

Infertility and the risk of adverse pregnancy outcomes: a systematic review and meta-analysis

Carmen Messerlian^{1,*}, Laura Maclagan¹, and Olga Basso^{1,2}

¹Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada H3A1 A2

²Department of Obstetrics and Gynecology, Royal Victoria Hospital, Montreal, Canada H3A 1A1

*Correspondence address. E-mail: carmen.messerlian@mail.mcgill.ca

Submitted on February 29, 2012; resubmitted on May 30, 2012; accepted on July 6, 2012

STUDY QUESTION: Do women who conceive without treatment after a long time to pregnancy (TTP) have an increased risk of preterm birth compared with women in the general obstetric population?

SUMMARY ANSWER: Based on this meta-analysis of 14 studies, women with a long TTP are at an increased risk of preterm birth: pooled crude odds ratio (OR): 1.38 (95% CI: 1.25–1.54).

WHAT IS KNOWN ALREADY: Several studies have shown that women who conceive without treatment after >12 months of trying have an elevated risk of poor pregnancy outcomes. To date, no systematic review or meta-analysis of this evidence has been published.

STUDY DESIGN, SIZE, DURATION: This systematic review identified literature from Embase, Medline and Popline published between January 1974 and October 2011, on the association between infertility in a non-treated population and the risk of preterm birth, low birthweight (LBW), small-for-gestational age and birthweight deficits.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Two authors independently conducted the searches, selected the studies and abstracted the data. A total of 89 full-text articles were assessed for eligibility and 17 met the inclusion criteria. The pooled analysis of the primary outcome led to a total sample size of 1 269 758 births: 19 983 in the exposed/infertile group and 1 249 775 in the unexposed/fertile group. There were a total 68 885 preterm births in the overall sample: 1644 (8.2%) and 67 241 (5.4%) among the infertile and reference groups, respectively.

MAIN RESULTS AND THE ROLE OF CHANCE: A moderate increase in the risk of preterm birth persisted irrespective of the type of pooling. The common OR of the pooled crude preterm birth data compared with the pooled regression-adjusted analysis was modestly attenuated: from 1.38 (95% CI: 1.25, 1.54) to 1.31 (95% CI: 1.21, 1.42), with I^2 decreasing from 53.2 to 3.9% in the crude to adjusted results, respectively. An association of a similar magnitude was seen between infertility and LBW, due in part to overlapping of outcomes.

LIMITATIONS, REASONS FOR CAUTION: Consistency of the estimates across various types of pooling, including the more restricted sensitivity analyses of higher quality studies, is reassuring. While it is possible that systematic error may have been present through misclassification of exposure and confounding, these findings suggest that it would need to be of the same magnitude across diverse studies, which seems unlikely.

WIDER IMPLICATIONS OF THE FINDINGS: A long TTP is only a symptom, research is needed to assess whether specific groups of infertile couples are at increased risk of adverse outcome, or whether the increased risk is due to characteristics common to most infertile couples. As long as the contribution of infertility is not clarified, the risks due to assisted reproductive technologies cannot be properly assessed.

STUDY FUNDING/COMPETING INTEREST(S): C.M. was supported by a Canadian Institutes of Health Research doctoral research award at the time of this study. No competing interests are declared.

Key words: infertility / preterm birth / birthweight / meta-analysis

Introduction

Infertility, generally defined as failure to conceive a clinically detectable pregnancy after >12 months of unprotected intercourse, is a common condition, reported by ~1 in 6 couples (Hull et al., 1985; Boivin et al., 2007). While it is well known that babies born after assisted reproductive technology (ART) are at increased risk of poor obstetric and perinatal outcomes (Helmerhorst et al., 2004; Allen et al., 2006; Reddy et al., 2007; McDonald et al., 2009), a substantial body of literature suggests that infertility itself, regardless of treatment, is also associated with an elevated risk of adverse pregnancy outcome (Joffe and Li, 1994; Henriksen et al., 1997; Basso and Baird, 2003; Thomson et al., 2005; Romundstad et al., 2008). In particular, several studies show that, compared with infants conceived within 12 months of trying, those conceived after a waiting time of >12 months have a higher risk of preterm birth, low birthweight (LBW) and small-for-gestational age (SGA) (Henriksen et al., 1997; Basso and Baird, 2003; Thomson et al., 2005; Zhu et al., 2007; Jaques et al., 2010; Wisborg et al., 2010). As infertility is a heterogeneous condition, caused by various underlying pathologies, it is possible that some of the mechanisms leading to infertility also play a role in the etiology of these outcomes (Saunders et al., 1988; Tan et al., 1992; Goldenberg et al., 2000; Kramer et al., 2001a,b; Basso and Baird, 2003; Goldenberg et al., 2008). The elevated risk observed in couples conceiving naturally after a long time to pregnancy (TTP) may thus reflect a high risk in a small group or an overall elevated risk common to most infertile couples (Basso et al., 2003). Attempting to understand the contribution of infertility on adverse pregnancy outcomes, separate from that of treatment, is recognized as a priority by several experts (Allen et al., 2006; Reddy et al., 2007). In this context, it is important to obtain a more precise estimate of the actual risk associated with a long TTP. Most of the studies examining the association between TTP and reproductive outcome in the general population are relatively recent and, to date, a systematic review of this literature has not been carried out. We performed a systematic review and meta-analysis of the published literature on infertility (measured by TTP) and preterm birth, LBW and SGA, as well as birthweight deficits.

Materials and Methods

Search strategy

We employed a three-stage search strategy and consulted with an experienced medical librarian throughout the process. First, we conducted a limited search of Medline and Embase to identify and generate all key words and subject headings (MeSH/Embase terms) for concept 1 (infertility) and concept 2 (adverse pregnancy outcomes). In the second stage, carried out after identifying all appropriate search terms (see Appendix 1), we conducted comprehensive searches of Embase, Medline and Popline from 1 January 1974 to 14 October, 2011 using the OVID interface. Stage three involved systematically hand-searching the reference lists of key articles and reports.

Study selection criteria

Type of study and participants

We considered any case-control or cohort study of pregnant women that examined the association between infertility and the outcomes of interest

(preterm birth, LBW, SGA and birthweight deficits) in either a clinical or population setting, provided that a non-treated infertile group was included as a reference.

Exposure

All types of infertility, subfertility or delayed conception were included. Studies that compared births from infertile couples conceiving through ART with births from the general obstetric population were excluded, as such comparisons would not allow to separate the effect of infertility from that of treatment.

Outcomes

Our primary outcome was preterm birth, defined as a birth occurring before 37 completed weeks of gestation. The secondary outcomes were indicators of infant growth: LBW (defined as <2500 g at birth), SGA [defined as the lowest 10% of birthweight by gestational age, except in one study (Zhu et al., 2007), where the cut-off was set at 5%], and continuous birthweight (in g).

We restricted our search to studies in humans and to publications in the English language. We excluded reviews, editorials, case reports, letters to the editor, unpublished data and duplicate publications.

Study process

Two independent reviewers (C.M. and L.M.) screened and reviewed the titles and abstracts of all identified citations. We retrieved full-text articles if either reviewer considered the publication relevant based on the initial abstract review. Full-text articles were carefully assessed independently by each reviewer and inclusion into the final set was based on the *a priori* selection criteria. Consensus was sought on the final set of articles to be included and disagreements were resolved through discussions with the senior author (O.B.). Using the data extraction form, the two reviewers independently abstracted relevant data from the full-text articles included in the analysis. Information on study population and exposure groups, characteristics of participants, confounding and the methods used to control for confounding, were obtained from each article. We reconstructed 2 × 2 tables by abstracting raw data presented in the primary studies for each relevant study outcome. Crude and adjusted measures of effect and 95% confidence intervals were also obtained from the tables of results.

Study quality

Study quality was assessed using the Newcastle–Ottawa Scale (NOS), a qualitative instrument designed to evaluate observational studies in three domains: selection of participants, comparability of study groups and ascertainment of outcome or exposure depending on study design (Wells et al., 2000). The instrument comprises eight scored items and the total scores can range from 0 to 9 points. While some controversy exists regarding the use of the NOS to evaluate observational studies for systematic reviews, with critics suggesting that its validity has not been sufficiently tested (Stang, 2010), we nevertheless chose to provide a qualitative assessment of the primary studies as a means of describing and comparing the sourced articles.

We further evaluated the quality through a checklist of assessment of bias and confounding that we specifically developed for the purpose of this review. Answers to the following yes/no questions were used to appraise each study: (i) was there a direct measure of TTP (as opposed to it being inferred from having sought help for infertility) and was the TTP cut-off set at 12 months? (ii) did the study adjust for age and/or parity through matching or modeling? (iii) was the infertile group unexposed to

pharmacological infertility treatment? and (iv) did the study restrict to singletons only? These four questions were used as a further criterion to describe and evaluate the quality of the included studies. Studies with three or more 'yes' responses were considered to be at low risk for major bias and confounding issues pertinent to this area of research and judged to be of higher quality.

Statistical analysis

All statistical analyses were performed using STATA version 11 (Stata-Corp, 2009). We calculated crude odds ratios (ORs) for the dichotomous study outcomes (preterm birth, LBW and SGA), using the data from the reconstructed 2×2 tables. As we wanted to compare the overall pooled crude data with the data from studies that controlled for at least age and/or parity through restriction or matching, we stratified our analysis by combining data according to whether the main confounders were adjusted for at the design stage, and termed these results 'pooled matched or stratified analyses'. We furthermore used the regression-adjusted effect estimates as reported in the results of the original studies and calculated the log ORs and the relative standard errors in order to weight and pool these data, and termed the results 'pooled regression-adjusted analyses'. The crude, matched and regression-adjusted data were pooled separately to allow for comparison of different groups of studies, according to whether confounding had been accounted for.

We examined birthweight as a continuous variable using the sample size, mean and standard deviation of birthweight for studies that reported such data. We then multiplied the standardized mean difference (SMD) produced in the pooled analysis by an estimate of the standard deviation of birthweight [440 g—the mean standard deviation of birthweight of males and females at 40 weeks as reported in the Canadian birthweight standard (Kramer *et al.*, 2001a,b)] in order to obtain an interpretable result of the overall mean difference in birthweight for pooled studies.

We performed two separate sensitivity analyses on the primary outcome. First, we restricted our pooling to include only those studies with a score of 8 or 9 on the NOS. We furthermore restricted the analysis to studies for which the answer to at least three out of the four questions concerning bias and confounding was 'yes', in order to examine whether we would obtain a different pooled estimate when using a more stringent criterion to select studies.

We assessed the possibility that our results were influenced by publication bias and conducted an Egger test by regressing the log OR of preterm birth against its standard error of the log OR (Egger and Davey Smith, 1997). We graphically depicted small-study reporting bias through the use of a contour-enhanced funnel.

All analyses pooled measures through a random effects model, using the DerSimonian-Laird method (Egger and Davey Smith, 1997). We made this choice because a random effects model incorporates both random error and between-study variability: given the anticipated heterogeneity in the data, these models produce more conservative estimates with wider confidence intervals (Egger and Davey Smith, 1997). We assessed the heterogeneity due to the potentially large variability in results across studies using the I^2 -squared (I^2) statistic in Stata's metan command. The I^2 value is the percentage of total variability of study estimates that are due to heterogeneity (Higgins and Thompson, 2002). I^2 values of 25, 50 and 75% correspond to low, moderate and high heterogeneity (Higgins and Thompson, 2002). We attempted to address some of the heterogeneity in the overall group of studies by pooling subgroups of studies (crude, matched, regression-adjusted analyses) that were potentially more similar.

Results

Search results

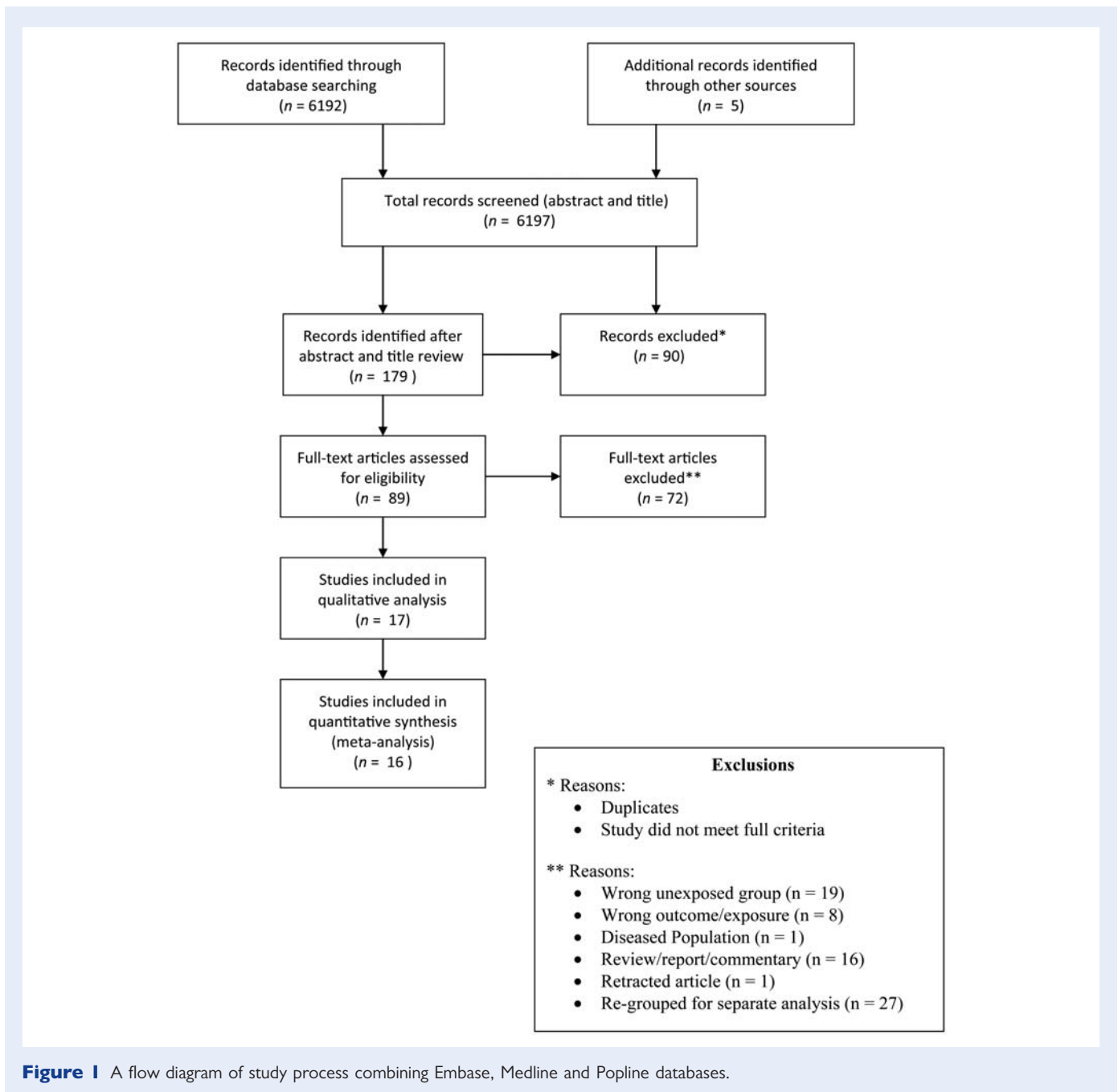
We identified 6192 citations through Medline, Embase and Popline database searches. An additional five records (Cooney *et al.*, 2006; Juang *et al.*, 2007; Romundstad *et al.*, 2008; Pritts *et al.*, 2009; Wisborg *et al.*, 2010) were retrieved by hand searching the reference list of key articles, resulting in a total of 6197 screened articles at the abstract and title level. Of these, 179 were identified by either reviewer as being eligible for full-text review. After excluding duplicates and records that clearly failed to meet the study selection criteria ($n = 90$), a total of 89 full-text articles were assessed for eligibility. Figure 1 shows the flow diagram of the literature search process and results. After careful review of all 89 full-text articles, a total of 17 met all inclusion criteria (Tuck *et al.*, 1988; Varma *et al.*, 1988; Hill *et al.*, 1990; Bhalla *et al.*, 1992; Joffe and Li, 1994; Henriksen *et al.*, 1997; McElrath and Wise, 1997; Wang *et al.*, 2002; Basso and Baird, 2003; Thomson *et al.*, 2005; Cooney *et al.*, 2006; Zhu *et al.*, 2007; Romundstad *et al.*, 2008; Jaques *et al.*, 2010; Raatikainen *et al.*, 2010a,b; Ranta *et al.*, 2010; Wisborg *et al.*, 2010) with one study (Romundstad *et al.*, 2008) included only in the qualitative synthesis and not in the meta-analysis. We excluded the Romundstad article from the quantitative synthesis as infertility in the index pregnancy was inferred based on previous history of conceiving a sibling with ART rather than on clinical data of current/index pregnancy. Characteristics of the included studies are summarized in Table 1.

Outcomes

The pooled analysis of the primary outcome led to a total sample size of 1 269 758 births: 19 983 in the exposed/infertile group and 1 249 775 in the unexposed/fertile group. There was a total of 68 885 preterm births in the overall sample: 1644 (8.2%) and 67 241 (5.4%) among the infertile and reference groups, respectively. Pooling of the 14 studies with crude preterm birth data yielded a common OR of 1.38 (95% CI: 1.25, 1.54), with moderate heterogeneity ($I^2 = 53.2%$) (Fig. 2).

The OR did not change when we performed subgroup analyses and pooled only studies that matched or stratified on age and/or parity, 1.39 (95%CI: 1.20, 1.62), however heterogeneity decreased to $I^2 = 33%$ (Supplementary data, Fig. S1). Pooling studies that adjusted for relevant confounders using regression models resulted in a modestly attenuated result, 1.31 (95% CI: 1.21, 1.42) with an I^2 of 3.9% (Fig. 3). Overall, the results did not materially change by type of pooling, with the regression-adjusted results producing estimates slightly closer to the null compared with the pooled crude and pooled matched analysis. Table II compares the three-study outcomes by type of pooling.

Pooled crude SGA data ($n = 9$ studies) suggests an increased odds of SGA among infertile women who conceive without treatment after a period of infertility compared with women who conceive within 12 months: 1.24 (95% CI: 1.12, 1.36) (Fig. 4). Pooled matched and pooled regression-adjusted results reduced the association slightly to 1.16 (95% CI: 0.97, 1.37) and 1.17 (95% CI: 1.03, 1.33), respectively (Table II). However, only a small number of studies were included in these subgroups (two and four studies in the matched and adjusted analyses, respectively).



A total of eight studies were pooled with LBW data, resulting in an estimated increased risk of the same magnitude as that seen for preterm birth. The crude analysis revealed an association between infertility and the risk of LBW when comparing untreated pregnancies with a history of infertility with pregnancies in the reference population: 1.30 (95% CI: 1.16, 1.45) (data not shown). The association was strengthened when pooling matched and regression-adjusted data: 1.50 (95% CI: 1.27, 1.78) and 1.34 (95% CI: 1.21, 1.48), respectively (Table II, Fig. 5).

A total of four studies reported sufficient continuous birthweight data to be included in the meta-analysis (Hill *et al.*, 1990; Thomson *et al.*, 2005; Cooney *et al.*, 2006; Zhu *et al.*, 2007). The overall

mean decrease in birthweight among the infertile group was 44 g (95% CI: -57.2, -35.2) compared with the reference population (data not shown). Further analyses of birthweight showed no clinically meaningful result; this end-point was not explored further.

Sensitivity analysis

When we restricted our results based on the NOS, seven studies with the highest scores were pooled (Tuck *et al.*, 1988; Wang *et al.*, 2002; Basso and Baird, 2003; Thomson *et al.*, 2005; Cooney *et al.*, 2006; Jaques *et al.*, 2010; Wisborg *et al.*, 2010). The association between infertility and preterm birth among this restricted group of studies was

Table 1 Details of studies included in the qualitative and quantitative analyses between infertility and adverse outcomes.

Authors and publication year	Location	Population versus clinic-based sample	Inclusion criteria	Exposure ascertainment	Outcome ascertainment	Relevant outcomes	Number exposed ^a	NOS quality score	Matching, stratification, adjustment
Basso <i>et al.</i> (2003)	Denmark	Population (Danish National Cohort)	Singletons ≥ 24 weeks of gestation with complete records	Interview: pregnancy planning and TTP	GA based on self-reported LMP and birth records	PTB, VPTB, LBW, BW	3826	9	Stratified by parity. Model adjusted for age, BMI, smoking, socio-economic status (SES)
Bhalla <i>et al.</i> (1992)	India	Clinic (Department of obstetrics and gynecology in the Nehru Hospital)	Exposed group: previous history of infertility (2 years of primary or 3 years of secondary)	Medical chart review	Medical chart review	PTB	112	6	Control group/unexposed cohort matched for age and parity. ^b No adjusted models
Cooney <i>et al.</i> (2006)	USA	Population (US Collaborative Perinatal Project)	Only women with complete data on TTP and other variables were included. Singleton pregnancies	Interview	Medical record abstraction	PTB, SGA, LBW	2654 ^c	8	Model adjusted for age, parity, SES
Henriksen <i>et al.</i> (1997)	Denmark	Population (two cohorts: <i>Aalborg-Odense cohort</i> and <i>Aarhus cohort</i>)	Women with planned pregnancies. Singletons only	Self-reported TTP	<i>Aalborg-Odense cohort</i> : birth certificates and medical records. <i>Aarhus cohort</i> : gestational age estimated by ultrasound and by LMP. Study outcomes cross-validated using hospital records	PTB	1321	6	Study results from adjusted analysis only include <i>Aarhus cohort</i> . Model adjusted for parity and smoking
Hill <i>et al.</i> (1990)	USA	Clinic (Infertile couples assessed at the Vanderbilt University Center, general obstetric population at same hospital)	Singletons > 24 weeks of gestation were included in the meta-analysis	Medical chart review	Clinic and hospital data	PTL, PPROM, fetal growth retardation, BW	66	5	Unexposed controls matched only to the treated group: crude data presented in the results are considered non-matched
Jaques <i>et al.</i> (2010)	Australia	Clinic (women registered at one of the four infertility clinics in Victoria)	Singleton > 20 weeks conceived without ART within 4 years	ART clinic databases and medical record abstraction	Victorian Birth Record Data: the Perinatal Data Collection Unit	PTB, VPTB, LBW, SGA	2171	8	Unexposed cohort matched for age and year of infant's birth. Model adjusted for age, parity, previous abortions, public/private hospital

Continued

Table 1 Continued

Authors and publication year	Location	Population versus clinic-based sample	Inclusion criteria	Exposure ascertainment	Outcome ascertainment	Relevant outcomes	Number exposed ^a	NOS quality score	Matching, stratification, adjustment
Joffe and Li (1994)	London, UK	Population (The National Child Development Study)	Excluded if women reported any contraception use around the time of pregnancy	Survey: retrospective, self-reported TTP reported at age 33	Survey: retrospective self-reported pregnancy outcomes	PTB, LBW	680	4	Only crude results included
McElrath <i>et al.</i> (1997)	USA	Population (National Maternal and Infant Health Survey)	Subfertile group were 'concerned' with infertility and untreated	Survey: self-reported subfertility and treatment information	Birthweight data obtained from birth records	VLBW	680 ^d	6	Model adjusted for age, prior miscarriage, multiple gestation ^b
Raatikainen <i>et al.</i> (2010a,b)	Finland	Clinic (Clinical birth database of infants born at Kuopio University Hospital from 1989 to 2007)	Singleton pregnancies with available TTP information, conceived spontaneously (without artificial insemination and/or IVF-treatments) were included	Questionnaire: self-reported TTP or if unavailable based on clinical data	Nurse or midwife responsible for delivery—entered in real time	PTB, SGA	1790	6	Adjusted analysis compares TTP 13–24 months versus 0–6 months for PTB and SGA. Model adjusted for age, BMI, prior miscarriage, smoking and other covariates
Ranta <i>et al.</i> (2010)	Finland	Clinic (maternity care patients in Finnish hospital)	Singleton pregnancies, >22 weeks of gestation with first trimester screening	Self-reported TTP	Data collected in maternity care units of two hospitals	PTB, SGA, BW	182	4	Only crude data presented. Stratification by TTP. No adjusted analysis presented
Romundstad <i>et al.</i> (2008)	Norway	Population (women with a history of infertility identified through all infertility clinics in Norway compared with the general obstetric population)	Included all singletons >22 weeks of gestation	Medical Birth Registry data on health of mother during pregnancy included including treatment for infertility	Medical Birth Registry data to obtain outcome information on gestation and birthweight	PTB, SGA, BW	2204	8	Only qualitative analysis included in this review as results compare groups that are inconsistent in comparison with other studies in our meta-analysis ^e
Thomson <i>et al.</i> (2005)	Scotland	Clinic/population subfertile untreated clinic patients versus general obstetric population delivering singletons during the study time period	Singleton pregnancies identified through database	Subfertility inferred by the Aberdeen Fertility Clinic: women are referred only if they fail to achieve a pregnancy after at least 1 year of attempt. Only untreated subfertile selected (treatment status verified by 3 sources)	Record linkage to the Aberdeen Maternity and Neonatal Databank	PTB, LBW, BW	632	9	Age and parity adjusted in model

Tuck <i>et al.</i> (1988)	England	Clinic infertile patients (≥ 35 years) versus general obstetric patients (≥ 35 years) from same hospital between 1978 and 1983	Restricted to primiparous and singletons. Women ≥ 35 years of age	Oxford Obstetric Data: women with a history of involuntary infertility	Oxford Obstetric Data System and case notes	PTB, SGA, BW	72	8	Matched on age and parity
Varma <i>et al.</i> (1988)	London/England	Clinic (women seen at infertility clinic at St-George's Hospital who conceived compared with general obstetric population in the same hospital)	Included multiples	Inferred by infertility clinic assessment	Ultrasound examination to date pregnancy	Premature rupture of membranes, preterm labor, BW, SGA	444 ^f	5	No regression analysis or matching
Wang <i>et al.</i> (2002)	Australia	Clinic (Obstetrics and gynecology patients in Adelaide between 1986 and 1998)	Singleton >20 weeks with complete records. Treatment group included intrauterine insemination or donor insemination with minimal gonadotrophin stimulation	Infertility clinic data collected for the Perinatal Outcome Statistics Unit	GA obtained through ultrasound	PTB	1015 ^f	8	Adjusted for age, parity, type of delivery, smoking, length of infertility and other covariates
Wisborg <i>et al.</i> (2010)	Denmark	Population (patients at the Department of Obstetrics and Gynecology at Aarhus University Hospital between 1989 and 2006)	Singleton, primiparous, Danish-speaking. Excluded women with chronic illnesses and women with missing TTP and ART data	Self-reported TTP	GA based on early ultrasound or LMP. Outcome data obtained through birth registration at time of delivery and cross-validated with medical records prior to entry	PTB, LBW	2009	9	Adjusted for age, parity, education, alcohol/caffeine use, cohabitation with partner, BMI
Zhu <i>et al.</i> (2007)	Denmark	Population (Danish National Birth Cohort)	Singletons pregnancies with available TTP and treatment data	Interview: questions on planning and TTP	Medical Birth Register data for birthweight and gestational age	SGA ^g , BW	5722	9	Results adjusted for age and parity

ART, assisted reproductive technology; IVF, *in vitro* fertilization; GA, gestational age; LMP, last menstrual period; TTP, time to pregnancy; BW, birthweight; LBW, low birthweight; PTL, preterm labor; PROM, premature rupture of membranes; VLBW, very low birthweight; SGA, small-for-gestational age; PTB, preterm birth.

^aUnless otherwise indicated, the infertile/exposed cohort is defined as having a TTP of ≥ 12 months. The fertile/unexposed cohort is defined as having a TTP of < 12 months.

^bEstimate for PTB and/or VLBW included multiples.

^cUsed a >6 months TTP definition for infertility.

^dNo direct measure of TTP was used. The subfertility group defined as being 'concerned' about infertility; consulted physician and/or having been tested for infertility but did not follow-up with treatment.

^eStudy not included in the meta-analysis as infertility in index pregnancy would be assumed based on previous history of conceiving sibling with ART.

^fInfertility is assumed based on previous history of ART or clinic visit.

^gSGA in the Zhu *et al.* paper was defined as lowest 5% of birthweight distribution by gestational age and sex.

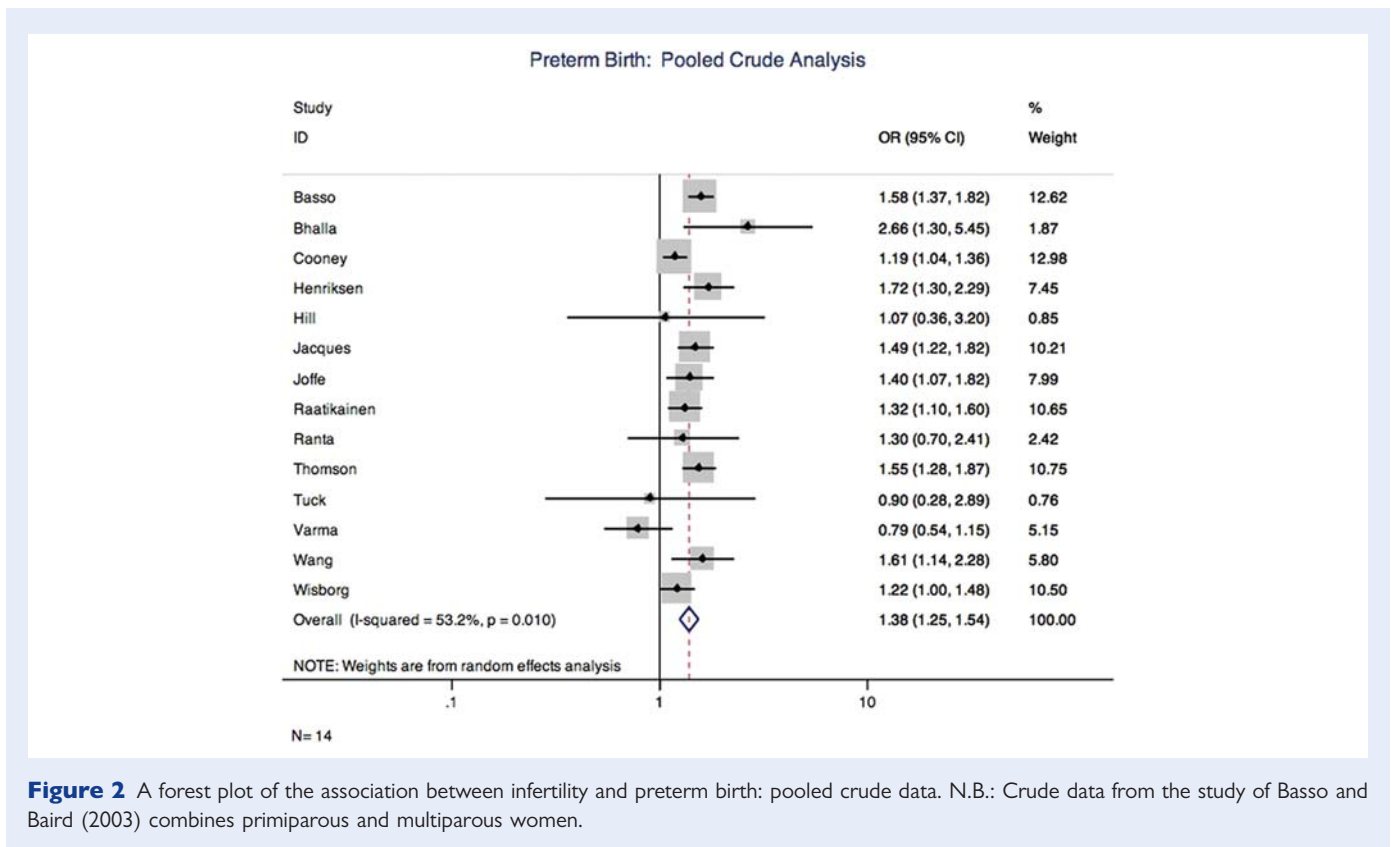


Figure 2 A forest plot of the association between infertility and preterm birth: pooled crude data. N.B.: Crude data from the study of Basso and Baird (2003) combines primiparous and multiparous women.

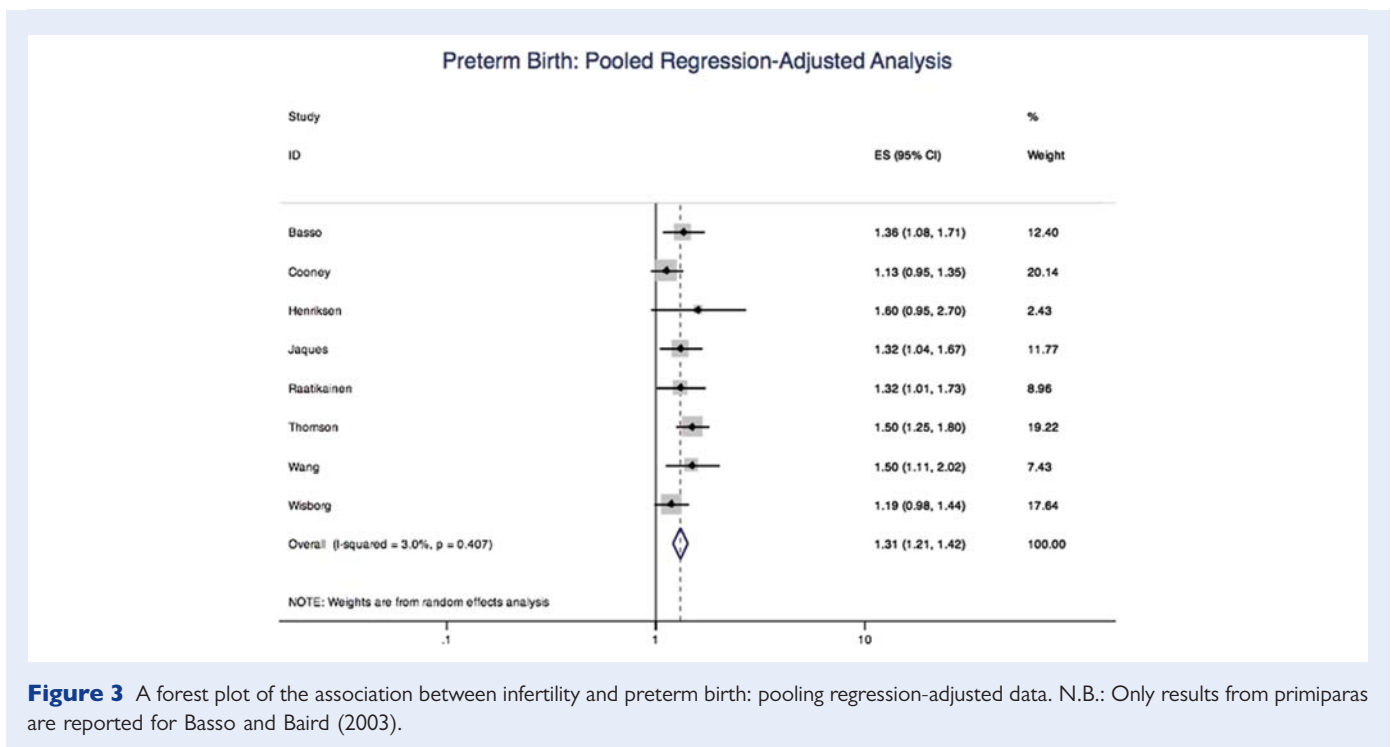


Figure 3 A forest plot of the association between infertility and preterm birth: pooling regression-adjusted data. N.B.: Only results from primiparas are reported for Basso and Baird (2003).

1.36 (95% CI: 1.23–1.50) (Fig. 6), almost unchanged from the pooled crude OR of 1.38 (95% CI: 1.25, 1.54). We furthermore excluded studies that were potentially more prone to bias or confounding according to our quality assessment (Varma et al., 1988; Hill et al., 1990; Bhalla et al., 1992; Joffe and Li, 1994; Wang et al., 2002; Romundstad et al., 2008) and pooled the nine higher quality studies (those with three or more 'yes' responses out of the four questions described above and shown in Supplementary data, Table I) (Tuck

Table II Summary of results by type of pooling; comparing study outcomes by pooling crude, matched and adjusted data.

Outcomes	Total number of studies	Pooled crude data			Pooled matched data			Pooled adjusted data		
		Number of studies	<i>I</i> ² value (%)	OR; 95% CI	Number of studies	<i>I</i> ² value (%)	OR; 95% CI	Number of studies	<i>I</i> ² value (%)	OR; 95% CI
PTB <37 weeks	14	14	53.2	1.38; 1.25–1.54	5	33.0	1.39; 1.20–1.62	8	3.0	1.31; 1.21–1.42
SGA	9	9	36.8	1.24; 1.12–1.36	2	0.0	1.16; 0.97–1.37	4	54.5	1.17; 1.03–1.33
LBW	8	8	19.8	1.30; 1.16–1.45	4	0.0	1.50; 1.27–1.78	6	0.0	1.34; 1.21–1.48

et al., 1988; Henriksen et al., 1997; Basso and Baird, 2003; Thomson et al., 2005; Cooney et al., 2006; Jaques et al., 2010; Raatikainen et al., 2010a,b; Ranta et al., 2010; Wisborg et al., 2010). This restriction resulted in a negligible increase in the estimated association between infertility and the risk of preterm birth compared with overall pooled crude results: 1.40 (95% CI: 1.27, 1.55) (Supplementary data, Fig. S2) versus 1.38 (95% CI: 1.25, 1.54), respectively. When we further restricted the analysis to studies with ‘yes’ response to all four questions (Henriksen et al., 1997; Basso and Baird, 2003; Raatikainen et al., 2010a,b; Wisborg et al., 2010), the results were materially unchanged from the less restrictive sensitivity analysis [1.43 (95% CI: 1.24, 1.66) (not shown)].

Publication bias

The Egger’s test for small-study effect failed to reject the null hypothesis of no small-study effects with a bias coefficient of 0.29 (95% CI: –1.43, 2.02), *P* = 0.72. The contour-enhanced funnel plot showing points for all 14 preterm birth studies (Fig. 7) provides a graphical display of an inverted funnel with little indication that small positive studies influenced the association.

Heterogeneity

Statistical heterogeneity among all pooled analyses ranged from low (*I*² = 0%) in the pooled matched analyses of LBW and SGA (likely owing to the small number of included studies) to moderate (*I*² = 54.5%) in the pooled regression-adjusted SGA analysis. The low-to-moderate heterogeneity results presented in Table II suggest that between-study variability was reasonable. This consistency in the overall group of included studies permits pooling with greater confidence, as the ORs presented in the pooled analysis are therefore likely not an average of extremes, but rather a more precise estimate of the overall effect of infertility on the adverse pregnancy outcomes examined in this review (Higgins et al., 2003).

Study quality

The quality scores using the NOS ranged from 4 to 9, and the mean overall score for all 17 studies was 7.0 (SD = 1.9). The main areas of quality concern were in the ascertainment of infertility/TTP and comparability of studies based on controlling for main covariates (see Supplementary data, Table S1).

Discussion

Saunders et al. (1988) first reported that infertility itself could contribute to problems during pregnancy. Using Australian *in vitro* fertilization registry data, the authors found that couples who conceived spontaneously while on a waiting list for evaluation and treatment of infertility had a higher risk of preterm birth compared with the general population. Moreover, among singleton births, the estimated increased risk was comparable to births conceived through IVF, suggesting an elevated baseline risk among infertile couples, irrespective of treatment.

Subsequently, numerous studies have examined this association by studying the untreated infertile population and comparing outcomes following various lengths of TTP (Bhalla et al., 1992; Joffe and Li,

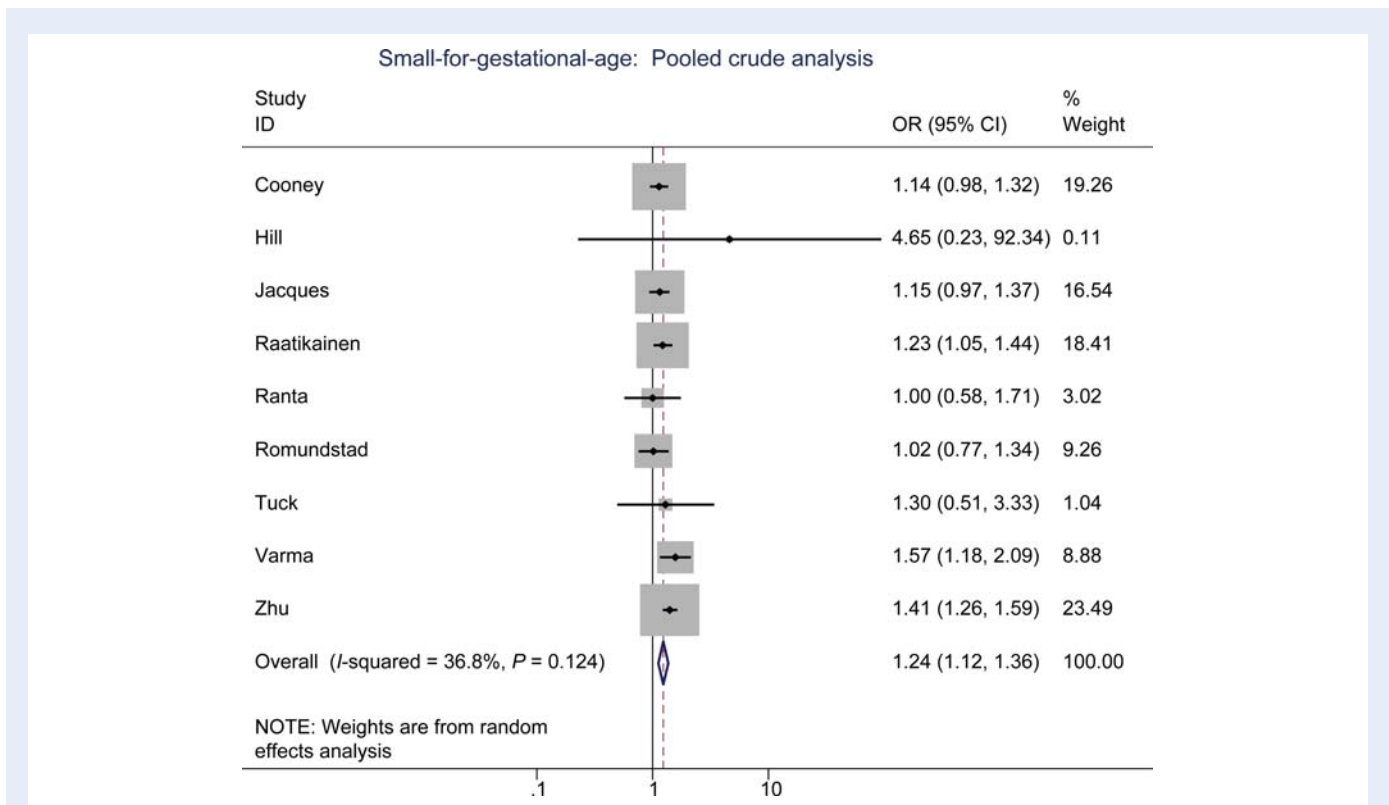


Figure 4 A forest plot of the association between infertility and SGA: pooled crude analysis. N.B.: Zhu et al. (2007) defined SGA as the lowest 5% of birth weight by gestational age.

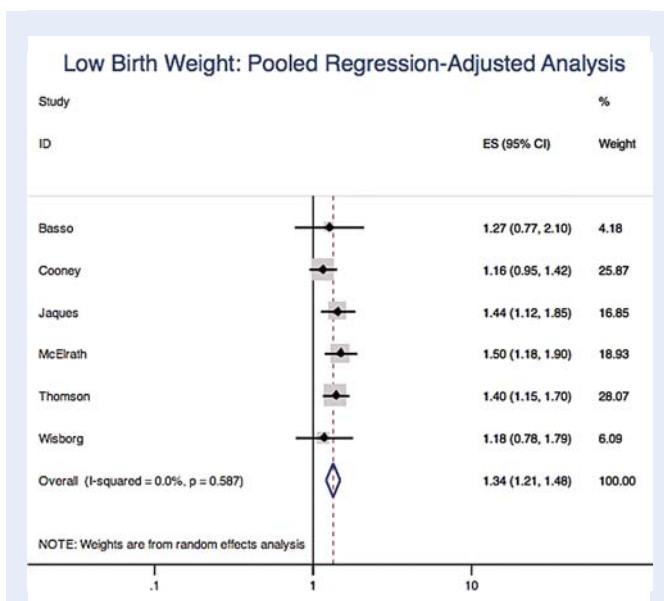


Figure 5 A forest plot of the association between infertility and LBW: pooled regression-adjusted analysis. N.B.: Study by McElrath et al. (1997) examined very low birth weight (<1500 g) as the study outcome.

1994; Basso and Baird, 2003; Thomson et al., 2005; Cooney et al., 2006; Romundstad et al., 2008; Jaques et al., 2010; Raatikainen et al., 2010a,b). We aimed to systematically identify the published literature in this field and produce a pooled estimate of the risk of selected adverse pregnancy outcomes associated with infertility itself. Thus, we selected studies that compared pregnancies conceived without treatment after a long TTP with pregnancies conceived within 12 months of trying.

In total, we identified 17 studies that met our criteria. According to our results, a moderate increase in the risk of preterm birth persisted irrespective of the type of pooling. The common OR of the pooled crude preterm birth data compared with the pooled regression-adjusted analysis was only modestly attenuated: from 1.38 (95% CI: 1.25, 1.54) to 1.31 (95% CI: 1.21, 1.42), with the pooled adjusted I^2 decreasing from 53.2% in the crude results to 3.9% in the adjusted results. Furthermore, we observed an association of a similar magnitude between infertility and LBW, likely due in part to overlapping of outcomes, as the estimates were not adjusted for gestational age. Similar to preterm birth, the estimates for LBW showed little variation between the crude pooled analyses and the pooled adjusted analyses.

The association between infertility and SGA appears to be more modest than the one observed for preterm birth. The study by Jaques et al. (2010) was the only one among those included in the pooled adjusted analysis that showed a null association with SGA,

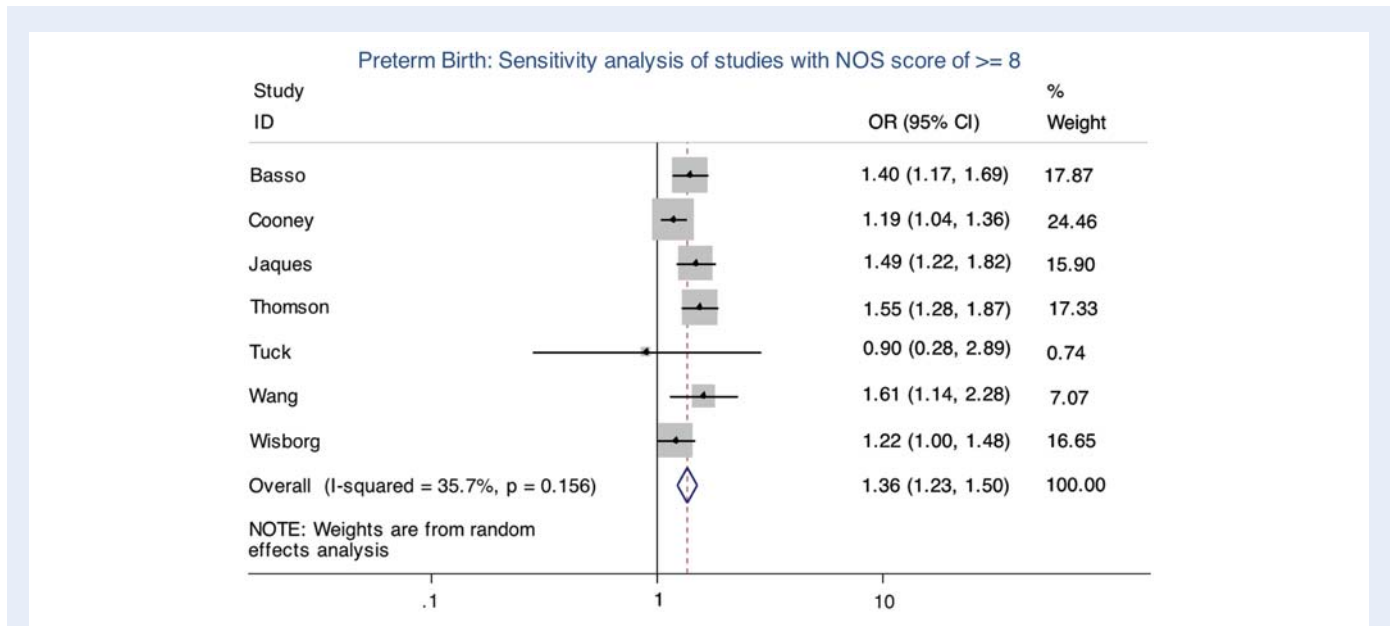


Figure 6 Sensitivity analysis of preterm birth using NOS including only higher quality studies (NOS score ≥ 8).

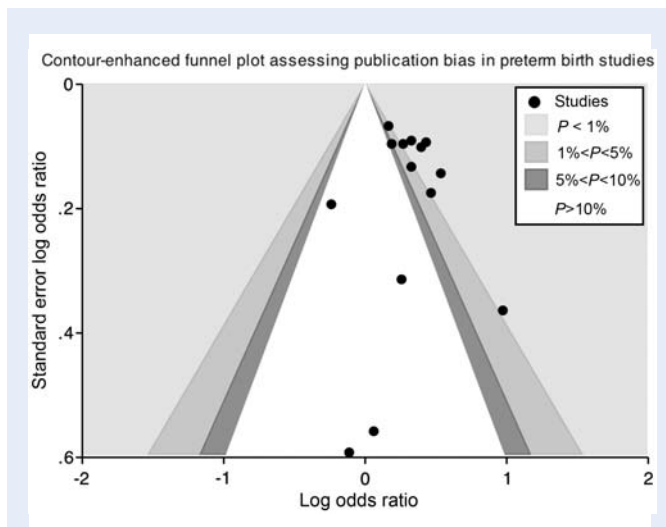


Figure 7 A contour-enhanced funnel plot of small-study bias for all 14 preterm birth studies.

while contributing almost 24% of the weight. It is possible that a selected group of infertile women were included among the exposed, as the requirement for inclusion in their analysis was that women had to have been registered in a fertility clinic and then give birth within 4 years without the use of ART or artificial insemination. This study also linked various databases in order to identify a group of women who were infertile, untreated and had given birth. However, no information on the length or severity of infertility was provided.

To our knowledge, this is the first systematic review reporting on the association between infertility and adverse pregnancy outcome. Numerous systematic reviews have been published suggesting an increased risk of a range of poor obstetric and neonatal outcomes

among infertile women undergoing different types of medically assisted reproductive treatment (Helmerhorst *et al.*, 2004; Reddy *et al.*, 2007; McDonald *et al.*, 2009). However, these studies could not differentiate between the risk associated with ART and that associated with infertility (or, more likely, its causes). Our results corroborate the hypothesis that the association between ART and adverse pregnancy outcome reported in several studies is in part due to the underlying infertility. Thus, the true causal effect of ART procedures on adverse outcome cannot be estimated until we achieve a better understanding of the contribution of infertility. Given that infertility is a heterogeneous condition, some of the mechanisms that lead to infertility may also be involved in the etiology of preterm birth. It is possible that an exposure that causes couples to experience difficulties conceiving also compromises the pregnancy. Baird *et al.* (1999), for example, have hypothesized that prenatal exposure to stress, pelvic infections and environmental contaminants may contribute to both infertility and preterm birth.

Our estimates must be considered in light of certain limitations inherent to some of the included studies. In several of these (Supplementary data, Table S1), women were selected based on a history of infertility, or on having been registered at an infertility clinic, which likely resulted in substantial heterogeneity as to the length and severity of infertility. In such studies, a direct measure of TTP was absent. Furthermore, the study by Cooney *et al.* (2006), included in the pooled preterm birth and LBW analyses, used a TTP cut-off of 6 months, which may partly explain the study's negative findings.

Infertility is not a dichotomous state, but rather a continuum ranging from complete sterility (which, by definition, would not be included in our study) to subtly reduced fertility (Habbema *et al.*, 2004). While infertility is defined as a TTP of > 12 months among couples who engage in regular unprotected intercourse (Evers, 2002), it is possible that some study participants included in the infertile group would not meet this general clinical definition (as explicitly mentioned in the study

by Cooney *et al.*). If the misclassification was non-differential, the estimated association with adverse outcomes would likely be attenuated, resulting in bias towards the null. Furthermore, there was heterogeneity in the methods by which exposure and outcome data were ascertained, with medical charts and clinical data likely producing the most valid estimates, followed by self-reported data through prenatal questionnaires, and data linkage of large administrative databases. In addition, not every study adjusted for age and parity, which would be considered a minimum requirement in estimating the effect of infertility on pregnancy outcomes.

Conclusion

Despite the above limitations, our findings produced consistent estimates with three different types of pooling. Furthermore, when excluding studies with potential limitations, the magnitude of the pooled OR was unaffected. Infertile couples who conceive spontaneously without treatment are at higher risk of preterm birth and LBW and also have a modestly elevated risk of having an SGA baby. Infertility, however, is only a symptom of underlying pathology, and a TTP of >1 year is an arbitrary definition. Future research is needed to assess whether specific groups of infertile couples have a substantially increased risk of adverse outcome, or whether characteristics common to most infertile couples confer a modestly elevated risk across the board.

Supplementary data

Supplementary data are available at <http://humrep.oxfordjournals.org/>.

Acknowledgements

The authors thank Dr Seang-Lin Tan (McGill University, Dept. of Obstetrics and Gynecology) for his helpful contributions to the manuscript and Lorie Kloda (McGill University Life Sciences Library) for her assistance throughout our database search process.

Authors' roles

C.M. conceived and designed the study, developed the search methods, analyzed and interpreted the results and drafted the manuscript. C.M. and L.M. executed the searches, assessed the abstracts and full-text articles, selected articles, abstracted the data and revised the manuscript. O.B. contributed to the design of the study and the search method, assessed selected articles, interpreted the results, critically revised the manuscript and contributed important intellectual content. All authors approved the final version of the manuscript.

Funding

C.M. was supported by a Canadian Institutes of Health Research doctoral research award at the time of this study.

Conflict of interest

None declared.

References

- Allen VM, Wilson RD, Cheung A. Pregnancy outcomes after assisted reproductive technology. *J Obstet Gynaecol* 2006;**3**:220–250.
- Baird DD, Wilcox AJ, Kramer MS. Why might infertile couples have problem pregnancies? *Lancet* 1999;**353**:1724–1725.
- Basso O, Baird D. Infertility and preterm delivery, birthweight, and caesarean section: a study within the danish national birth cohort. *Hum Reprod* 2003;**11**:2478–2484.
- Basso O, Weinberg C, Baird D, Wilcox A, Olsen J. Subfecundity as a correlate of preeclampsia: a study within the Danish National Birth Cohort. *Am J Epidemiol* 2003;**3**:195–202.
- Bhalla AK, Sarala G, Dhaliwal L. Pregnancy following infertility. *Aust N Z J Obstet Gynaecol* 1992;**3**:249–251.
- Boivin J, Bunting L, Collins JA, Nygren KG. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. *Hum Reprod* 2007;**6**:1506–1512.
- Cooney MA, Buck Louis GM, Sun W, Rice MM, Klebanoff MA. Is conception delay a risk factor for reduced gestation or birthweight? *Paediatr Perinat Epidemiol* 2006;**3**:201–209.
- Egger M, Davey Smith G. Meta-analysis: potentials and promise. *Br Med J* 1997;**315**:1371–1374.
- Evers JLH. Female subfertility. *Lancet* 2002;**9327**:151–159.
- Goldenberg RL, Hauth JC, Andrews WW. Mechanisms of disease— intrauterine infection and preterm delivery. *N Engl J Med* 2000;**20**:1500–1507.
- Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet* 2008;**9606**:75–84.
- Habbema JDF, Collins J, Leridon H, Evers JLH, Lunenfeld B, te Velde ER. Towards less confusing terminology in reproductive medicine: a proposal. *Hum Reprod* 2004;**7**:1497–1501.
- Helmerhorst FM, Perquin DAM, Donker D, Keirse MJNC. Perinatal outcome of singletons and twins after assisted conception: a systematic review of controlled studies. *Br Med J* 2004;**7434**:261–264.
- Henriksen TB, Baird DD, Olsen J, Hedegaard M, Secher NJ, Wilcox AJ. Time to pregnancy and preterm delivery. *Obstet Gynecol* 1997;**4**:594–599.
- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–1558.
- Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Br Med J* 2003;**7414**:557–560.
- Hill GA, Bryan S, Herbert ICM, Shah DM, Wentz AC. Complications of pregnancy in infertile couples: routine treatment versus assisted reproduction. *Obstet Gynecol* 1990;**5**:790–794.
- Hull MGR, Glazener CMA, Kelly NJ, Conway DI, Foster PA, Hinton RA, Coulson C, Lambert PA, Watt EM, Desai KM. Population study of causes, treatment, and outcome of infertility. *Br Med J* 1985;**6510**:1693–1697.
- Jaques AM, Amor DJ, Baker HWG, Healy DL, Ukoumunne OC, Breheny S, Garrett C, Halliday JL. Adverse obstetric and perinatal outcomes in subfertile women conceiving without assisted reproductive technologies. *Fertil Steril* 2010;**7**:2674–2679.
- Joffe M, Li Z. Association of time to pregnancy and the outcome of pregnancy. *Fertil Steril* 1994;**1**:71–75.
- Juang CM, Chou P, Yen MS, Twu NF, Horng HC, Hsu WL. Adenomyosis and risk of preterm delivery. *Br J Obstet Gynaecol* 2007;**2**:165–169.
- Kramer MS, Goulet L, Lydon J, Séguin L, McNamara H, Dassa C, Platt RW, Fong Chen M, Gauthier H, Genest J *et al.* Socio-economic disparities in preterm birth: causal pathways and mechanisms. *Paediatr Perinat Epidemiol* 2001a;**5**:104–123.
- Kramer MS, Platt RW, Wen SW, Joseph KS, Allen A, Abrahamowicz M, Blondel B, Breart G, for the Fetal/Infant Health Study Group of the

- Canadian Perinatal Surveillance System. A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics* 2001b;**2**:e35.
- McDonald SD, Han Z, Mulla S, Murphy KE, Beyene J, Ohlsson A. Preterm birth and low birth weight among in vitro fertilization singletons: a systematic review and meta-analyses. *Eur J Obstet Gynecol Reprod Biol* 2009;**2**:138–148.
- McElrath TF, Wise PH. Fertility therapy and the risk of very low birth weight. *Obstet Gynecol* 1997;**4**:600–605.
- Pritts EA, Parker WH, Olive DL. Fibroids and infertility: an updated systematic review of the evidence. *Fertil Steril* 2009;**4**:1215–1223.
- Raatikainen K, Harju M, Hippelainen M, Heinonen S. Prolonged time to pregnancy is associated with a greater risk of adverse outcomes. *Fertil Steril* 2010a;**3**:1148–1151.
- Raatikainen K, Harju M, Hippelainen M, Heinonen S. Prolonged time to pregnancy is associated with a greater risk of adverse outcomes. *Fertil Steril* 2010b;**3**:1148–1151.
- Ranta JK, Raatikainen K, Romppanen J, Pulkki K, Heinonen S. Increased time-to-pregnancy and first trimester Down's syndrome screening. *Hum Reprod* 2010;**2**:412–417.
- Reddy UM, Wapner RJ, Rebar RW, Tasca RJ. Infertility, assisted reproductive technology, and adverse pregnancy outcomes: executive summary of a national institute of child health and human development workshop. *Obstet Gynecol* 2007;**4**:967–977.
- Romundstad LB, Romundstad PR, Sunde A, von Doring V, Skjaerven R, Gunnell D, Vatten LJ. Effects of technology or maternal factors on perinatal outcome after assisted fertilisation: a population-based cohort study. *Lancet* 2008;**9640**:737–743.
- Saunders DM, Mathews M, Lancaster PAL. The Australian register: current research and future role. A preliminary report. *Ann NY Acad Sci* 1988; **21**:7–21.
- Stang A. Critical evaluation of the Newcastle-Ottawa Scale for the assessment of the quality of nonrandom studies in meta-analyses. *Eur J Epidemiol* 2010;**9**:603–605.
- StataCorp. *Stata Statistical Software: Release 11*. College station, TX: Statacorp LP, 2009.
- Tan S, Doyle P, Campbell S, Beral V, Rizk B, Brinsden P, Mason B, Edwards R. Obstetric outcome of in vitro fertilization pregnancies compared with normally conceived pregnancies. *Am J Obstet Gynecol* 1992;**3**:778–784.
- Thomson F, Shanbhag S, Templeton A, Bhattacharya S. Obstetric outcome in women with subfertility. *Br J Obstet Gynaecol* 2005; **5**:632–637.
- Tuck S, Yudkin P, Turnbull A. Pregnancy outcome in elderly primigravidae with and without a history of infertility. *Br J Obstet Gynaecol* 1988; **3**:230–237.
- Varma TR, Patel RH, Bhathenia RK. Outcome of pregnancy after infertility. *Acta Obstet Gynecol Scand* 1988;**2**:115–119.
- Wang JX, Norman RJ, Kristiansson P. The effect of various infertility treatments on the risk of preterm birth. *Hum Reprod* 2002; **4**:945–949.
- Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Department of Epidemiology and Community Medicine, University of Ottawa, Canada. http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm.
- Wisborg K, Ingerslev H, Henriksen T. IVF and stillbirth: a prospective follow-up study. *Hum Reprod* 2010;**5**:1312–1316.
- Zhu JL, Obel C, Hammer Bech B, Olsen J, Basso O. Infertility, infertility treatment, and fetal growth restriction. *Obstet Gynecol* 2007;**6**: 1326–1334.