>						
Low	Low	MLDA of 18 x mother $\leq$ 17 years of age interaction; 0.50% (SE 0.18) <sup>#</sup>						

(Continued)

Table 2. Continued

Study (year)	Country	Sample size	Exposure definition (proxy for PAE)	Exposure categories	Age at outcome assessment (child)	Outcomes	Summary of results & conclusions as presented in the paper
		Native Black women 3 032 108					MLDA of 18 x mother 18–20 years of age interaction; 0.26% (SE 0.10) <sup>§</sup> MLDA of 18 main effect; -0.35% (SE 0.09) <sup>§</sup> MLDA of 18 x mother ≤17 years of age interaction; 0.86% (SE 0.28) <sup>§</sup> MLDA of 18 x mother 18–20 years of age interaction; 0.26% (SE 0.12) <sup>§</sup> MLDA of 18 main effect; -0.18% (SE 0.12) <sup>§</sup> MLDA of 18 x mother ≤17 years of age interaction; -0.04% (SE 0.05) <sup>§</sup> MLDA of 18 x mother 18–20 years of age interaction; -0.03% (SE 0.02) <sup>§</sup> Beer tax ( $\beta$ 0.931, SE 0.003) Wine tax ( $\beta$ 0.340, SE 0.006) Liquor tax ( $\beta$ 0.072, SE 0.027) Beer tax ( $\beta$ -0.023, SE 0.001) Wine tax ( $\beta$ -0.006, SE 0.0004) Liquor tax ( $\beta$ -0.001, SE 0.0001) Beer tax ( $\beta$ -0.002, SE 0.0004) Wine tax ( $\beta$ -0.002, SE 0.001) Liquor tax ( $\beta$ -0.001, SE 0.001) Beer tax ( $\beta$ -0.0002, SE 0.000001) Wine tax ( $\beta$ 0.0002, SE 0.0002) Liquor tax ( $\beta$ -0.0001, SE 0.0000)
Zhang (2010) <sup>37</sup>	USA	Infants with birthweight 71 501 237 Infants with APGAR scores 55 054 916	Raising alcohol taxes	Higher vs lower alcohol tax	At birth	Birthweight (g) Low birthweight (<2.500 g) Extremely low birthweight (<1.500 g) Low APGAR scores (<7)	
Zhang and Caine (2011) <sup>38</sup>	USA	All women (<21 years) 26 743 White women (<21 years) 16 596 Black women (<21 yrs) 11 147	State-specific MLDA when woman is 14 years as proxy of alcohol availability	Effects of different MLDA (18–21.21 years), when woman is 14 (e.g. proxying for different alcohol availability)	At Birth	Low birthweight (<2.500 g) Low APGAR scores (<7) Preterm birth (<37 weeks)	MLDA of 18 (vs higher); 0.14% ( $P$ < 0.0001) <sup>h</sup> MLDA of 19 (vs 18); -0.16% ( $P$ = 0.002) <sup>h</sup> MLDA of 20 (vs 18); -0.05% ( $P$ = 0.217) <sup>h</sup> MLDA of 21 (vs 18); -0.24% ( $P$ < 0.0001) <sup>h</sup> MLDA of 18 (vs higher); 1.12% ( $P$ < 0.0001) <sup>h</sup> MLDA of 19 (vs 18); -1.80% ( $P$ < 0.0001) <sup>h</sup> MLDA of 20 (vs 18); -1.03% ( $P$ < 0.0001) <sup>h</sup> MLDA of 21 (vs 18); -1.82% ( $P$ < 0.0001) <sup>h</sup> MLDA of 18 (vs higher); 0.02% ( $P$ = 0.051) <sup>i</sup> MLDA of 19 (vs 18); -0.02% ( $P$ = 0.051) <sup>i</sup> MLDA of 20 (vs 18); 0.01% ( $P$ = 0.215) <sup>i</sup> MLDA of 21 (vs 18); -0.04% ( $P$ = 0.0002) <sup>i</sup> MLDA of 18 main effect; -0.19% (SE 0.08) <sup>j</sup> MLDA of 18 x mother 14–17 years of age interaction; -0.02% (SE 0.07) <sup>j</sup> MLDA of 18 x mother 18–20 years of age interaction; 0.10% (SE 0.05) <sup>j</sup> Mean of outcome: 7.5
Barreca and Page (2015) <sup>39</sup>	USA	14–17 years at conception 3 314 000 18–20 years at conception 6 287 000 21–24 years at conception 10 178 000	Differences in MLDA	Lower (18 years) vs higher (19–21) MLDA	At birth	Low birthweight (<2500 g) Preterm birth (<37 weeks) Low Apgar score (<7) Congenital anomaly Female	MLDA of 18 main effect; -0.04% (SE 0.10) <sup>j</sup> MLDA of 18 x mother 14–17 years of age interaction; 0.05% (SE 0.12) <sup>j</sup>

(Continued)





Table 2. Continued

Study (year)	Country	Sample size	Exposure definition (proxy for PAE)	Exposure categories	Age at outcome assessment (child)	Outcomes	Summary of results & conclusions as presented in the paper
Randomized controlled trial Tzilos <i>et al.</i> (2011) <sup>43</sup>	USA	Intervention group 27 Control group 23	Computer-delivered brief intervention for reduced prenatal alcohol use (33 days)	Children born to mothers receiving intervention on prenatal alcohol use vs to mothers receiving standard care	At birth	Birthweight	A reduction in years of schooling $-0.31$ (SE 0.09) years, males $-0.52$ (SE 0.17) years, females $-0.21$ (SE 0.12) Were less likely to complete high school $-0.63$ (0.02), males $-0.1$ (SE 0.02), females $-0.03$ (SE 0.03) Had lower (log) earnings $-0.24$ (SE 0.09), males $-0.24$ (SE 0.11), females $-0.17$ (0.14) An increased risk of no labour income 0.07 (SE 0.03), males 0.08 (SE 0.02), females 0.06 (SE 0.04) A higher proportion were on welfare 0.04 (0.01), males 0.05 (SE 0.02), females 0.03 (SE 0.01) Intervention group: mean = 3189.6, SD = 328.0 Control group: mean = 2965.3, SD = 387.7

LBW, low birth weight; ELBW, extremely low birthweight; GA, gestational age; APCR, appearance, pulse, grimace, activity, respiration; MLDA, minimum legal drinking age; NICU, neonatal intensive care unit; OR, odds ratio; CI, confidence interval; MD, mean difference; SD, standard deviation; SGA, small for gestational age; IQ: intelligence quotient; RR, relative risk.

<sup>a</sup>Unweighted allele score composed of the four fetal SNPs.

<sup>b</sup>Adjustments: mother's ancestry principle components from Genome wide association studies (GWAS) analysis.

<sup>c</sup>Adjustments: sex, other parent's alcohol consumption, maternal age, parity, socio-economic position, ethnicity, and, maternal and paternal education and smoking.

<sup>d</sup>Test for trend maternal and paternal alcohol intake in full model:  $P = 0.267$  and  $P = 0.201$ .

<sup>e</sup>Test for trend maternal and paternal alcohol intake in full model:  $P = 0.390$  and  $P = 0.124$ .

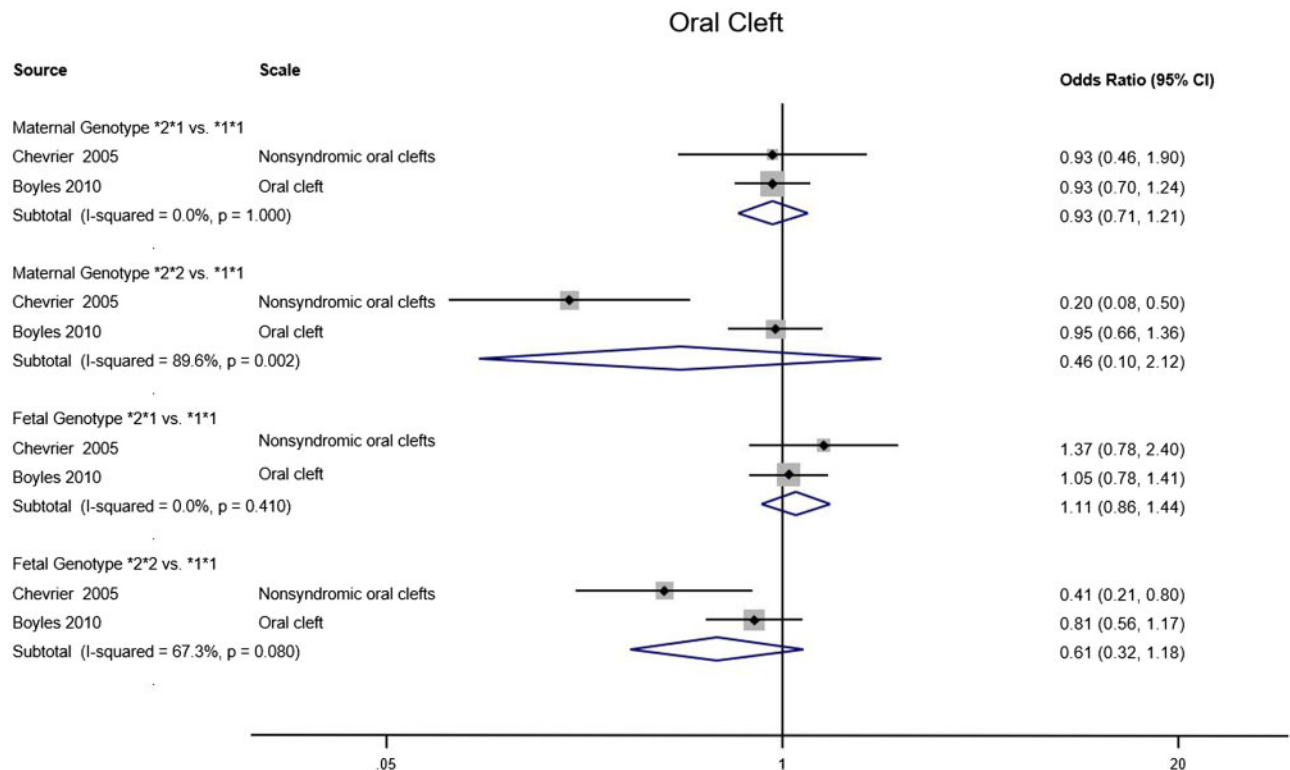
<sup>f</sup>Test for trend maternal and paternal alcohol intake in full model:  $P = 0.545$  and  $P = 0.056$ .

<sup>g</sup>Test for trend maternal and paternal alcohol intake in full model:  $P = 0.178$  and  $P = 0.090$ .

<sup>h</sup>Adjustments: state fixed effects, year-month fixed effects, maternal age fixed effects, state-specific time trends and birth characteristic controls.

<sup>i</sup>Adjustments: state fixed effects, year fixed effects, mother's education, age, marital status, smoking during pregnancy, real income per capita and real beer taxes (federal plus state level).

<sup>j</sup>Adjustments: state fixed effects, year-by-month fixed effects, age fixed effects, state-specific trends, state-by-year fixed effects, state-by-age fixed effects and state-by-year fixed effects.



**Figure 3.** Pooled odds ratios for outcomes of oral cleft in two MR studies.

evidence to support an association of prenatal binge drinking with offspring IQ at age 8 years.<sup>32</sup> In a large Norwegian cohort, there was no evidence of association between maternal or paternal alcohol use during or before pregnancy and head circumference at birth or 3 months.<sup>34</sup> In the same study, odds of microcephaly increased with higher paternal but not maternal alcohol consumption prior to pregnancy and in the first trimester.<sup>34</sup> A recent Australian study showed no consistent evidence of association between maternal alcohol use in different trimesters of gestation and cognitive function in children aged 1 year (Bayley Scales of Infant Development), and even scatter evidence for partner alcohol intake.<sup>35</sup>

### Natural experiments

Seven reports analysed data from natural experiments involving changes in government laws that effected the availability or affordability of alcohol,<sup>36–42</sup> or required point-of-sale warnings about the risks of drinking alcohol during pregnancy<sup>41</sup> (Table 2).

### Pregnancy outcomes

Three US-based studies used reductions in the minimum legal drinking age (MLDA) to proxy for prenatal alcohol exposure, under the assumption that a lower MLDA would increase alcohol availability to young women<sup>36,38,39</sup>

(Table 2). The studies by Fertig and Watson<sup>36</sup> and Barreca and Page<sup>39</sup> were based on US-wide birth data and estimated the association between MLDA and low birthweight (<2500 g), preterm delivery (<37 weeks) and congenital anomalies, with the latter additionally examining Apgar scores. Both used a triple difference approach (Supplementary Material, available as Supplementary data at *IJE* online) and substantially the same data, although the latter study ran additional analyses with more covariates and interaction terms to check the robustness of the model to some of its assumptions. When running similar age-specific analyses, the second study replicated the first study's results of an increase in both preterm deliveries and low birthweight corresponding to a lowering of MLDA, more marked for babies conceived to younger (<18 year old) compared with older (18–20 year old) women.<sup>36,39</sup> In more fully adjusted analyses, the negative association with birthweight was still found to be robust for younger mothers (<18 years). However, no consistent evidence of association was found for other age groups in the main effects analyses, or for other adverse fetal outcomes including gestational age, congenital abnormalities and Apgar score.<sup>39</sup> Neither study reported data on actual population-level alcohol use. The third study, by Zhang and Caine (2011),<sup>38</sup> investigated the same outcomes (low birthweight, preterm delivery and Apgar scores) in relation to a State's MLDA at the time a woman is 14 years old. The difference with

**Table 3.** Summary of direction of association of prenatal alcohol exposure with selected outcomes (cognitive/brain development and birthweight), in the context of expected and observed differences in prenatal alcohol exposure in each study

Outcome	Study	Direction PAE→outcome	Direction	
			Exposure proxy→PAE	
			Expected	Observed
Cognition, brain development	Lewis <i>et al.</i> (2012) <sup>26</sup>	↓	NA (Not Applicable)	NA
	Zuccolo <i>et al.</i> (2013) <sup>27</sup>	↑	↓	↓
	von Hinke Kessler Scholder <i>et al.</i> (2014) <sup>28</sup>	↑	↓	↓
	Zuccolo <i>et al.</i> (2016) <sup>34</sup>	↔	↑	↑
	McCormack <i>et al.</i> (2018) <sup>35</sup>	↔	↑	↑
	Alati <i>et al.</i> (2008) <sup>32</sup>	↔	↑	↑
	Alati <i>et al.</i> (2013) <sup>33</sup>	↓	↑	↑
	Nilsson (2017) <sup>42</sup>	↓	↑	NA
	Evans <i>et al.</i> (2016) <sup>40</sup>	↑	↓	NA
Birthweight	Arfsten <i>et al.</i> (2004) <sup>23</sup>	↔	NA	NA
	Fertig and Watson (2009) <sup>36</sup>	↓	↑	NA
	Barreca and Page (2015) <sup>39</sup>	↓	↑	NA
	Zhang and Caine (2011) <sup>38</sup>	↓	↑	NA
	Zhang (2010) <sup>37</sup>	↑	↓	↓
	Cil (2017) <sup>41</sup>	↓	↑	↑
	Tzilos <i>et al.</i> (2011) <sup>43</sup>	↑	↓	↔

respect to the two previous studies was that the ‘exposed’ status is assigned based on MLDA at the time the women are 14 years, regardless of what it is when she is older and pregnant. The authors hypothesize that the drinking environment at age 14 sets a woman’s future ‘drinking propensity’ including binge drinking behaviour, but no data were reported to confirm this. The estimates were derived from difference-in-difference specifications, but with additional controls for State-specific effects. The authors presented evidence that women who lived in a State where the MLDA was 18 years at the time they themselves were 14 years, compared with those in States with higher MLDA, had higher chances of giving birth to low birthweight babies with lower Apgar scores, but no association with prematurity.<sup>38</sup>

A fourth paper examined the effect of within-State changes in alcohol taxation in the US and within-State variation in birthweight and Apgar scores<sup>37</sup> (Table 2). The authors found evidence that increases in alcohol taxes are associated with increases in birthweight and Apgar scores. The authors also tried to validate their assumptions that changes in taxation are a valid proxy for alcohol consumption and therefore prenatal alcohol exposure, by regressing several alcohol drinking variables from a federal behavioural survey on alcohol taxation. They found some evidence of reduced binge drinking behaviour among pregnant women, corresponding to increases in alcohol taxes,

however no evidence that the quantity consumed was sensitive to alcohol pricing.<sup>37</sup>

Another US-based study explored the impact of State laws requiring point-of-sale warnings about the risks of drinking alcohol during pregnancy on outcomes including birthweight, pre-term birth, FAS and Apgar scores.<sup>41</sup> There was evidence that the warnings reduced the chances of very low birth weight babies (<1500 g), but no evidence of association with the other outcomes. The authors validated their assumption that alcohol warning signs would reduce prenatal alcohol exposure by regressing several alcohol drinking variables on whether the State prescribed health warnings or not, using both individual birth and national survey data. They found that adoption of the law was associated with a reduction in alcohol consumption and binge drinking among pregnant women.

#### Features of FASD

Two studies looked at long-term offspring outcomes (Table 2). Based on data from World War II US enlistees, the first study used different timings of prohibition implementation in different States to proxy for reduced likelihood of prenatal alcohol exposure as a result of reduced availability to women, and examined attained education and height in adult offspring.<sup>40</sup> The authors report an increase in years of education associated with the introduction of prohibition, but no evidence of an effect on height.

However, there were no estimates of actual alcohol consumption in States introducing prohibition.<sup>40</sup>

A Swedish study compared earnings, education and welfare dependency rates in children born in counties that did and did not relax the regulation of alcohol sales in 1967.<sup>42</sup> The relaxation of alcohol policy, used as a proxy for increased prenatal alcohol exposure, was shown to be related to reduced earnings, years of schooling and high school completion rates, as well as to a higher proportion of individuals on welfare.<sup>42</sup> The author reported some evidence of increased consumption of alcohol for the counties during the period where the more liberal policy applied, but no results specifically for pregnant women.

### Randomized controlled trial

We included one RCT<sup>43</sup> feasibility study with a small sample size (control group 23 women, intervention group 27 women; [Table 2](#)).

### Pregnancy outcomes

In the RCT feasibility study, 50 pregnant women who screened positive for risky drinking were randomized: 27 pregnant women in the intervention group received a 20-min computer-based, self-administered program intended to motivate them to reduce their drinking, whereas 23 pregnant women in the control group received a questionnaire about television preferences. Follow-up after 1 month (average 33 days) showed no difference in alcohol use between the intervention and control groups but some evidence of higher birthweight for infants born to women in the intervention group compared with the control group. As there was no strong evidence of a difference in alcohol consumption between the randomized groups this does not support any causal effect of alcohol on birthweight but may suggest bias in the RCT, some pathways (other than change in alcohol) from the intervention to birthweight that might counter any effect of alcohol and/or too little power to detect effects on alcohol robustly.

## Discussion

### Summary of the evidence

Our systematic review of the literature found a limited number of studies addressing the effects of prenatal alcohol exposure using experimental designs or alternative analytical strategies to improve causal inference in observational studies, which we described in narrative format. Twenty-three reports were included, representing five types of study design, with MR and natural experiments the most common designs (9 and 7 studies, respectively). Cognitive

outcomes were the most commonly reported (by 9 studies), followed by birthweight (7 studies). The overall picture that emerges from this review is that moderately strong evidence exists for detrimental effects of prenatal alcohol exposure on cognitive outcomes ([Table 3](#)). For cognitive outcomes and birth weight outcomes, we found the highest degree of consistency across study types (MR,<sup>26–28</sup> parental comparisons<sup>33</sup> and natural experiments exploiting different policy changes<sup>40,41</sup>) as well as with the direction of association predominantly reported in conventional epidemiological studies.<sup>47,48</sup> Based on natural experiments<sup>36–39</sup> and one feasibility RCT,<sup>43</sup> some evidence was also found for reduced birthweight following higher prenatal alcohol exposure ([Table 3](#)), in line with recent reviews<sup>6</sup> and pooled analyses of observational studies.<sup>49</sup>

Only one outcome-study design combination had more than one result that could be combined into a meta-analysis. For the rest, we described results in narrative format. We also developed and deployed customized risk of bias (RoB) assessment tools for the different types of study design. None of the studies scored 'low' RoB in all domains, therefore we recommend caution in interpreting the results of any one study as 'causal', since it is impossible to predict the overall direction of bias affecting each result.

Results of our co-citation analysis showed that the field of (health) economics is relatively isolated compared with the other clusters. It also shows a limited number of studies in public health. This shows that the findings published in health economics journals are not well recognised in the fields of epidemiology and public health, although the evidence they contribute should be considered alongside that from more traditional epidemiological studies when updating public health guidance on alcohol use, as evidenced by our reviewing efforts.

### Strengths and limitations of alternative study designs

An extensive literature exists exploring the strengths and limitations of the observational study designs and analytical strategies<sup>19</sup> included in this review, especially when applied to the study of intergenerational effects such as here.<sup>50,51</sup> In theory, all study types attempt to minimize confounding by shared genetic and environmental factors by design, all but MR and some of the natural experiments address the specificity of the effect to the intrauterine period (i.e. not confounded by postnatal alcohol use), and MR and natural experiments avoid reverse causality (Box 1). In practice, sources of bias varied both across and within each study-type category, as evidenced by our customized RoB tools showing some of the included studies

being at higher risk of bias than others. For example, data availability may restrict the extent to which one can test and/or account for potential differential trends in studies exploiting natural experiments such as MLDA. Similarly, data availability may restrict the extent to which one can explore whether (in particular historic) policies affected prenatal alcohol consumption. This is also true for many of the (particularly older) MR studies that did not report genotype associations with maternal alcohol use. Furthermore, ensuring that the analytical strategy identifies effects that are specific to the intrauterine period may be difficult. For example, a reduction in the MLDA in the year of birth is likely to be related to alcohol exposure in that year, but potentially also in the year after. This is less of an issue in studies that exploit temporary changes in alcohol exposure, such as Nilsson, as temporary policies are more likely to only affect alcohol exposure at that point in time only.<sup>42</sup> In MR studies, one analytical strategy that improves specific attribution of effects to the intrauterine period is using alcohol metabolizing genotypes in the offspring (not just the mothers) as proxy for prenatal alcohol exposure. This is because maternal genotype in theory predisposes to lower or higher alcohol use in pregnancy as well as before and after (therefore it is not specific to the intrauterine period). Additionally, MR studies of intrauterine exposures that do not account for both offspring and maternal genotype can suffer from bias because of violation of the exclusion restriction assumption.<sup>52</sup> On the other hand, offspring genotype (conditional on maternal genotype) is more specific, since children do not consume alcohol themselves and the only time in early life where they are exposed to alcohol is *in utero*. Therefore, different alcohol metabolizing genotypes in the offspring could modulate prenatal alcohol exposure, independently of maternal alcohol use. This strategy of presenting results for offspring genotype adjusted for maternal genotype was only adopted by a couple of the included MR studies and has the additional advantage of minimizing dynastic effects bias.

An additional strength of some of the natural experiments included here is that they investigated possible mechanisms for the observed effects of prenatal alcohol exposure, in particular through a postulated increase in unplanned pregnancies (also known as 'compositional changes'). This was explored through, e.g. sensitivity analyses to test whether MLDA changes resulted in more unplanned pregnancies. The idea is that, if MLDA led to an increase in unplanned pregnancies, this may have particularly affected mothers with a systematically different e.g. socio-economic position, whose children also have systematically different outcomes. But these effects are then driven by socio-economic confounding, not (necessarily) only by intrauterine toxicity. This was done by Fertig and

Watson<sup>36</sup> examining the percentage of births recorded with missing paternal information, with the analysis confirming evidence of effect for this in black women, and stronger effects in younger girls (<18 years), thus providing a possible partial explanation for the birthweight effects in their study. Compositional changes or changes in the demographics of mothers giving birth, are also thought to play a role in explaining some of the effect on adverse pregnancy outcomes observed in the study by Zhang<sup>37</sup> Specifically, since an increase in alcohol taxes appeared to lead to a reduction in pregnancies amongst younger and less educated mothers, who are more likely to experience adverse pregnancy outcomes, maternal age and education (over and above alcohol consumption per se), may explain some of the apparent effect of alcohol. The study by Nilsson<sup>42</sup> was able to avoid potential bias due to possible compositional changes by focusing on children who were conceived prior to the start of the relaxation of alcohol policy. Hence, his study did not include children who were conceived due to the change in alcohol policy.

Another study by Barreca and Page<sup>39</sup> additionally investigated the presence of an early selection effect that intrauterine alcohol exposure could have on the least healthy fetuses, by examining gender ratio of live births as a marker of early fetal loss. The authors' interpretation, although highly speculative, is that this selection indeed is present and could explain the unexpected direction of effect for the main effect analyses in their study.

Small sample sizes in many of the studies (especially for the earlier studies) means that estimates were often imprecise. This was particularly true for the MR studies, some of which were among the first ever to be conducted, and none of which adopted a multi-cohort approach to increase sample size, or multiple genetic variants to improve the variance explained in alcohol consumption, as is recommended and customary in recent times.<sup>53</sup>

Another limitation of the MR and natural experiment studies that were included in this review is the inability to provide dose-response estimates. Instead, they provide estimates of the effect of prenatal alcohol exposure around mean levels of consumption in the study sample. This falls short of the most interesting research question which is whether the effects are linear or whether there is a threshold at low levels of drinking under which alcohol is not harmful to the fetus.

Additionally, for MR studies, only ADH variants have been used and there is a possibility that acetaldehyde is both the deterrent to drinking and the cause of damage, which could lead to null results. Many more loci affecting alcohol intake are now available for future studies,<sup>54</sup> although their effect on prenatal alcohol use will require validation in studies of pregnant women.



## Strengths and limitations of this systematic review

This systematic review is the first of its kind to explicitly search for and integrate the evidence from different study designs and analytical approaches in a true triangulation framework.<sup>55</sup> Efforts were made to include studies from different disciplines for the first time, as evidenced by the results of our co-citation analysis (Fig. 2). Alongside this triangulation approach, strengths of this review include the pre-registered protocol [http://www.crd.york.ac.uk/PROSPERO/display\\_record.php?ID=CRD42015015941](http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42015015941), and the thorough assessment of RoB through deployment of customized RoB tools.

This literature could be affected by publication bias. Given the lack of sufficient numbers of studies to meta-analyse, we could not investigate publication bias through funnel plots, so it remains speculative whether further unpublished (negative) studies could exist. On the other hand, we notice a trend for a number of studies to attempt replication of certain (positive) seminal papers (e.g. Eilertsen *et al.*<sup>31</sup> replicating D'Onofrio *et al.*,<sup>30</sup> Barreca and Page<sup>39</sup> replicating Fertig and Watson,<sup>36</sup> Boyles *et al.*<sup>25</sup> replicating Chevrier *et al.*<sup>24</sup>) and occasionally failing to replicate the original results. Being able to capture these failures to replicate is a strength of the current review.

The main limitation of the review derives from the nature of the evidence we found, the paucity and heterogeneity of which prevented us from pooling effects through meta-analysis. Instead, we systematically grouped results by outcome and study type and examined them for consistency.

Another important limitation is that we cannot infer causal dose–response relationships based on this body of evidence. For example, the effect estimates from MR studies using maternal genotypes effectively refer to the average difference in alcohol to which the offspring are exposed. Therefore, what we can infer is that (often small) increases in prenatal alcohol exposure are associated with lower neurocognitive outcomes and to a lesser extent lower birthweight, based on studies that minimize confounding. The most pressing question of relevance to public health remains whether the recommendation to abstain from alcohol in pregnancy is backed by solid evidence, as opposed to being purely precautionary. We have already explored this extensively with our previous review of observational studies on the effects of low levels of drinking in pregnancy, which concluded that the abstinence advice was mainly a precaution.<sup>6</sup> The research studies brought together by the current review add to this in a significant way, tipping the balance towards a more solid evidence-base, in particular for neurocognitive and behavioural outcomes.

## Public health/policy and research implications

This review seeks to address an area of great public health impact. Alcohol use in pregnancy is still widespread worldwide,<sup>56</sup> despite claims that it causes the most common neurodevelopmental impairments, included under the umbrella diagnosis of FASD.<sup>56</sup> The claims of causality implicit in the diagnostic definition of FASD, however, have occasionally been disputed (e.g. McLennan *et al.*<sup>57</sup>) due to lack of robust evidence of specific alcohol effects on different domains in the child, and whether thresholds apply. This review highlights the need for more studies using a variety of analytical approaches to establish the extent to which prenatal alcohol exposure causes specific neurobehavioral outcomes in the offspring. Studies simultaneously addressing multiple sources of bias are particularly needed (e.g. MR exploiting trio data from fathers, mothers and offspring, to conduct both negative control MR with paternal effects, and analyses accounting for transmitted and un-transmitted alleles,<sup>52</sup> and 2-samples MR using recently developed approaches to study intrauterine effects),<sup>58</sup> as are studies allowing for more sophisticated dose–response estimation (e.g. Silverwood *et al.*<sup>59</sup>).

This evidence will then feed into a revised and improved definition of FASD. Results from this review will also inform future reviews of guidelines on alcohol use in pregnancy. The current UK guidelines for example, revised down to abstinence in 2016,<sup>60</sup> are heavily based on the precautionary principle. The present review of alternative study designs to improve causal inference will strengthen the evidence base for the abstinence recommendation, as well as highlighting the considerable gaps in evidence and quality of studies needed to move the field forward and draw firm conclusions.

## Conclusion

Our understanding of the specific causal effects of alcohol in pregnancy, especially at low levels of exposure, is limited due to biases affecting traditional observational methods and the practical and ethical obstacles to conducting an RCT. Alternative study designs such as MR and natural experiments make an important contribution to our understanding of these effects, as they overcome some of the limitations of traditional methods. Currently, these comprise a modest body of evidence suggestive of a detrimental effect on cognitive outcomes and infant birthweight, which corroborate findings of conventional epidemiological studies. The studies included in this review do not provide evidence on whether the effect of alcohol exposure is linear, or whether there is a safe threshold for drinking in pregnancy, although many of them compare groups of

offspring with at most small differences in their prenatal alcohol exposure. Although it remains true that the only way to avoid alcohol-related risks to the fetus is to abstain from alcohol during pregnancy, it is also important to communicate both to mothers-to-be and healthcare professionals that there remains uncertainty in the evidence base for this recommendation,<sup>61</sup> although we welcome the fact that more and more studies with complementary strengths and weaknesses are emerging in this field.

## Supplementary data

Supplementary data are available at *IJE* online.

## Funding

This work was jointly supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West (NIHR CLAHRC West), now recommissioned as NIHR Applied Research Collaboration West (ARC West), the NIHR Bristol Biomedical Research Centre (BRC), and the Medical Research Council (MRC) Integrative Epidemiology Unit. L.M., S.J.L., J.S., H.B.E., V.L., T.J., T.H.M.M. and J.L.D. are fully or partially funded by the NIHR CLAHRC West. The UK MRC (MC\_UU\_00011/1 and MC\_UU\_00011/6) and the University of Bristol provide core funding for the MRC Integrative Epidemiology Unit: L.M., D.A.L., G.D.S., A.F. and L.Z. are partially or fully funded through this. A.F., L.Z. and S.vH. are, or were funded, through UK Medical Research Council personal fellowships (Grant refs: MR/M009351/1, G0902144 and G1002345, respectively). D.A.L.'s contribution to this work was supported by the US National Institute of Health (R01 DK10324) and European Research Council (Advanced Grant 669545). D.A.L. (NF-0616-10102) and J.L.D. (NF-SI-0512-10119) are NIHR Senior Investigators. The views expressed here are those of the authors and not necessarily those of the NHS, the MRC, NIHR or the Department of Health and Social Care.

## Acknowledgements

We thank Louise Waters for assistance with article collection and Alison Richards for developing the search strategy for this review.

## Authors' contributions

L.M.: drafting the protocol, drafting the risk of bias assessment tools, preparing the data extraction database, data screening and collection (lead reviewer), data analysis, figures, tables, data interpretation, and drafting the manuscript. T.J.: data extraction, data quality checking, and full text screening. S.L.: risk of bias reviewer, data quality checking, and provided methodological advice for the review and meta-analysis conduct. H.B.E.: screening, data collection, data quality checking. J.S.: protocol development/study design (including the development of searching, article screening and data collection strategies), oversaw and managed the review process and the review team, provided methodological advice for the review and meta-analysis conduct, and critically revised the manuscript. T.H.M.M.: advising on the protocol, preparing the protocol,

screening title and abstract. G.D.S.: obtained funding, formulation of project plan, review of the manuscript. J.L.D.: obtained funding and contributed to drafting the manuscript and approved the final version. V.L.: data extraction. D.A.L.: risk of bias reviewer, contributed to data interpretation and review of earlier manuscript drafts, approved the final version that has been submitted. S.J.L.: contributed to data interpretation and review of earlier manuscript drafts, approved the final version that has been submitted. S.vH.: risk of bias reviewer, revised one of the risk of bias assessment tools, completed risk of bias reviews and review of manuscript draft. A.F.: risk of bias reviewer, contributed to the study design, reviewed manuscript draft. L.Z.: contributed to the study protocol, study design, data interpretation, drafted parts of the manuscript, designed and revised the risk of bias assessment tools, completed risk of bias reviews. L.M. has full access to all the data in the study and had final responsibility for the decision to submit for publication. The manuscript is an honest, accurate and transparent account of the study being reported, no important aspects of the study have been omitted and any discrepancies from the study as planned (and, if relevant, registered) have been explained.

**Conflict of interest:** D.A.L. has received support from several national and international government and charitable funders and Medtronic and Roche Diagnostics in relation to research that is not presented here. Other authors declare no conflicts.

## References

1. Patra J, Bakker R, Irving H, Jaddoe VW, Malini S, Rehm J. Dose-response relationship between alcohol consumption before and during pregnancy and the risks of low birthweight, preterm birth and small for gestational age (SGA)—a systematic review and meta-analyses. *BJOG* 2011;118:1411–21.
2. Simoes HO, Zanchetta S, Furtado EF. What we know of the central auditory disorders in children exposed to alcohol during pregnancy? Systematic review. *Codas* 2016;28:640–45.
3. Bell JC, Raynes-Greenow C, Turner RM, Bower C, Nassar N, O'Leary CM. Maternal alcohol consumption during pregnancy and the risk of orofacial clefts in infants: a systematic review and meta-analysis. *Paediatr Perinat Epidemiol* 2014;28:322–32.
4. O'Keeffe LM, Greene RA, Kearney PM. The effect of moderate gestational alcohol consumption during pregnancy on speech and language outcomes in children: a systematic review. *Syst Rev* 2014;3:1.
5. Henderson J, Gray R, Brocklehurst P. Systematic review of effects of low-moderate prenatal alcohol exposure on pregnancy outcome. *BJOG* 2007;114:243–52.
6. Mamluk L, Edwards HB, Savović J *et al*. Low alcohol consumption and pregnancy and childhood outcomes: time to change guidelines indicating apparently 'safe' levels of alcohol during pregnancy? A systematic review and meta-analyses. *BMJ Open* 2017;7:e015410.
7. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;27:1133–63.
8. Lahey BB, D'Onofrio BM. All in the family: comparing siblings to test causal hypotheses regarding environmental influences on behavior. *Curr Dir Psychol Sci* 2010;19:319–23.

9. Dinardo J. Natural Experiments and Quasi-Natural Experiments. In: Durlauf SN, Blume LE (eds). *Microeconometrics. The New Palgrave Economics Collection*. London: Palgrave Macmillan 2010, pp. 139–53.
10. Richmond RC, Al-Amin A, Davey Smith G., Relton CL. Approaches for drawing causal inferences from epidemiological birth cohorts: a review. *Early Hum Dev* 2014;**90**:769–80.
11. Trinquart L, Johns DM, Galea S. Why do we think we know what we know? A metaknowledge analysis of the salt controversy. *Int J Epidemiol* 2016;**45**:251–60.
12. Barlow P, Reeves A, McKee M, Galea G, Stuckler D. Unhealthy diets, obesity and time discounting: a systematic literature review and network analysis. *Obes Rev* 2016;**17**:810–19.
13. Barlow P, McKee M, Reeves A, Galea G, Stuckler D. Time-discounting and tobacco smoking: a systematic review and network analysis. *Int J Epidemiol* 2017;**46**:860–69.
14. Glymour MM, Tchetgen EJT, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol* 2012;**175**:332–39.
15. Centre for Reviews and Dissemination. *CRD's Guidance for Undertaking Reviews in Health Care*. York: York Publishing Services Ltd, 2009, p. 32.
16. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–60.
17. Deeks JJ, Higgins JP, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT GS (ed). *Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011)*. The Cochrane Collaboration, 2011, pp. 243–96. [www.handbook.cochrane.org](http://www.handbook.cochrane.org).
18. Higgins JPT, Altman GD, Sterne JAC (eds). Chapter 8: Assessing risk of bias in included studies. *Cochrane Handbook for Systematic Reviews of Interventions Version 510 (Updated March 2011)*. The Cochrane Collaboration, 2011. Oxford. [www.handbook.cochrane.org](http://www.handbook.cochrane.org).
19. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;**45**:1866–86.
20. Taylor AE, Davies NM, Ware JJ, VanderWeele T, Davey Smith G., Munafò MR. Mendelian randomization in health research: using appropriate genetic variants and avoiding biased estimates. *Econ Hum Biol* 2014;**13**:99–106.
21. Viljoen DL, Carr LG, Foroud TM, Brooke L, Ramsay M, Li TK. Alcohol dehydrogenase-2\*2 allele is associated with decreased prevalence of fetal alcohol syndrome in the mixed-ancestry population of the Western Cape Province, South Africa. *Alcoholism Clin Exp Res* 2001;**25**:1719–22.
22. Stoler JM, Ryan LM, Holmes LB. Alcohol dehydrogenase 2 genotypes, maternal alcohol use, and infant outcome. *J Pediatr* 2002;**141**:780–85.
23. Arfsten DP, Silbergeld EK, Loffredo CA. Fetal ADH2\*3, maternal alcohol consumption, and fetal growth. *Int J Toxicol* 2004;**23**:47–54.
24. Chevrier C, Perret C, Bahuau M *et al*. Interaction between the ADH1C polymorphism and maternal alcohol intake in the risk of nonsyndromic oral clefts: an evaluation of the contribution of child and maternal genotypes. *Birth Defect Res A* 2005;**73**: 114–22.
25. Boyles AL, Deroo LA, Lie RT *et al*. Maternal alcohol consumption, alcohol metabolism genes, and the risk of oral clefts: a population-based case-control study in Norway, 1996–2001. *Am J Epidemiol* 2010;**172**:924–31.
26. Lewis SJ, Zuccolo L, Davey Smith G. *et al*. Fetal alcohol exposure and IQ at age 8: evidence from a population-based birth-cohort study. *PLoS One* 2012;**7**:e49407.
27. Zuccolo L, Lewis SJ, Smith GD *et al*. Prenatal alcohol exposure and offspring cognition and school performance. A 'Mendelian randomization' natural experiment. *Int J Epidemiol* 2013;**42**:1358–70.
28. von Hinke Kessler Scholder S, Wehby GL, Lewis S, Zuccolo L. Alcohol exposure in utero and child academic achievement. *Econ J* 2014;**124**:634–67.
29. Murray J, Burgess S, Zuccolo L, Hickman M, Gray R, Lewis SJ. Moderate alcohol drinking in pregnancy increases risk for children's persistent conduct problems: causal effects in a Mendelian randomisation study. *J Child Psychol Psychiatr* 2016;**57**:575–84.
30. D'Onofrio BM, Van Hulle CA, Waldman ID, Rodgers JL, Rathouz PJ, Lahey BB. Causal inferences regarding prenatal alcohol exposure and childhood externalizing problems. *Arch Gen Psychiatry* 2007;**64**:1296–304.
31. Eilertsen EM, Gjerde LC, Reichborn-Kjennerud T *et al*. Maternal alcohol use during pregnancy and offspring attention-deficit hyperactivity disorder (ADHD): a prospective sibling control study. *Int J Epidemiol* 2017;**46**:1633–40.
32. Alati R, Macleod J, Hickman M *et al*. Intrauterine exposure to alcohol and tobacco use and childhood IQ: findings from a parental-offspring comparison within the Avon Longitudinal Study of Parents and Children. *Pediatr Res* 2008;**64**:659–66.
33. Alati R, Davey Smith G., Lewis SJ *et al*. Effect of prenatal alcohol exposure on childhood academic outcomes: contrasting maternal and paternal associations in the ALSPAC study. *PLoS One* 2013;**8**:e74844.
34. Zuccolo L, DeRoo LA, Wills AK *et al*. Pre-conception and prenatal alcohol exposure from mothers and fathers drinking and head circumference: results from the Norwegian Mother-Child Study (MoBa). *Sci Rep* 2016;**7**:39535.
35. McCormack C, Hutchinson D, Burns L *et al*. Maternal and partner prenatal alcohol use and infant cognitive development. *Drug Alcohol Depend* 2018;**185**:330–38.
36. Fertig AR, Watson T. Minimum drinking age laws and infant health outcomes. *J Health Econ* 2009;**28**:737–47.
37. Zhang N. Alcohol taxes and birth outcomes. *Int J Environ Res Public Health* 2010;**7**:1901–12.
38. Zhang N, Caine E. Alcohol policy, social context, and infant health: the impact of minimum legal drinking age. *Int J Environ Res Public Health* 2011;**8**:3796–809.
39. Barreca A, Page M. A pint for a pound? Minimum drinking age laws and birth outcomes. *Health Econ* 2015;**24**:400–18.
40. Evans MF, Helland E, Klick J, Patel A. The developmental effect of state alcohol prohibitions at the turn of the twentieth century. *Econ Inq* 2016;**54**:762–77.
41. Cil G. Effects of posted point-of-sale warnings on alcohol consumption during pregnancy and on birth outcomes. *J Health Econ* 2017;**53**:131–55.
42. Nilsson JP. Alcohol availability, prenatal conditions, and long-term economic outcomes. *J Polit Econ* 2017;**125**:1149–207.
43. Tzilos GK, Sokol RJ, Ondersma SJ. A randomized phase I trial of a brief computer-delivered intervention for alcohol use during pregnancy. *J Women's Health* 2011;**20**:1517–24.

44. Yin SJ, Bosron WF, Magnes LJ, Li TK. Human liver alcohol dehydrogenase: purification and kinetic characterization of the beta 2 beta 2, beta 2 beta 1, alpha beta 2, and beta 2 gamma 1 "Oriental" isoenzymes. *Biochemistry* 1984;23:5847–53.
45. Holmes MV, Dale CE, Zuccolo L *et al.*; on behalf of The InterAct Consortium. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ* 2014;349:g4164.
46. Lawlor DA, Benn M, Zuccolo L *et al.* ADH1B and ADH1C genotype, alcohol consumption and biomarkers of liver function: findings from a Mendelian randomization study in 58, 313 European origin Danes. *PLoS One* 2014;9:e114294.
47. Huang J, Zhu T, Qu Y, Mu D. Prenatal, perinatal and neonatal risk factors for intellectual disability: a systemic review and meta-analysis. *PLoS One* 2016;11:e0153655.
48. Mattson SN, Bernes GA, Doyle LR. Fetal alcohol spectrum disorders: a review of the neurobehavioral deficits associated with prenatal alcohol exposure. *Alcohol Clin Exp Res* 2019;43:1046–62.
49. Strandberg-Larsen K, Poulsen G, Bech BH *et al.* Association of light-to-moderate alcohol drinking in pregnancy with preterm birth and birth weight: elucidating bias by pooling data from nine European cohorts. *Eur J Epidemiol* 2017;32:751–64.
50. Richmond RC, Timpson NJ, Felix JF *et al.* Using genetic variation to explore the causal effect of maternal pregnancy adiposity on future offspring adiposity: a Mendelian randomisation study. *PLoS Med* 2017;14:e1002221.
51. 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.
52. Brumpton B, Sanderson E, Hartwig FP *et al.* Within-family studies for Mendelian randomization: avoiding dynastic, assortative mating, and population stratification biases. *bioRxiv* 2019; 602516.
53. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018;362:k601.
54. Liu M, Jiang Y, Wedow R *et al.*; 23andMe Research Team. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019;51:237–44.
55. Munafo MR, Davey Smith G. Robust research needs many lines of evidence. *Nature* 2018;553:399–401.
56. Popova S, Lange S, Probst C, Gmel G, Rehm J. Estimation of national, regional, and global prevalence of alcohol use during pregnancy and fetal alcohol syndrome: a systematic review and meta-analysis. *Lancet Glob Health* 2017;5:e290–99.
57. McLennan JD, Braunberger P. A critique of the new Canadian fetal alcohol spectrum disorder guideline. *J Can Acad Child Adolesc Psychiatry* 2017;26:179–83.
58. Evans DM, Moen G-H, Hwang L-D, Lawlor DA, Warrington NM. Elucidating the role of maternal environmental exposures on offspring health and disease using two-sample Mendelian randomization. *Int J Epidemiol* 2019;1:15.
59. Silverwood RJ, Holmes MV, Dale CE *et al.* Testing for non-linear causal effects using a binary genotype in a Mendelian randomization study: application to alcohol and cardiovascular traits. *Int J Epidemiol* 2014;43:1781–90.
60. Department of Health. *Alcohol Guidelines Review—Report from the Guidelines Development Group to the UK Chief Medical Officers*. Department of Health London, 2016.
61. Freeman A, Galvao A, Zaval L, Spiegelhalter D. Communicating uncertainty about facts, numbers, and science. *R Soc Open Sci* 2019;6:42.