

# Assisted reproductive technologies and the risk of birth defects—a systematic review

Michèle Hansen<sup>1,3</sup>, Carol Bower<sup>1</sup>, Elizabeth Milne<sup>1</sup>, Nicholas de Klerk<sup>1</sup>  
and Jennifer J.Kurinczuk<sup>2</sup>

<sup>1</sup>Centre for Child Health Research, The University of Western Australia Telethon Institute for Child Health Research, West Perth, Western Australia 6872, Australia and <sup>2</sup>The National Perinatal Epidemiology Unit, University of Oxford, OX3 7LF, UK

<sup>3</sup>To whom correspondence should be addressed at: Telethon Institute for Child Health Research, PO Box 855, West Perth, Western Australia 6872. E-mail: michele@ichr.uwa.edu.au

**BACKGROUND:** The risk of birth defects in infants born following assisted reproductive technology (ART) treatment is a controversial question. Most publications examining the prevalence of birth defects in ICSI and IVF infants compared to spontaneously conceived infants have serious methodological limitations; despite this, most researchers have concluded that there is no increased risk. **METHODS:** We carried out a systematic review to identify all papers published by March 2003 with data relating to the prevalence of birth defects in infants conceived following IVF and/or ICSI compared with spontaneously conceived infants. Independent expert reviewers used criteria defined *a priori* to determine whether studies were suitable for inclusion in a meta-analysis. Fixed effects meta-analysis was performed for all studies and reviewer-selected studies. **RESULTS:** Twenty-five studies were identified for review. Two-thirds of these showed a 25% or greater increased risk of birth defects in ART infants. The results of meta-analyses of the seven reviewer-selected studies and of all 25 studies suggest a statistically significant 30–40% increased risk of birth defects associated with ART. **CONCLUSIONS:** Pooled results from all suitable published studies suggest that children born following ART are at increased risk of birth defects compared with spontaneous conceptions. This information should be made available to couples seeking ART treatment.

*Key words:* assisted reproductive technology/congenital malformations/IVF/meta-analysis/systematic review

## Introduction

It is well established that infants conceived following *in vitro* fertilization (IVF) and intracytoplasmic sperm injection (ICSI) are more likely to be born preterm, of low birth weight and to be a twin or higher order multiple than spontaneously conceived infants (Beral and Doyle, 1990; Helmerhorst *et al.*, 2004; Jackson *et al.*, 2004). The evidence relating to the risk of birth defects is less clear.

The publication of our paper (Hansen *et al.*, 2002) reporting a statistically significant 2-fold increased risk of major birth defects in children conceived following IVF and ICSI in Western Australia generated much discussion (Barlow, 2002; Lambert, 2002; Mitchell, 2002; Schultz and Williams, 2002; Winston and Hardy, 2002; Kovalevsky *et al.*, 2003; Powell, 2003). Despite other reports of an increased risk of birth defects following assisted reproductive technologies (ART), most authors have been reassuring, often dismissing increased risk estimates because they were not statistically significant (Morin *et al.*, 1989; Sutcliffe *et al.*, 1995; Verlaenen *et al.*, 1995; Isaksson *et al.*, 2002; Zadori *et al.*, 2003).

In order to evaluate published data on birth defects and ART systematically, we carried out an extensive literature

search to identify all papers with data relating to the prevalence of birth defects in infants conceived following IVF and/or ICSI compared to spontaneously conceived infants.

Our aims were first, to summarize the results of each study, and to independently identify those studies considered to have used sound epidemiological methods; and second, to calculate pooled estimates of the risk of birth defects, using quantitative meta-analysis of methodologically sound studies, and secondarily, of all studies. We have followed the MOOSE guidelines (Stroup *et al.*, 2000) for reporting meta-analyses of observational studies.

## Materials and methods

### Literature search strategy

Using a broad combination of search terms (Table I), we performed a computerized literature search of Medline, Embase and Current Contents databases. Medline searches were restricted to literature published from 1978 (since the first IVF child was born in that year). Embase and Current Contents searches were limited to the coverage years of these databases (Embase from 1988 and Current Contents from 1993). The search strategy was written in Ovid, then

**Table I.** Literature search strategy

All combinations of terms in the first column with terms in the second column (e.g. IVF AND birth defect?<sup>a</sup>; IVF AND record linkage etc.)

IVF	Birth defect?
<i>In vitro</i> fertili?ation <sup>b</sup>	Congenital malformation?
<i>In-vitro</i> fertili?ation	Congenital abnormalit? <sup>d</sup>
ICSI	Hospital? <sup>c</sup>
Intracytoplasmic sperm injection	Follow-up
Assisted reproduction	Health and child
Assisted reproductive techn? <sup>c</sup>	Record linkage
Infertility treatment?	

<sup>a</sup>'defect?' will find defect or defects;

<sup>b</sup>fertili?ation will find fertilisation or fertilization;

<sup>c</sup>'techn?' will find technology, technologies, technique(s) etc.;

saved and run in each database at the end of March 2003 (see Table I). We also reviewed the reference lists of all identified studies and review articles to search for additional references.

The criteria for inclusion in the review stage were kept broad so that crude birth defect data were acceptable, as was an absence of statistical analysis. Fifty-one papers were identified using the search strategy. These were all reports of individual studies and did not include any review papers.

We specifically searched for papers that compared birth defects in IVF or ICSI infants and a spontaneously conceived comparison group. Papers such as the large Belgian series (e.g. Bonduelle *et al.*, 2002) that assessed children born following one ART technique compared to another were therefore not included, as were papers that compared birth defect data for an ART group to data from birth defect registries in other countries (e.g. Friedler *et al.*, 1992). Also not included were papers that reported comparisons based on a single type of birth defect.

### Exclusions

Where more than one paper dealt with essentially the same group of infants, the paper with the more detailed birth defect information was selected. Papers that included a larger group of ART infants

were selected in preference to those containing only a subset of the same infants. Twenty-six of the 51 studies were excluded prior to expert review for the reasons shown in Table II.

Papers in languages other than English were not excluded in the search strategy; however, we found only one such paper with relevant data (Berg *et al.*, 2001), which in fact reported the same information as an English language paper (Ericson and Kallen, 2001).

### Independent assessment

Seven independent expert reviewers with postgraduate qualifications in epidemiology reviewed the studies identified. The reviewers were blinded to identifying information for each study, and were asked to abstract information onto a standard data extraction sheet and to complete a questionnaire relating to study methodology. These forms (available from the authors on request) were based on those used previously in a large Australian meta-analysis (English *et al.*, 1995). Six reviewers assessed between three and five papers, and one reviewer assessed twelve.

Reviewers were asked to extract both crude and adjusted odds ratio estimates from each study. Where an odds ratio estimate was not provided, reviewers recorded the number of infants with and without birth defects by method of conception. They also recorded information about the study design, methods, birth defect definition and adjustment for confounders. Finally, reviewers were asked whether they thought the paper was of adequate quality to be included in a meta-analysis. In making their decision, reviewers were asked in particular to consider sample size; whether the same method of assessment of birth defects had been used in exposed and unexposed infants; whether the investigators were blinded to conception status; whether the intensity of surveillance differed between the groups; and whether data were matched or adjusted for potential confounders in the analysis. We made the decision not to contact study authors as we thought it more appropriate that the results of our review were based on the available published information about each study. This is the information on which practitioners have been basing their advice to potential patients.

A subset of 11 randomly selected papers was reviewed by two reviewers to allow assessment of inter-reviewer variation. All reviewers extracted the same birth defect data from each paper;

**Table II.** List of studies excluded from the systematic review ( $n = 26$ )

Reason for exclusion	Reference(s)
Overlapping data	(Lambalk and Van Hooff, 2001) substantial overlap with (Anthony <i>et al.</i> , 2002) (Leslie <i>et al.</i> , 1998) overlap with (Bowen <i>et al.</i> , 1998) (Bergh <i>et al.</i> , 1999; Ericson <i>et al.</i> , 2002; Stromberg <i>et al.</i> , 2002; Wennerholm <i>et al.</i> , 1997, 1998, 2000) overlap with (Ericson and Kallen, 2001) (Berg <i>et al.</i> , 2001) same data as (Ericson and Kallen, 2001) (Dhont <i>et al.</i> , 1997) overlap with (Dhont <i>et al.</i> , 1999) (Sutcliffe <i>et al.</i> , 1999) overlap with (Sutcliffe <i>et al.</i> , 2001) (Rizk <i>et al.</i> , 1991) excerpt of (Beral and Doyle, 1990) (Ludwig and Diedrich, 2002) data are excerpt of (Ludwig and Katalinic, 2002) (Cadman <i>et al.</i> , 1999) same data as (D'Souza <i>et al.</i> , 1997) (Yeh <i>et al.</i> , 1990; FIVNAT, 1995; Minakami <i>et al.</i> , 1998; Harrison <i>et al.</i> , 1995; Cederblad <i>et al.</i> , 1996; Chou <i>et al.</i> , 2002; Nuojua-Huttunen <i>et al.</i> , 1999) (Lahat <i>et al.</i> , 1999) (10 infants); (Ron-El <i>et al.</i> , 1994) (32 infants); (Brandes <i>et al.</i> , 1992) (116 infants) (Petersen <i>et al.</i> , 1995)
Insufficient data to calculate an odds ratio estimate	
No birth defects found in either group <sup>a</sup>	
Inappropriate comparison group—included both spontaneously conceived infants and those born to infertile patients following fertility treatments other than IVF/ICSI	
Total number of malformations reported rather than number of children with malformations	(Saunders <i>et al.</i> , 1996)

<sup>a</sup>It is not possible to calculate an odds ratio estimate with zero cells, therefore these studies could not contribute to the meta-analysis—see formula for pooled estimate in Materials and methods.

however, some disagreement arose over the suitability of three papers for inclusion in the meta-analysis. These papers were reassessed by a final independent arbiter, again blinded to study authors. His decision was considered final.

### Calculating a pooled estimate: meta-analysis

If effect measures were not reported in a paper, we calculated odds ratios and their 95% confidence intervals from the raw data. Where more than one odds ratio was available from a particular study (e.g. an adjusted odds ratio estimate as well as a crude estimate), all of these were extracted and used in relevant subgroup analyses (e.g. of crude data only). However, for the main analysis involving all studies, we used adjusted odds ratio estimates in preference to crude estimates; estimates of major birth defect risk in preference to major and minor defects combined; major and minor defects combined in preference to minor defects only; estimates relating to all infants in preference to singletons only; and singletons only in preference to twins only.

Where a study provided a number of odds ratio estimates adjusted for different factors, we used an *a priori* list of rules to determine which odds ratio to include. For example, estimates adjusted for maternal age and parity were used in preference to estimates also adjusted for plurality, as we consider it inappropriate to adjust for factors that may lie on the causal pathway (Rothman and Greenland, 1998). We also felt it was inappropriate to adjust for duration of involuntary childlessness, as it is almost synonymous with exposure.

Where a study provided birth defect data for ICSI and IVF infants separately compared to a single spontaneous conception comparison group (e.g. Bowen *et al.*, 1998; Hansen *et al.*, 2002), the data were pooled to form one odds ratio for ICSI + IVF vs spontaneous conception, to avoid double counting of the spontaneous conception comparison group.

We used precision-based weighting and a fixed effects model to obtain pooled estimates of the odds ratio (OR) for all studies, and for those studies assessed by the independent reviewers to be suitable for inclusion in a meta-analysis (Kleinbaum *et al.*, 1982).

The formula for the pooled estimate of the OR from  $N$  studies using this method is:

$$OR = \exp \left[ \frac{\sum_{i=1}^N W_i \ln(OR_i)}{\sum_{i=1}^N W_i} \right] \quad \text{where } W_i = 1/\text{Var}(\ln(OR_i))$$

where  $\text{Var}(\ln(OR))$  was calculated from the published or calculated confidence intervals.

A 95% confidence interval around the pooled estimate is given by:

$$\left( \exp \left( \ln(OR) - 1.96/\sqrt{\sum W_i} \right), \exp \left( \ln(OR) + 1.96/\sqrt{\sum W_i} \right) \right)$$

and a test for heterogeneity in the OR estimates across studies using the chi-square statistic is given by:

$$\chi^2_{N-1} = \sum_{i=1}^N W_i [\ln(OR_i) - \ln(OR)]^2$$

which has  $N - 1$  degrees of freedom.

We chose to use a fixed effects model since random effects models tend to give more importance to smaller studies within a set, and smaller studies are more likely to suffer methodological limitations (Elwood, 1998). However, where the results of studies used

to estimate a pooled estimate were significantly heterogeneous ( $P < 0.10$ ), the pooled estimate from a random effects model is also reported for comparison purposes. As birth defects are rare, we assumed equivalence of the odds ratio and the relative risk.

### Sensitivity analyses and publication bias

In order to investigate heterogeneity between studies, we plotted the odds ratio estimate with its 95% confidence interval for each study, together with the pooled estimate, in forest plots. We then examined the effect on the pooled odds ratio estimate of excluding obvious outliers. We also examined the relative weights attributed to different studies. Recalculating a pooled estimate excluding studies with high weight allowed us to determine how sensitive the combined estimate was to any one study or group of studies.

Sub-group analyses were used to investigate differences in study design and their effect on the pooled odds ratio estimate. For example, studies that included adjusted or matched data were included in one sub-group analysis, while studies that included only crude unadjusted data were included in another. A funnel plot was used to assess publication bias.

### Number needed to harm

The number of patients needed to be treated for one additional patient to be harmed (NNTH) refers to a method of converting the odds ratio estimates derived from case-control and cohort studies into a more intuitively understandable quantity (Bjerre and LeLorier, 2000). It is an analogous concept to the more widely known 'number needed to treat' (NNT) and 'number needed to harm' (NNH) developed for randomized controlled trials. In the context of this study, the NNTH relates to the number of children that would need to be conceived by ART for one additional child to be born with a birth defect. We have calculated the NNTH for a range of baseline birth defect prevalences based on the pooled estimates derived from this study.

The formula for calculating the NNTH is:

$$NNTH = \frac{1}{(OR - 1)UER}$$

where  $OR$  is the odds ratio provided by the case-control, cohort study (or meta-analysis) and  $UER$  is the unexposed event rate (in this case the baseline prevalence of birth defects in a given population).

### Results

Twenty-five papers with no data overlap were assessed by external reviewers for possible inclusion in a meta-analysis. Of the 25 studies reviewed, 18 originated from Europe, four from Australia, two from the Middle East and one from the United States. The earliest reviewed study was published in 1989 and the latest in 2003, with over half the studies published in the last 4 years. The size of the ART group in each study ranged from 32 to 9111 infants. Seventeen of 25 papers (68%) had an ART group comprising <500 infants. Most studies included children conceived using standard IVF or a mixture of standard IVF and ICSI or GIFT. Only five studies reported results separately for children conceived by ICSI (Bowen *et al.*, 1998; Sutcliffe *et al.*, 2001, 2003; Hansen *et al.*, 2002; Ludwig and Katalinic, 2002).

The presence of birth defects was assessed only at birth in the majority of the studies reviewed (64%). Birth defects were the primary outcome measure (20%) (Ericson and Kallen, 2001; Anthony *et al.*, 2002; Hansen *et al.*, 2002; Ludwig and Katalinic, 2002; Zadori *et al.*, 2003), or one of a number of main outcome measures (48%) (Morin *et al.*, 1989; Beral and Doyle, 1990; Sutcliffe *et al.*, 1995, 2001, 2003; Nassar *et al.*, 1996; D'Souza *et al.*, 1997; Fisch *et al.*, 1997; Bowen *et al.*, 1998; Dhont *et al.*, 1999; Westergaard *et al.*, 1999; Koivurova *et al.*, 2002) in 68% of studies reviewed. The remainder (32%) were not designed specifically to assess birth defect risk.

Only seven studies were considered by the external reviewers to be appropriate for inclusion in a meta-analysis (Morin *et al.*, 1989; Dhont *et al.*, 1999; Westergaard *et al.*, 1999; Ericson and Kallen, 2001; Hansen *et al.*, 2002; Isaksson *et al.*, 2002; Koivurova *et al.*, 2002). The majority of these (5/7) were population-based studies with a clear definition of a birth defect. Most had a large sample size and birth defects were ascertained without knowledge of conception status in all seven studies. All reviewer-selected studies included data adjusted or matched for maternal age and parity. A number of other factors such as infant sex (Morin *et al.*, 1989; Dhont *et al.*, 1999; Hansen *et al.*, 2002; Koivurova *et al.*, 2002), year of birth (Morin *et al.*, 1989; Westergaard *et al.*, 1999; Ericson and Kallen, 2001; Isaksson *et al.*, 2002; Koivurova *et al.*, 2002), and plurality (Morin *et al.*, 1989; Dhont *et al.*, 1999; Westergaard *et al.*, 1999; Ericson and Kallen, 2001; Isaksson *et al.*, 2002; Koivurova *et al.*, 2002) were adjusted or matched for in some of these studies. As stated earlier, when a study provided a number of odds ratio estimates adjusted for different factors (e.g. Ericson and Kallen, 2001), we chose odds ratios that were not adjusted for plurality, as plurality may lie on the causal pathway. The methodological limitations identified in the remaining 18 studies are listed in Table III.

Throughout this paper, we present results for the reviewer-selected studies ( $n = 7$ ) and all studies combined ( $n = 25$ ). A total of 28 638 ART children were included in the 25 studies; over half (56%) came from the seven reviewer-selected studies.

For the seven reviewer-selected studies, the pooled odds ratio was 1.40 (95% CI 1.28–1.53), indicating a significantly increased risk of birth defects in children born following assisted reproductive technologies (Table IV). The individual point estimates for these studies ranged from 1.04 to 2.27 (Figure 1). The pooled odds ratio for all 25 studies was 1.29 (95% CI 1.21–1.37) (Table IV). The range of point estimates was 0.67 to 15.39 (Figure 1). The odds ratio estimate was  $\geq 1.25$  in 16 of the 25 studies (64%), although most of these were not statistically significant.

### Sensitivity analyses and publication bias

The heterogeneity statistic for the pooled estimate of the seven reviewer-selected studies was not statistically significant ( $P = 0.12$ ), indicating that the pooled odds ratio of 1.40 (95% CI 1.28–1.53) is an adequate representation of this set of studies (Table IV). However, a large Swedish study (Ericson and Kallen, 2001) contributed 72.8% of the total weight. Removal of this study from the analysis had very little effect on the pooled odds ratio (OR = 1.42; 95% CI 1.20–1.69). The smallest reviewer-selected study (Morin *et al.*, 1989) contributed the largest odds ratio estimate (2.27) (top of Figure 1). Removal of this study had no effect on the pooled estimate.

The heterogeneity statistic for the pooled estimate of all 25 studies was statistically significant at the conservative level of  $P < 0.10$ , indicating greater between-study heterogeneity than for the reviewer-selected studies. We examined the effect of excluding the three studies with the largest odds ratio estimates. One of these (D'Souza *et al.*, 1997) was excluded by the independent reviewers because

**Table III.** Methodological limitations of papers identified by external reviewers leading to their exclusion from the meta-analysis

Methodological limitations	Reference(s)
Inadequate or absent birth defect definition	(Fisch <i>et al.</i> , 1997; Tanbo <i>et al.</i> , 1995; Anthony <i>et al.</i> , 2002)
Too few details of study methodology: e.g. absent or inadequate information on how, when, by whom birth defects assessed and/or not able to tell whether ascertainment blind	(Koudstaal <i>et al.</i> , 2000a,b; Zadori <i>et al.</i> , 2003; Addor <i>et al.</i> , 1998; D'Souza <i>et al.</i> , 1997; Fisch <i>et al.</i> , 1997; Nassar <i>et al.</i> , 1996; Tanbo <i>et al.</i> , 1995; Zuppa <i>et al.</i> , 2001; Wang <i>et al.</i> , 2002)
Selection process for inclusion in study poorly described or raises doubts as to validity of groups included	(Zadori <i>et al.</i> , 2003; Verlaenen <i>et al.</i> , 1995; Sutcliffe <i>et al.</i> , 1995, 2001, 2003; Ludwig and Katalinic, 2002; Anthony <i>et al.</i> , 2002)
Overmatching such that cases discarded because sufficient matches not found	(Koudstaal <i>et al.</i> , 2000a,b)
Different method of birth defect assessment used for exposed and unexposed	(Beral and Doyle, 1990; Verlaenen <i>et al.</i> , 1995)
Birth defect assessors not blind to conception status	(Verlaenen <i>et al.</i> , 1995; Sutcliffe <i>et al.</i> , 1995, 2001, 2003; Bowen <i>et al.</i> , 1998; Ludwig and Katalinic, 2002)
Birth defects reported only for survivors to infancy/childhood	(Sutcliffe <i>et al.</i> , 1995, 2001, 2003; D'Souza <i>et al.</i> , 1997)
Lack of suitable comparison data	(Beral and Doyle, 1990; D'Souza <i>et al.</i> , 1997)
Ascertainment bias (greater surveillance of IVF group)	(Verlaenen <i>et al.</i> , 1995)
Large loss to follow-up (10–33%)	(Koudstaal <i>et al.</i> , 2000a,b; Bowen <i>et al.</i> , 1998; Sutcliffe <i>et al.</i> , 2001)
Length of follow-up differs between groups	(Beral and Doyle, 1990; Bowen <i>et al.</i> , 1998; D'Souza <i>et al.</i> , 1997; Ludwig and Katalinic, 2002)
No information provided about those who refused to participate in study	(Sutcliffe <i>et al.</i> , 1995, 2001, 2003; Bowen <i>et al.</i> , 1998; Ludwig and Katalinic, 2002)
No information on record linkage, e.g. insufficient information about methods, completeness of linkage etc.	(Wang <i>et al.</i> , 2002)

**Table IV.** Study characteristics for (a) reviewer-selected studies ( $n = 7$ ) and (b) remaining studies ( $n = 18$ ). Pooled odds ratio estimates for reviewer-selected studies and all studies combined ( $n = 25$ )

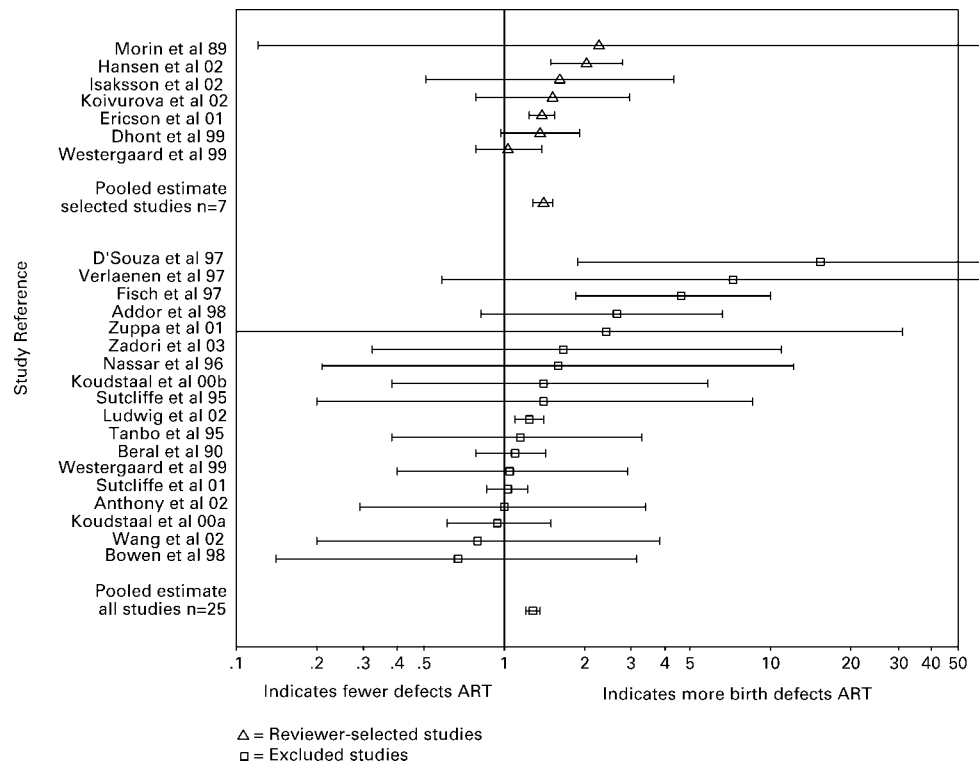
Authors and publication year	Location	Population vs clinic-based sample	Number exposed	ART treatment	Plurality	Defects assessed	Age /Time assessment	Adjusted, matched or crude data	OR	95% CI	Weight
(a) Reviewer-selected studies ( $n = 7$ ): OR = 1.40 (95% CI 1.28–1.53). $P$ -value for heterogeneity test = 0.120.											
Dhont <i>et al.</i> , 1999	Belgium	Population	3048	IVF, GIFT	Singletons	All	Birth	Matched	1.36	0.97–1.93	32.47
Ericson and Kallen, 2001	Sweden	Population	9111	IVF, ICSI	All <sup>a</sup>	All	Birth	Adjusted	1.39	1.25–1.54	353.01
Hansen <i>et al.</i> , 2002	Australia	Population	1138	IVF, ICSI	All	Maj	12 months	Adjusted	2.04	1.50–2.77	40.84
Isaksson <i>et al.</i> , 2002	Finland	Clinic	109	IVF, ICSI	All <sup>b</sup>	Maj	Birth	Matched	1.61	0.51–4.33	3.36
Koivurova <i>et al.</i> , 2002	Finland	Population	304	IVF	All <sup>b</sup>	All	36 months	Matched	1.53	0.79–2.93	8.94
Morin <i>et al.</i> , 1989	USA	Clinic	83	IVF	All <sup>b</sup>	Maj	12–30 mths	Matched	2.27	0.12–135.45	0.31
Westergaard <i>et al.</i> , 1999	Denmark	Population	2245	IVF, ICSI	All <sup>b</sup>	All	Birth	Matched	1.04	0.78–1.39	46.03
(b) Remaining studies ( $n = 18$ ). OR for all studies combined ( $n = 25$ ) = 1.29 (95% CI 1.21–1.37). $P$ -value for heterogeneity test = 0.060.											
Addor <i>et al.</i> , 1998	Switzerland	Population	82	IVF + other	All	All	Birth	Crude	2.64	0.82–6.56	3.55
Anthony <i>et al.</i> , 2002	The Netherlands	Population	4224	IVF, ICSI	All	All	Birth	Adjusted	1.03	0.86–1.23	120.01
Beral and Doyle, 1990 <sup>c</sup>	United Kingdom	Population	1581	IVF, GIFT	All	Maj	Birth	Adjusted	1.11	0.79–1.43	43.64
Bowen <i>et al.</i> , 1998	Australia	Clinic	173	IVF, ICSI	All	Maj	12 months	Crude	0.80	0.20–3.85	1.76
D'Souza <i>et al.</i> , 1997	United Kingdom	Clinic	278	IVF	All	Maj	48 months	Matched	15.39	1.90–infinity	0.09
Fisch <i>et al.</i> , 1997	Israel	Clinic	100	IVF	All	Maj	Birth	Crude	4.60	1.86–9.95	5.46
Koudstaal <i>et al.</i> , 2000a	The Netherlands	Clinic	307	IVF	Singletons	All	Birth	Matched	1.00	0.29–3.39	2.54
Koudstaal <i>et al.</i> , 2002	The Netherlands	Clinic	192	IVF	Twins	All	Birth	Matched	1.42	0.38–5.76	2.08
Ludwig <i>et al.</i> , 2002	Germany	Population	3372	ICSI	All	Maj	Birth <sup>d</sup>	Crude	1.25	1.11–1.40	285.22
Nassar <i>et al.</i> , 1996	Egypt	Clinic	128	IVF	All	Maj	Birth	Crude	1.60	0.21–12.12	0.93
Sutcliffe <i>et al.</i> , 1995	United Kingdom	Clinic	91	IVF	All	Maj	8–48 months	Matched	1.40	0.20–8.50	1.09
Sutcliffe <i>et al.</i> , 2001	United Kingdom	Clinic	208	ICSI	Singletons	Maj	12–24 months	Matched	1.06	0.39–2.92	3.92
Sutcliffe <i>et al.</i> , 2003	Australia	Clinic	56	ICSI	Singletons	Maj	13–15 months	Matched	0.67	0.14–3.15	1.59
Tanbo <i>et al.</i> , 1995	Norway	Clinic	355	IVF, GIFT, tubal transfer, insemination	Singletons	All	Birth	Matched	1.16	0.38–3.30	3.29
Verlaenen <i>et al.</i> , 1995	Belgium	Clinic	140	IVF	Singletons	Min	Birth	Matched	7.15	0.59–infinity	0.07
Wang <i>et al.</i> , 2002	Australia	Clinic	1019	IVF, ICSI, GIFT	Singletons	All	Birth	Matched	0.95	0.61–1.49	19.27
Zadori <i>et al.</i> , 2003	Hungary	Clinic	262	IVF	All <sup>b</sup>	Maj	1 month	Matched	1.68	0.32–10.92	1.23
Zuppa <i>et al.</i> , 2001	Italy	Clinic	32	IVF, GIFT, clomiphene etc.	Twins	Maj	Birth	Crude	2.42	0.04–31.08	0.35

<sup>a</sup>This study provided an OR estimate adjusted for plurality, however this estimate was not selected for inclusion in the meta-analysis.

<sup>b</sup>Matched for plurality

<sup>c</sup>Comparison data from the Liverpool Congenital Malformations Registry were used for OR calculation.

<sup>d</sup>ICSI group assessed at ~1 month, spontaneous conception group assessed at birth.



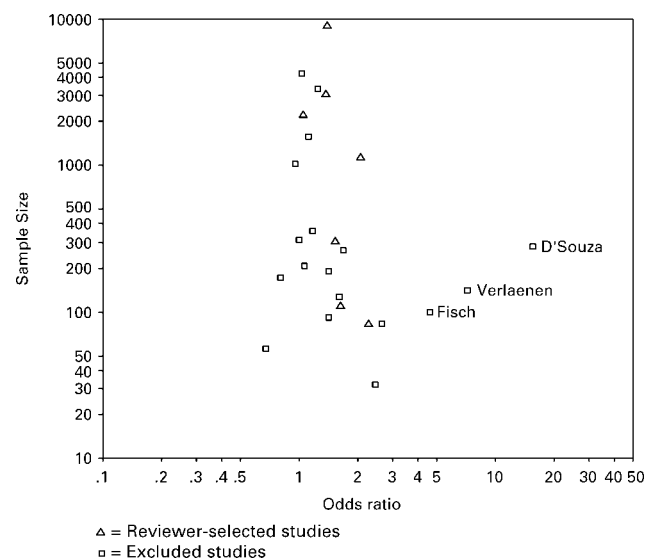
**Figure 1.** Individual odds ratio estimates from reviewer-selected studies (top portion of graph) and remaining studies (lower portion of graph) together with fixed pooled odds ratio estimates from meta-analyses combining reviewer-selected studies ( $n = 7$ ) and all studies ( $n = 25$ ).

the comparison group comprised only full-term singletons and were therefore less likely to have birth defects than the IVF group, which also included preterm infants and multiples. Furthermore, birth defects were assessed at 4 years of age, and neonatal deaths were excluded. A small Belgian study (Verlaenen *et al.*, 1995) included only 140 IVF singletons, and was not selected by the independent reviewers for inclusion in a meta-analysis because of the potential for bias in ascertainment of birth defects in the IVF group who underwent a series of ultrasound examinations that the comparison group did not. Finally, the independent reviewers excluded a study from Israel (Fisch *et al.*, 1997) because of its small sample size, inadequate birth defect definition and crude analyses. Further, the authors did not report whether the birth defect assessors were blinded to conception status. When these three studies were removed from the meta-analysis of all studies, the pooled odds ratio barely changed, but the between-study heterogeneity was much reduced ( $P = 0.21$ ).

The funnel plot of all 25 studies (Figure 2) was not symmetrical due to the three outlying studies described above.

### Sub-group analyses

The pooled odds ratio estimates remained elevated and statistically significant when we restricted our analyses to those studies that assessed major birth defects separately (reviewer-selected studies OR = 2.01; 95% CI 1.49–2.69; and all studies OR = 1.32; 95% CI 1.20–1.45); or defects in singleton births (reviewer-selected studies OR = 1.35; 95% CI 1.20–1.51; and all studies OR = 1.31; 95% CI 1.17–1.46).



**Figure 2.** Funnel plot of sample size against effect size.

In fact, the majority of sub-groups had odds ratios that were similar to or greater than the overall summary odds ratio estimates and all remained statistically significant (Table V). Since not all studies contributed data for sub-group analyses, the number of studies included varies.

The random effects model—used for sub-group analyses including crude data and all infants—gave rise to slightly increased pooled estimates with wider confidence intervals

Table V. Sub-group analyses: pooled odds ratios for reviewer-selected studies and all studies combined

	Number of studies	Pooled OR (fixed effects model)	95% CI	P-value for test of heterogeneity	Pooled OR (random effects model)	95% CI
Major defects						
Reviewer-selected	3	2.01	1.49–2.69	0.915		
All studies	15	1.32	1.20–1.45	0.104		
All infants (singletons + multiples)						
Reviewer-selected	7	1.50	1.38–1.62	0.008	1.53	1.26–1.85
All studies	17	1.36	1.28–1.45	<0.001	1.41	1.21–1.64
Singletons only						
Reviewer-selected	6	1.35	1.20–1.51	0.114		
All studies	15	1.31	1.17–1.46	0.477		
Adjusted/matched data						
Reviewer-selected	7	1.40	1.28–1.53	0.120		
All studies	19	1.29	1.19–1.39	0.153		
Crude data						
Reviewer-selected	3	1.60	1.48–1.72	0.003	1.76	1.41–2.18
All studies	10	1.49	1.40–1.58	<0.001	1.63	1.36–1.95
IVF only						
Reviewer-selected	3	1.90	1.41–2.54	0.770		
All studies	12	1.94	1.50–2.50	0.736		
ICSI only						
Reviewer-selected	1 <sup>a</sup>	[2.0]	1.3–3.2			
All studies	5	1.28	1.14–1.43	0.297		

<sup>a</sup>A single reviewer-selected study provided birth defect data for ICSI infants separately.

Table VI. Number needed to harm (NNTH) for different combinations of pooled odds ratio and baseline birth defect prevalence

Baseline prevalence <sup>a</sup> of birth defects in a given population	Primary meta-analysis, all defects		Sub-group analysis, major defects	
	All studies (pooled OR = 1.3)	Reviewer-selected studies (pooled OR = 1.4)	All studies (pooled OR = 1.3)	Reviewer-selected studies (pooled OR = 2.0)
1%	333	250	333	100
2%	167	125	167	50
3%	111	83	111	33
4%	83	62	83	25

<sup>a</sup>The baseline prevalence refers to the prevalence of all birth defects in a given population for the primary meta-analysis and the prevalence of major birth defects only for the sub-group analysis.

than the fixed effects models, but did not materially alter the inferences.

Number needed to harm

Table VI provides estimates of the number of children that need to be conceived by ART for one additional child to be born with a birth defect, based on the pooled estimates derived from the primary meta-analyses and the sub-group analyses involving major birth defects only. The NNTH is expressed for a range of different baseline birth defect prevalences. The results show that for a given odds ratio estimate, the number of children that need to be conceived by ART for one additional child to be born with a birth defect decreases with increasing prevalence of birth defects in the baseline population. For example, given the pooled odds ratio estimate from reviewer-selected studies (OR = 1.4), the NNTH ranges from 250 if the baseline prevalence of birth defects is 1% to 62 if the baseline prevalence is 4%. When the meta-analysis is restricted to those studies reporting major birth defects separately (OR = 2.0 for reviewer-selected studies), the NNTH decreases to 100 given a baseline prevalence of

major birth defects of 1%, or 25 given a baseline prevalence of 4%.

Discussion

To our knowledge, this is the first report to systematically review and pool epidemiological data assessing the risk of birth defects following assisted reproductive technologies. Although our literature search identified 25 papers with birth defect data in IVF/ICSI and spontaneously conceived infants, many had serious methodological limitations and nearly a third were not specifically designed to assess birth defect risk. A panel of independent, expert reviewers considered only seven of the 25 studies to be appropriate for inclusion in a meta-analysis. In conducting this systematic review and meta-analysis, our goal was to identify whether the published evidence suggests that infants born following ART treatment have a significantly increased risk of birth defects compared to spontaneously conceived infants. In doing this, we were very aware of the many differences in study design and methodology complicating this task. We have therefore presented

the results of meta-analyses for the seven reviewer-selected studies, for all 25 studies combined, and for a range of sub-group analyses, to allow the reader to assess all the available independent data. Our results suggest there is a statistically significant increased risk of birth defects in infants conceived using assisted reproductive technologies of the order of 30–40%. Two-thirds of the studies reviewed suggest an increased risk of birth defects of at least 25%. However, authors' conclusions have not always reflected this. Authors have generally placed too much emphasis on statistical significance, and have ignored or dismissed repeatedly raised odds ratio estimates (Kurinczuk, 2003; Kurinczuk *et al.*, 2004).

It has been argued that odds ratios are not intuitively understandable estimates of risk, and that the results of epidemiological studies need to be expressed in comprehensible terms if they are to be of practical use to clinicians and policy makers (Bjerre and LeLorier, 2000). We have therefore also expressed our results in terms of the number needed to harm (NNTH) which in this case equates to the number of children that need to be conceived by ART for one additional child to be born with a birth defect. For the purposes of counselling their patients, clinicians should calculate the NNTH based on a 30–40% increased risk of birth defects compared to the baseline birth defect prevalence for their population. Our pooled odds ratio from reviewer-selected studies suggests a NNTH of between 250 and 62, allowing for an underlying prevalence of birth defects between 1% and 4%.

### Validity and objectivity

Given that we have ourselves published in this area, and that our study included data relevant to this systematic review, our objectivity and thus the validity of our review may be questioned. Cognisant of this, we included other researchers in our study team who were experienced in the field of meta-analysis and were not involved in our previous study or in any research concerning assisted reproductive technology. We also sought external, blinded reviewers to assess each paper for inclusion in the meta-analysis. These reviewers were all trained in epidemiology with a Master and/or PhD qualification. The systematic nature of the review was ensured through the use of specifically designed data collection forms adapted from a previous study in an unrelated field (English *et al.*, 1995). Despite this, a reviewer's decision on whether to include or exclude a particular paper was, as in any systematic review, an opinion, based on their overall impression of the study as well as their answers to the structured questionnaire. Although we made a division between those papers considered by the independent reviewers to be more scientifically rigorous ( $n = 7$ ) and the remainder ( $n = 18$ ), we included all the studies ( $n = 25$ ) in our review and meta-analysis. Of those papers assessed by two reviewers, there was 73% agreement regarding which papers should be included/excluded from the meta-analysis. In a similar study using the same method of paper review, there was 83% agreement on inclusion/exclusion (English *et al.*, 1995), and in an experiment assessing epidemiologists' assessment of whether a particular exposure was likely to be

a cause of a condition, there was agreement in 63% of cases (Holman *et al.*, 2001). Thus, some level of disagreement between reviewers is not unexpected. Importantly, the results and inferences were consistent between the analyses of the smaller group of seven reviewer-selected papers and the whole group of 25. Excluding studies with high weight, or obvious outliers, had little impact on the pooled estimates, leading us to conclude that our results are robust.

The pooled odds ratio estimates derived from this study incorporate estimates of birth defect risk calculated from studies that variously collected information on major defects only, or major and minor defects combined, and examined the presence of defects in singletons only or in singletons and multiples combined, as well as in children conceived by IVF only, ICSI only or a combination of different ART techniques. However, the results of our sub-group analyses which pool estimates from studies including only major birth defects, singleton infants, or children conceived following IVF or ICSI show that the pattern of increased risk of birth defects in ART infants remains, regardless of the way in which these data are grouped.

The increased risk of birth defects in the sub-group analysis of major birth defects is of particular note given that major defects are less subject to problems of definition and under-reporting than minor defects. Similarly, the persistence of an elevated odds ratio when singletons only are considered indicates that the elevated pooled odds ratio estimates from our primary analyses, which include many studies where birth defects are reported for singletons and multiples combined, are not due to the presence of multiple births alone.

Although we acknowledge that there are significant differences between the ART techniques of ICSI and IVF, and have provided results for the subgroup analyses examining each of these techniques, we caution against inferring from these results that the ICSI technique involves a lower risk of birth defects than standard IVF. The sub-group analysis comprising those studies that examined birth defects in IVF infants includes many of the smaller studies in our review. The three outliers discussed in relation to the funnel plot also appear in this subgroup analysis. The ICSI sub-group included data on <4000 children, 85% of which were contributed by a single study (Ludwig and Katalinic, 2002). There is only one reviewer-selected study in this subgroup (Hansen *et al.*, 2002), and this study has a markedly increased odds ratio estimate (2.0) compared to the unselected studies (OR range 0.67 to 1.25). Although not the subject of this review, a number of studies have compared birth defect risk in IVF and ICSI infants and have not found significant differences (Bowen *et al.*, 1998; Bonduelle *et al.*, 2002; Hansen *et al.*, 2002; Place and Englert, 2003). We do not believe that this review, which specifically required data comparisons with spontaneously conceived infants, is the most appropriate for inferring differences between the two techniques. The most that can be said from these data are that they support the general trend of increased birth defect risk in ART infants.

The results of our meta-analyses are unlikely to be explained by significant residual confounding. All



reviewer-selected studies included data adjusted or matched for maternal age and parity. A number of other factors such as infant sex ( $n = 4$ ), year of birth ( $n = 5$ ), and plurality ( $n = 5$ ) were adjusted or matched for in some of these studies. Information on other potential confounders such as maternal exposure to toxins and socio-economic status were not available in the majority of studies. However, it has been suggested that women undertaking ART treatment are likely to have lower exposure to toxins such as alcohol and cigarettes, and to be of higher socio-economic status, than the general population of pregnant women (Simpson, 1996; Bergh *et al.*, 1999; Buitendijk, 1999). Therefore, a lack of adjustment for these factors is more likely to have led to an under rather than overestimate of the risk.

Whilst the funnel plot shows three small outlying studies with large odds ratio estimates, it seems unlikely that they would have been published because they had a large estimate of risk, since the authors themselves concluded there was no increase in birth defect risk in the ART group (Verlaenen *et al.*, 1995; D'Souza *et al.*, 1997; Fisch *et al.*, 1997). Aside from publication bias, another possible explanation for an asymmetrical funnel plot is the exaggeration of observed treatment effects in small studies of low quality (Sterne *et al.*, 2001). The methodological limitations of the three outlying studies have been described in the Results. They were not selected by the independent reviewers as appropriate for inclusion in our primary meta-analysis, and their exclusion from the meta-analysis of all studies had very little effect on the pooled odds ratio estimate.

### **Biological plausibility**

An excess risk of birth defects in IVF and ICSI infants is biologically plausible. Factors associated with ART treatment that may increase the risk of birth defects include the underlying causes of infertility in the couples seeking treatment; and factors associated with the IVF/ICSI procedures themselves, such as the freezing and thawing of embryos, the delayed fertilization of oocytes, culture media composition and the medications used to induce ovulation or for luteal phase support (Lancaster, 1985; Rizk *et al.*, 1991; Simpson, 1998; Buitendijk, 1999). Some researchers have argued that the excess risk of birth defects found in infants born following ART treatment may be due to the underlying infertility of the couples seeking treatment, rather than the treatments themselves (Ericson and Kallen, 2001; Ludwig and Diedrich, 2002; Lambert, 2003). It has recently been suggested, to address this question, that an appropriate comparison group for infants born following ART treatment would include children born to infertile couples who do eventually conceive spontaneously without IVF treatment (Kovalevsky *et al.*, 2003). In practice this comparison group would be difficult to identify (Schisterman *et al.*, 2003). An alternative may be to assess the prevalence of birth defects in the children of couples seeking ART treatment following failed vasectomy reversal or tubal ligation reversal, since these couples are not infertile due to an underlying disease process.

### **Implications of our results**

In order to counsel prospective patients effectively, IVF clinicians must assess all the available data on birth defect risk in infants born following ART treatment. This systematic review of published data has highlighted some important methodological issues and difficulties in comparing data across different studies in this field. In particular, the reader should consider the source of data on birth defects used in each study. Hospital notes are likely to underestimate birth defect risk, as are birth defect registries that do not actively promote and seek notifications beyond birth. Registers of assisted conception births that rely on clinic reports of birth defects may also underestimate birth defect risk, since many clinics do not follow patients beyond the immediate birth period or in some instances even to the end of a pregnancy. The pooling of major and minor defects may lead to less precise estimates of risk since the notification of minor defects is often incomplete. Despite this, only 60% of the studies in this review provided separate data on major birth defects. Finally, when assessing individual studies, the reader should consider whether the authors have used the same method of ascertaining birth defects for the groups being compared. If the two groups were followed for different lengths of time, underwent different birth defect assessments, or the defects were classified according to different birth defect classification systems, then the study results may not reflect true differences between the groups.

Our findings also have implications for future research in this field. Since it appears there is an increased risk of birth defects in infants born following ART treatment and we cannot yet identify the cause, it is now very important to collect detailed and accurate information about all treatments that couples have undergone and their underlying causes of infertility; and to be able to identify children born following ART procedures so they can be followed. Registers of ART births such as the statutory Reproductive Technology Register in Western Australia (The WA RTC, 1997) or similar collections in Sweden (Bergh *et al.*, 1999; Ericson and Kallen, 2001) and Denmark (Westergaard *et al.*, 1999) have obvious advantages. They enable record linkage research that is cost effective, does not require patient contact, minimizes losses to follow-up and provides larger sample sizes than are available through clinic-based studies.

### **Conclusions**

The results of our systematic review and meta-analyses suggest that infants born following ART treatment are at increased risk of birth defects, compared to spontaneously conceived infants. This information should be made available to couples seeking ART treatment. Larger, population-based studies are now needed to address questions of aetiology so we can provide better information for counselling prospective patients.

### **Postscript**

Since completion of our systematic review and meta-analysis, a second paper (Kozinszky *et al.*, 2003) has been published

by a group whose short communication was included in our systematic review (Zadori *et al.*, 2003). This paper includes a larger sample size and explains study methodology in more detail. It appears from this second paper that the 'IVF-ET' group referred to in the short communication comprise a mixture of infants conceived following standard IVF (59%), ovulation induction (32%) and intra-uterine insemination (9%). The results of this paper suggest a higher birth defect risk for the mixed exposure group (OR = 2.03, 95% CI 0.76–6.01) than the short communication included in our systematic review (OR = 1.68, 95% CI 0.32–10.92).

### Conflict of interest

None declared.

### Contributors

Michèle Hansen conducted the literature search, data analysis and drafted the report. All authors contributed to study design, analysis and interpretation of the data and all were involved in critical revision of the paper.

### Acknowledgements

We are greatly indebted to the independent expert reviewers and final independent arbiter who assessed papers for inclusion in our meta-analysis. Our thanks also to the peer reviewers whose comments helped us to improve our paper. Michèle Hansen is supported by a research grant (211930) from the National Health and Medical Research Council of Australia. Carol Bower is supported by a research fellowship (172303) from the National Health and Medical Research Council of Australia. Jennifer Kurinczuk is funded by a National Public Health Career Scientist Award from the Department of Health and NHS R&D (PHCS 022).

### References

- Addor V, Santos-Eggimann B, Fawer CL, Paccaud F and Calame A (1998) Impact of infertility treatments on the health of newborns. *Fertil Steril* 69, 210–215.
- Anthony S, Buitendijk SE, Dorrepaal CA, Lindner K, Braat DD and Den Ouden AL (2002) Congenital malformations in 4224 children conceived after IVF. *Hum Reprod* 17,2089–2095.
- Barlow DH (2002) The children of assisted reproduction—The need for an ongoing debate. *Hum Reprod* 17,1133–1134.
- Beral V and Doyle P (1990) Births in Great Britain resulting from assisted conception, 1978–87 (MRC Working Party on Children Conceived by In Vitro Fertilisation). *Br Med J* 300,1229–1233.
- Berg I, Finnstrom O and Nygren KG (2001) [Children born after in vitro fertilization in Sweden 1982–1997. Small, but significant increase of risks among test-tube children]. *Lakartidningen* 98,4020–4021, 4024–4025.
- Bergh T, Ericson A, Hillensjö T, Nygren KG and Wennerholm UB (1999) Deliveries and children born after in-vitro fertilisation in Sweden 1982–95: a retrospective cohort study. *Lancet* 354,1579–1585.
- Bjerre LM and LeLorier J (2000) Expressing the magnitude of adverse effects in case-control studies: 'the number of patients needed to be treated for one additional patient to be harmed'. *Br Med J* 320,503–506.
- Bonduelle M, Liebaers I, Deketelaere V, Derde MP, Camus M, Devroey P and Van Steirteghem A (2002) Neonatal data on a cohort of 2889 infants born after ICSI (1991–1999) and of 2995 infants born after IVF (1983–1999). *Hum Reprod* 17,671–694.
- Bowen JR, Gibson FL, Leslie GI and Saunders DM (1998) Medical and developmental outcome at 1 year for children conceived by intracytoplasmic sperm injection. *Lancet* 351,1529–1534.
- Brandes JM, Scher A, Itzkovits J, Thaler I, Sarid M and Gershoni-Baruch R (1992) Growth and development of children conceived by in vitro fertilization. *Pediatrics* 90,424–429.
- Buitendijk SE (1999) Children after in vitro fertilization. An overview of the literature. *Int J Technol Assess Health Care* 15,52–65.
- Cadman J, Richards B, D'Souza S, Lieberman BA, Buck P and Rivlin E (1999) A computerised study on the results of in-vitro fertilisation. *Stud Health Technol Inform* 68,343–346.
- Cederblad M, Friberg B, Ploman F, Sjöberg NO, Stjernqvist K and Zackrisson E (1996) Intelligence and behaviour in children born after in-vitro fertilization treatment. *Hum Reprod* 11,2052–2057.
- Chou HC, Tsao PN, Yang YS, Tang JR and Tsou KI (2002) Neonatal outcome of infants born after in vitro fertilization at National Taiwan University Hospital. *J Formosan Med Assoc* 101,203–205.
- Dhont M, De Neubourg F, Van der Elst J and De Sutter P (1997) Perinatal outcome of pregnancies after assisted reproduction: a case-control study. *J Assist Reprod Genet* 14,575–580.
- Dhont M, De Sutter P, Ruysinck G, Martens G and Bekaert A (1999) Perinatal outcome of pregnancies after assisted reproduction: a case-control study. *Am J Obstet Gynecol* 181,688–695.
- D'Souza SW, Rivlin E, Cadman J, Richards B, Buck P and Lieberman BA (1997) Children conceived by in vitro fertilisation after fresh embryo transfer. *Arch Dis Child* 76,F70–F74.
- Elwood JM (1998) Critical appraisal of epidemiological studies and clinical trials, 2nd edn, Oxford University Press, Oxford, UK, pp. 198–217.
- English DR, Holman CDJ, Milne E *et al.* (1995) The quantification of drug caused morbidity and mortality in Australia. In Commonwealth Department of Human Services and Health, Canberra, Australia.
- Ericson A and Kallen B (2001) Congenital malformations in infants born after IVF: a population-based study. *Hum Reprod* 16,504–509.
- Ericson A, Nygren KG, Otterblad Olausson P and Kallen B (2002) Hospital care utilization of infants born after IVF. *Hum Reprod* 17,929–932.
- Fisch B, Harel L, Kaplan B, Pinkas H, Amit S, Ovadia J, Tadir Y and Merlob P (1997) Neonatal assessment of babies conceived by in vitro fertilization. *J Perinatol* 17,473–476.
- FIVNAT (1995) Pregnancies and births resulting from in vitro fertilization: French national registry, analysis of data 1986 to 1990. *Fertil Steril* 64, 746–756.
- Friedler S, Mashiach S and Laufer N (1992) Births in Israel resulting from in-vitro fertilization/embryo transfer, 1982–1989: National Registry of the Israeli Association for Fertility Research. *Hum Reprod* 7, 1159–1163.
- Hansen M, Kurinczuk JJ, Bower C and Webb S (2002) The risk of major birth defects after intracytoplasmic sperm injection and in vitro fertilization. *N Engl J Med* 346,725–730.
- Harrison RF, Hennelly B, Woods T, Lowry K, Kondaveeti U, Barry-Kinsella C and Nargund G (1995) Course and outcome of IVF pregnancies and spontaneous conceptions within an IVF setting. *Eur J Obstet Gynecol Reprod Biol* 59,175–182.
- Helmerhorst FM, Perquin DAM, Donker D and Keirse MJNC (2004) Perinatal outcome of singletons and twins after assisted conception: a systematic review of controlled studies. *Br Med J*. doi:10.1136/bmj.37957.560278.EE.
- Holman CD, Arnold-Reed DE, de Klerk N, McComb C and English DR (2001) A psychometric experiment in causal inference to estimate evidential weights used by epidemiologists. *Epidemiology* 12,246–255.
- Isaksson R, Gissler M and Tiitinen A (2002) Obstetric outcome among women with unexplained infertility after IVF: a matched case-control study. *Hum Reprod* 17,1755–1761.
- Jackson RA, Gibson KA, Wu YW and Croughan MS (2004) Perinatal outcomes in singletons following in vitro fertilization: a meta-analysis. *Obstet Gynecol* 103,551–563.
- Kleinbaum DG, Kupper LL and Morgenstern H (1982) Epidemiologic research. Principles and quantitative methods. Wadsworth Inc, Belmont, California.
- Koivuova S, Hartikainen AL, Gissler M, Hemminki E, Sovio U and Jarvelin MR (2002) Neonatal outcome and congenital malformations in children born after in-vitro fertilization. *Hum Reprod* 17,1391–1398.
- Koudstaal J, Braat DD, Bruinse HW, Naaktgeboren N, Vermeiden JP and Visser GH (2000a) Obstetric outcome of singleton pregnancies after IVF: a matched control study in four Dutch university hospitals. *Hum Reprod* 15,1819–1825.
- Koudstaal J, Bruinse HW, Helmerhorst FM, Vermeiden JP, Willemsen WN and Visser GH (2000b) Obstetric outcome of twin pregnancies after in-vitro fertilization: a matched control study in four Dutch university hospitals. *Hum Reprod* 15,935–940.

- Kovalevsky G, Rinaudo P and Coutifaris C (2003) Do assisted reproductive technologies cause adverse fetal outcomes? *Fertil Steril* 79,1270–1272.
- Kozinszky Z, Zadori J, Orvos H, Katona M, Pal A and Kovacs L (2003) Obstetric and neonatal risk of pregnancies after assisted reproductive technology: a matched control study. *Acta Obstet Gynecol Scand* 82,850–856.
- Kurinczuk JJ (2003) Safety issues in assisted reproduction technology—From theory to reality—just what are the data telling us about ICSI offspring health and future fertility and should we be concerned? *Hum Reprod* 18,925–931.
- Kurinczuk JJ, Hansen M and Bower C (2004) The risk of birth defects in children born after assisted reproductive technologies. *Curr Opin Obstet Gynecol* 16,201–209.
- Lahat E, Raziel A, Friedler S, Schieber-Kazir M and Ron-El R (1999) Long-term follow-up of children born after inadvertent administration of a gonadotrophin-releasing hormone agonist in early pregnancy. *Hum Reprod* 14,2656–2660.
- Lambalk CB and Van Hooff M (2001) Natural versus induced twinning and pregnancy outcome: A Dutch nationwide survey of primiparous dizygotic twin deliveries. *Fertil Steril* 75,731–736.
- Lambert RD (2002) Safety issues in assisted reproduction technology: The children of assisted reproduction confront the responsible conduct of assisted reproductive technologies. *Hum Reprod* 17,3011–3015.
- Lambert RD (2003) Safety issues in assisted reproductive technology: Aetiology of health problems in singleton ART babies. *Hum Reprod* 18, 1987–1991.
- Lancaster PA (1985) Obstetric outcome. *Clin Obstet Gynaecol* 12,847–864.
- Leslie GI, Gibson FL, McMahon C, Tennant C and Saunders DM (1998) Infants conceived using in-vitro fertilization do not over-utilize health care resources after the neonatal period. *Hum Reprod* 13,2055–2059.
- Ludwig M and Diedrich K (2002) Follow-up of children born after assisted reproductive technologies. *RBM online* 5,317–322.
- Ludwig M and Katalinic A (2002) Malformation rate in fetuses and children conceived after ICSI: results of a prospective cohort study. *RBM online* 5,171–178.
- Minakami H, Sayama M, Honma Y *et al.* (1998) Lower risks of adverse outcome in twins conceived by artificial reproductive techniques compared with spontaneously conceived twins. *Hum Reprod* 13,2005–2008.
- Mitchell AA (2002) Infertility treatment—More risks and challenges. *N Engl J Med* 346,769–770.
- Morin NC, Wirth FH, Johnson DH, Frank LM, Presburg HJ, Van De Water VL, Chee EM and Mills JL (1989) Congenital malformations and psychosocial development in children conceived by in vitro fertilization. *J Pediatr* 115,222–227.
- Nassar S, Boutros J, Aboulghar H, Mansour R, Hussein M and Aboulghar M (1996) Perinatal outcome after in vitro fertilization and spontaneous pregnancy: A comparative study. *Middle East Fertil Soc J* 1,151–158.
- Nuojua-Huttunen S, Gissler M, Martikainen H and Tuomivaara L (1999) Obstetric and perinatal outcome of pregnancies after intrauterine insemination. *Hum Reprod* 14,2110–2115.
- Petersen K, Hornnes PJ, Ellingsen S, Jensen F, Brocks V, Starup J, Jacobsen JR and Andersen AN (1995) Perinatal outcome after in vitro fertilisation. *Acta Obstet Gynecol Scand* 74,129–131.
- Place I and Englert Y (2003) A prospective longitudinal study of the physical, psychomotor, and intellectual development of singleton children up to 5 years who were conceived by intracytoplasmic sperm injection compared with children conceived spontaneously and by in vitro fertilization. *Fertil Steril* 80,1388–1397.
- Powell K (2003) Fertility treatments: Seeds of doubt. *Nature* 422,656–658.
- Rizk B, Doyle P, Tan SL, Rainsbury P, Betts J, Brinsden P and Edwards R (1991) Perinatal outcome and congenital malformations in in-vitro fertilization babies from the Bourn-Hallam group. *Hum Reprod* 6,1259–1264.
- Ron-El R, Lahat E, Golan A, Lerman M, Bukovsky I and Herman A (1994) Development of children born after ovarian superovulation induced by long-acting gonadotropin-releasing hormone agonist and menotropins, and by in vitro fertilization. *J Pediatr* 125,734–737.
- Rothman KJ and Greenland S (1998) *Modern Epidemiology*. Lippincott-Raven, Philadelphia.
- Saunders K, Spensley J, Munro J and Halasz G (1996) Growth and physical outcome of children conceived by in vitro fertilization. *Pediatrics* 97, 688–692.
- Schisterman EF, Buck GM and Lynch CD (2003) Trying to avoid bias in case-control and case-cohort studies. *Fertil Steril* 80,1537–1538.
- Schultz RM and Williams CJ (2002) The science of ART. *Science* 296, 2188–2190.
- Simpson JL (1996) Registration of congenital anomalies in ART populations: pitfalls. *Hum Reprod* 11,81–88.
- Simpson JL (1998) Are anomalies increased after ART and ICSI? In Kempers RD, Cohen J, Haney AF, and Younger JB (eds) *Fertility and reproductive medicine*. Elsevier Science B.V., Amsterdam, pp. 199–209.
- Sterne JAC, Egger M and Davey Smith G (2001) Investigating and dealing with publication and other biases in meta-analysis. *Br Med J* 323, 101–105.
- Stromberg B, Dahlquist G, Ericson A, Finnstrom O, Koster M and Sjöernqvist K (2002) Neurological sequelae in children born after in-vitro fertilisation: a population-based study. *Lancet* 359,461–465.
- Stroup DF, Berlin JA, Morton SC *et al.* (2000) Meta-analysis of observational studies in epidemiology. A proposal for reporting. *JAMA* 283, 2008–2012.
- Sutcliffe AG, D'Souza SW, Cadman J, Richards B, McKinlay IA and Leiberman B (1995) Minor congenital anomalies, major congenital malformations and development in children conceived from cryopreserved embryos. *Hum Reprod* 10,3332–3337.
- Sutcliffe AG, Taylor B, Li J, Thornton S, Grudzinskas JG and Leiberman BA (1999) Children born after intracytoplasmic sperm injection: population control study. *Br Med J* 318,704–705.
- Sutcliffe AG, Taylor B, Saunders K, Thornton S, Leiberman BA and Grudzinskas JG (2001) Outcome in the second year of life after in-vitro fertilisation by intracytoplasmic sperm injection: a UK case-control study. *Lancet* 357,2080–2084.
- Sutcliffe AG, Saunders K, McLachlan R, Taylor B, Edwards P, Grudzinskas G, Leiberman B and Thornton S (2003) A retrospective case-control study of developmental and other outcomes in a cohort of Australian children conceived by intracytoplasmic sperm injection compared with a similar group in the United Kingdom. *Fertil Steril* 79,512–516.
- Tanbo T, Dale PO, Lunde O, Moe N and Abyholm T (1995) Obstetric outcome in singleton pregnancies after assisted reproduction. *Obstet Gynecol* 86,188–192.
- Verlaenen H, Cammu H, Derde MP and Amy JJ (1995) Singleton pregnancy after in vitro fertilization: expectations and outcome. *Obstet Gynecol* 86,906–910.
- Wang JX, Norman RJ and Kristiansson P (2002) The effect of various infertility treatments on the risk of preterm birth. *Hum Reprod* 17,945–949.
- Wennerholm UB, Hamberger L, Nilsson L, Wennergren M, Wikland M and Bergh C (1997) Obstetric and perinatal outcome of children conceived from cryopreserved embryos. *Hum Reprod* 12,1819–1825.
- Wennerholm UB, Albertsson-Wikland K, Bergh C, Hamberger L, Niklasson A, Nilsson L, Thiringer K, Wennergren M, Wikland M and Borres MP (1998) Postnatal growth and health in children born after cryopreservation as embryos. *Lancet* 351,1085–1090.
- Wennerholm UB, Bergh C, Hamberger L, Lundin K, Nilsson L, Wikland M and Kallen B (2000) Incidence of congenital malformations in children born after ICSI. *Hum Reprod* 15,944–948.
- Westergaard HB, Johansen AM, Erb K and Andersen AN (1999) Danish National In-Vitro Fertilization Registry 1994 and 1995: a controlled study of births, malformations and cytogenetic findings. *Hum Reprod* 14, 1896–1902.
- Western Australian Reproductive Technology Council (WA RTC) (1997) *The Human Reproductive Technology Act 1991—Directions*. Perth, Australia.
- Winston RML and Hardy K (2002) Are we ignoring potential dangers of in vitro fertilization and related treatments? *Nat Med* 8,S14–S18.
- Yeh J, Leipzig S, Friedman EA and Seibel MM (1990) Results of in vitro fertilization pregnancies: experience at Boston's Beth Israel Hospital. *Int J Fertil* 35,116–119.
- Zadori ZI, Kozinszky Z, Orvos H, Katona M, Kaali SG and Pal A (2003) The incidence of major birth defects following in vitro fertilization. *J Assist Reprod Genet* 20,131–132.
- Zuppa AA, Maragliano G, Scapillati ME, Crescimbeni B and Tortorolo G (2001) Neonatal outcome of spontaneous and assisted twin pregnancies. *Eur J Obstet Gynecol Reprod Biol* 95,68–72.

Submitted on May 17, 2004; accepted on October 6, 2004