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

Loubaba Mamluk,1,2,3 Hannah B Edwards,2,3 Jelena Savović,2,3 Verity Leach,2,3 Timothy Jones,2,3 Theresa H M Moore,2,3 Sharea Iljaz,2,3 Sarah J Lewis,2 Jenny L Donovan,2,3 Debbie Lawlor,1,2,3 George Davey Smith,1,2 Abigail Fraser,1,2 Luisa Zuccolo1,2

ABSTRACT

Objectives To determine the effects of low-to-moderate levels of maternal alcohol consumption in pregnancy on pregnancy and long-term offspring outcomes.

Search strategy Medline, Embase, Web of Science and PsycINFO from inception to 11 July 2016.

Selection criteria Prospective observational studies, negative control and quasiexperimental studies of pregnant women estimating effects of light drinking in pregnancy (≤32g/week) versus abstaining. Pregnancy outcomes such as birth weight and features of fetal alcohol syndrome were examined.

Data collection and analysis One reviewer extracted data and another checked extracted data. Random effects meta-analyses were performed where applicable, and a narrative summary of findings was carried out otherwise.

Main results 24 cohort and two quasiexperimental studies were included. With the exception of birth size and gestational age, there was insufficient data to meta-analyse or make robust conclusions. Odds of small for gestational age (SGA) and preterm birth were higher for babies whose mothers consumed up to 32 g/week versus none, but estimates for preterm birth were also compatible with no association: summary OR 1.08, 95% CI (1.02 to 1.14), I² 0%, (seven studies, all estimates were adjusted) OR 1.10, 95% CI (0.95 to 1.28), I² 60%, (nine studies, includes one unadjusted estimates), respectively. The earliest time points of exposure were used in the analysis.

Conclusion Evidence of the effects of drinking ≤32 g/week in pregnancy is sparse. As there was some evidence that even light prenatal alcohol consumption is associated with SGA and preterm delivery, guidance could advise abstention as a precautionary principle but should explain the paucity of evidence.

INTRODUCTION

Alcohol is a known teratogen1 and the evidence about the risks of heavy alcohol consumption during pregnancy on intellectual ability, birth defects, behaviour, fine motor skills and mental health (comprising fetal alcohol spectrum disorder (FASD))2 is clear and compelling.3 Internationally, clinical guidelines recommend that pregnant women should abstain from heavy or ‘binge’ drinking.4 However, until recently UK guidelines advised women to avoid drinking alcohol while trying to conceive, and in the first trimester, but at the same time indicated that consumption should be restricted to within ‘1 to 2 UK units, once or twice a week’.5 The UK Chief Medical Officer commissioned a review.
of guidelines on alcohol consumption during pregnancy. Based on a review of reviews, the Guidelines Development Expert Group has recently proposed a change to guidelines such that women should be advised to abstain from alcohol when pregnant and/or trying to conceive, based on the precautionary principle (ie, ‘better safe than sorry’), in the absence of robust evidence.

Our aim was to conduct a comprehensive systematic review and meta-analysis of the literature to determine the effects of low-to-moderate levels of maternal alcohol consumption on pregnancy and longer term offspring outcomes. Here, we report on alcohol consumption of up to two UK units of alcohol up to twice a week (the equivalent of ~32 g/week) compared with no alcohol. In the absence of evidence from randomised controlled trials, we examine observational studies of pregnant women from the general population with prospective assessment of alcohol exposure to reduce recall bias. In particular, we specifically seek out quasi-experimental studies, negative control comparisons and Mendelian randomisation analyses to reduce the impact of confounding and measurement error on the effect estimates.

METHODS

Selection strategy and selection criteria

A full protocol of this systematic review carried out using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (see online supplementary document) is available from the PROSPERO systematic review register (registration number CRD4201501594; http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015015941).

In brief, eligible studies were defined as epidemiological studies of pregnant women or women trying to conceive with prospective assessment of prenatal alcohol exposure (ie, before birth), sampled from general population. The protocol specifically included studies using standard analytical approaches (eg, multivariable regression analysis), as well as studies that used innovative analytical methods to improve causal inference, such as (1) quasi-experimental studies (for example comparing outcomes before and after implementation of new guidelines on alcohol consumption), (2) negative control studies (eg, comparing the association of offspring outcomes with maternal alcohol consumption to the association of the same outcomes with consumption among fathers, under the assumption that confounding is likely to be similar but that if there was a direct causal effect of fetal alcohol consumption on outcomes, maternal associations would be stronger) and Mendelian randomisation studies (using genetic variants associated with alcohol consumption and metabolism). We considered these analytical approaches to be the most appropriate in terms of their ability to minimise bias from confounding and other sources. Our original protocol included studies exploring the effects of prenatal alcohol consumption up to 83 g/week (the commonly used threshold for moderate consumption) versus abstinence. Here, we have focused specifically on low alcohol consumption, that is, up to 32 g/week as this was the cut-off specified by the UK guidelines at the time of writing this review as being an implicitly ‘safe’ threshold. The specific cut-off value has not been reviewed and is the main point of discussion as the guideline change from low consumption (equating to 1 to 2 UK units, once or twice a week or 32 g/week) to abstinence (reference group).

Outcomes included (1) pregnancy outcomes: stillbirth (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 SD 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. (2) Features of FASD: childhood growth restriction; cranium size and head circumference; developmental delays; behaviour problems; cognitive impairment and IQ; facial malformations. We adopted study-specific definitions for all outcomes.

Studies were excluded if: there was no quantitative measure of alcohol consumption that could be converted to grams of alcohol/week; there was insufficient data to estimate the effect size of the association of our predefined low-consumption categories versus abstinence with any outcome, including studies that analysed alcohol as a continuous variable (ie, assuming the same linear or log-linear effect across the entire alcohol distribution); the lowest exposure category (compared with non-drinkers) had an upper bound exceeding 32 g/week or was unspecified; they were cohort studies of pregnant women with alcohol abuse/dependency; they were case–control studies or cohort studies with retrospective alcohol consumption assessment (eg, after birth).

The following databases were searched: Medline, PsycINFO, Embase on Ovid; the Cochrane Library including 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 11 July 2016 (see online supplementary table 1). We limited the search to English-language papers and excluded animal studies, letters, editorials and conference proceedings for which there were no full-text papers. Searches were tailored to each database by investigators. The search focused on published medical literature and did not include grey literature. We additionally performed manual searches of the reference lists of (1) all papers included in recent systematic reviews of the effects of prenatal alcohol exposure on the outcomes of interest and (2) all recent papers citing those reviews.

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

We assessed potential for bias in included studies by assessing how well the study adjusted for several main confounders known to impact on the exposure–outcome associations (socioeconomic positioning as measured by the individual study, smoking during pregnancy, maternal age and ethnicity). We considered the potential for confounding and bias across studies included in the analyses and described it narratively alongside summary results.

Data extraction
Data were extracted into a custom-built Microsoft Access database. We extracted the following information from each study: title, authors, publication year, country/region, study design, population characteristics (sample size, methods of sampling, age distribution and ethnicity), measures of exposure (assessment method including timing and quantification of alcohol consumption, reference group (abstinence), exposure (eg, 1–2 units or 2–4 units) and information on unit equivalence if stated), outcome assessment methods (including whether this was abstracted from medical records, obtained via a research interview and the person reporting the outcome eg, 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 (eg, relative to different timing of exposure, model adjustments, etc.). Information from each included paper was extracted by one reviewer (LM) and subsequently checked for accuracy and completeness by another reviewer (HE). Extraction errors were minimal and were resolved through discussion between extractor and checker.

Alcohol unit conversion
Alcohol consumption in drinks/week was converted into grams/week based on the pure ethanol equivalent of one drink, as stated in each individual article, or otherwise inferred based on the definition of standard drinks in the country where the study took place.

Data analysis
The association of low alcohol use with pregnancy and related outcomes was investigated comparing the highest category within the range of 0–32 g/week to abstention (during pregnancy). In studies providing data across several categories of intake within the 0–32 g/week range, we used the effect estimate for the highest category of intake. If studies reported on exposure to alcohol during different trimesters, we included estimates relative to the earliest exposure. This is because for some outcomes, the first trimester tends to be the most critical timing/window of exposure and because most studies that only reported on one time point reported on exposure in early gestation. Similarly, if results were available from both unadjusted and adjusted regressions, we prioritised fully adjusted results, as a way of minimising the impact of confounding by important factors such as maternal smoking, age, socioeconomic position and ethnicity. In case of multiple results from the same cohort (relative to the same outcome), we analysed those pertaining to the largest population size (ie, conducted on the least ‘selected’ population as result of exclusions to minimise selection bias). Results from all studies that fulfilled our inclusion criteria were summarised, together with information about the study. Where appropriate, we additionally pooled results for each outcome. Authors were not contacted for extra data.

Results from different study designs have been reviewed separately. Individual study estimates were pooled using random effects meta-analysis. Where only two studies were available to meta-analyse, results were pooled unless they were very different from each other (I^2 ≥50%). In this case, a narrative summary of findings was carried out. Where a study only reported unadjusted results, we kept these separate in the forest plots (subgroup analysis) but then also showed overall pooled estimates combining all results.

Planned subgroup analyses by trimester could not be performed due to insufficient number of included studies with this information.

Risk of bias assessment
The Newcastle-Ottawa Scale (NOS) was used to assess risk of bias for included reports. This is an eight-item questionnaire assessing the following: representativeness of the exposed cohort; selection of the non-exposed cohort; exposure assessment methods; absence of outcome (of interest) at the start of the study; comparability of exposed and non-exposed groups (with regard to confounding variables); blind assessment of exposure and outcome and length and adequacy of follow-up. NOS allocates ‘stars’ for adequate methods but does not specifically advise calculating the sum of allocated stars to give an overall score. Scores for quality are not helpful in assessing the effect of risk of bias on a meta-analysis so we report each item separately in line with recommended methods. To be assessed as adequate for comparability of cohorts (risk of confounding) a study had to control for the following four prespecified potential confounding factors related to foetal development: maternal age, socioeconomic status, ethnicity and smoking.

The likelihood of small study bias, such as publication bias, could not be assessed through visual inspection of funnel plots for pooled analyses as no outcome was assessed by 10+ studies. Statistical analyses were carried out using Stata V.13.1 (StataCorp).

RESULTS
A flowchart of the article review process is shown in figure 1. A total of 4680 citation records were identified from searching the four relevant databases. A manual search of recent systematic reviews identified 33 additional articles. After exclusions, 24 prospective studies
analysed using standard approaches and two quasi-experimental studies were included, reporting on 30 outcomes in total.

Risk of bias

Six studies20–25 had low risk of bias for all eight NOS items and were therefore considered at low risk of bias overall. All studies were judged to be at a low risk of bias for the following three NOS items (see online supplementary table 2): selection of the non-exposed cohort (always from the same source population as the exposed cohort); the absence of outcome at the start of the study and adequate length of follow-up for outcome to have occurred. Fourteen studies had adequate ascertainment of exposure as these were all based on structured interviews or validated records. Objective outcome assessments (assessor unaware of the exposure status) were reported in 16 of the studies. For five others either parent self-report was used (high risk), and for the remaining three the method of outcome assessment was not reported (unclear risk). Eleven studies did not report enough detail to decide if cohorts were representative of the population, therefore only 10 could be judged as low risk. Only four studies did not control for the prespecified potential confounding factors, and one did not report enough detail to permit judgement. Thus, in the majority of studies (19) the compared groups were similar. Nineteen studies had adequate follow-up of the cohort (small loss to follow-up). Only three were judged high risk for this item and two studies presented insufficient information to make a clear judgement.

Studies included in the meta-analysis are presented in table 1. Standard analytical approaches pooled estimates for continuous and binary outcomes are presented in figures 2 and 3, respectively. Figure 2 presents results for birth weight (seven studies). Figure 3A presents results for preterm delivery (nine studies). Figure 3B presents results for SGA (seven studies), and results for low birth weight (six studies) are given in figure 3C.

The meta-analysis yielded a summary OR of 1.10 (95% CI 0.95 to 1.28) for preterm delivery, but there
<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Country</th>
<th>Event</th>
<th>Total number</th>
<th>Outcomes</th>
<th>Timing of exposure</th>
<th>Age at outcome assessment (child)</th>
<th>Adjusted (yes/no)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lundsberg (2015)</td>
<td>USA</td>
<td>191</td>
<td>3907</td>
<td>Birth weight (&lt;2500 g) SGA (&lt;10th percentile) Preterm birth (&lt;37 weeks)</td>
<td>First trimester</td>
<td>Birth</td>
<td>Yes</td>
</tr>
<tr>
<td>Nykjaer (2014)</td>
<td>UK</td>
<td>535</td>
<td></td>
<td>Birth weight (g) Customised birth centile SGA (&lt;10th centile) Low birth weight (&lt;2500 g) Preterm birth (&lt;37 weeks)</td>
<td>First trimester</td>
<td>Birth</td>
<td>Yes</td>
</tr>
<tr>
<td>Niclasen (2014)</td>
<td>Denmark</td>
<td>−</td>
<td>10649</td>
<td>10% with highest problem scores, SDQ, parent report Boys internalising behaviour Girls internalising behaviour Boys externalising behaviour Girls externalising behaviour</td>
<td>Entire pregnancy</td>
<td>7 years</td>
<td>Yes</td>
</tr>
<tr>
<td>Miyake</td>
<td>Japan</td>
<td>126</td>
<td>1493</td>
<td>Preterm birth (&lt;37 weeks) SGA (&lt;10th percentile) Adjusted mean birth weight Low birth weight (&lt;2500 g)</td>
<td>Between 5th and 39th weeks of pregnancy</td>
<td>0–11 months</td>
<td>Yes</td>
</tr>
<tr>
<td>McCarthy</td>
<td>New Zealand, Australia, UK and Ireland</td>
<td>–</td>
<td>3166</td>
<td>Birth weight SGA (&lt;10th percentile) Preterm birth (&lt;37 weeks)</td>
<td>First trimester</td>
<td>Birth</td>
<td>Yes</td>
</tr>
<tr>
<td>Salihu</td>
<td>USA</td>
<td>−</td>
<td>463</td>
<td>Preterm birth (&lt;37 weeks) SGA (&lt;10th percentile)</td>
<td>Within the gestational age range 20–44 weeks</td>
<td>Birth</td>
<td>Yes</td>
</tr>
<tr>
<td>Robinson</td>
<td>Australia</td>
<td>202</td>
<td>1335</td>
<td>Clinically significant problems (T score ≥60), CBCL Parent report Internalising behaviour Externalising behaviour</td>
<td>Second trimester</td>
<td>2–14 years inclusive</td>
<td>Yes</td>
</tr>
<tr>
<td>Jaddoe</td>
<td>Holland; UK</td>
<td>–</td>
<td>4132</td>
<td>Low birth weight SGA (weight &lt;−2 SD scores) Preterm birth (&lt;37 weeks) Birth weight</td>
<td>Late pregnancy</td>
<td>Birth</td>
<td>Yes</td>
</tr>
<tr>
<td>Saya</td>
<td>UK</td>
<td>−</td>
<td>10323</td>
<td>10% with highest problem scores, SDQ parent report Hyperactivity—externalising</td>
<td>First trimester</td>
<td>7.75 years</td>
<td>Yes</td>
</tr>
<tr>
<td>Saya</td>
<td>UK</td>
<td>−</td>
<td>10323</td>
<td>10% with highest problem scores, SDQ parent report Conduct problems—externalising</td>
<td>First trimester</td>
<td>7.75 years</td>
<td>Yes</td>
</tr>
<tr>
<td>Albertsen</td>
<td>Denmark</td>
<td>1488</td>
<td>29463</td>
<td>Preterm birth (&lt;37 weeks) Moderate preterm birth (32–37 weeks) Very preterm birth (&lt;32 weeks)</td>
<td>During pregnancy</td>
<td>Birth</td>
<td>Yes</td>
</tr>
<tr>
<td>Lundsberg</td>
<td>USA</td>
<td>2062</td>
<td></td>
<td>SGA (lowest 10th percentile) Low birth weight Preterm delivery (&lt;37 weeks)</td>
<td>First trimester</td>
<td>Birth</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 1: Prospective studies examining the effects of prenatal alcohol exposure on pregnancy outcomes (studies included in the meta-analysis)
was substantial statistical heterogeneity between studies ($I^2=60\%$) due to a large Danish study reporting a protective effect (figure 3A). Additionally, most studies assessing preterm birth had corrected for main confounders known to be associated with preterm birth, with the exception of the study by Peacock et al$^{26}$ that did not correct for any. There was also modest evidence for an increased risk of being SGA (OR 1.08, 95% CI 1.02 to 1.14) for a total of 288,512 participants, although this was almost entirely driven by a single US study contributing 95% of the participants to this meta-analysis (figure 3B). The birth weight meta-analysis yielded a summary effect of $-13.49$ g (95% CI $-30.28$ to $+3.31$) for offspring of light drinkers versus non-drinkers (figure 2). Summary effect for birth weight <2500 g was OR 1.00, 95% CI 0.82 to 1.22 (figure 3C).

Other outcomes were typically reported by a limited number of studies and mostly could not be meta-analysed due to clinical heterogeneity in outcome assessment or incompleteness of published data (see online supplementary figure 1 and table 2). Based on two studies with data on behavioural outcomes, there was little evidence of any effect for internalising symptoms but a suggestion that light drinking in pregnancy protected against high externalising behaviour scores (OR 0.97, (95% CI 0.93 to 1.01, see onlinesupplementary figure 1)). However, an additional study assessing conduct problems and hyperactivity (in the same externalising domain) reported results in the opposite direction, which could not be meta-analysed due to different outcome definitions.

Table 2 presents results of included studies that did not contribute to the meta-analyses for various reasons. There was no strong evidence of association between consuming up to 32 g/week of alcohol and any of the remaining outcomes excluded from meta-analyses, with three exceptions: a very large US study showing increased risk of placental abruption and decreased risk of pre-eclampsia (OR 1.24, 95% CI 1.05; 1.46 and OR 0.82, 95% CI 0.74; 0.90, respectively),$^{27}$ and a single British study reporting better cognitive outcomes in children exposed to light maternal drinking in pregnancy.$^{28}$

We did not include funnel plots as no outcome was assessed by 10+ studies. Of all included results, only two were unadjusted,$^{26,29}$ and most of the others were adjusted for maternal smoking, age and socio-economic position (see online supplementary table 3). Studies that did not adjust for ethnicity were generally conducted in homogeneous populations. Due to the small number of studies for each outcome, we could not further investigate the effect of adjusting for all or some of these confounders. Similarly, there was insufficient data to examine the effect of timing of exposure on outcomes.

**Alternative analytical approaches**

Two negative control publications$^{30}$ based on the same UK cohort met our inclusion criteria (table 3).$^{29,31}$ They investigated the effects of maternal alcohol consumption...
on childhood educational achievement\textsuperscript{41} and IQ\textsuperscript{29} offspring exposed to maternal consumption of <12g/week of alcohol in the first trimester did not have worse outcomes compared with those of mothers who abstaining from alcohol, and a similar pattern was found for paternal alcohol consumption.

One further quasiexperimental study, one natural experiment and five Mendelian randomisation studies

**Figure 2** Pooled mean difference for birth weight comparing low alcohol consumption (up to 32g/week) with no alcohol consumption (seven studies). ‘Adjusted’ refers to adjusted for both smoking and a measure of socioeconomic status.

**Figure 3** (A) ORs for preterm birth comparing low alcohol consumption (up to 32g/week) with no alcohol consumption (nine studies); (B) ORs for small for gestational age comparing low alcohol consumption (up to 32g/week) with no alcohol consumption (seven studies); (C) ORs for low birth weight comparing low alcohol consumption (up to 32g/week) with no alcohol consumption (six studies). Pooled OR includes both adjusted and unadjusted estimates from studies, ‘Adjusted’ refers to adjusted for both smoking and a measure of socioeconomic status.
Table 2  Prospective studies examining the effects of prenatal alcohol exposure on pregnancy outcomes (results not included in the meta-analysis)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study (year)</th>
<th>Country (total no)</th>
<th>Outcome details</th>
<th>Reason not included in meta-analysis</th>
<th>Age at outcome assessment (child)</th>
<th>Results*</th>
<th>Adjusted (yes/no)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pregnancy outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stillbirth</td>
<td>Andersen25</td>
<td>Denmark (89 322)</td>
<td>Stillbirth</td>
<td>Estimates are in different directions</td>
<td>Outcome confirmed via national/hospital registries or from maternal report</td>
<td>OR 0.90 (95% CI 0.73 to 1.12) Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Salihu27</td>
<td>USA (1 226 685)</td>
<td>Stillbirth</td>
<td>Estimates are in different directions</td>
<td>Birth</td>
<td>OR 1.10 (95% CI 0.88 to 1.39) Yes</td>
<td></td>
</tr>
<tr>
<td>Miscarriage</td>
<td>Andersen25</td>
<td>Denmark (89 322)</td>
<td>Miscarriage in first trimester</td>
<td>Only study with this outcome</td>
<td>(Outcome confirmed via national/hospital registries or from maternal report)</td>
<td>OR 1.05 (95% CI 0.94 to 1.18) Yes</td>
<td></td>
</tr>
<tr>
<td>Windham</td>
<td>USA (5324)</td>
<td></td>
<td>Spontaneous abortion occurring ≤20 weeks gestation</td>
<td>Only study with this outcome</td>
<td>(Outcome confirmed via medical records)</td>
<td>OR 1.0 (95% CI 0.7 to 1.5) Yes</td>
<td></td>
</tr>
<tr>
<td><strong>Gestational age and preterm</strong></td>
<td>Passaro58</td>
<td>UK (10 539)</td>
<td>Gestational age</td>
<td>Only study with this outcome</td>
<td>Birth</td>
<td>Mean gestational age 40.1 (SD 1.9) for those exposed, compared with 40.1 (SD 2.0) for those unexposed</td>
<td>Yes</td>
</tr>
<tr>
<td>Birth weight</td>
<td>Ogston8</td>
<td>The Netherlands, Scotland, France, Spain, Denmark, Germany, Portugal (8453)</td>
<td>Per cent of infants with gestational age &lt;37 weeks</td>
<td></td>
<td>Birth</td>
<td>3.7% in those exposed compared with 3.9% in those unexposed</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Ogston8</td>
<td>The Netherlands, Scotland, France, Spain, Denmark, Germany, Portugal (8453)</td>
<td>Mean birth weight (g) in non-smokers</td>
<td>SE/SD not reported</td>
<td>Birth</td>
<td>3414 g in those exposed, compared with 3363 g in those unexposed</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Ogston8</td>
<td>The Netherlands, Scotland, France, Spain, Denmark, Germany, Portugal (8453)</td>
<td>Mean birth weight (g) in smokers</td>
<td>SE/SD not reported</td>
<td>Birth</td>
<td>3225 g in those exposed, compared with 3225 g in those unexposed</td>
<td>Yes</td>
</tr>
<tr>
<td>Placenta related</td>
<td>Salihu27</td>
<td>USA (1 226 685)</td>
<td>Placental abruption</td>
<td>Only study with this outcome</td>
<td>Birth</td>
<td>OR 1.24 (95% CI 1.05 to 1.46) Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Salihu27</td>
<td>USA (1 226 685)</td>
<td>Placenta previa</td>
<td>Only study with this outcome</td>
<td>Birth</td>
<td>OR 1.11 (95% CI 0.87 to 1.43) Yes</td>
<td></td>
</tr>
<tr>
<td>Pre-eclampsia</td>
<td>Salihu27</td>
<td>USA (1 226 685)</td>
<td>Pre-eclampsia</td>
<td>Only two studies with this outcome</td>
<td>Birth</td>
<td>OR 0.82 (95% CI 0.74 to 0.90) Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>McCarthy23</td>
<td>New Zealand, Australia, Ireland, UK</td>
<td>Pre-eclampsia</td>
<td>Only two studies with this outcome</td>
<td>Birth</td>
<td>OR 0.59 (95% CI 0.35 to 0.90) Yes</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>Lundsberg22</td>
<td>USA (4496)</td>
<td>Lowest 10th percentile of birth length</td>
<td>Only study with this outcome</td>
<td>Birth</td>
<td>OR 1.10 (95% CI 0.78 to 1.54) Yes</td>
<td></td>
</tr>
<tr>
<td>Head circumference</td>
<td>Lundsberg22</td>
<td>USA (4496)</td>
<td>Lowest 10th percentile of head circumference</td>
<td>Only study with this outcome</td>
<td>Birth</td>
<td>OR 1.08 (95% CI 0.83 to 1.42) Yes</td>
<td></td>
</tr>
<tr>
<td>Apgar score</td>
<td>Lundsberg22</td>
<td>USA (4496)</td>
<td>Apgar &lt;7 at 5 min</td>
<td>Only study with this outcome</td>
<td>Birth</td>
<td>OR 1.61 (95% CI 0.67 to 3.84) Yes</td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Study (year)</th>
<th>Country (total no)</th>
<th>Outcome details</th>
<th>Reason not included in meta-analysis</th>
<th>Age at outcome assessment (child)</th>
<th>Results*</th>
<th>Adjusted (yes/no)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admission to NICU</td>
<td>Lundsberg</td>
<td>USA (4496)</td>
<td>Admission to NICU</td>
<td>Only study with this outcome</td>
<td>Birth</td>
<td>OR 1.12 (95% CI 0.86 to 1.46)</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>FASD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malformations</td>
<td>Lundsberg</td>
<td>USA (4496)</td>
<td>Major malformations</td>
<td>Only study with this outcome</td>
<td>Birth</td>
<td>OR 0.78 (95% CI 0.40 to 1.50)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Bille</td>
<td>Denmark (1020)</td>
<td>Oral clefts</td>
<td>Only study with this outcome</td>
<td>Birth</td>
<td>OR 1.06 (95% CI 0.74 to 1.50)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Emhart</td>
<td>USA (873)</td>
<td>Craniofacial anomalies</td>
<td>SE/SD not reported</td>
<td>Birth</td>
<td>Mean craniofacial anomalies was 1.92 for those exposed, compared with 1.85 for those unexposed (p=0.26)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Emhart</td>
<td>USA (873)</td>
<td>Total anomalies</td>
<td>SE/SD not reported</td>
<td>Birth</td>
<td>Mean total anomalies was 2.60 for those exposed, compared with 2.53 for those unexposed (p=0.28)</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Motor</strong></td>
<td>Faedo Larsen</td>
<td>Denmark (32 097)</td>
<td>DCD</td>
<td>Only study with this outcome</td>
<td>7 years</td>
<td>OR 0.85 (95% CI 0.70 to 1.03)</td>
<td>Yes</td>
</tr>
<tr>
<td>development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Behaviour</strong></td>
<td>Sayal</td>
<td>UK (967)</td>
<td>10% with highest problem scores, SDQ parent report Conduct problems</td>
<td>Later age used in analysis</td>
<td>3.9 years</td>
<td>OR 1.14 (95% CI 0.98 to 1.32)</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>/development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sayal</td>
<td>UK (1077)</td>
<td>10% with highest problem scores, SDQ parent report Hyperactivity</td>
<td>Later age used in analysis</td>
<td>3.9 years</td>
<td>OR 0.99 (95% CI 0.86 to 1.16)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Sayal</td>
<td>UK (257)</td>
<td>10% with highest problem scores, SDQ teacher report Conduct problems</td>
<td>Parent report used in analysis instead</td>
<td>7.75–9 years</td>
<td>OR 1.41 (95% CI 1.02 to 1.94)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Sayal</td>
<td>UK (525)</td>
<td>10% with highest problem scores, SDQ teacher report Hyperactivity</td>
<td>Parent report used in analysis instead</td>
<td>7.75–9 years</td>
<td>OR 1.20 (95% CI 0.96 to 1.51)</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Parry</td>
<td>The Netherlands, Scotland, France, Spain, Denmark, Germany, Portugal (8453)</td>
<td>Mental Development Index, Bayley Scales of Infant Development</td>
<td>No SE/SD reported for reference group</td>
<td>18 months</td>
<td>Mean increase of 1.80 points (SE 1.1) for those exposed compared with unexposed</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Parry</td>
<td>The Netherlands, Scotland, France, Spain, Denmark, Germany, Portugal (8453)</td>
<td>Psychomotor Development Index, Bayley Scales of Infant Development</td>
<td>No SE/SD reported for reference group</td>
<td>18 months</td>
<td>Mean increase of 0.81 points (SE 0.8) for those exposed compared with unexposed</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Cognition</strong></td>
<td>Sayal</td>
<td>UK (10 558)</td>
<td>KS2 scores</td>
<td>Only study with this outcome</td>
<td>11 years</td>
<td>Mean increase of 0.38 (95% CI −0.02 to 0.78) on KS2 score for those exposed compared with unexposed</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Continued
### Table 2  Continued

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Country</th>
<th>Total number</th>
<th>Study type</th>
<th>Gene</th>
<th>Age at outcome assessment (child)</th>
<th>Outcomes</th>
<th>Results*</th>
<th>Adjusted (yes/no)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alati (2013)</td>
<td>UK</td>
<td>7062</td>
<td>Maternal–paternal comparison</td>
<td>NA</td>
<td>11 years</td>
<td>Academic achievement: KS2 scores (standardised)</td>
<td>Maternal alcohol consumption abstainers: mean 102.0 (SD 9.1); &lt;1 glass/week: mean 102.8 (SD 8.7) Paternal alcohol consumption abstainers: mean 98.9 (SD 11); &lt;1 glass/week: mean 101.1 (SD 9.1)</td>
<td>Different confounding structures for the association of maternal vs paternal alcohol with the outcome</td>
</tr>
<tr>
<td>Alati (2008)</td>
<td>UK</td>
<td>4332</td>
<td>Maternal–paternal comparison</td>
<td>NA</td>
<td>8 years</td>
<td>IQ: WISC</td>
<td>Maternal alcohol consumption abstainers: mean 105.7 (SD 16.2); &lt;1 glass/week: mean 106.4 (SD 16.3) Paternal alcohol consumption abstainers: mean 102.2 (SD 16.8); &lt;1 glass/week: mean 104.0 (SD 16.7)</td>
<td>Different confounding structures for the association of maternal vs paternal alcohol with the outcome</td>
</tr>
</tbody>
</table>

*OR results compare the odds of outcome in those exposed to ≤32 g AA (absolute alcohol) per week.

KS2, Key Stage 2 score; FASD, fetal alcohol spectrum disorder; NICU, neonatal intensive care unit; WISC, Wechsler Intelligence Scale for Children.

### Table 3  Quasi-experimental studies examining the effects of prenatal alcohol exposure on pregnancy outcomes

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Country</th>
<th>Total number</th>
<th>Study type</th>
<th>Gene</th>
<th>Age at outcome assessment (child)</th>
<th>Summary of results and conclusions as presented in the paper</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alati (2013)</td>
<td>UK</td>
<td>7062</td>
<td>Maternal–paternal comparison</td>
<td>NA</td>
<td>11 years</td>
<td>Maternal alcohol consumption abstainers: mean 102.0 (SD 9.1); &lt;1 glass/week: mean 102.8 (SD 8.7) Paternal alcohol consumption abstainers: mean 98.9 (SD 11); &lt;1 glass/week: mean 101.1 (SD 9.1)</td>
<td>Different confounding structures for the association of maternal vs paternal alcohol with the outcome</td>
</tr>
<tr>
<td>Alati (2008)</td>
<td>UK</td>
<td>4332</td>
<td>Maternal–paternal comparison</td>
<td>NA</td>
<td>8 years</td>
<td>Maternal alcohol consumption abstainers: mean 105.7 (SD 16.2); &lt;1 glass/week: mean 106.4 (SD 16.3) Paternal alcohol consumption abstainers: mean 102.2 (SD 16.8); &lt;1 glass/week: mean 104.0 (SD 16.7)</td>
<td>Different confounding structures for the association of maternal vs paternal alcohol with the outcome</td>
</tr>
</tbody>
</table>

KS2, Key Stage 2 score; NA, not applicable; WISC, Wechsler Intelligence Scale.
were excluded from the present review because they did not specifically test the effect of consuming up to 32 g/week in pregnancy versus abstaining. These will be included in a forthcoming review focused on estimating the causal effects of prenatal alcohol exposure based on alternative study designs and analytical approaches to strengthen causal evidence.

**DISCUSSION**

**Main findings**

In this comprehensive systematic review of the literature on the effects of low levels of alcohol drinking in pregnancy, the two main findings are: (1) a surprisingly limited number of prospective studies specifically addressing the question of whether light maternal alcohol consumption (ie, up to 32 g/week (or 4 UK units) has any causal effect (adverse or beneficial) on infant and later offspring outcomes and pregnancy outcomes, and, as a result, (2) a paucity of evidence demonstrating a clear detrimental effect, or safe limit, of light alcohol consumption on outcomes. The upper limit that we chose to examine here is that of the current version of the UK National Institute for Health and Care Excellence guidelines.32 The question we have attempted to address is very important given the mixed advice that women are given with regards to whether they should abstain completely or be allowed light alcohol consumption in pregnancy. The lack of research evidence to address this question is notable.

**Strengths and limitations**

Strengths of this review include the completeness of searches with a focused research question aimed at informing alcohol in pregnancy guidelines. In addition to observational studies’ biases minimised by only including those with prospective assessment of exposure and prioritising results adjusted for main confounders. Another strength of this review is the unique effort to include alternative study designs to further improve causal inference alongside standard analytical approaches. The main limitation of results on the effects of light drinking in pregnancy from standard analytical approaches is bias due to residual confounding. SE position is a complex, multifaceted entity. Several studies have attempted to adjust for SE position by collecting information on, for example, maternal education, family-level SE position around the time of the pregnancy, home address-based deprivation index and so on. Few studies included more than one of these measured.25 26 33 34 Whereas we consider attempting to adjust for at least one of these characteristics to be a minimum requirement to account for some of the confounding introduced by SE position, there remains scope for residual confounding.35 Given the strong relationship between SE position and both the exposure (alcohol use in pregnancy) and outcomes in this review, any degree of residual confounding is of course an issue when interpreting the effect estimates from the observational studies included in this review.

Women who drink low amounts of alcohol may be more likely to be of higher socioeconomic position, compared with abstainers, at least in developed settings in recent years,36 and both of these characteristics are associated with better pregnancy and cognitive outcomes.37 Maternal smoking and ethnicity are also known correlates of maternal alcohol use and risk factors for, for example, low birth weight.38 Most studies included in this review adjusted for at least some of these factors. However, due to the small number of studies included for any given outcome, it was impossible to formally investigate the effect of incomplete adjustment for some (or all) of these confounders. Additionally, for most outcomes, we could not pool eligible studies for various reasons (eg, too few studies, lack of standard errors) and that we limited our review to a number of prespecified outcomes including the most common pregnancy-related outcomes and childhood outcomes related to FASD. This also was the case for identifying effects based on time of exposure, which is also a limitation.

The inclusion of only English-language studies may have led to missing some studies, however there is little evidence that exclusion of non-English-language studies leads to systematic bias in systematic reviews of conventional medicine.39–42

**Interpretation**

This review demonstrates the paucity and poor quality of evidence addressing this important public health question and the difficulty of designing studies that can effectively evaluate the causal impact of low alcohol consumption while minimising bias and confounding. It also shows the value of reporting measures of effect for meaningful categories of the exposure. While many studies reported that associations did not differ from linearity prior to providing a single coefficient for the dose–response association, it is possible that statistical power limited the ability to detect non-linear associations in single studies. Such detail is especially important when there are controversies about the shape of the association of interest (linear, U or J-shaped) and/or the existence of safe thresholds.

Outstanding questions also remain about the effects of maternal alcohol consumption at different stages of conception and pregnancy. Alternative analytical approaches such as sibling comparisons43 and the use of instrumental variable approaches44 as well as triangulating the totality of evidence from multiple study types45 (formally or informally) are needed to strengthen confidence in the direction and size of any potential causal relationships.

The recently proposed change in the guidelines for alcohol use in pregnancy in the UK to complete abstinence would be an application of the precautionary principle. This review confirmed some increased risk of babies being born SGA but little direct evidence of any other detrimental effect for maternal drinking up to 32 g/week. However, there have been few well-conducted
studies examining this specific category of exposure. This issue remains of great public health importance, with alcohol consumption during pregnancy prevalent in the UK, Ireland, New Zealand and Australia with up to 80% of women consuming some alcohol during pregnancy. For some, the evidence of the potential for harm—mostly coming from animal experiments and human studies of effects due to higher levels of exposure will be sufficient to advocate that guidelines should advise women to avoid all alcohol in pregnancy, while others will wish to retain the existing wording of guidelines. Here we found that maternal alcohol consumption of up to 32 g/week was associated with an 10% increased risk of preterm birth (95% CI 0.95 to 1.28). In comparison, light to moderate smoking (<20 cigarettes per day) is associated with a 22% increased risk of preterm birth (95% CI 1.13 to 1.32).

In conclusion, we found limited evidence for a causal role of light drinking in pregnancy, compared with abstaining, on most of the outcomes examined. Despite the distinction between light drinking and abstinence being the point of most tension and confusion for health professionals and pregnant women and contributing to inconsistent guidance and advice now and in the past, our extensive review shows that this specific question is not being researched thoroughly enough, if at all. In addition, there has been no evidence regarding possible benefits of light alcohol consumption versus absence. Further studies, including those using designs that improve causal inference, are required to provide further evidence and a better estimation of the likely effects. Formulating guidance on the basis of the current evidence is challenging. However, describing the paucity of current research and explaining that ‘absence of evidence is not evidence of absence’, appears warranted.

Acknowledgements We thank our colleagues from NIHR CLAHRC West who provided support that greatly assisted the research. We thank Louise Waters for assistance with article collection and Alison Richards for developing the search strategy for this review.

Contributors LM: drafting the protocol, preparing the data extraction database, data screening and collection (lead reviewer), data analysis, figures, tables, data interpretation and drafting the manuscript. HBE: screening, data collection, data quality checking, data interpretation and preparation of final manuscript. THMM: advising on the protocol, preparing the protocol, screening title and abstract. TJ: data extraction and full-text screening. SI: abstract and full-text screening. JS: 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. GDS: obtained funding, formulation of project plan, review of progress and of the manuscript. JLD: obtained funding and contributed to drafting the manuscript and approved the final version. VL: data extraction. DL: contributed to the study design, data interpretation and review of earlier manuscript drafts; approved the final version that has been submitted. AF: contributed to the study design, data analysis and interpretation and drafting of manuscript. LZ: contributed to the study design, data analysis and interpretation and drafting of manuscript. I, LM, the corresponding author of this manuscript, certify that I have full access to all the data in the study and had final responsibility for the decision to submit for publication.

Funding The UK Medical Research Council (MC, UU, 12013/1 and MC, UU, 12013/5) and the University of Bristol provide core funding for the MRC Integrative Epidemiology Unit; LM, DL, GDS, AF and LZ are partially or fully funded through this. AF and LZ are UK Medical Research Council research fellowships (Grant refs: MR/M009351/1 and 60902144, respectively). LM, JS, HBE, VL, TJ, SI, THMM and JLD are fully or partially partly funded by National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care West (CLAHRC West) at University Hospitals Bristol NHS Foundation Trust. DL (NI-FS-I1-101196 and JLD (NI-FS-I1-10119) are NIHR Senior Investigators. SJL is a senior lecturer funded by HEFCE. The views expressed here are those of the author(s) and not necessarily those of the NHS, the MRC, NIHR or the Department of Health.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES


5. (Nice) NICEFATCE. Antenatal care 2008. CG62. NICE.


41. Pham B, Klassen TP, Lawson ML, et al. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. J Clin Epidemiol 2005;58:769−76. e2.


