Caffeine intake during pregnancy and adverse birth outcomes: a systematic review and dose-response meta-analysis.

Darren C Greenwood[1], Natalie J Thatcher[2], Ye J[1], Garrard L[1], Keogh G[1], King LG[1], Cade JE[3].

[1] Division of Epidemiology & Biostatistics, School of Medicine, University of Leeds, Leeds. LS2 9JT. UK.

[2] Food Standards Agency, Aviation House, 125 Kingsway, London. WC2B 6NH.

[3] Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds, Leeds. LS2 9JT. UK.

Corresponding author:

Dr Darren C Greenwood, Division of Epidemiology & Biostatistics, University of Leeds, Leeds. LS2 9JT. UK.

Email: d.c.greenwood@leeds.ac.uk
Tel: 0113 343 1813
Fax: 0113 343 4877

Abstract

Background: Caffeine is commonly consumed during pregnancy, crosses the placenta, with fetal serum concentrations similar to the mother's, but studies of birth outcome show conflicting findings.

Methods: We systematically searched Medline and Embase for relevant publications. We conducted meta-analysis of dose-response curves for associations between caffeine intake and spontaneous abortion, stillbirth, preterm delivery, low birth weight and small for gestational age (SGA) infants.

Results: Meta-analyses included 60 unique publications from 53 cohort and case-control studies. An increment of 100g caffeine was associated with a 14% (95% CI: 10 to 19%) increase in risk of spontaneous abortion, 19% (5 to 35%) stillbirth, 2% (-2 to 6%) preterm delivery, 7% (1 to 12%) low birth weight, and 10% (95%CI: 6 to 14%) SGA. There was substantial heterogeneity in all models, partly explained by adjustment for smoking and previous obstetric history, but not by prospective assessment of caffeine intake. There was evidence of small-study effects such as publication bias.

Conclusion: Greater caffeine intake is associated with an increase in spontaneous abortion, stillbirth, low birth weight, and SGA, but not preterm delivery. There is no identifiable threshold below which the associations are not apparent, but the size of the associations are generally modest within the range of usual intake and are potentially explained by bias in study design or publication. There is therefore insufficient evidence to support further reductions in the maximum recommended intake of caffeine, but maintenance of current recommendations is a wise precaution.

Keywords: Caffeine; pregnancy; miscarriage; stillbirth; preterm birth; birth weight; small for gestational age infant; meta-analysis

Introduction

Caffeine is present in many drinks and foods consumed during pregnancy, most notably in tea, coffee, colas, energy drinks and chocolate. Although amounts vary due to brand and brewing methods, and vary from country to country, on average one 260ml (10oz) mug of coffee contains around 100mg caffeine. It is also a constituent of several common medications available over the counter or on prescription [1]. Given its widespread presence, most pregnant women consume at least small amounts. However, its effects on the developing fetus are still not fully understood [2].

Animal studies suggest possible adverse effects on reproductive outcomes, including fetal growth, but may not be relevant for humans because caffeine metabolism varies greatly across species [2;3]. In humans, studies have reached mixed conclusions, in part because of difficulties measuring caffeine intake, but also because of other clinical influences on fetal growth and birth outcomes [4-6]. Only one randomised controlled trial of caffeine reduction during pregnancy has been conducted to date, concluding that "moderate reductions" in caffeine intake (of around 200mg) do not substantially alter birthweight or length of gestation [7]. Given that the intervention came after the 1st trimester, did not investigate the more important outcome of spontaneous abortion, and only intervened on one source of caffeine (coffee), questions over the safe consumption of caffeine in pregnancy remain.

Because of the possible associations with restricted fetal growth, birth defects, miscarriage and stillbirth in humans, guidance in several countries, including the US and UK, has continued to consider it a wise precaution to limit caffeine intake to less than 200mg immediately before and during pregnancy [2;8-10]. In the UK, guidance has recently changed to reduce the recommended maximum intakes to this level [11]. However, because of the differing conclusions reached by several major studies, the strength of any association, and the possibility of a threshold effect, has not been fully quantified [2;8;12].

To inform guidance, and in the light of recent large prospective studies, we aimed to pool information from the different observational studies that present information on the association between caffeine and the adverse birth outcomes of spontaneous abortion, stillbirth, preterm birth, low birth weight and small for gestational age. We aimed to avoid problems present in many systematic reviews of observational studies that compare extreme

categories by estimating the dose-response slope over all categories of intake [13]. In particular, we aimed to quantify the degree of any association and identify any possible threshold effects through modelling nonlinear dose-response curves.

Methods

Data sources and searches

We conducted a comprehensive systematic literature search covering all case-control and cohort studies providing evidence on dietary caffeine intake and adverse pregnancy outcomes, including spontaneous abortion or miscarriage, stillbirth, preterm birth, low birth weight and small for gestational age (SGA) infants. We searched MEDLINE and EMBASE online databases for all studies published in any language up to 15th May 2014 (detailed search strategy in online table 1). We also hand searched the reference lists of included studies and relevant review articles. The guidelines for conducting meta-analysis of observational studies in epidemiology were used throughout the design, conduct, analysis, and reporting of this review [14]. A detailed protocol for this review was produced, but is not yet available to download. Instead, methods are provided in detail in this paper.

Study selection

We screened titles and abstracts to remove publications when it was immediately apparent they were not relevant, such as editorials and single case-study reports. We obtained full-text versions of potentially relevant articles. The process of identifying relevant articles was conducted independently by five members of the review team (DCG, JY, LG, GK, LGK). The first author made the final decision where there were any differences. Only case-control and cohort studies were eligible for inclusion in the review, including nested case-control and case-cohort studies.

Inclusion criteria were studies based on dietary assessment of maternal caffeine intake during pregnancy (not before), published in any language, with assessment of caffeine or coffee with more than two categories of exposure, with outcomes including at least one of spontaneous abortion (fetal loss under 24 weeks gestation or less), stillbirth (fetal loss after 24 weeks gestation), preterm birth (before 37 weeks), low birth weight (less than 2500g), or small for gestational age infant (defined as a weight below the 10th percentile for a given gestational

age), as an outcome. Studies had to include some estimate of relative risk (RR) with a measure of uncertainty such as 95% CIs.

Where results from the same study were presented in several papers, the results based on the larger sample, with the most complete assessment of caffeine intake (i.e. total caffeine rather than just caffeine from coffee), or that with the most appropriate adjustment for confounding, were used.

Where caffeine-containing beverages were not presented as mg of caffeine per day, a serving of coffee was assumed to contain approximately 100mg caffeine and any caffeine-containing beverage not separated into different sources (e.g. coffee, tea or cola) was assumed to contain 60mg of caffeine on average [15]. These are broadly in line with the assumptions made in the studies included that assessed caffeine intake on the basis of dietary intake.

To be included in linear dose–response meta-analyses, studies needed to present estimates of RR with associated CIs for at least three categories of intake, alongside a quantified measure of caffeine or coffee intake, with sufficient detail regarding the numbers of cases and non-cases. Studies already presenting results as linear dose-response (e.g. relative risk per 100mg/day) were included as well. Where studies presented results in both ways, the latter was taken as being the most accurate. To be included in nonlinear dose-response meta-analyses, more than three categories of intake were required, and results presented only as linear dose-response curves could not be used.

Data extraction and quality assessment

From the publications identified, we extracted the following information: authors, publication year, geographical region of the study, numbers of cases and non-cases, whether dietary assessment was prospective (before delivery) or retrospective, whether assessment of caffeine was based on multiple dietary sources or just coffee, level of dietary exposure (either as mean, median, midpoint or range for each category or unit of increment for continuous estimates), estimated relative risks with confidence intervals, and characteristics controlled for either by modelling, matching or stratification. Data extraction was carried out by YJ and DCG and its accuracy checked by DCG and a sample by JC.

We assessed the methodological quality of studies using the Newcastle-Ottowa scale for either case-control or cohort studies, as appropriate, presented as a risk of bias table [16]. For cohorts, stars were awarded in the "selection" category for participants being representative pregnancies in terms of caffeine intake, sampling of unexposed participants from the same community, detailed caffeine assessment, and demonstration that adverse pregnancy outcomes were not present at the start of the study. Comparability stars were awarded for adjustment for age and for smoking. These should be interpreted as risk of bias in the estimate, and not the study itself, as some estimates were derived from unadjusted descriptive statistics of secondary exposures. Outcome stars were awarded for outcomes based on medical records rather than self-report, for follow-up to end of pregnancy, and for at least 70% follow-up. For case-control studies, stars were awarded for similar criteria, selection stars being awarded for independently validated cases, cases representative of all incident adverse outcomes, appropriately sourced controls, and controls demonstrably free from the adverse outcome. Comparability stars were awarded as for cohorts. For case-control studies, the remaining quality assessment is based on the exposure definition rather than the outcome definition, with exposure stars awarded for caffeine assessment blind to case-control outcome, the same method of caffeine assessment for cases and controls, and the same response rate in both groups (within 10 percentage points of each other). All studies were included, regardless of perceived quality.

Data synthesis and analysis

So that we could legitimately pool results from studies using different exposure categorisation, we derived a linear dose-response trend for each study using Greenland and Longnecker's method [17]. This method allows estimation of study-specific linear dose-response slopes and their associated confidence intervals, based on the results presented for each category of caffeine intake. These dose-response slopes, now all representing estimates of the same quantity, can then be combined into an overall pooled estimate using standard methods for meta-analysis.

So that we could derive the dose-response trend, we used the mean or median caffeine intake for each category if this was presented. We used the category midpoints when caffeine intake was only reported as a range. When the highest or lowest categories of intake were unbounded, we estimated the midpoint by assuming the width of the category was the same as the next adjacent category. If the distribution of cases was not presented in the publication,

we estimated these initial numbers from definitions of the quantiles, assuming equal numbers in each category.

When studies already presented a linear dose-response trend based on a continuous measure of caffeine intake, alongside a measure of precision such as a confidence interval or a standard error, then we used this instead of deriving the trend indirectly. Where results were presented for two distinctly separate subgroups of women[18], we combined the separate subgroup results first using a fixed effects meta-analysis before they were combined with the other studies. In this way the between-study heterogeneity was estimated appropriately. Where results were presented for two distinct sub-outcomes on the same women [19], these were first combined using an approach suggested by Hamling et al. [20]. This combines the fitted cell counts for the two sub-outcomes, maintaining the same exposure category totals, combining the two results into one with no double counting. We then pooled all the estimated dose-response trends for each study using a random effects model that takes into account the anticipated between-study heterogeneity [21].

We also explored potential nonlinear associations between caffeine and birth outcome. For each study that presented results for more than three categories of intake (the minimum required to model a nonlinear curve) we calculated a restricted cubic spline. This was based on three knots fixed at 10%, 50% and 90% through the total distribution of intake [22]. We then combined these using multivariate meta-analysis [23].

We quantified the percentage of total variation in study estimates attributable to between-study heterogeneity (I^2) and tested between-study heterogeneity using Cochran's Q [24]. We assessed the methodological quality of studies using the Newcastle-Ottowa scale, presented as a risk of bias table. We also performed a limited number of pre-defined subgroup analyses to explore aspects of study quality that may have contributed to the differences in the results seen across the studies, such as retrospective or prospective assessment of intake, source of caffeine, geographic location, and adjustment for pre-specified potential confounders. Though caffeine metabolism may be moderated by CYP1A2 activity, smoking status, or nausea [25], numbers of studies reporting stratified analyses were too few to enable investigation of effect modification by these factors. We conducted sensitivity analysis excluding potentially highly influential studies and studies that might be considered materially different. Any such analyses are included in the relevant section of the results. We

investigated potential small study effects, such as publication bias, using contour-enhanced funnel plots. However, with small numbers of included studies, exploration of sources of heterogeneity and of small study effects lacked power. All analyses were conducted using Stata version 13.1 [26].

Results

Literature search

In total, 91 relevant publications were identified, of which 14 contained replicated information, and 17 did not contain sufficient information for inclusion in the meta-analyses (figure 1). This left 60 unique publications from 53 studies (26 cohort and 27 case-control) with sufficient information for inclusion in dose-response meta-analyses. Of these 53 studies, 29 were from the US, 21 from Europe, 2 from South America, and 1 from Cuba. Characteristics of the included studies are shown in online tables 2a to 2e alongside an assessment of risk of bias for each study.

One publication only presented miscarriage combined with stillbirth [27]. These two outcomes could have very different risk factors, but because of the low incidence of stillbirth relative to miscarriage, this publication was included in the meta-analysis of miscarriage, with sensitivity analysis to confirm this did not materially influence the pooled estimate. One study of low birth weight also included some small for gestational age infants [28]. This was included in the low birth weight analysis as these formed the majority of events, but with sensitivity analysis to confirm that this did not substantially change the results. One study provided additional information on miscarriage and stillbirth separately that was incorporated into the meta-analyses [29]. One publication provided results for the association between preterm delivery and caffeine separately for small for gestational age infants and normal for gestational age infants [18]. These two subgroups were first combined using a fixed effects meta-analysis before they were pooled with the other studies. One study presented the association between caffeine and preterm delivery with premature rupture of the membranes separately from those without premature rupture of the membranes [19]. These were first combined using an approach suggested by Hamling et al. [20] before pooling with the other studies. One publication was based on women with type 1 diabetes [30]. Because of the non-general population, a sensitivity analysis was conducted to confirm that this did not

materially influence the pooled estimate. Where a publication presented results for different definitions of small for gestational age, the method most consistent with the other studies was used [31]. However, one publication with a combined outcome of ultrasound-based diagosis of intrauterine growth restriction and low birth weight, could not be included in either category [32].

Miscarriage

Our literature search identified 35 unique publications from 32 studies investigating the association between caffeine and miscarriage. Data were extracted from 26 of these studies (14 cohort, 12 case-control), 13 from Europe and 13 from the US. The pooled estimate of relative risk from linear dose–response meta-analysis was 1.14 (95% CI: 1.10 to 1.19) per 100 mg/day of caffeine (P < .001) (figure 2A). Five of the six studies that could not be included in meta-analyses supported this direction of association, with all but one statistically significant [33-37], and the remaining study finding no evidence of any association [38].

There was substantial heterogeneity between the studies ($I^2 = 89\%$; 95% CI: 85 to 92%; Q = 230; df = 25; P < .001). The estimated between-study variance was unchanged by excluding the study that had combined miscarriage and stillbirth into a single outcome [27], and this resulted in a negligible change to the pooled estimate (RR = 1.14; 95% CI: 1.09 to 1.18; P < .001). Excluding the study with greatest influence [39] decreased the estimated between-study variance by more than half, but resulted in only a marginally reduced estimate (RR = 1.11; 95% CI: 1.07 to 1.14; P < .001), so was subsequently excluded from the subgroup analyses.

There was no evidence that any study characteristics investigated, such as retrospective caffeine assessment or adjustment for specific potential confounders, were associated with higher or lower estimates in subgroup analyses (online table 3a). However, there was evidence of considerable asymmetry in the funnel plot (online figure 1a), to the extent that small-study effects such as publication bias cannot be ruled out for this outcome.

The nonlinear dose–response meta-analysis showed a small but consistently increasing incidence of miscarriage associated with increased daily caffeine intake (figure 2B). There was little evidence of any nonlinear association, such as a threshold effect, in the plot.

Stillbirth

Our literature search identified 8 unique publications from 5 studies investigating the association between caffeine and stillbirth. Data were extracted from all 5 of these studies (3 cohort, 2 case-control), 3 from Europe, 1 from the US and 1 from Uruguay. The pooled estimate of relative risk from linear dose–response meta-analysis was 1.19 (95% CI: 1.05 to 1.35) per 100 mg/day of caffeine (P = .007) (figure 2C). All the identified studies were included in meta-analyses.

There was substantial heterogeneity between the studies (I^2 = 82%; 95% CI: 59 to 92%; Q = 22; df = 4; P < .001). Studies that only considered coffee had substantially lower estimates than those that considered caffeine from multiple sources (P = .02). Stratifying on this study characteristic reduced the percentage of total variation in estimates attributable to between-study heterogeneity to less than 50% in both groups (online table 3b). Again there was evidence of asymmetry in the funnel plot (online figure 1b), but the number of studies was small.

The nonlinear dose–response meta-analysis showed a small but consistently increasing incidence of stillbirth associated with increased daily caffeine intake (figure 2D). There was little evidence of any nonlinear association.

Preterm delivery

The literature search identified 21 unique publications from 20 studies investigating the association between caffeine and preterm delivery. Data were extracted from 15 of these studies (8 cohort, 7 case-control), 6 from Europe and 9 from the US. The pooled estimate of relative risk from linear dose–response meta-analysis was 1.02 (95% CI: 0.98 to 1.06) per 100 mg/day of caffeine (P = .42) (figure 2E). Of the five studies that could not be included in meta-analyses, four suggested some positive association [40-43], and one did not [44].

There was substantial heterogeneity between the studies (I^2 = 63%; 95% CI: 34 to 79%; Q = 38; df = 14; P = .001). Studies with prospective dietary assessment of caffeine intake tended to have more positive associations with preterm delivery than those with retrospective assessment (P = .04) (online table 3c). There was no evidence of any asymmetry in the funnel plot (online figure 1c).

The nonlinear dose–response meta-analysis showed a generally straight flat line with no evidence of any nonlinear association (figure 2F).

Low birth weight

We identified 14 unique publications from 14 studies investigating the association between caffeine and low birth weight. Data were extracted from 11 of these studies (6 cohort, 5 case-control), 3 from Europe, 7 from the US and 1 from Brazil. The pooled estimate of relative risk from linear dose–response meta-analysis was 1.07 (95% CI: 1.01 to 1.12) per 100 mg/day of caffeine (P = .01) (figure 3A). Of the three studies not included in meta-analysis, one found a strong association between caffeine and low birth weight [45], whilst the other two found no evidence of any association [40;46].

There was substantial heterogeneity between the studies ($I^2 = 75\%$; 95% CI: 55 to 86%; Q = 40; df = 10; P < .001). Excluding the study with additional small for gestational age infants [28], had a negligible impact on the pooled estimate or the estimated heterogeneity (RR = 1.06; 95% CI: 1.01 to 1.12; P = .02). Studies that adjusted for maternal education or socio-economic factors had substantially lower estimates than those that did not (P = .02) (online table 3d). There was some evidence of asymmetry in the funnel plot (online figure 1d), leaving open the possibility of small-study effects such as publication bias.

The nonlinear dose–response meta-analysis showed a small but consistently increasing incidence of low birth weight associated with increased daily caffeine intake (figure 3B). There was little evidence of any nonlinear association.

Small for gestational age

A total of 18 unique publications were identified investigating the association between caffeine and small for gestational age infants, based on 18 studies. Data were extracted from 15 studies (10 cohort, 5 case-control) that could be included in meta-analysis, 6 from Europe and 9 from the US. The pooled estimate of relative risk from linear dose–response meta-analysis was 1.10 (95% CI: 1.06 to 1.14) per 100 mg/day of caffeine (P < .001) (figure 3C). The three studies that could not be included in meta-analysis provided mixed results, with one suggesting an association [47], one not [40], and a third suggesting no overall association, but some evidence of effect modification by smoking [48].

There was substantial heterogeneity between the studies ($I^2 = 64\%$; 95% CI: 38 to 80%; Q = 39; df = 14; P < .001). Studies that adjusted for smoking (P = .04) and studies that adjusted for previous adverse pregnancy outcomes (P = .05) tended to have lower estimates than those that did not adjust for these potential confounders (online table 3e). There was some evidence of asymmetry in the funnel plot (online figure 1e), leaving open the possibility of small-study effects such as publication bias.

The nonlinear dose–response meta-analysis showed a small but consistently increasing incidence of small for gestational age infants associated with increased daily caffeine intake (figure 3D). There was little evidence of any nonlinear association.

Discussion

We have, for the first time, quantified with precision the association between caffeine and adverse birth outcomes, based on 60 publications from 53 separate cohort and case-control studies of adverse pregnancy outcomes. Meta-analysis of the associations between caffeine intake contain nearly 15000 cases of miscarriage from 180000 women, 700 still births from 120000 women, 8000 preterm deliveries from nearly 110000 women, 5000 low birth weight infants from nearly 78000 women, and nearly 12000 small for gestational age infants from 160000 women. The evidence covers a variety of countries with different levels of intake, including non-consumers and categories consuming over 1000 mg/day. This pooled evidence allows the associations between caffeine intake during pregnancy and these adverse outcomes to be described in greater detail than previously possible, and in a manner that allows the shape of the dose-response curve to be described.

A small but quantifiable association was observed between caffeine intake during pregnancy and incidence of miscarriage, stillbirth and low birth weight. There is also a similar sized association between caffeine intake during pregnancy and small for gestational age. There was no evidence of an association between caffeine intake and preterm delivery. For all outcomes the dose-response curves are fairly linear, with no evidence of any "threshold effect" or "plateau" in the dose-response curves. Heterogeneity is generally high, with little between-study heterogeneity being explained by aspects of study design or analysis investigated.

12

The size of the associations are relatively modest within the range of intakes consumed by the majority of women in the included studies, and within the range of intake currently recommended in most countries during pregnancy. In addition, the size of the associations are small relative to some established risk factors such as maternal smoking, but similar to others such secondhand smoke [49]. It is therefore important to interpret any public health implications regarding caffeine intake in the context of known lifestyle risk factors.

It is also important to interpret these results alongside the clinical implications of the outcomes. Whilst the consequences of small for gestational age infants are less severe than miscarriage or stillbirth, small for gestational age has been associated with an increased risk of perinatal mortality and morbidity, including perinatal asphyxia. There is also a body of literature suggesting that it is associated with adverse effects in adult life [50;51], such as increased incidence of obesity, hypertension, hypercholesterolemia, cardiovascular disease, and type 2 diabetes [52-54], If shown to be causal, a small association could therefore still be of importance from a public health perspective.

Given the observational nature of the evidence, we cannot draw inferences on the causal nature of the association identified in this review. Meta-analyses of observational studies are prone to the same biases as the studies they pool evidence across. Therefore the evidence from case-control studies are particularly susceptible to selection bias and recall bias, and all the studies are susceptible to uncontrolled confounding. In addition, all these observational studies are liable to bias from measurement error in using self-report measures to estimate the dietary intake of caffeine. One particular issue common to the majority of studies was the lack of an objective measure of exposure to tobacco smoke. Smoking is a potentially very strong confounder: smokers both consume more caffeine than non-smokers (because smokers' altered CYP1A2 activity leads to faster caffeine clearance) and have much higher rates of adverse birth outcomes [55]. It is therefore important to measure smoking objectively, using a repeated biomarker such as cotinine, to avoid measurement error bias, which in this case could lead to exaggerated associations from only partially controlling for its confounding effects [56].

The large heterogeneity observed in the meta-analyses also requires caution to be exercised in the interpretation of the results. Whilst earlier meta-analyses have also observed substantial

heterogeneity [4;5;57], these may be explained in part by their pooling studies using different categories of intake. We tried to avoid this by placing each study on the same scale, pooling dose-response trends instead [13]. Beyond this, we investigated other potential sources of heterogeneity through a small number of a subgroup analyses specified in advance. We used this to explore whether different study characteristics were associated with the observed differences in the results. These included study design, the method of caffeine intake assessment and adjustment for pre-specified potential confounders such as smoking. Whilst each of these were associated with some of the heterogeneity, it was not consistent across the outcome groups. Though heterogeneity was generally high, it mostly reflected variation in the size of the association, rather than whether there was an association. Heterogeneity associated with small-study effects such as publication bias was also observed for the meta-analyses of miscarriage, stillbirth, low birth weight and small for gestational age.

It is possible that reduction in caffeine may be a marker for a healthier pregnancy and that caffeine is not the cause of the adverse outcomes [58;59]. None of the studies reviewed in this paper have adequately addressed this issue; simply adjusting for nausea does not correct for this potential bias and subgroup analysis suggested it made little difference to the estimates. This potential bias therefore remains the most prominent argument against a causal role of caffeine. Neither are pragmatic trials immune to this potential bias, where greater compliance with the intervention may be associated with healthier pregnancy.

Only one large double-blind randomised controlled trial of caffeine reduction during pregnancy and subsequent birthweight has been conducted to date [7]. Over one thousand Danish women were recruited, each consuming over three cups of coffee a day, and randomised to either caffeinated or decaffeinated coffee. However, the trial did not assess the important outcomes of miscarriage or stillbirth, ignored caffeine intake during the first trimester when caffeine consumption changes markedly and the majority of fetal deaths occur [7;60], and did not measure compliance through objective biomarkers of caffeine intake. In addition, the intervention focussed on coffee intake rather than caffeine as a whole, whilst there is evidence from other countries that cola drinks, tea and chocolate may all contribute at least as much caffeine to the diet during pregnancy [6;60]. These features of the trial limit the extent to which its results can contribute towards discussion of caffeine intake as a whole, or the association with miscarriage and stillbirth. In the absence of any other substantive trial data, our meta-analysis of observational studies provides a valuable resource.

14

If the observed association is causal, it is possible that it may be due to caffeine itself, to one of its metabolites, or a combination of them. Of the four primary routes of caffeine metabolism in humans, 3-demethylation is quantitatively the most important, the caffeine being converted to paraxanthine by CYP1A2. Studies have shown there to be varying levels of CYP1A2 activity in humans and there is considerable inter-individual variation in caffeine metabolism [61]. Measures of caffeine consumption therefore do not necessarily indicate the levels of caffeine and caffeine metabolites in the maternal or fetal circulation. A small number of studies have measured levels of caffeine and its metabolites in maternal or umbilical cord blood rather than assessing caffeine consumption [55;62;63], though given the range of possible exposures, this was not the focus of this review. Linking phenotype with genotype is, however, an area for possible future research.

Given the observational nature of the studies, the heterogeneity and small-study effects, it is not possible to conclude that these associations are causal. The modest sizes of the associations are such that it is possible they could be explained by any or all of these potential biases. However, the plausible biological mechanisms, the evidence from animal studies, the mounting evidence from different observational human studies, and the dose-response slopes, provide some evidence to support the current recommendations limiting caffeine intake during pregnancy, such as restricting to less than 200 mg/day, as a precaution in case the associations really are causal. Whilst the associations are modest in size, they are potentially important at a public health level, and for infants already at elevated risk of adverse outcomes.

In summary, combining results from a large number of studies has allowed associations between caffeine intake and adverse pregnancy outcomes to be quantified with precision and discern a modest but significant association with caffeine intake that could only be adequately quantified by pooling results. A number of questions still remain to be answered. These include confirming causality, such as identifying whether caffeine is the causal agent, one of its metabolites, or whether the associations are completely explained by publication bias or caffeine being a marker of healthy pregnancy. Whilst these issues are unresolved, our results confirm the precautionary guidance adopted by countries recommending limiting caffeine consumption during pregnancy.

Acknowledgements

Competing interests

The authors have no competing interests.

Figure 1. Article retrieval and screening process flow chart.

Total number of references retrieved from electronic searches:
    Medline (n=180)
    Embase (n=150)
	Total number of references retrieved from hand searches (n=13)

Total number of references retrieved (n=343)

→ Duplicate references removed (n=101)

Total number of unique references identified (n=242)

→ Title/abstracts excluded as not relevant (n=129)

Unique references identified as potentially relevant and full text obtained (n=113)

→ Full text references excluded as not relevant (n=22)

Relevant full text publications (n=91)

→ Relevant references excluded from review:
    Data replicated in another publication (n=14)

→ Relevant references excluded from meta-analysis:
    Insufficient information for meta-analysis (binary exposure) (n=14)
    Insufficient information for meta-analysis (p-value only) (n=2)
    Insufficient information for meta-analysis (transformed data) (n=1)

References included in meta-analyses (n=60)

Figure 2. Dietary caffeine intake and estimated relative risk of miscarriage, stillbirth, and preterm birth.

A,C,E: Forest plots of linear dose–response trends with pooled estimates from random-effects meta-analysis per 100mg/day caffeine intake for miscarriage (A), stillbirth (C), and preterm birth (E). B,D,F: Summary nonlinear dose–response curves from multivariate random effects meta-analysis of restricted cubic spline curves for miscarriage (B), stillbirth (D), and preterm birth (F), using zero intake as the reference intake. Tickmarks on the horizontal axis indicate the location of category medians, means, or midpoints for included studies.

a

Study		Estimated RR (95% CI)
Srisuphan et al, 1986		1.32 (1.01, 1.72)
Hansteen et al, 1990		1.04 (1.01, 1.07)
Wilcox et al, 1990		1.18 (0.74, 1.89)
Fenster et al, 1991		1.06 (0.97, 1.17)
Kline et al, 1991		1.07 (1.03, 1.11)
Armstrong et al, 1992		1.02 (1.00, 1.03)
Infante-Rivard et al, 1993		1.32 (1.15, 1.52)
Mills et al, 1993		1.15 (0.88, 1.51)
Dominguez-Rojas, 1994		2.04 (1.80, 2.30)
Dlugosz et al, 1996		1.11 (0.94, 1.31)
Agnesi et al, 1997		1.36 (1.06, 1.74)
Fenster et al, 1997		1.07 (0.97, 1.19)
Parazzini et al, 1998		1.37 (1.27, 1.49)
Cnattingius et al, 2000		1.07 (1.00, 1.15)
Wen et al, 2001		1.26 (1.02, 1.55)
Giannelli et al, 2003		1.18 (1.04, 1.34)
Rasch et al, 2003		1.25 (1.13, 1.38)
Tolstrup et al, 2003		1.04 (1.00, 1.07)
Wisborg et al, 2003		1.11 (1.04, 1.18)
Khoury et al, 2004		1.86 (0.98, 3.54)
Bech et al, 2005		1.04 (1.01, 1.08)
George et al, 2006		1.16 (0.93, 1.44)
Maconochie et al, 2007		1.02 (0.98, 1.07)
Savitz et al, 2008		0.92 (0.79, 1.07)
Weng et al, 2008		1.29 (1.10, 1.51)
Greenwood et al, 2010		1.08 (0.86, 1.37)
Pooled estimate		1.14 (1.10, 1.19)

RR per 100 mg/day of caffeine

c

Study		Estimated RR (95% CI)
Linn et al, 1982		1.08 (0.88, 1.32)
Wisborg et al, 2003		1.10 (1.02, 1.18)
Bech et al, 2005		1.03 (0.94, 1.14)
Matijasevich et al, 2006		1.32 (1.17, 1.49)
Greenwood et al, 2011		1.53 (1.29, 1.83)
Pooled estimate		1.19 (1.05, 1.35)

RR per 100 mg/day of caffeine

e

Study		Estimated RR (95% CI)
Berkowitz, 1982		0.98 (0.83, 1.14)
Olsen, 1991		1.02 (0.98, 1.07)
McDonald, 1992		1.02 (0.99, 1.04)
Williams, 1992		1.13 (1.06, 1.19)
Fortier et al, 1993		0.98 (0.86, 1.11)
Pastore, 1995		0.92 (0.82, 1.04)
Peacock, 1995		0.96 (0.85, 1.09)
Bicalho et al, 2002		0.79 (0.67, 0.93)
Bracken et al, 2003		1.19 (0.99, 1.44)
Khoury et al, 2004		1.53 (0.97, 2.40)
de Souza et al, 2005		1.75 (0.74, 4.14)
Chiaffarino et al, 2006		0.95 (0.78, 1.15)
Boylan et al, 2008		1.12 (1.01, 1.24)
Bakker et al, 2010		1.01 (0.93, 1.11)
Sengpiel et al, 2013		0.98 (0.92, 1.05)
Pooled estimate		1.02 (0.98, 1.06)

RR per 100 mg/day of caffeine

b

Best fitting cubic spline
95% confidence interval

Estimated RR

Caffeine intake (mg/day)

d

Best fitting cubic spline
95% confidence interval

Estimated RR

Caffeine intake (mg/day)

f

Best fitting cubic spline
95% confidence interval

Caffeine intake (mg/day)

19

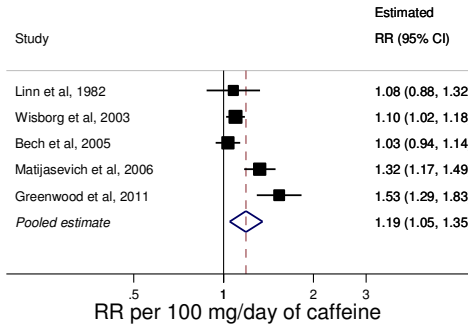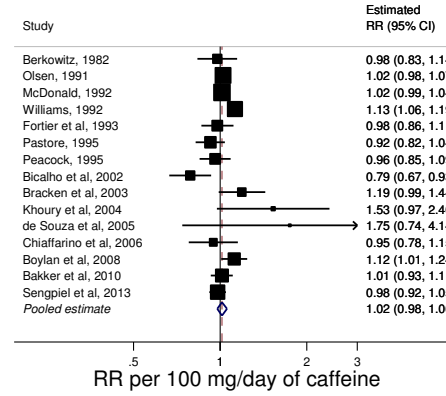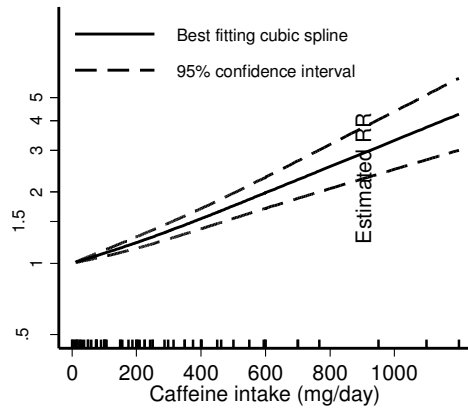Figure 3. Dietary caffeine intake and estimated relative risk of low birth weight and small for gestational age.

A,C: Forest plots of linear dose–response trends with pooled estimates from random-effects meta-analysis per 100mg/day caffeine intake for low birth weight (A) and small for gestational age (C). B,D: Summary nonlinear dose–response curves from multivariate random effects meta-analysis of restricted cubic spline curves for low birth weight (B) and small for gestational age (D), using zero intake as the reference intake. Tickmarks on the horizontal axis indicate the location of category medians, means, or midpoints for included studies.
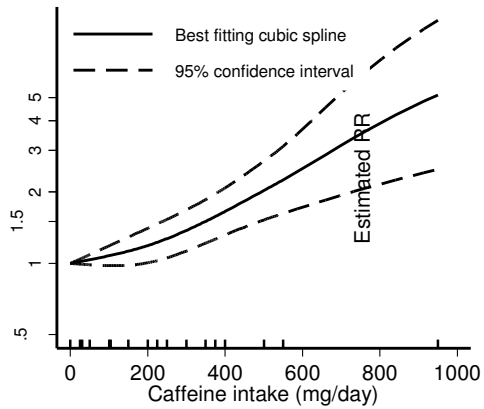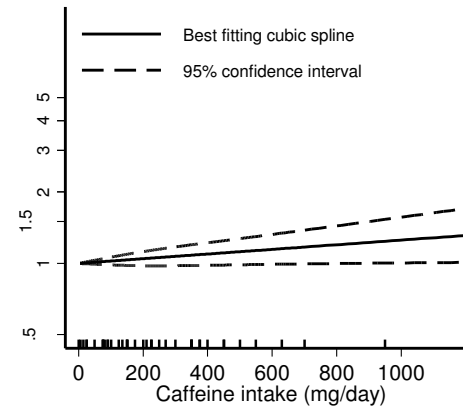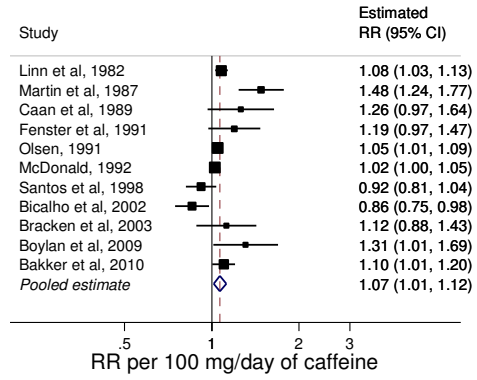
a

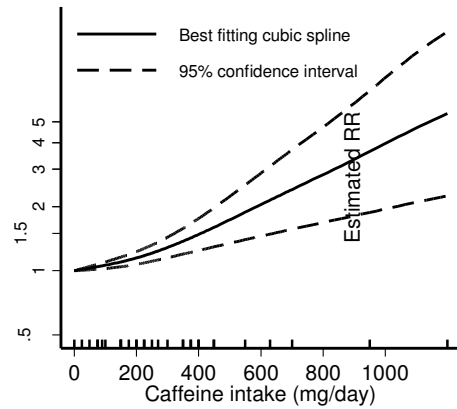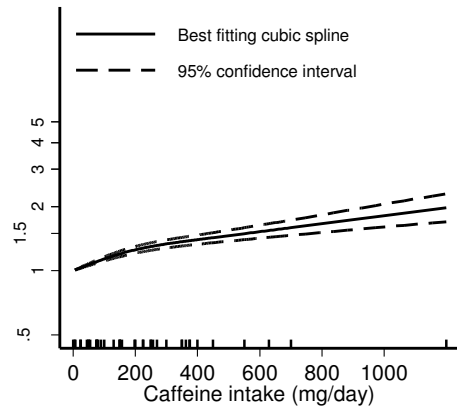| Study		Estimated RR (95% CI)
Linn et al, 1982		1.08 (1.03, 1.13)
Martin et al, 1987		1.48 (1.24, 1.77)
Caan et al, 1989		1.26 (0.97, 1.64)
Fenster et al, 1991		1.19 (0.97, 1.47)
Olsen, 1991		1.05 (1.01, 1.09)
McDonald, 1992		1.02 (1.00, 1.05)
Santos et al, 1998		0.92 (0.81, 1.04)
Bicalho et al, 2002		0.86 (0.75, 0.98)
Bracken et al, 2003		1.12 (0.88, 1.43)
Boylan et al, 2009		1.31 (1.01, 1.69)
Bakker et al, 2010		1.10 (1.01, 1.20)
Pooled estimate		1.07 (1.01, 1.12)

RR per 100 mg/day of caffeine

c

Study		Estimated RR (95% CI)
Fenster et al, 1991		1.27 (1.03, 1.55)
McDonald, 1992		1.04 (1.02, 1.06)
Fortier et al, 1993		1.13 (1.03, 1.23)
Mills et al, 1993		1.73 (1.25, 2.40)
Rondo et al, 1996		1.19 (1.02, 1.38)
Grosso et al, 2001		0.98 (0.83, 1.15)
Bicalho et al, 2002		0.87 (0.69, 1.09)
Bracken et al, 2003		1.12 (0.93, 1.34)
Vik et al, 2003		1.31 (1.09, 1.56)
Parazzini et al, 2005		1.02 (0.94, 1.10)
Xue et al, 2007		1.09 (1.05, 1.13)
Boylan et al, 2009		1.14 (1.02, 1.27)
Bakker et al, 2010		1.14 (1.05, 1.25)
Sengpiel et al, 2013		1.07 (1.04, 1.11)
van den Berg et al, 2013		1.15 (1.03, 1.29)
Pooled estimate		1.10 (1.06, 1.14)

RR per 100 mg/day of caffeine

b

Best fitting cubic spline
95% confidence interval

Estimated RR

Caffeine intake (mg/day)

d

Best fitting cubic spline
95% confidence interval

Caffeine intake (mg/day)

21

References

1. Ministry of Agriculture Fisheries and Food. Survey of caffeine and other methylxanthines in energy drinks and other caffeine-containing products (updated). London: Ministry of Agriculture, Fisheries and Food; 1998. Report No.: 144.

2. Committee on Toxicity of Chemicals in Food Consumer Products and the Environment. Statement on the reproductive effects of caffeine. Food Standards Agency; 2001.

3. Wilcox A, Weinberg C, Baird D. Caffeinated beverages and decreased fertility. The Lancet 1988;332:1453-6.

4. Santos IS, Victora CG, Huttly S, Morris S. Caffeine intake and pregnancy outcomes: a meta-analytic review. Cad Saude Publica 1998;14:523-30.

5. Fernandes O, Sabharwal M, Smiley T, Pastuszak A, Koren G, Einarson T. Moderate to heavy caffeine consumption during pregnancy and relationship to spontaneous abortion and abnormal fetal growth: a meta-analysis. Reproductive Toxicology 1998;12:435-44.

6. Boylan SM, Cade JE, Kirk SFL, Greenwood DC, White KLM, Shires S, et al. Assessing caffeine exposure in pregnant women. Br J Nutr 2008;100:875-82.

7. Bech BH, Obel C, Henriksen TB, Olsen J. Effect of reducing caffeine intake on birth weight and length of gestation: randomised controlled trial. BMJ 2007;334:409.

8. Committee on Toxicity of Chemicals in Food Consumer Products and the Environment. Statement on the review of caffeine. Department of Health; 1984.

9. Committee on Toxicity of Chemicals in Food Consumer Products and the Environment. Statement on the reproductive effects of caffeine. Food Standards Agency; 2008.

10. Organisation of teratology information specialists. Caffeine and pregnancy. 2006. http://www.otispregnancy.org/pdf/caffeine. Accessed 13 June 2014.

11. NHS choices. Caffeine and pregnancy. http://www.nhs.uk/news/2008/11November/Pages/Newcaffeineadviceinpregnancy.aspx. Accessed 13 June 2014.

12. American College of Obstetricians and Gynecologists. Education Pamphlet AP001 - Nutrition During Pregnancy. Washington DC: American College of Obstetricians and Gynecologists; 2008.

13. Yu WW, Schmid CH, Lichtenstein AH, Lau J, Trikalinos TA. Empirical evaluation of meta-analytic approaches for nutrient and health outcome dose-response data. Res Syn Meth 2013;4:256-68.

14. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. J Am Med Assoc 2000;283:2008-12.

15. NHS choices. Should I limit caffeine during pregnancy? http://www.nhs.uk/chq/Pages/limit-caffeine-during-pregnancy.aspx. Accessed 13 June 2014.

16. Wells GA, Shea B, O'Connell D, Peterson J, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed 13 June 2014.

17. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. American Journal of Epidemiology 1992;135:1301-9.

18. Chiaffarino F, Parazzini F, Chatenoud L, Ricci E, Tozzi L, Chiantera V, et al. Coffee drinking and risk of preterm birth. European journal of clinical nutrition 2006;60:610-3.

19. Williams MA, Mittendorf R, Stubblefield PG, Lieberman E, Schoenbaum SC, Monson RR. Cigarettes, coffee, and preterm premature rupture of the membranes. American Journal of Epidemiology 1992;135:895-903.

20. Hamling J, Lee P, Weitkunat R, Ambuhl M. Facilitating meta-analyses by deriving relative effect and precision estimates for alternative comparisons from a set of estimates presented by exposure level or disease category. Stat Med 2008;27:954-70.

21. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials 1986;7:177-88.

22. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst 1988;80:1198-202.

23. White IR. Multivariate random-effects meta-analysis. Stata Journal 2009;9:40-56.

24. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539-58.

25. Boylan SM, Greenwood DC, Alwan N, Cooke MS, Dolby VA, Hay AW, et al. Does nausea and vomiting of pregnancy play a role in the association found between maternal caffeine intake and fetal growth restriction? Maternal & Child Health Journal 2013;17:601-8.

26. StataCorp. Stata statistical software: Release 13.1. College Station, TX: Stata Corporation; 2013.

27. Infante-Rivard C, Fernandez A, Gauthier R, David M, Rivard GE. Fetal loss associated with caffeine intake before and during pregnancy. JAMA 1993;270:2940-3.

28. Caan BJ, Goldhaber MK. Caffeinated beverages and low birthweight: a case-control study. American Journal of Public Health 1989;79:1299-300.

29. Boylan S, Cade JE, Dolby VA, Greenwood DC, Hay AWM, Kirk SFL, et al. Maternal caffeine intake during pregnancy and risk of fetal growth restriction: a large prospective observational study. BMJ 2008;337:a2332.

30. Khoury JC, Miodovnik M, Buncher CR, Kalkwarf H, McElvy S, Khoury PR, et al. Consequences of smoking and caffeine consumption during pregnancy in women with type 1 diabetes. J Matern Fetal Neonatal Med 2004;15:44-50.

31. Sengpiel V, Elind E, Bacelis J, Nilsson S, Grove J, Myhre R, et al. Maternal caffeine intake during pregnancy is associated with birth weight but not with gestational length: results from a large prospective observational cohort study. BMC Medicine 2013;11:42.

32. Spinillo A, Capuzzo E, Nicola SE, Colonna L, Egbe TO, Zara C. Factors potentiating the smoking-related risk of fetal growth retardation. British Journal of Obstetrics & Gynaecology 1994;101:954-8.

33. Al-Ansary LA, Babay ZA. Risk factors for spontaneous abortion: a preliminary study on Saudi women. Journal of the Royal Society of Health 1994;114:188-93.

34. Axelsson G, Rylander R, Molin I. Outcome of pregnancy in relation to irregular and inconvenient work schedules. Br J Ind Med 1989;46:393-8.

35. Furuhashi N, ata S, uzuki M, iruta M, anaka M, akahashi T. Effects of caffeine ingestion during pregnancy. Gynecologic and Obstetric Investigation 1985;19:187-91.

36. Parazzini F, Bocciolone L, Fedele L, Negri E, La VC, Acaia B. Risk factors for spontaneous abortion. International Journal of Epidemiology 1991;20:157-61.

37. Windham GC, Von BJ, Waller K, Fenster L. Exposure to environmental and mainstream tobacco smoke and risk of spontaneous abortion. American Journal of Epidemiology 1999;149:243-7.

38. Pollack AZ, Buck Louis GM, Sundaram R, Lum KJ. Caffeine consumption and miscarriage: a prospective cohort study. Fertility & Sterility 2010;93:304-6.

39. Dominguez-Rojas V, De Juanes-Pardo JR, Astasio-Arbiza P, Ortega-Molina P, Gordillo-Florencio E. Spontaneous abortion in a hospital population: are tobacco and coffee intake risk factors? European Journal of Epidemiology 1994;10:665-8.

40. Eskenazi B, Stapleton AL, Kharrazi M, Chee WY. Associations between Maternal Decaffeinated and Caffeinated Coffee Consumption and Fetal Growth and Gestational Duration. Epidemiology 1999;10:242-9.

41. Tough SC, Newburn-Cook CV, White DE, Fraser-Lee NJ, Faber AJ, Frick C, et al. Do maternal characteristics and past pregnancy experiences predict preterm delivery among women aged 20 to 34? Journal of Obstetrics and Gynaecology Canada 2003;25:656-66.

42. Kesmodel U, Olsen SF, Secher NJ. Does alcohol increase the risk of preterm delivery? Epidemiology 2000;11:512-8.

43. Mikkelsen TB, Osterdal ML, Knudsen VK, Haugen M, Meltzer HM, Bakketeig L, et al. Association between a Mediterranean-type diet and risk of preterm birth among Danish women: a prospective cohort study. Acta Obstetricia et Gynecologica Scandinavica 2008;87:325-30.

44. Jarosz M, Wierzejska R, Siuba M. Maternal caffeine intake and its effect on pregnancy outcomes. European Journal of Obstetrics, Gynecology, & Reproductive Biology 2012;160:156-60.

45. Stusser R, Paz G, Ortega M, Pineda S, Infante O, Martin P, et al. Risk of low birth weight in the Plaza de la Habana region. Boletin de la Oficina Sanitaria Panamericana 1993;114:229-41.

46. Jadsri S, Jadsri C. Parental smoking habits and infant birth weight. Asia-Pacific Journal of Public Health 1995;8:134-7.

47. Alonso J, Sosa C, Verde ME, Balsamo A, Moraes M, Zolessi M, et al. Risk factors for term small for gestational age. A case-control study in an uruguayan population. International Journal of Gynecology and Obstetrics 2012;119:S276.

48. Infante-Rivard C. Caffeine intake and small-for-gestational-age birth: modifying effects of xenobiotic-metabolising genes and smoking. Paediatric and Perinatal Epidemiology 2007;21:300-9.

49. Leonardi-Bee J, Britton J, Venn A. Secondhand Smoke and Adverse Fetal Outcomes in Nonsmoking Pregnant Women: A Meta-analysis. Pediatrics 2011;127:734-41.

50. Barker DJ. Childhood causes of adult diseases. Arch Dis Child 1988;63:867-9.

51. Nathanielsz PW. Fetal Programming: How the Quality of Fetal Life Alters Biology for a Lifetime. NeoReviews 2000;1:e126-e131.

52. Eriksson M, Wallander MA, Krakau I, Wedel H, Svardsudd K. The impact of birth weight on coronary heart disease morbidity and mortality in a birth cohort followed up for 85 years: a population-based study of men born in 1913. J Intern Med 2004;256:472-81.

53. Ong KK, Dunger DB. Birth weight, infant growth and insulin resistance. Eur J Endocrinol 2004;151 Suppl 3:U131-U139.

54. Levy-Marchal C, Jaquet D. Long-term metabolic consequences of being born small for gestational age. Pediatr Diabetes 2004;5:147-53.

55. Grosso LM, Triche EW, Belanger K, Benowitz NL, Holford TR, Bracken MB. Caffeine Metabolites in Umbilical Cord Blood, Cytochrome P-450 1A2 Activity, and Intrauterine Growth Restriction. American Journal of Epidemiology 2006;163:1035-41.

56. Cook DG, Peacock JL, Feyerabend C, Carey IM, Jarvis MJ, Anderson HR, et al. Relation of caffeine intake and blood caffeine concentrations during pregnancy to fetal growth: prospective population based study. BMJ 1996;313:1358-62.

57. Maslova E, Bhattacharya S, Lin SW, Michels KB. Caffeine consumption during pregnancy and risk of preterm birth: a meta-analysis. The American Journal of Clinical Nutrition 2010;92:1120-32.

58. Leviton A, Cowan L. A review of the literature relating caffeine consumption by women to their risk of reproductive hazards. Food Chem Toxicol 2002;40:1271-310.

59. Peck JD, Leviton A, Cowan LD. A review of the epidemiologic evidence concerning the reproductive health effects of caffeine consumption: a 2000-2009 update. Food Chem Toxicol 2010;48:2549-76.

60. Boylan S, Cade JE, Dolby VA, Greenwood DC, Hay AWM, Kirk SFL, et al. Maternal caffeine intake during pregnancy and risk of fetal growth restriction: a large prospective observational study. BMJ 2008;337:a2332.

61. Landi MT, Sinha R, Lang NP, Kadlubar FF. Human cytochrome P4501A2. IARC scientific publications 1999. Issue 148.

62. Klebanoff MA, Levine RJ, DerSimonian R, Clemens JD, Wilkins DG. Maternal serum paraxanthine, a caffeine metabolite, and the risk of spontaneous abortion. N Engl J Med 1999;341:1639-44.

63. Klebanoff MA, Levine RJ, Clemens JD, Wilkins DG. Maternal Serum Caffeine Metabolites and Small-for-Gestational Age Birth. American Journal of Epidemiology 2002;155:32-7.