**human
reproduction
update**

Assisted reproductive technology and birth defects: a systematic review and meta-analysis

Michèle Hansen[1,*], Jennifer J. Kurinczuk[2], Elizabeth Milne[1], Nicholas de Klerk[3], and Carol Bower[1,4]

[1]Division of Population Sciences, Telethon Institute for Child Health Research and the Centre for Child Health Research, University of Western Australia, PO Box 855, Perth, Western Australia 6872, Australia [2]National Perinatal Epidemiology Unit, University of Oxford, Oxford, UK [3]Department of Biostatistics and Bioinformatics, Telethon Institute for Child Health Research and the Centre for Child Health Research, University of Western Australia, Perth, Western Australia, Australia [4]Western Australian Register of Developmental Anomalies, King Edward Memorial Hospital, Subiaco, Western Australia, Australia

Correspondence address. Tel: +61-8-9489-7771; Fax: +61-8-9489-7700; E-mail: michele@ichr.uwa.edu.au

TABLE OF CONTENTS

BACKGROUND: It has been 10 years since we carried out a systematic search of the literature on birth defect risk in infants born following assisted reproductive technology (ART) compared with non-ART infants. Because of changes to ART practice since that review and the publication of more studies the objective of this review was to include these more recent studies to estimate birth defect risk after ART and to examine birth defect risk separately in ART singletons and multiples.

METHODS: We searched Medline, Embase and Current Contents databases (1978–2012). We used the same data extraction sheet and questionnaire we had used previously with the addition of a quality score to the questionnaire. Pooled relative risk (RR) estimates were calculated using a random effects model. All data were analysed using Comprehensive Meta-Analysis V2.

RESULTS: There were 45 cohort studies included in this review. ART infants ($n = 92\,671$) had a higher risk of birth defects [RR 1.32, 95% confidence interval (CI) 1.24–1.42] compared with naturally conceived infants ($n = 3\,870\,760$). The risk further increased when data were restricted to major birth defects (RR 1.42, 95% CI 1.29–1.56) or singletons only (RR 1.36, 95% CI 1.30–1.43). The results for ART multiples were less clear. When all data for multiples were pooled the RR estimate was 1.11 (95% CI 0.98–1.26) but this increased to 1.26 (0.99–1.60) when the analysis was restricted to studies of ART twins where some adjustment was made for differences in zygosity distribution between ART and non-ART multiples.

CONCLUSIONS: Birth defects remain more common in ART infants. Further research is required to examine risks for important subgroups of ART exposure.

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

Introduction

It has been almost 10 years since we conducted our last systematic review and meta-analysis examining the risk of birth defects in infants born following assisted reproductive technology (ART) treatment (Hansen *et al.*, 2005). In 2003, we identified 25 papers for inclusion in a meta-analysis, however, many more studies have since been published. The field of ART has undergone rapid change since the birth of Louise Brown in 1978 with new techniques or adjustments to laboratory conditions allowing a broader group of subfertile patients to access treatment. Pregnancy rates have improved over time and there has been a shift towards the transfer of fewer embryos resulting in a marked decline in the multiple pregnancy rate in many countries together with important improvements in many perinatal outcomes. Against this changing backdrop, we were interested to see whether more recent publications also support an increased birth defect risk, whether the pooled estimate may have changed, and whether we could use more stringent selection criteria for study inclusion now that more studies are available for analysis. We were hopeful that there would be sufficient numbers of studies to examine birth defect risk within subgroups with greater confidence. In particular, we were keen to look at whether there were sufficient papers to estimate birth defect risk in ART singletons and multiples separately.

The specific aims of our study were to quantify the risk of birth defects in ART infants compared with non-ART infants overall, and when data on singleton and multiple births were considered separately. We followed the PRISMA guidelines for reporting of systematic reviews and meta-analyses of observational studies (http://www.prisma-statement.org/).

Methods

Literature search strategy and inclusion criteria

We performed an extensive literature search of Medline, Embase and Current Contents databases (1978–2012) using a broad combination of search terms (Table I). The search strategy was written in Ovid, then saved and run in each database in early September 2012. 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 very broad, as in our previous review, so that crude birth defect data were acceptable, as was an absence of statistical analysis. We specifically searched for papers that compared birth defects in IVF or ICSI infants with a non-ART comparison group. Papers that compared children born following one ART technique to another were not reviewed. We also did not include papers that reported comparisons based on a single type or group of birth defects, or conference abstracts where more detailed papers describing the same study were unavailable.

Exclusions

We excluded non-English language papers; overlapping data (even from different authors within the same country); papers with mixed exposure groups [for example, including children born following ovulation induction (OI) or intrauterine insemination (IUI) within the 'ART' group]; and all studies with essentially cross-sectional design, that is, studies where infants were not followed from birth but recruited and assessed at one or more years of age (Table II).

Quality assessment

We used the same structured data extraction sheets and questionnaires relating to study methodology that we had used in our previous systematic review (Supplementary data, Fig. S1). However, for this review, we added a scoring system to the data extraction sheet to assess study quality in terms of birth defect assessment. We examined checklists from the Critical Appraisal Skills Programme (CASP; http://www.caspinternational.org) in order to assign scores to appropriate sections of our data extraction sheets; however, we retained our own questions which had been created specifically to examine how well the studies were designed to assess birth defect risk in particular rather than a more generic set of questions. A study was penalized, for example, where no definition of birth defects was given and also if no distinction was made between major and minor defects. In contrast to our previous review where we sent out the papers to external reviewers for assessment, a single researcher (M.H.) was responsible for extracting all of the data and completing the questionnaires. Both crude and adjusted relative risk (RR) estimates were extracted from each study. Where an RR was not provided we recorded the number of infants with and without birth defects by method of conception. We also recorded information about the study design, methods, birth defect definition and adjustment for confounders. Each paper was given a quality score out of a possible total of 17 points. Studies were considered of higher methodological quality if a score of ≥12 out of 17 was achieved, medium quality if the score was between

Table I Literature search strategy.

All combinations of terms in the first column with terms in the second column (For example: IVF AND birth defect?[a]; IVF AND record linkage etc.)	
IVF	Birth defect?
In vitro fertilisation[b]	Congenital malformation?
In-vitro fertili?ation	Congenital abnormalit$
ICSI	Health AND child
Intracytoplasmic sperm injection	Record linkage
Assisted reproduction	
Assisted reproductive techn$[c]	
Infertility treatment?	

[a]'Defect?' will find defect or defects.
[b]Fertili?ation will find fertilisation or fertilization.
[c]'Techn$' will find technology, technologies, technique(s) etc.

Table II List of studies excluded from the systematic review (n = 83).

Reason for exclusion	Reference(s)
Overlapping data (All papers published from the same country were examined for data overlap.[a] Papers with partial data overlap that could still contribute several years of unique data were included in order not to lose any information.)	Lambalk and van Hooff (2001) substantial overlap with Anthony et al. (2002) Leslie et al. (1998) overlap with Bowen et al. (1998) Bergh et al. (1999), Ericson and Kallen (2001), Ericson et al. (2002), Stromberg et al. (2002), Wennerholm et al. (1997), Wennerholm et al. (1998), Wennerholm et al. (2000) overlap with Kallen et al. (2005) Berg et al. (2001) same data as Ericson and Kallen (2001) Dhont et al. (1997) overlap with Dhont et al. (1999) Sutcliffe et al. (1999) overlap with Sutcliffe et al. (2001) Rizk et al. (1991) excerpt of Beral and Doyle (1990) Ludwig and Diedrich (2002) data are excerpt of Ludwig and Katalinic (2002) and ICSI group the same as Katalinic et al. (2004) Cadman et al. (1999) same data as D'Souza et al. (1997) Hansen et al. (2002), Hansen et al. (2008), Hansen et al. (2009) overlap with Hansen et al. (2012) Fisch et al. (1997) overlap with Merlob et al. (2005) Zhu et al. (2006) overlap with Pinborg et al. (2010a)—all singletons from Zhu would be included in larger Pinborg et al.
Insufficient data to calculate a RR estimate (e.g. data reported per pregnancy rather than per child)	Yeh et al. (1990), FIVNAT (1995), Minakami et al. (1998), Harrison et al. (1995), Cederblad et al. (1996), Chou et al. (2002), Nuojua-Huttunen et al. (1999), Al-Fifi et al. (2009), Caserta et al. (2008), Daniel et al. (2000), Ezechi et al. (2008), Gojnic et al. (2005), Kanat-Pektas et al. (2008), Szymankiewicz et al. (2004), Baxi and Kaushal (2008), Ozer et al. (2011), Gupta et al. (2012)
No birth defects found in either group[b]	Lahat et al. (1999) (10 infants), Ron-El et al. (1994) (32 infants) Brandes et al. (1992) (116 infants), Tomic and Tomic (2011) (283 infants)
Inappropriate comparison group—included both spontaneously conceived infants and those born to infertile patients following fertility treatments other than IVF/ICSI. Or compared birth defect data for an ART group to general population results covering different birth years, or different regions than the ART group.	Petersen et al. (1995), Allen et al. (2008), Beral and Doyle (1990), Yan et al. (2011)
Total number of malformations reported rather than number of children with malformations	Saunders et al. (1996), Al-Fifi et al. (2009), Caserta et al. (2008)
Cross-sectional design at late age (≥12 months) or other. (Excluded on the basis that children with severe defects may have died)	Morin et al. (1989), D'Souza et al. (1997), Sutcliffe et al. (1995), Sutcliffe et al. (2001), Sutcliffe et al. (2003), Agarwal et al. (2005), Banerjee et al. (2008), Belva et al. (2007), Bonduelle et al. (2004), Bonduelle et al. (2005), Desmyttere et al. (2009), Knoester et al. (2008), Place and Englert (2003), Sanchez-Albisua et al. (2007), Shu-Chi et al. (2006)
Mixed exposure group (for example, those including children born by OI or IUI within the 'ART' group)[c]	Addor et al. (1998), Tanbo et al. (1995), Zadori et al. (2003), Kozinszky et al. (2003), Zadori et al. (2004), Zuppa et al. (2001), Kanat-Pektas et al. (2008), Aslan et al. (2005), Kor-anantakul et al. (2007), Morcel et al. (2010), Sperling et al. (2007), Welmerink et al. (2010), Ozer et al. (2011)
Foreign language papers	Wang et al. (2009), Sebastiani et al. (2009), Gonzalez et al. (2006), Sanchis Calvo et al. (2009), Rodrigues et al. (2005), Sipek et al. (2004), Snajderova et al. (2008)
Other	Schimmel et al. (2006)—infants restricted to VLBW only—excluding terminations, stillbirths and deaths in delivery room. Data not comparable to rest of meta-analysis. Reefhuis et al. (2009)—case–control study of select birth defects only. Sala et al. (2011)—comparison of birth defects assessed in pregnancies at a referral prenatal diagnosis centre (i.e. comparison group for ART are high-risk non-ART pregnancies). Data reported per pregnancy rather than per fetus/baby

[a]For example, the 1881 ICSI and IVF singleton infants included in the study by Zhu et al. (2006) would all have been included in the larger Pinborg et al. (2010a) study of all IVF and ICSI singletons born in Denmark between 1995 and 2006 (n = 10 329), therefore only the latter study was included.
[b]It is not possible to calculate an RR estimate with zero cells, therefore these studies could not contribute to the meta-analysis.
[c]If possible we extracted information about IVF and/or ICSI infants separately, even if this meant using crude data. For example, El-Chaar et al. (2009) include an adjusted estimate of birth defect risk in a mixed exposure group (ICSI + IVF + OI + IUI) but it was possible to extract crude data for the ICSI and IVF infants for use in our review.

8.5 to 11.5, and lower quality where a score of ≤8 was assigned. These categories were chosen to reflect scores of >70%, 50–70% and <50%. Any papers where decisions on study quality or scoring were less certain were discussed with a second author (C.B.) and a consensus reached.

Statistical analysis

Since the prevalence of birth defects is <10% in all studies, we have assumed that the (adjusted) odds ratio (OR) is equal to the (adjusted) relative risk (RR; Greenland and Thomas, 1982; McNutt et al., 2003), and have used the term RR throughout, even if the term used in the original paper was OR.

If effect measures were not reported in a paper, we calculated RRs and their 95% confidence interval (CI) from the raw data. Where more than one RR was available from a particular study (e.g. an adjusted RR estimate as well as a crude estimate or estimates for singletons and twins separately as well as for all infants combined) all of these were extracted and used in relevant subgroup analyses (e.g. of crude data only).

We used the same hierarchy of RR inclusion as in our previous meta-analysis: adjusted RR estimates in preference to crude; estimates of major birth defect risk in preference to major and minor defects combined; and estimates relating to all infants in preference to singletons or twins only. However, since the pooled estimate of all studies included studies which had assessed birth defects in singletons only, twins only or all infants combined which might mask differences in birth defect risk by plurality, we also examined the results of studies that had looked at birth defects in ART singletons and multiples separately. We included estimates for ART twins where some adjustment had been made for differences in zygosity distribution in preference to estimates including all twin infants.

Where a study provided birth defects data for ICSI and IVF infants separately compared with a single non-ART conception comparison group (e.g. Bowen et al., 1998; Kuwata et al., 2004; Buckett et al., 2007; Davies et al., 2012a) the data were pooled to avoid double counting of the non-ART group. We used a random effects model to obtain pooled estimates of the RR for all studies, and for the different subgroups assessed.

Sensitivity analyses and publication bias

In order to investigate heterogeneity between studies, we plotted the RR estimate with its 95% CI for each study, together with the pooled estimate, in forest plots. We then examined the effect on the pooled 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.

Finally, heterogeneity among studies was formally tested using the Q-statistic (with χ^2 distribution and $n-1$ degrees of freedom where n refers to the number of studies that are combined). P-values <0.10 were considered statistically significant. We also examined the I^2 statistic which reflects the proportion of the observed dispersion between studies that is due to true variation rather than random error. Values of I^2 near zero reflect that almost all the observed variation in study estimates is due to random error whereas when I^2 is large we may want to consider subgroup analyses that might help explain the dispersion. I^2 values <25% were considered low, around 50% moderate and >75% high (Borenstein et al., 2009).

Subgroup analyses were used to investigate differences in study design and their effect on the pooled estimate. For example studies that included major birth defects only were compared with those that included any defects and studies considered to be of higher methodological quality based on our scoring system (score ≥ 12/17) were compared with those of medium and lower quality.

Funnel plots were used to assess publication bias together with the Begg and Mazumdar rank correlation test. A P-value <0.10 was considered to indicate the presence of publication bias.

All data were analysed using Comprehensive Meta-Analysis V2.

Results

Characteristics of included studies

The extensive literature search yielded 2316 citations. Of these 2193 were excluded based on the title and abstract. The full text of 77 new articles (not included in our previous meta-analysis) was obtained together with the full text of the 51 papers examined in our previous meta-analysis ($n = 128$ articles studied). These were all reports of individual studies and did not include any review papers. Following a careful review, we excluded 83 papers for the reasons shown in Table II. Fourteen papers that were included in our previous meta-analysis were excluded from this review for the following reasons: overlapping data with more recent studies ($n = 4$; Fisch et al., 1997; Ericson and Kallen, 2001; Hansen et al., 2002; Ludwig and Katalinic, 2002); more stringent exclusion criteria such as the exclusion of studies with mixed exposure groups ($n = 4$; Tanbo et al., 1995; Addor et al., 1998; Zuppa et al., 2001; Zadori et al., 2003) or cross-sectional design ($n = 5$; Morin et al., 1989; Sutcliffe et al., 1995, 2001, 2003; D'Souza et al., 1997); inappropriate comparison data ($n = 1$; Beral and Doyle, 1990). A flowchart showing the steps involved in study selection can be found at Supplementary data, Fig. S2.

We included 45 papers with no data overlap; 34 new papers and 11 that were included in our previous meta-analysis published in 2005 (Table III). The earliest included study was published in 1995 and the latest in 2012. A total of 92 671 ART infants were included in the 45 studies; 68 026 (or 73%) were born in Europe and almost half of these were born in Sweden (31 850; Fig. 1). The size of the ART group in each study ranged from 76 to 16 280 infants. The presence of birth defects was assessed only at birth in the majority of studies reviewed (78%).

Only 13 studies (29%) were considered of higher methodological quality (with respect to birth defect assessment) achieving a quality score ≥ 12 out of 17 (Table III); however, these 13 studies contributed 78% of the ART infants in the meta-analysis. The majority of the higher quality studies were population-based with a clear definition of a birth defect. All had a large sample size and all included data that were either adjusted or matched for at least maternal age and parity.

Primary outcomes

Table IV shows the pooled estimates of birth defect risk in ART compared with non-ART infants generated by meta-analyses for all studies combined ($n = 45$); and studies that have provided estimates of birth defect risk in ART singletons ($n = 23$) and multiples ($n = 27$) separately. The final column includes studies of twin infants where some adjustment was made for the different zygosity distributions seen in ART and non-ART twins (e.g. by comparing only unlike-sex twins; $n = 12$).

Birth defect risk in all studies of ART compared with non-ART infants
The pooled estimate for all 45 studies combined was 1.32 (95% CI 1.24–1.42) indicating a significant 30% increased risk of birth defects in children born following ART (Table IV). The individual point

Table III Study characteristics for all 45 studies included in meta-analysis.

Authors and publication year (birth years included in study)	Location	Population versus clinic-based sample	Total number (n) ART (% birth defect) Total n non-ART (% birth defect)	ART treatment	Plurality	Defects assessed	Age/time assessment	Adjusted, matched or crude data	RR	95% CI	Quality Score[a]
Adler-Levy et al. (2007) (1988–2002)	Israel	Clinic	558 (12.0) 3694 (7.4)	IVF, ICSI	Twins	All	Birth (≥24 weeks)	Adjusted	1.20	0.83–1.72	3
Anthony et al. (2002) (1995–1996)	The Netherlands	Population	4224 (3.2) 314 605 (2.7)	IVF, ICSI	All	All	Birth	Adjusted	1.03	0.86–1.23	2
Apantaku et al. (2008) (9/1999–3/2004)	UK	Clinic	88 (9.1) 88 (3.4)	IVF, ICSI	Singletons	All	Birth (≥24 weeks)	Matched	3.5	0.6–34.0	2
Bowen et al. (1998) (1993–1995)	Australia	Clinic	173 (4.0) 80 (5.0)	IVF, ICSI	All[d]	Major	12 months	Matched	0.80	0.20–3.85	3
Buckett et al. (2007) (1998–2003)	Canada	Clinic	377 (9.0) 350 (6.6)	IVF, ICSI	All	All	Birth	Matched			2
Davies et al. (2012a, b) (1986–2002)	Australia	Population	3708 (8.2) 300 662 (5.8)	IVF, ICSI	All	All	5 years	Adjusted	1.24	1.09–1.41	1
Dhont et al. (1999) (1992–1997)	Belgium	Population	4196 (2.9) 4196 (2.3)	IVF, GIFT	Singletons+ ULS Twins	All	Birth	Matched	1.25	0.96–1.64	1
El Hage et al. (2006) (1996–2001)	Lebanon	Clinic	780 (2.4) 2168 (1.1)	IVF, ICSI	All	All	Birth	Crude	2.33	1.27–4.26	3
El-Chaar et al. (2009) (2005)	Canada	Population	319 (3.4) 43 462 (1.9)	IVF, ICSI	All	All	Birth	Crude[b]	1.88	0.93–3.44	3
Fujii et al. (2010) (2006)	Japan	Population	1396 (2.3) 53 566 (2.0)	IVF, ICSI, GIFT, ZIFT	Singletons	All	Birth	Adjusted	1.17	0.81–1.69	2
Halliday et al. (2010) (1991–2004)	Australia	Population	6946 (6.4) 20 838 (4.8)	IVF, ICSI	Singletons	All	Not stated[c]	Adjusted	1.36	1.19–1.55	1
Hansen et al. (2012) (1994–2002)	Australia	Population	2911 (8.2) 207 260 (5.4)	IVF, ICSI	Singletons Twins	Major	6 years	Adjusted	1.53 1.08	1.30–1.79 0.77–1.51	1
Ho et al. (2005) (2002–2003)	Taiwan	Clinic	139 (3.6) 52 (7.7)	IVF	Twins— dichorionic	All	Birth	Crude	0.45	0.09–2.36	3
Isaksson et al. (2002) (1993–3/1999)	Finland	Clinic	109 (5.5) 545 (3.5)	IVF, ICSI	All[d]	Major	Birth	Matched	1.61	0.51–4.33	2
Joy et al. (2008) (2002–2003)	Ireland	Clinic	76 (0) 170 (1.2)	IVF, ICSI	Twins— dichorionic	All	Birth	Crude	0.56[h]	0.02–34.49	3
Kallen et al. (2005) (1982–2001)	Sweden	Population	16 280 (3.3) 2 039 943 (2.2)	IVF, ICSI	All	Major ('weeded')[e]	Birth	Adjusted	1.44	1.32–1.57	1
Kallen et al. (2010a) (2001–2007)	Sweden	Population	15 570 (3.7) 689 157 (3.0)	IVF, ICSI	All	Major ('Rel. severe')	Birth	Adjusted	1.25	1.15–1.37	1
Kanyo and Konc (2003) (12/1998–12/1999)	Hungary	Clinic	134 (1.5) 894 (3.0)	IVF, ICSI	All	Major	Birth[f]	Crude	0.48	0.06–1.98	3

Katalinic et al. (2004) (8/1998–8/2000 ICSI; 1993–2001 non-ART)	Germany	Population	3372 (8.7) 8016 (6.1)	ICSI	All	Major	Birth[g]	Adjusted	1.24	1.02–1.50	2
Klemetti et al. (2005) (10/1996–9/1999)	Finland	Population	4459 (4.4) 27 078 (2.9)	IVF, ICSI	All	Major	1 year?	Adjusted	1.31	1.10–1.57	1
Koivurova et al. (2002) (1990–1995)	Finland	Population	304 (6.6) 569 (4.4)	IVF	All	All	36 months	Matched	1.53	0.79–2.93	2
Koudstaal et al. (2000a) (1992)	The Netherlands	Clinic	307 (2.3) 307 (2.3)	IVF	Singletons	All	Birth	Matched	1.00	0.29–3.39	3
Koudstaal et al. (2000b) (1992)	The Netherlands	Clinic	192 (3.6) 192 (2.6)	IVF	Twins	All	Birth	Matched	1.42	0.38–5.76	3
Kuwata et al. (2004) (1990–7/2001)	Japan	Clinic	232 (9.5) 188 (2.1)	IVF, ICSI	Twins – dichorionic	All	Birth	Adjusted (mat age only)	4.91	2.14–11.26	3
Lehnen et al. (2011) (1/2000–4/2009)	Germany	Clinic	142 (0) 506 (0.6)	IVF, ICSI	Twins – dichorionic	All	Birth	Crude	0.59[i]	0.03–11.90	3
Manoura et al. (2004) (1994–7/2002)	Greece	Clinic	139 (3.6) 288 (1.7)	IVF	Twins	All	Birth	Crude	2.11	0.48–9.33	3
Merlob et al. (2005) (1986–2002)	Israel	Clinic	1910 (9.1) 82 305 (4.8)	IVF, ICSI	All	Major	Birth	Crude	1.99	1.68–2.33	2
Nassar et al. (1996) (year not stated)	Egypt	Clinic	128 (2.3) 203 (1.5)	IVF	All	Major	Birth	Crude	1.60	0.21–12.12	3
Nassar et al. (2003) (1995–2000)	Lebanon	Clinic	112 (5.4) 224 (4.9)	IVF	Twins	All	Birth	Matched	1.10	0.32–3.34	3
Olson et al. (2005) (1989–2002)	USA- Iowa	Clinic	1462 (6.2) 8422 (4.4)	IVF, ICSI, ZIFT	All[d]	Major	1 year	Matched and adjusted	1.30	1.00–1.67	1
Ombelet et al. (2005) (1997–2003)	Belgium	Population	2125 (2.4) 4185 (2.2)	ICSI	Singletons+ ULS Twins	All	Birth	Matched	1.11	0.77–1.58	2
Palermo et al. (2008) (9/1993–6/2006)	USA – New York	Clinic	229 (6.6) 194 (6.2)	ICSI	Singletons	All	Birth	Matched (mat age only)	1.06	0.45–2.56	3
Pinborg et al. (2004) (1995–2000)	Denmark	Population	1650 (not stated) 3546 (not stated)	IVF, ICSI	ULS Twins	All	1–6 years	Adjusted	1.24	0.97–1.58	1
Pinborg et al. (2010a) (1995–2006)	Denmark	Population	10 329 (5.9) 4800 (4.7)	IVF, ICSI	Singletons	Major ('weeded')[e]	1–13years	Adjusted	1.27	1.09–1.43	1
Sagot et al. (2012) (2000–6/2009)	France	Population	1071 (4.7) 4594 (2.1)	IVF, ICSI	Singletons+ULS Twins	Major	Birth	Matched and adjusted	2.00 3.70	1.30–3.10, 1.10–16.90	1
Saygan-Karamursel et al. (2006) (1999–2003)	Turkey	Clinic	274 (4.4) 348 (0.9)	ICSI	Twins	Major	Birth	Adjusted (mat age only)	3.89	0.65–23.07	3
Shebl et al. (2008a, b) (1996–2005)	Austria	Clinic	432 (3.7) 754 (3.3)	IVF, ICSI	Twins	All	Birth	Crude	1.12	0.55–2.21	3
Shevell et al. (2005) (1999–2002)	USA	Population	554 (3.5) 34 286 (1.9)	IVF, ICSI, GIFT, ZIFT	Singletons	All	Birth	Adjusted	0.90	0.40–2.00	3
Smithers et al. (2003) (1991–1999)	Australia	Population	514 (5.6) 2147 (5.2)	IVF, ICSI, GIFT	ULS Twins	All	Not stated[c]	Crude	1.09	0.69–1.67	2

Continued

Table III *Continued*

Authors and publication year (birth years included in study)	Location	Population versus clinic-based sample	Total number (n) ART (% birth defect) Total n non-ART (% birth defect)	ART treatment	Plurality	Defects assessed	Age/time assessment	Adjusted, matched or crude data	RR	95% CI	Quality Score[a]
Vasario et al. (2010) (9/2004−9/2008)	Italy	Clinic	168 (11.3) 278 (6.1)	IVF, ICSI	Twins − dichorionic	All	Birth	Adjusted	1.30	0.60−2.82	3
Verlaenen et al. (1995) (1988−6/1994)	Belgium	Clinic	140 (2.1) 140 (0)	IVF	Singletons	Minor	Birth	Matched	6.11[h]	0.30−321.98	2
Wang et al. (2002) (1986−1998)	Australia	Clinic	1019 (4.3) 1019 (4.5)	IVF, ICSI, GIFT	Singletons	All	Birth	Matched	0.95	0.61−1.49	2
Wen et al. (2010) (1996−2005)	Canada	Clinic	1044 (7.7) 1910 (4.4)	IVF, ICSI	All[d]	Major	Birth	Adjusted and matched	1.58	1.10−2.27	1
Westergaard et al. (1999) (1994−1995)	Denmark	Population	2245 (4.8) 2245 (4.6)	IVF, ICSI	All[d]	All	Birth	Matched	1.04	0.78−1.39	1
Yang et al. (2011) (1995−2008)	South Korea	Clinic	134 (7.5) 286 (8.0)	IVF	Twins—dichorionic	All	Birth	Crude	0.92	0.38−2.09	3

GIFT, gamete intrafallopian transfer; ICSI, intracytoplasmic sperm injection; ZIFT, zygote intrafallopian transfer.
[a]Each study was assigned a score out of 17; 1 = higher quality studies with scores ≥12, 2 = medium quality with scores 8.5−11.5, 3 = low-quality with scores ≤8.
[b]This study did include adjusted estimates of birth defect risk for a mixed exposure group including OI and IUI but we extracted crude data for IVF and ICSI infants separately.
[c]The Victorian Birth Defects Register collects information on birth defects diagnosed to 15 years of age.
[d]Matched for plurality.
[e]'Weeded' refers to the removal of a number of more common minor conditions (e.g. preauricular tags, tongue tie, undescended testes, patent ductus arteriosus etc.).
[f]We excluded one defect diagnosed in the ART group post-birth via telephone interview—this follow-up was not performed for non-ART group.
[g]ICSI group assessed at ∼1 month, spontaneous conception group assessed at birth.
[h]There were no birth defects found in the non-ART group in this study. We added 0.5 to allow estimation of RR and 95% CI. We excluded defects in the ART group that were diagnosed due to increased surveillance of the ART group (ultrasound scan of heart and kidneys).
[i]There were no birth defects found in the ART group in this study. We added 0.5 to allow estimation of RR and 95% CI.

Figure 1 Number of ART infants included in meta-analysis by region of birth.
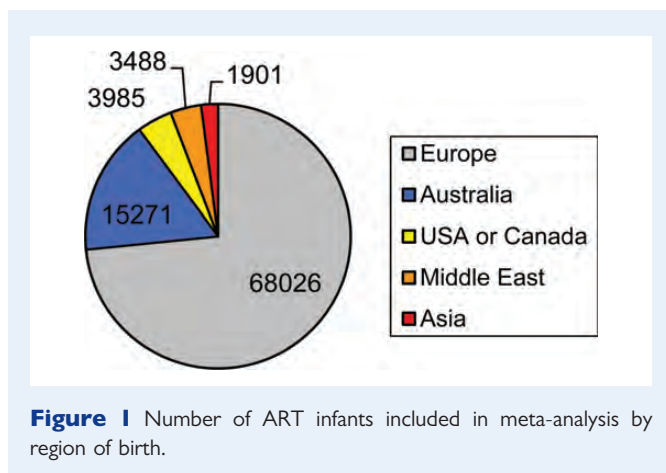
estimates for these studies ranged from 0.45 to 6.11 (see Fig. 2 forest plot). This estimate includes studies that have grouped all infants together, studies of singletons only and studies of twins only, some of which have adjusted for differences in zygosity distribution between ART and non-ART twins and some of which have not. Not surprisingly, the heterogeneity statistic was highly significant ($P = 0.000$) and the I^2 statistic moderately high (47%).

Birth defect risk for ART singletons and ART multiples separately
When we restricted our analysis to studies where the prevalence of birth defects had been reported in ART singletons or ART multiples separately the between-study heterogeneity was reduced and no longer significant at the $P < 0.1$ level. The pooled estimate comparing birth defects in ART and non-ART singletons was 1.36 (95% CI 1.30–1.43) with no evidence of heterogeneity, $I^2 = 0$ (Fig. 3). The pooled estimate for ART compared with non-ART multiples was 1.11 (95% CI 0.98–1.26) with low heterogeneity, $I^2 = 24\%$ (Fig. 4). The pooled estimate for singletons was significantly greater than the pooled estimate for multiples ($\chi^2 = 8.74$, 1 df, $P = 0.003$).

Birth defect risk for ART twins where some adjustment made for differences in zygosity distribution between ART and non-ART twins
When we further restricted our pooled estimate to summarize only the results of the 12 studies comparing ART and non-ART twins where some adjustment had been made for the different zygosity distributions between these two groups the pooled risk of birth defects for ART twins was 1.26 (95% CI 0.99–1.60; Fig. 5).

Sensitivity analyses and publication bias
When we looked at studies reporting results for singletons separately there were two obvious outliers in the forest plot (Verlaenen et al., 1995; Apantaku et al., 2008; Fig. 3). Removal of these studies had no effect on the pooled estimate which remained at 1.36 (95% CI 1.30–1.43; Supplementary data, Table SI). Similarly, removal of the study with the highest weight (Kallen et al., 2005) had no material effect on the pooled risk ratio [RR 1.35 (95% CI 1.28–1.44)]. We also ran an analysis where each study was removed in turn to examine whether any one study had a large effect on the pooled estimate. The pooled estimate varied from 1.35 to 1.38 when we did this and all 95% CIs excluded unity.

When we looked at the results for multiples separately there were a few more outliers (Bowen et al., 1998; Kuwata et al., 2004; Manoura et al., 2004; Ho et al., 2005; Saygan-Karamursel et al., 2006; Sagot et al., 2012; Fig. 4). Removal of these six studies gave a slightly lower pooled estimate [RR 1.07 (95% CI 0.97–1.18); Supplementary data, Table SI]. In contrast, removal of studies with the highest weight (Pinborg et al., 2004; Kallen et al., 2005; Davies et al., 2012a) led to small increases in the pooled estimate, none of which reached statistical significance.

The funnel plot (of study precision against log OR) for all studies combined, singletons and multiples are shown at Supplementary data, Figs S3–5. All plots are fairly symmetrical apart from the two outliers described for singleton studies, the removal of which had no effect on the pooled estimate for singletons. The *P*-value of the Begg's test for all analyses (i.e. all studies, singletons only, multiples only) was >0.1 suggesting there was no obvious publication bias (Supplementary data, Table SI).

Secondary outcomes: subgroup analyses
We performed a series of subgroup analyses within the different plurality groups comparing studies assessing major birth defects versus any birth defects; population versus clinic-based studies; crude versus adjusted or matched data; large versus smaller studies; higher quality versus lower quality studies; studies that had used different methods of birth defect classification; studies comparing IVF versus non-ART and ICSI versus non-ART infants, and finally grouping studies by region (Table IV).

Crude versus adjusted/matched data
Pooled estimates incorporating only matched or adjusted data were lower than those incorporating crude data only. In all cases where the results of adjusted/matched studies were combined (that is for all studies, singletons only, multiples and twins with zygosity adjustment) the pooled estimate was statistically significantly increased for ART compared with non-ART infants.

Major birth defects versus any birth defects
Only 36% of studies had compared major birth defects in ART and non-ART infants, the remainder reporting results for any birth defect (major or minor). The pooled estimates were generally higher for those studies assessing major birth defects compared with any birth defects with the exception of multiples. The pooled estimates of major birth defects were almost identical across studies of singletons only and all studies combined (RR 1.41 and 1.42), but close to one for all multiples (RR 1.06) and 1.73 for twins where some adjustment was made for zygosity. This last estimate pools the results of only two studies, one of which had a very small sample size ($n = 168$) and a very large RR estimate (RR 3.70; Sagot et al., 2012).

Population versus clinic-based studies
Less than half of all studies (42%) were population-based, however, this increased to 61% of studies reporting results for singletons separately. When all studies were combined, population-based studies had a lower pooled estimate (RR 1.29) compared with clinic-based studies (RR 1.41) but the reverse was true for singletons (population-based pooled RR 1.37 versus clinic-based 1.30).

Table IV Subgroup analyses.

Overall	All studies (92 671 infants)	ART Singletons (48 944 singletons)	ART Multiples[a] (19 361 multiples)	Twins only, where adjustment made for differences in zygosity (5780 twins)
	1.32 (1.24–1.42) (n = 45) $\chi^2 = 82.26$, P = 0.000, $I^2 = 47\%$	1.36 (1.30–1.43) (n = 23) $\chi^2 = 19.61$, P = 0.607, $I^2 = 0\%$	1.11 (0.98–1.26) (n = 27) $\chi^2 = 34.06$, P = 0.134, $I^2 = 24\%$	1.26 (0.99–1.60) (n = 12) $\chi^2 = 18.43$, P = 0.072, $I^2 = 40\%$
Subgroup analyses				
Grouped by adjusted/matched versus crude data				
Adjusted/matched	1.29 (1.22–1.37) (n = 33) $\chi^2 = 44.99$, P = 0.063, $I^2 = 29\%$	1.35 (1.28–1.42) (n = 22) $\chi^2 = 17.86$, P = 0.658, $I^2 = 0\%$	1.16 (1.00–1.35) (n = 19) $\chi^2 = 28.49$, P = 0.055, $I^2 = 37\%$	1.44 (1.05–1.98) (n = 7) $\chi^2 = 14.76$, P = 0.022, $I^2 = 59\%$
Crude[b]	1.49 (1.38–1.61) (n = 28) $\chi^2 = 67.47$, P = 0.000, $I^2 = 60\%$	1.44 (1.33–1.56) (n = 8) $\chi^2 = 10.74$, P = 0.150, $I^2 = 35\%$	1.24 (1.00–1.57) (n = 15) $\chi^2 = 29.20$, P = 0.010, $I^2 = 52\%$	1.28 (0.90–1.81) (n = 8) $\chi^2 = 11.35$, P = 0.124, $I^2 = 38\%$
Major versus any defect				
Major	1.42 (1.29–1.56) (n = 16) $\chi^2 = 38.06$, P = 0.001, $I^2 = 61\%$	1.41 (1.33–1.50) (n = 10) $\chi^2 = 7.95$, P = 0.539, $I^2 = 0\%$	1.06 (0.83–1.34) (n = 9) $\chi^2 = 12.06$, P = 0.149, $I^2 = 34\%$	1.73 (0.54–5.60) (n = 2) $\chi^2 = 3.68$, P = 0.055, $I^2 = 73\%$
Any defect[c]	1.23 (1.13–1.33) (n = 30) $\chi^2 = 34.78$, P = 0.212, $I^2 = 17\%$	1.28 (1.17–1.39) (n = 13) $\chi^2 = 8.00$, P = 0.785, $I^2 = 0\%$	1.15 (1.00–1.33) (n = 18) $\chi^2 = 20.54$, P = 0.247, $I^2 = 17\%$	1.24 (0.95–1.63) (n = 10) $\chi^2 = 14.72$, P = 0.099, $I^2 = 39\%$
Population versus clinic-based				
Population	1.29 (1.21–1.36) (n = 19) $\chi^2 = 28.73$, P = 0.052, $I^2 = 37\%$	1.37 (1.30–1.44) (n = 14) $\chi^2 = 12.61$, P = 0.478, $I^2 = 0\%$	1.04 (0.93–1.18) (n = 12) $\chi^2 = 12.84$, P = 0.304, $I^2 = 14\%$	1.19 (1.01–1.39) (n = 6) $\chi^2 = 4.38$, P = 0.496, $I^2 = 0\%$
Clinic	1.41 (1.18–1.69) (n = 26) $\chi^2 = 42.71$, P = 0.015, $I^2 = 41\%$	1.30 (1.04–1.62) (n = 9) $\chi^2 = 6.81$, P = 0.557, $I^2 = 0\%$	1.35 (1.03–1.76) (n = 15) $\chi^2 = 17.35$, P = 0.238, $I^2 = 19\%$	1.25 (0.56–2.78) (n = 6) $\chi^2 = 13.30$, P = 0.021, $I^2 = 62\%$
Grouped by sample size				
<500	1.41 (1.14–1.73) (n = 22) $\chi^2 = 22.44$, P = 0.375, $I^2 = 6\%$	1.32 (0.85–2.04) (n = 7) $\chi^2 = 3.87$, P = 0.694, $I^2 = 0\%$	1.41 (1.07–1.87) (n = 17) $\chi^2 = 43.99$, P = 0.000, $I^2 = 66\%$	1.49 (0.84–2.66) (n = 8) $\chi^2 = 15.46$, P = 0.030, $I^2 = 55\%$
500–2000[d]	1.41 (1.21–1.64) (n = 12) $\chi^2 = 30.69$, P = 0.001, $I^2 = 64\%$	1.36 (1.17–1.58) (n = 9) $\chi^2 = 11.48$, P = 0.176, $I^2 = 30\%$	1.05 (0.94–1.16) (n = 10) $\chi^2 = 8.52$, P = 0.483, $I^2 = 0\%$	1.15 (0.97–1.36) (n = 4) $\chi^2 = 0.689$, P = 0.876, $I^2 = 0\%$
>2000	1.26 (1.19–1.34) (n = 11) $\chi^2 = 17.05$, P = 0.073, $I^2 = 41\%$	1.35 (1.28–1.43) (n = 7) $\chi^2 = 3.89$, P = 0.691, $I^2 = 0\%$		
Grouped by study quality—score out of 17				
High (quality score ≥12)	1.33 (1.26–1.41) (n = 13) $\chi^2 = 17.52$, P = 0.131, $I^2 = 31\%$	1.37 (1.30–1.45) (n = 11) $\chi^2 = 7.82$, P = 0.646, $I^2 = 0\%$	1.08 (0.93–1.25) (n = 9) $\chi^2 = 12.37$, P = 0.135, $I^2 = 35\%$	1.20 (0.95–1.51) (n = 4) $\chi^2 = 4.02$, P = 0.259, $I^2 = 25\%$
Medium (quality score 8.5–11.5)	1.29 (1.06–1.58) (n = 12) $\chi^2 = 39.98$, P = 0.000, $I^2 = 72\%$	1.28 (1.03–1.58) (n = 8) $\chi^2 = 9.71$, P = 0.206, $I^2 = 28\%$	0.99 (0.74–1.32) (n = 5) $\chi^2 = 2.34$, P = 0.674, $I^2 = 0\%$	1.16 (0.80–1.66) (n = 2) $\chi^2 = 0.33$, P = 0.564, $I^2 = 0\%$
Low (quality score ≤8)	1.34 (1.06–1.69) (n = 20) $\chi^2 = 24.68$, P = 0.171, $I^2 = 23\%$	0.98 (0.61–1.56) (n = 4) $\chi^2 = 0.09$, P = 0.993, $I^2 = 0\%$	1.34 (0.97–1.84) (n = 13) $\chi^2 = 16.66$, P = 0.163, $I^2 = 28\%$	1.25 (0.56–2.78) (n = 6) $\chi^2 = 13.30$, P = 0.021, $I^2 = 62\%$
Grouped by method of birth defect classification				
Birth defect register classification system	1.40 (1.29–1.52) (n = 15) $\chi^2 = 36.46$, P = 0.001, $I^2 = 62\%$	1.37 (1.29–1.45) (n = 11) $\chi^2 = 8.55$, P = 0.575, $I^2 = 0\%$	1.05 (0.92–1.20) (n = 9) $\chi^2 = 6.91$, P = 0.546, $I^2 = 0\%$	1.22 (0.81–1.83) (n = 3) $\chi^2 = 3.77$, P = 0.152, $I^2 = 47\%$
ICD codes	1.53 (1.02–2.31) (n = 5) $\chi^2 = 16.76$, P = 0.002, $I^2 = 76\%$	1.43 (0.95–2.16) (n = 1) $\chi^2 = 0.00$, P = 1.00, $I^2 = 0\%$	1.33 (0.76–2.33) (n = 4) $\chi^2 = 16.35$, P = 0.001, $I^2 = 82\%$	1.69 (0.75–3.82) (n = 3) $\chi^2 = 10.71$, P = 0.005, $I^2 = 81\%$

Functional impairment/surgical correction	1.29 (1.10–1.52) (n = 8) $\chi^2 = 6.46$, $P = 0.487$, $I^2 = 0\%$	1.51 (1.28–1.78) (n = 5) $\chi^2 = 2.14$, $P = 0.710$, $I^2 = 0\%$	1.18 (0.61–2.28) (n = 4) $\chi^2 = 6.07$, $P = 0.108$, $I^2 = 51\%$	–
None/not stated	1.18 (1.02–1.37) (n = 15) $\chi^2 = 9.28$, $P = 0.812$, $I^2 = 0\%$	1.06 (0.85–1.33) (n = 5) $\chi^2 = 2.46$, $P = 0.651$, $I^2 = 0\%$	1.19 (0.92–1.53) (n = 9) $\chi^2 = 3.52$, $P = 0.897$, $I^2 = 0\%$	1.12 (0.70–1.82) (n = 5) $\chi^2 = 2.61$, $P = 0.625$, $I^2 = 0\%$
Other[e]	1.11 (0.92–1.33) (n = 2) $\chi^2 = 1.42$, $P = 0.233$, $I^2 = 30\%$	1.36 (0.98–1.90) (n = 1) $\chi^2 = 0.00$, $P = 1.00$, $I^2 = 0\%$	1.06 (0.66–1.69) (n = 1) $\chi^2 = 0.00$, $P = 1.00$, $I^2 = 0\%$	1.06 (0.66–1.69) (n = 1) $\chi^2 = 0.00$, $P = 1.00$, $I^2 = 0\%$
Grouped by treatment type				
IVF versus non-ART	1.36 (1.22–1.51) (n = 17) $\chi^2 = 20.36$, $P = 0.205$, $I^2 = 21\%$	1.28 (1.15–1.41) (n = 8) $\chi^2 = 7.83$, $P = 0.348$, $I^2 = 11\%$	1.08 (0.88–1.34) (n = 10) $\chi^2 = 7.67$, $P = 0.567$, $I^2 = 0\%$	1.15 (0.65–2.02) (n = 4) $\chi^2 = 5.80$, $P = 0.122$, $I^2 = 48\%$
ICSI versus non-ART	1.37 (1.22–1.53) (n = 11) $\chi^2 = 14.72$, $P = 0.142$, $I^2 = 32\%$	1.42 (1.30–1.54) (n = 8) $\chi^2 = 6.77$, $P = 0.454$, $I^2 = 0\%$	1.36 (0.86–2.15) (n = 7)[f] $\chi^2 = 15.01$, $P = 0.020$, $I^2 = 60\%$	1.80 (0.69–4.65) (n = 3)[f] $\chi^2 = 9.15$, $P = 0.010$, $I^2 = 78\%$
Grouped by region				
Asia	1.34 (0.60–2.95) (n = 4) $\chi^2 = 13.32$, $P = 0.004$, $I^2 = 77\%$	1.17 (0.81–1.69) (n = 1) $\chi^2 = 0.00$, $P = 1.00$, $I^2 = 0\%$	1.35 (0.34–5.28) (n = 3) $\chi^2 = 12.37$, $P = 0.002$, $I^2 = 84\%$	1.35 (0.34–5.28) (n = 3) $\chi^2 = 12.37$, $P = 0.002$, $I^2 = 84\%$
Australia	1.31 (1.17–1.46) (n = 6) $\chi^2 = 8.06$, $P = 0.153$, $I^2 = 38\%$	1.34 (1.19–1.50) (n = 5) $\chi^2 = 6.12$, $P = 0.191$, $I^2 = 35\%$	1.11 (0.93–1.33) (n = 4) $\chi^2 = 0.52$, $P = 0.914$, $I^2 = 0\%$	1.08 (0.83–1.41) (n = 2) $\chi^2 = 0.00$, $P = 0.983$, $I^2 = 0\%$
Europe	1.27 (1.20–1.36) (n = 24) $\chi^2 = 27.66$, $P = 0.229$, $I^2 = 17\%$	1.38 (1.29–1.47) (n = 13) $\chi^2 = 10.74$, $P = 0.551$, $I^2 = 0\%$	1.04 (0.91–1.19) (n = 17) $\chi^2 = 16.52$, $P = 0.417$, $I^2 = 3\%$	1.25 (1.02–1.52) (n = 7) $\chi^2 = 4.15$ $P = 0.657$ $I^2 = 0\%$
Middle East	1.67 (1.23–2.28) (n = 5) $\chi^2 = 7.78$, $P = 0.100$, $I^2 = 49\%$	–	1.19 (0.84–1.67) (n = 2) $\chi^2 = 0.03$, $P = 0.870$, $I^2 = 0\%$	–
USA or Canada	1.38 (1.16–1.64) (n = 6) $\chi^2 = 3.30$, $P = 0.655$, $I^2 = 0\%$	1.36 (1.04–1.78) (n = 4) $\chi^2 = 1.90$, $P = 0.593$, $I^2 = 0\%$	1.55 (0.92–2.62) (n = 1) $\chi^2 = 0.00$, $P = 1.00$, $I^2 = 0\%$	–

[a]This column includes four studies where all multiples are grouped together and data for twins cannot be extracted separately. Based on information provided in three of these publications we estimate that higher order multiples represent <5% of infants included in this pooled estimate.

[b]Any one study may provide both an adjusted RR estimate and a crude RR estimate so could appear in both these subgroups.

[c]One study provides both an estimate for major defects only and all defects combined.

[d]For twins the sample size groupings are <500 and >500.

[e]Birth defects in these two studies are classified into categories but no reference is given to a particular system of classification that would allow replication in another study.

[f]Pooled estimate for ICSI versus non-ART multiples is heavily influenced by two smaller studies with very large RR estimates; analysis restricted to twins only influenced by one small study with very large RR estimate.
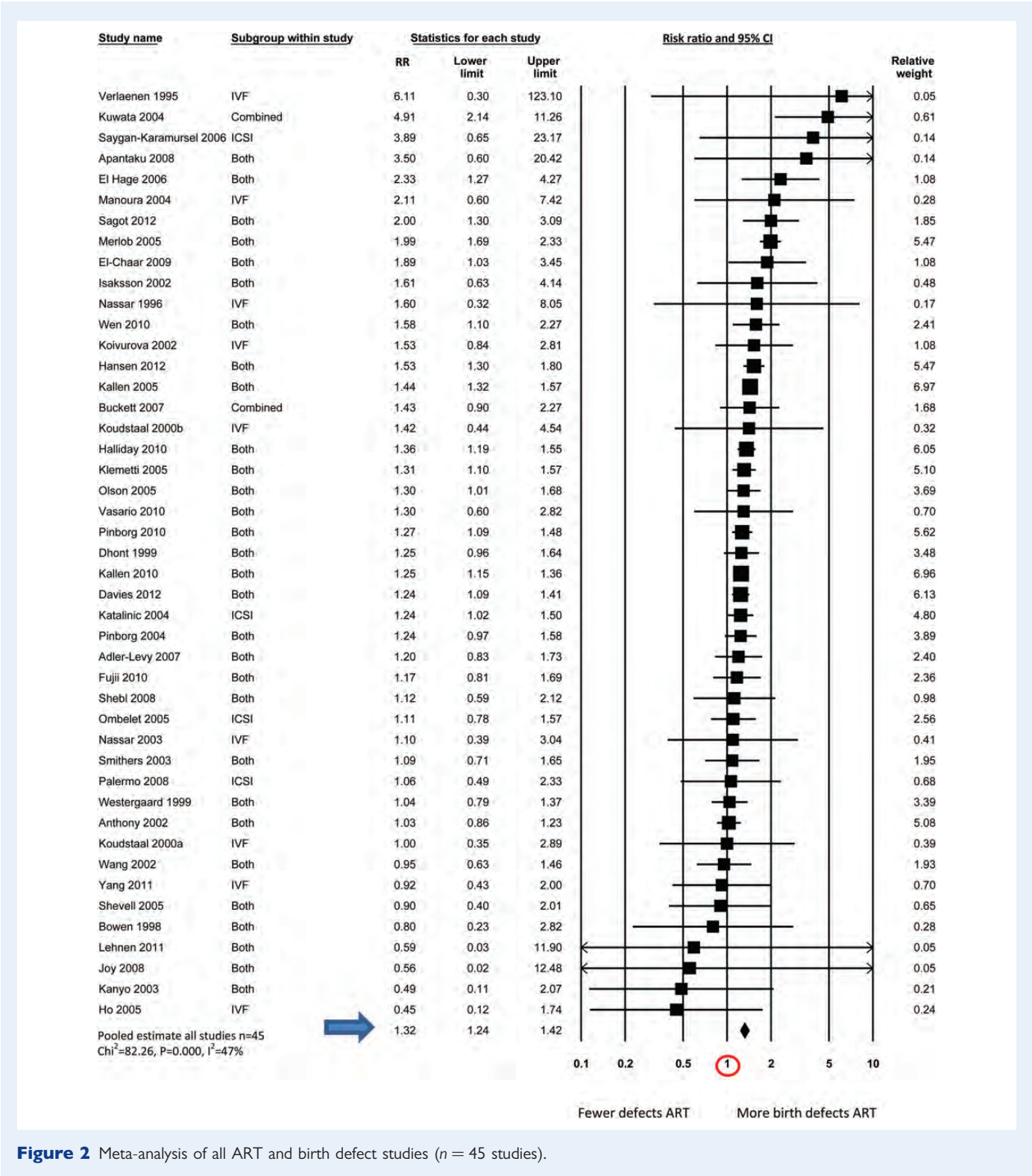
Figure 2 Meta-analysis of all ART and birth defect studies (*n* = 45 studies).

Large versus smaller studies

When we grouped all studies or studies of multiples according to sample size, larger studies had a lower pooled estimate compared with smaller studies. The results for singletons were fairly consistent across small, medium and large studies; however, the pooled estimate combining the results of the smaller studies did not reach statistical significance.

Higher quality versus lower quality studies

When studies were grouped according to methodological quality, the higher quality studies (with a quality score ≥12 out of 17) tended to show increased RR estimates for singletons but lower RR estimates for multiples. A greater proportion of singleton studies were considered high quality (48%) than studies of multiples or twins (33%).
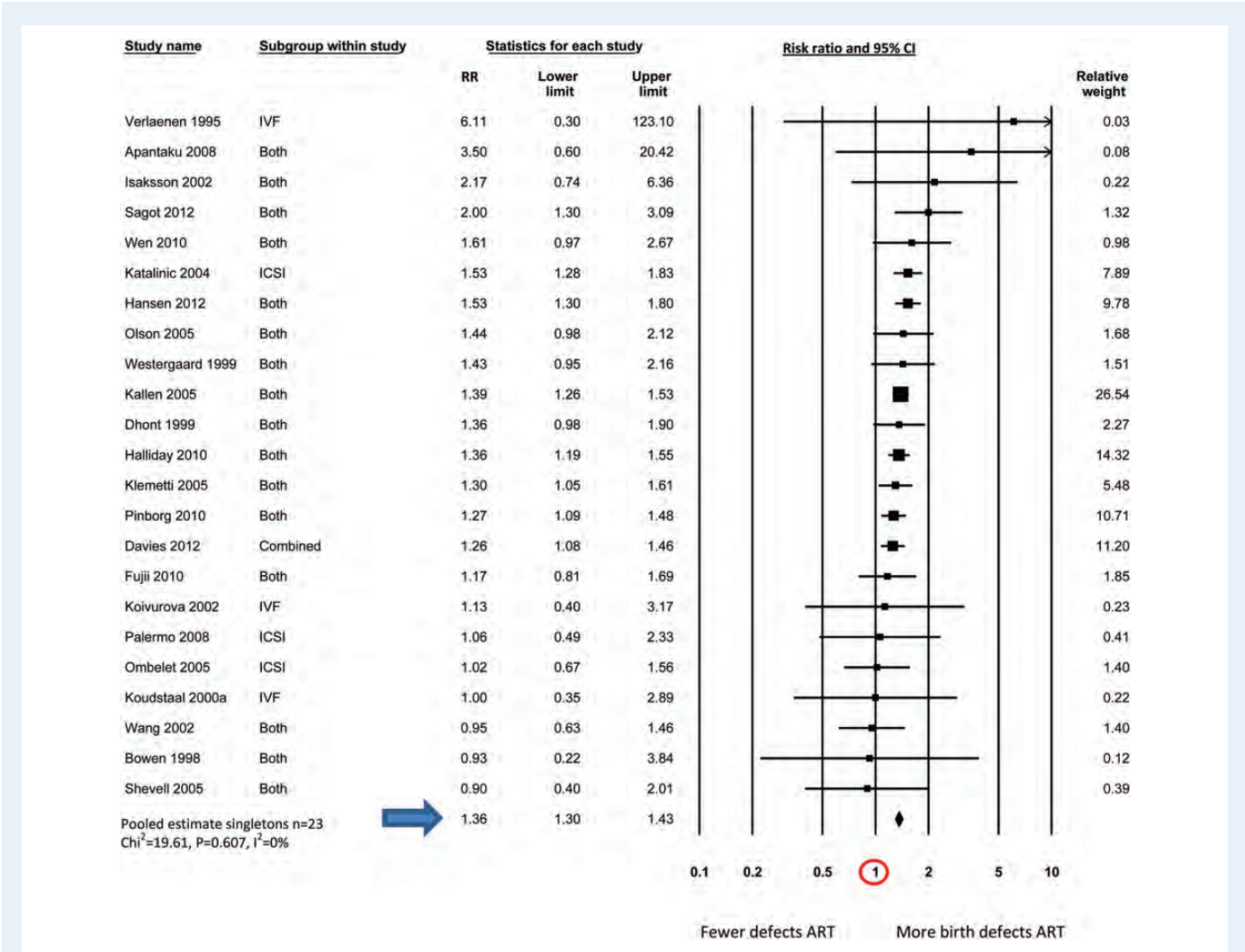
Figure 3 Meta-analysis of ART singletons and birth defects (n = 23 studies).

Studies grouped according to method of birth defect classification
One-third of studies did not provide any birth defect definition and these studies had lower pooled estimates when all such studies were grouped together (RR 1.18) and particularly when singletons were examined separately (RR 1.06). For the three defined methods of birth defect classification the results were fairly similar, showing higher pooled birth defect estimates for singletons than multiples. Many of the largest studies in the meta-analysis had used a dedicated birth defect registry classification system and the individual risk estimates from these studies ranged from 1.00 to 2.00. The pooled estimates from individual studies that reported using the 'major defects are those that cause functional impairment or require surgical correction' definition included many smaller studies where ART infants numbered <300 (75%) and the individual risk estimates ranged from 0.49 to 6.11. The studies reliant on international classification of diseases (ICD) codes showed high heterogeneity. When a small study with a very large RR estimate (Kuwata *et al.*, 2004) was removed, the pooled estimate for all studies in this ICD-based group dropped from 1.53 to 1.25 (95% CI 0.94–1.66). Similarly, the RR estimate for multiples dropped from 1.33 to 1.00 (95% CI 0.71–1.41) and

for twins where some adjustment was made for zygosity the pooled estimate dropped from 1.69 to 1.21 (95% CI 0.96–1.52), results that were more in line with the pooled estimates for studies that had used a birth defect registry classification system.

IVF versus non-ART and ICSI versus non-ART
There was little difference in the pooled estimate for studies examining IVF versus non-ART births (RR 1.36) compared with studies examining ICSI versus non-ART births (RR 1.37) when all studies were pooled. When we looked at singletons and multiples separately the pooled estimates for ICSI versus non-ART infants were higher than those for IVF versus non-ART infants although none of these differences reached statistical significance. The pooled estimate of ICSI versus non-ART multiples was influenced by two small studies with very large RR estimates (Kuwata *et al.*, 2004; Saygan-Karamursel *et al.*, 2006). Removal of these two studies gave a pooled estimate for ICSI versus non-ART multiples of 1.09 (95% CI 0.82–1.44). The pooled estimate for ICSI versus non-ART twins where some adjustment was made for zygosity also decreased (from 1.80 to 1.06)
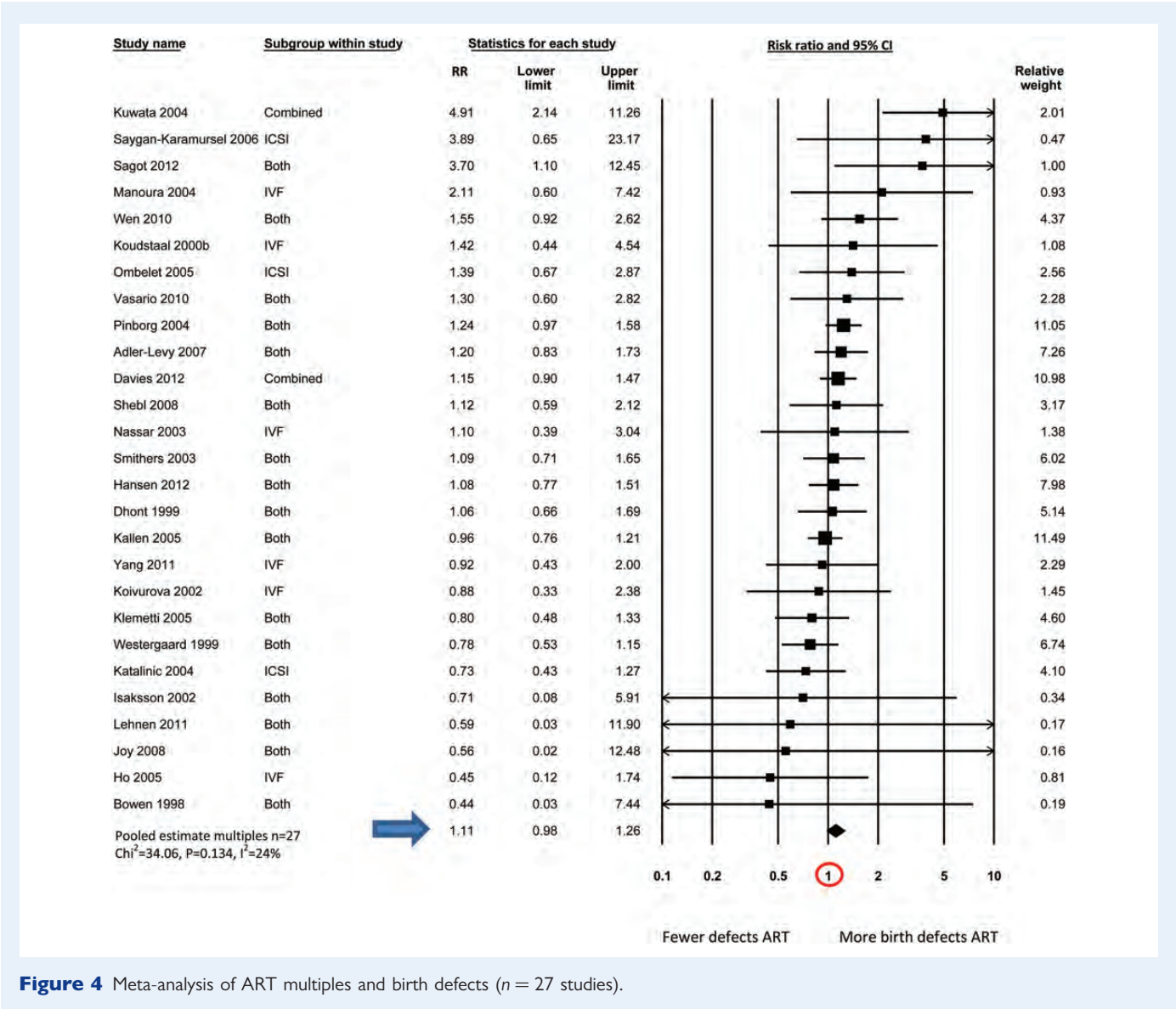
Figure 4 Meta-analysis of ART multiples and birth defects ($n = 27$ studies).

when the study by Kuwata *et al.* (2004) was removed from the analysis.

Studies grouped by region

When we grouped the results of all studies by region, the pooled estimates were similar across regions except for the Middle East (RR 1.67). The risk of birth defects in ART compared with non-ART infants was statistically significantly increased in all regions except Asia when all studies were grouped together and when studies reporting results for singletons separately were examined. The pooled RR estimates for ART versus non-ART multiples did not reach statistical significance in any region, although when we restricted the studies to those making some adjustment for differences in zygosity distribution the pooled estimate of birth defect risk in European twin studies increased from 1.04 to 1.25 (with 95% CI 1.02–1.52).

We found that grouping population-based studies and adjusted or matched data reduced heterogeneity. Heterogeneity was also reduced when we grouped studies of higher quality ($\geq 12/17$)

compared with the remainder of studies of medium or lower quality combined (data not shown). When we grouped studies by region, those from Asia showed the highest heterogeneity ($I^2 = 77\%$) and those from the USA or Canada the lowest.

Discussion

Our systematic review and meta-analysis address the question of whether an increased birth defect risk exists in ART compared with non-ART infants and whether this risk differs when singletons and multiples are examined separately. Our results suggest a 32% increased risk of birth defects in children born following ART compared with non-ART infants and this risk increases slightly when singleton births are examined separately (36%) and when the results are restricted to studies examining major birth defects only (42%). For multiple births the findings are less clear. When all multiples are grouped together, larger studies of higher quality or those assessing major birth defects only suggest no increased birth defect risk compared with
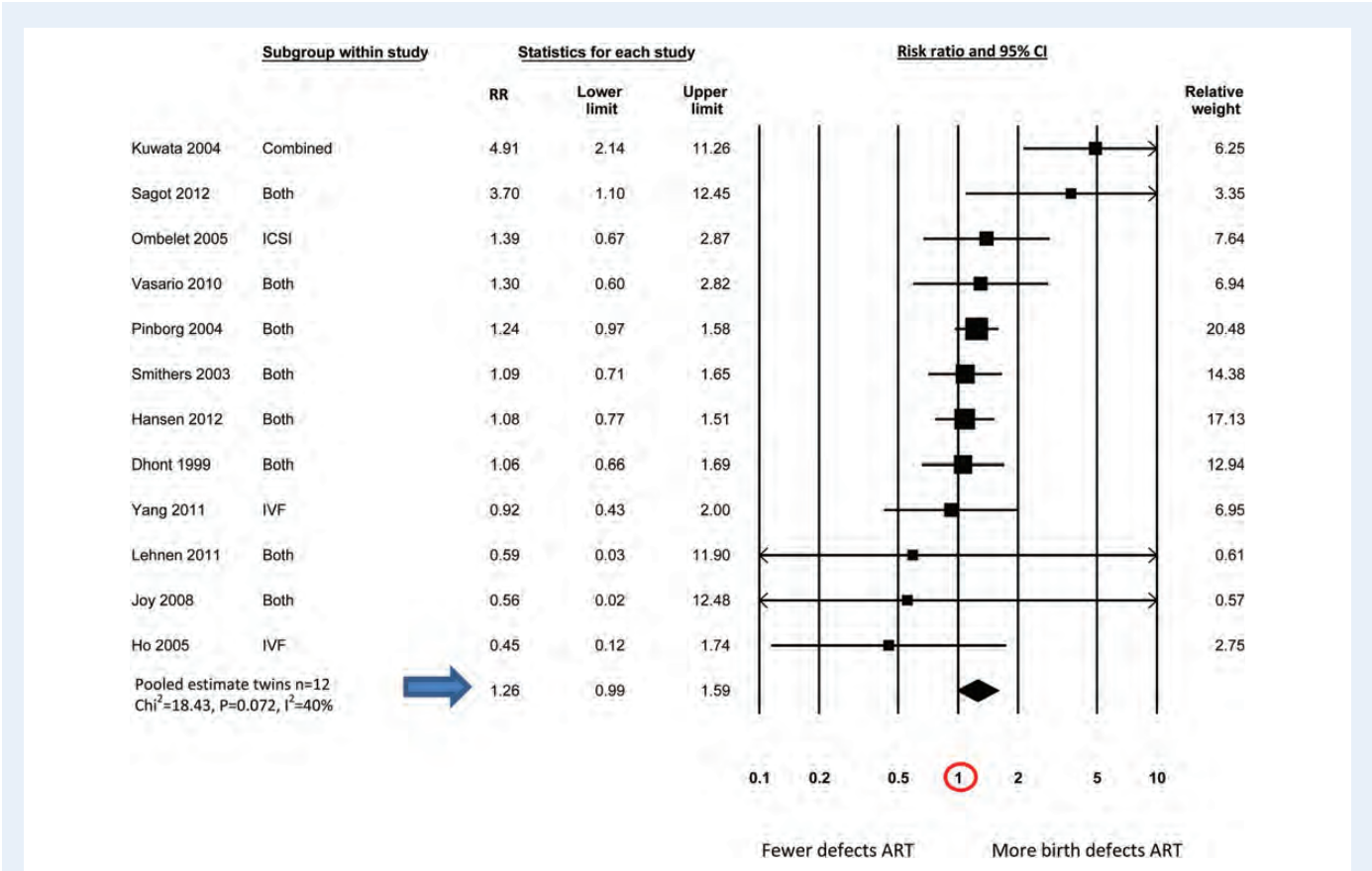
Figure 5 Meta-analysis of ART twins and birth defects where adjustment made for differences in zygosity distribution between ART and non-ART twins ($n = 12$ studies).

non-ART multiples. However, when the pooled estimate is restricted to studies of twins where some adjustment has been made for the differing proportions of monochorionic placentation in ART and non-ART twins the RR estimate increases to 1.26 (95% CI 0.99– 1.60). This finding requires confirmation when more data from larger twin studies become available since 67% of the studies included in this subgroup analysis had a small sample size (<500) and only 33% were considered of higher methodological quality.

We have pooled data from 92 671 ART infants; over 64 000 more than in our previous meta-analysis (Hansen et al., 2005) and 78% came from the 13 studies we considered of highest methodological quality for assessing birth defect risk. Half of the studies came from Europe which represents a shift from our last meta-analysis where 72% of studies were European. However, in terms of actual numbers of babies, the results are still highly dominated by Europe (73% of included infants) because of the large population-based registers available for record linkage in Sweden, Denmark and Finland. There were no studies from Asia in our previous meta-analysis.

There were fewer small studies ($n < 500$) in this meta-analysis (49%) compared with our previous meta-analysis (68%). Twin studies in general were smaller than singleton studies. In our previous meta-analysis where we had access to fewer studies that were often small in size, we chose to use a fixed effect model for combining the data because random effects models give more weight to small studies that may have extreme RR estimates. For this analysis, we

chose to use a random effects model which is more appropriate for combining the results of studies that have used a variety of different methodologies in different countries that may themselves have different underlying population rates of birth defects. It may therefore not be appropriate to assume a common effect size which is the basis of a fixed effect analysis.

Although our search strategy was similar to the one used for our previous systematic review, we tightened up our exclusion criteria so that studies were excluded if the 'ART' group included children born following OI and/or IUI. Our aim was to include only papers comparing birth defects in children born following IVF or ICSI to non-ART infants; however, we did not exclude studies where a small proportion of children were also conceived by gamete intrafallopian transfer (GIFT; $n = 7$). The other major difference in exclusion criteria was that we excluded all studies where birth defects were assessed in ART survivors aged ≥ 1 year where no information was available on stillbirths or infant deaths. These studies are essentially cross-sectional in design and often have the main aim of assessing neuro-developmental outcomes (e.g. Morin et al., 1989; Agarwal et al., 2005; Shu-Chi et al., 2006; Sanchez-Albisua et al., 2007; Banerjee et al., 2008; Knoester et al., 2008) or growth (e.g. Bonduelle et al., 2004) in ART infants rather than birth defects. Unfortunately many authors continue to publish birth defects data from these studies when it is clear that some of the most severely affected children will be missing from their sample (those that died in the neonatal

period or in the first year after birth, for example). ART infants are often identified from clinical records for these studies whilst comparison groups of non-ART infants are often selected from local schools or child care centres. Whilst this may be considered appropriate for assessing growth or neuro-developmental outcomes in 'healthy' children, it introduces the possibility of a healthy control bias for the assessment of birth defects. Children with severe birth defects who have survived up to 12 months or more of age may be unable to attend mainstream schools or child care centres and may therefore be excluded from the comparison group (Kurinczuk *et al.*, 2006).

Defining birth defects

Despite the fact that major birth defects are more likely to be consistently reported, only 36% of studies compared major birth defects, the remainder reporting results for any birth defect (major or minor). We believe it is preferable to include information on only major birth defects in studies of ART infants because: (i) these estimates avoid the further difficulties inherent in assessing and classifying minor anomalies; (ii) the notification of minor anomalies is often incomplete; (iii) it is possible that if ART infants are more closely examined than non-ART infants there will be more minor anomalies detected in this group than the natural conception group. It is less likely that this will be the case for major malformations, however, as they are more likely to be detected regardless of conception status; (iv) major birth defects are of greater clinical significance than minor defects and the inclusion of more common minor defects may 'swamp' associations of ART and more clinically important but rarer major defects (Lancaster, 1996; Simpson, 1996; Hansen *et al.*, 2005).

We understand, however, that when birth defect prevalence is assessed through record linkage to population-based registers of hospital admission, as in some of the large Nordic register studies, birth defects information is only available as ICD diagnoses and it is difficult to distinguish major defects from minor for some of these codes (e.g. ventricular septal defect, hypospadias). However, if a study involves physical examination of children or record linkage to a birth defects register where a detailed classification system is used to classify major versus minor defects then we believe reporting only major birth defects will limit the problems associated with agreement of what constitutes a birth defect.

In our subgroup analysis of birth defect classification methods we identified the three most commonly used classification systems in ART studies:

(i) 'Major birth defects are those that generally cause functional impairment or require surgical correction'. This is the birth defects definition used by Belgian researchers in their follow-up studies of ART infants (Bonduelle *et al.*, 1996, 2002). The definition is problematic in that it cannot be replicated with any confidence. There is no standard list of exclusions and since several minor defects may also undergo surgical correction (e.g. tongue tie and polydactyly) this may lead to uncertainty about whether these defects should be included. The defects that are included, however, may generally be of greater clinical significance and thus avoid some of the problems associated with under-reporting of minor birth defects.

(ii) The inclusion of any condition listed in the ICD chapter titled 'congenital anomalies'. Studies using this classification method have not linked to a dedicated birth defects register but have instead generally obtained birth defects information from hospital notes (either through direct examination of hospital records or through record linkage to a hospital register). Since the ICD code alone does not differentiate major or minor defects these studies generally report pooled major and minor results. Although these studies can be more easily replicated, they may suffer from problems associated with under- and variable reporting of minor defects. In addition the inclusion of common minor defects may 'swamp' associations of more clinically important but rarer major defects.

(iii) Classification systems used by dedicated birth defects registers. These include studies that have either linked to birth defects data collected on a dedicated birth defects register or have specified that they classified birth defects in their study according to a system used by a particular birth defects register. This would include reference to the classification system used by EUROCAT for example (http://www.eurocat-network.eu/). We have also included the Swedish register-based studies and the 2010 Pinborg study from Denmark here since they present birth defect information defined according to ICD diagnosis codes but they also present 'weeded' data where a number of more common minor conditions have been removed (e.g. pre-auricular tags, tongue tie, undescended testicles, unstable hips, single umbilical artery, and patent ductus arteriosus; Kallen *et al.*, 2005, 2010a; Pinborg *et al.*, 2010a). In general, it should be easy to replicate studies included in this category with reference to the particular classification method used, and their results should be less influenced by variation in the reporting of minor birth defects due to an explicit list of exclusions.

The results of our subgroup analysis (Table IV) indicate that although individual birth defects may have been included in some studies and excluded from others in our meta-analysis depending upon the classification method chosen, the use of these three different classification systems led to similar pooled estimates of birth defect risk. The most important 'within-study' criterion is that the 'same' classification system is used for both groups (ART and non-ART) in the study. Problems arise, for example, where a group of ART infants may be assessed using a more restrictive birth defect definition and the results compared with birth defect data for populations classified according to a more inclusive definition (Kurinczuk and Bower, 1997); such studies are not included in our meta-analysis.

Comparison with existing reviews

There are four early (Rimm *et al.*, 2004; Hansen *et al.*, 2005; McDonald *et al.*, 2005a, b) and three recently published meta-analyses (Rossi and D'Addario, 2011; Pandey *et al.*, 2012; Wen *et al.*, 2012) that compare birth defects in ART and non-ART infants. Just as for individual studies, the meta-analyses have used different methodologies resulting in different groups of studies being selected for inclusion.

The meta-analyses of McDonald *et al.* (2005a, b), Pandey *et al.* (2012) and Rossi (Rossi and D'Addario, 2011) were all interested in assessing a range of obstetric and perinatal outcomes in ART compared with non-ART infants with a preference (in the McDonald

and Pandey meta-analyses) for matched cohort studies. Although they include pooled estimates of birth defect data—birth defects were not a primary outcome so their search for birth defects data was not exhaustive and their results include only a subset of the available literature. McDonald et al. report a pooled estimate for seven studies of ART singletons (4031 infants) of 1.41 (95% CI 1.06−1.88) and a pooled estimate for four ART twins studies (n = 2681 infants) of 1.14 (95% CI 0.85−1.52). Pandey et al. report a slightly higher pooled estimate for seven studies of singletons (4382 IVF/ICSI pregnancies) RR 1.67 (95% CI 1.33−2.09), however, it is not clear why birth defects data were not also extracted from at least three other studies included in their meta-analysis (Westergaard et al., 1999; Katalinic et al., 2004; Pinborg et al., 2010a). Rossi and D'Addario report a pooled estimate for four 'controlled' ART twin studies of 1.15 (95% CI 0.80−1.63). They also report a pooled estimate for two studies including unlike-sex twins of 1.38 (95% CI 0.99−1.93; Rossi and D'Addario, 2011).

The meta-analysis by Rimm et al. (2004) combined the results of 19 studies (n = 35 578 infants) assessing major birth defects only. They found a pooled estimate of 1.29 (1.01−1.67), identical to our pooled estimate of 25 studies published soon after (Hansen et al., 2005). Our earlier meta-analysis included 12 studies that were not in the Rimm et al. meta-analysis [10 because of different exclusion criteria (Morin et al., 1989; Tanbo et al., 1995; Fisch et al., 1997; Addor et al., 1998; Dhont et al., 1999; Koudstaal et al., 2000a, b; Koivurova et al., 2002; Wang et al., 2002; Sutcliffe et al., 2003) and two that may have been missed in their search (Nassar et al., 1996; Zuppa et al., 2001)]. They included six studies that we did not [two published after our literature search (Place and Englert, 2003; Pinborg et al., 2004), two that we excluded based on overlapping data with Ericson and Kallen (2001) (Wennerholm et al., 1998, 2000), one we considered to have an inappropriate comparison group (Palermo et al., 2000), and a published letter (Merlob and Fisch, 2002) where we preferred to use the earlier more detailed published paper from the same group (Fisch et al., 1997)].

The recently published meta-analysis by Wen et al. (2012) includes 46 studies (124 468 infants), 25 of which are excluded from our current meta-analysis for the following reasons: 10 papers were essentially of cross-sectional design assessing birth defects in survivors to a year or more (Morin et al., 1989; D'Souza et al., 1997; Sutcliffe et al., 2003; Bonduelle et al., 2004, 2005; Agarwal et al., 2005; Belva et al., 2007; Sanchez-Albisua et al., 2007; Knoester et al., 2008), or in one case in full-term singletons with birthweight >2500 g (Place and Englert, 2003); three studies examined specific types of birth defects rather than all birth defects [Tararbit et al., 2011 examined heart defects only, Silver et al. (1999) examined hypospadias only, and Reefhuis et al. (2009) performed a case−control study of select birth defects only]; five papers had overlapping or identical data to other included papers (Zhu et al., 2006 overlaps completely with Pinborg et al., 2010a; Ludwig and Katalinic, 2002 includes exactly the same ICSI data as Katalinic et al., 2004; Lambalk and van Hooff, 2001 overlaps with Anthony et al., 2002; Bergh et al., 1999 overlaps with Kallen et al., 2005; Hansen et al., 2002 overlaps with Hansen et al., 2012) three papers where the total number of birth defects were assessed rather than birth defects per child (Saunders et al., 1996) or data were presented per woman rather than per child (Al-Fifi et al., 2009; Sala et al., 2011); two papers with mixed exposure groups

(Zadori et al., 2003; Welmerink et al., 2010); one paper with an inappropriate comparison group (Allen et al., 2008); and a Chinese reference that we did not pick up in our literature search (Liu and He, 2010). Our meta-analysis includes 24 papers that were not included in the Wen paper—12 of which concerned twins only, suggesting perhaps their search strategy was not picking up studies of ART twins. The pooled estimate for all 46 studies included in the Wen meta-analysis was 1.37 (95% CI 1.26−1.48) with a high heterogeneity (Wen et al., 2012). Heterogeneity remained high for most of their subgroup analyses probably because their inclusion criteria were broader than ours and they did not pool data according to plurality.

Table V shows that despite the many differences in inclusion/exclusion criteria particularly in the more recent meta-analyses, the results have been remarkably consistent. This may in part be explained by the inclusion of the large population-based series from Sweden, Denmark, Finland and Germany in the larger meta-analyses (Rimm et al., 2004; Hansen et al., 2005; Wen et al., 2012). The Swedish data (Kallen et al., 2005, 2010a) contribute 34% of all ART infants in the current meta-analysis for example.

Cause of increased birth defect risk: ART treatment and underlying infertility

An excess risk of birth defects in ART infants is biologically plausible. Factors associated with treatment that may increase the risk of birth defects include the underlying causes of infertility; and aspects of the ART procedures themselves such as the medications used, culture media composition, length of time in culture, freezing and thawing of embryos, altered hormonal environment at the time of implantation, the manipulation of gametes and embryos or a combination of these. The complexity of ART treatment renders the identification of individual risk factors extremely challenging. It is very difficult to obtain information about an individual aspect of ART treatment and its association with a rare outcome, such as birth defects, whilst holding all the other variables constant in sufficiently large studies. There is growing evidence, however, that some of these factors such as the transfer of frozen−thawed embryos or the use of different culture media can have an impact on other more common adverse perinatal outcomes including low-birthweight or preterm birth (Dumoulin et al., 2010; Pinborg et al., 2012a). A recent Australian study also suggested a difference in the risk of severe birth defects that arise during blastogenesis depending upon whether fresh or frozen−thawed embryos were transferred (Halliday et al., 2010).

Subfertility is known to increase risks of adverse perinatal outcome (Ghazi et al., 1991; Thomson et al., 2005). In an attempt to tease out the contribution of subfertility versus ART treatment to adverse perinatal outcomes in ART singletons, two studies have examined preterm birth in ART and naturally conceived singletons born to the same parents (Romundstad et al., 2008; Henningsen et al., 2011a). When meta-analysis was used to combine the results, the adjusted risk of preterm birth in ART versus non-ART siblings was 1.27 (95% CI 1.08−1.49) demonstrating a significantly higher risk of preterm birth in the ART sibling (Pinborg et al., 2012a). In addition a meta-analysis of studies comparing preterm birth in ART singletons to naturally conceived singletons born to subfertile couples [time to pregnancy (TTP) >1 year] showed a significantly higher risk of preterm birth in the ART singletons [OR 1.55 (95% CI 1.30−1.85)], thus indicating a risk

Table V Comparison of our results with other published meta-analyses.

	Number of studies	Number of ART infants included	Pooled estimate (95% CI)	Differences in inclusion criteria—shown for Rimm *et al.* and Wen *et al.* only	
				Number of studies excluded from our meta-analyses	Number of studies included in our meta-analyses but not in other
Overall					
Early meta-analyses					
Rimm *et al.* (2004)	19	35 578	1.29 (1.01–1.67)	6	
Hansen *et al.* (2005)	25	28 638	1.29 (1.21–1.37)		12
Recent meta-analyses					
Wen *et al.* (2012)	46	124 468	1.37 (1.26–1.48)	25	
Current study	45	92 671	1.32 (1.24–1.42)		24
Singletons					
Early meta-analyses					
Rimm *et al.* (2004)					
IVF	8	2064	1.51 (0.85–2.7)		
ICSI	6	3948	1.33 (0.90–1.95)		
Hansen *et al.* (2005)	15	13 059	1.31 (1.17–1.46)		
McDonald *et al.* (2005a, b)	7	4031	1.41 (1.06–1.88)		
Recent meta-analyses					
Pandey *et al.* (2012)	7	4382	1.67 (1.33–2.09)		
Current study	23	48 944	1.36 (1.30–1.43)		
Multiples					
Early meta-analyses					
Rimm *et al.* (2004)					
IVF	7	5561	0.92 (0.75–1.12)		
ICSI	4	3197	1.18 (0.60–2.37)		
McDonald *et al.* (2005a, b)	4	2681	1.14 (0.85–1.52)		
Recent meta-analyses					
Rossi and D'Addario (2011)					
Adjusted	4	1556	1.15 (0.80–1.63)		
Current study	27	19 361	1.11 (0.98–1.26)		
Twins—some adjustment for zygosity					
Rossi and D'Addario (2011)	2	1170	1.38 (0.99–1.93)		
Current study	12	5780	1.26 (0.99–1.60)		

attributable to the ART treatment itself (Pinborg *et al.*, 2012a). These results suggest that subfertility, although important, is not the only contributor to poorer perinatal outcome.

There is growing evidence that subfertile couples who manage to conceive without the use of ART also have an increased risk of birth defects (Zhu *et al.*, 2006; Jaques *et al.*, 2010; Davies *et al.*, 2012a), but again it is not known whether subfertility explains all of the increased risk seen in ART compared with non-ART infants or whether the treatments themselves further increase that risk. A single study has directly compared birth defect prevalence in ART infants to infants born to subfertile couples (TTP > 1 year) who conceived naturally (Zhu *et al.*, 2006). The results of this study indicate that a mixed ART exposure group including IVF, ICSI, but also IUI,

hormonal treatment, surgery and 'alternatives' (treatments not usually pooled under the heading 'ART') have an increased risk of 1.17 (95% CI 1.00–1.36) of being diagnosed with a birth defect, and the risk for ICSI infants separately was further increased [adjusted hazard ratio 1.57 (95% CI 1.11–2.23)].

The recent study by Davies *et al.* (2012a) included a group of 'infertile' couples who had conceived naturally and this group had an increased risk of birth defects compared with the fertile naturally conceiving group. The authors caution, however, that the subfertile group may have included women who received OI treatment outside the ART clinic setting (Marino *et al.*, 2011; Davies *et al.*, 2012a). In response to a letter from Rimm and Katayama (2012a) suggesting that there was no crude increase in birth defect risk in their study if the

data for ART infants were compared with this 'infertile' naturally conceiving subgroup, Davies et al. suggested that due to uncertainty about the proportion of women in this subgroup that had truly conceived naturally, a far better comparison would involve birth defects diagnosed in ART and non-ART infants born to the same mother. An initial look at this comparison yielded a crude OR for ART siblings compared with non-ART siblings of 1.50 (95% CI 1.08–2.09), suggesting that when children born to 'the same' subfertile mother are considered, an excess risk associated with ART remains (Davies et al., 2012b).

Contrary to the conclusions of Rimm et al. (2011) and Rimm and Katayama (2012b) that ascribe all increase in birth defect risk in ART infants to patient subfertility we suggest that the evidence from these two studies indicates that 'both' infertility and ART treatment factors may contribute to birth defect risk, just as they do for other adverse perinatal outcomes in ART infants.

Many authors have argued that it is inappropriate to compare children born to couples undergoing ART treatment with naturally conceived children and there have been various attempts at examining alternative comparison groups such as couples undergoing less invasive IUI or OI treatments; comparing health outcomes in ART infants to those in children born to parents who have taken a long time to conceive; or making some adjustment for years of involuntary childlessness.

Unfortunately none of these alternatives are themselves problem-free. Data on years of involuntary childlessness, for example, are often missing for large proportions of the population. In a recent Swedish study (Sazonova et al., 2012), years of involuntary childlessness was missing for up to 31% of ART infants and in a Finnish study examining obstetric outcomes in women with a long TTP, this information was only available for 47% (Raatikainen et al., 2012). In addition to the issue of missing data, subfertile couples who conceive naturally may be quite different from couples requiring ART treatment to conceive. Many couples who proceed to ART treatment, such as those where the female partner has blocked or absent fallopian tubes or is unable to ovulate, or where the male partner has very low or no sperm in his ejaculate, would have little or no chance of conceiving naturally. Other differences may also exist. A recent study by Raatikainen et al. (2012) found that women who conceived naturally with a long TTP (≥2 years) were more likely to be primiparous, and to have had a miscarriage or an induced abortion compared with women conceiving with ART. They were also more likely to be overweight, to have smoked before and also during pregnancy and to have drunk alcohol before pregnancy than women who conceived with ART.

Rather than discounting all efforts at research in this field that do not include a subfertile comparison group (Rimm and Katayama, 2012b), we think it is more constructive to identify questions that may be answered with the types of data that are more commonly available. We need to know how different types of ART exposure and different types of underlying infertility contribute to birth defect risk. If we find that more children with birth defects are born following the use of one technique compared with another there may be options to reduce this risk. For example, if vitrification gave rise to more birth defects than slow-freezing because of exposure to high levels of cryoprotectant we could revert to slow-freezing. If ICSI performed in cases of severe male factor infertility leads to greater birth

defect risk then we may not be able to influence this risk except through patient counselling, the use of preimplantation genetic diagnosis (PGD) and antenatal screening. But if the technique itself appears to increase birth defect risk (across a broad section of underlying infertility types), then we can at least prevent its use in situations where it is not clinically indicated. To answer these questions subgroup analyses within ART cohorts will become increasingly important rather than comparisons to non-ART births. Very large studies are needed to examine birth defect risk in multiple subgroups and perhaps information from the combined Nordic databases may be used to address some of these issues in the future (Henningsen et al., 2011b). Data from smaller studies should not be discounted, however, as they can be combined using meta-analysis provided the results are clearly reported according to similarly defined subgroups.

Counselling prospective patients

When counselling couples prior to ART treatment we have previously suggested that the relevant information is not birth defect risk once the effect of subfertility has been removed (if indeed this can be accurately quantified), but the overall risk of having a child with a birth defect if they do use ART (Hansen et al., 2011). The results of this and other recent meta-analyses confirm the earlier reported increase of 30–40%. For a population with a background birth defect prevalence of 5% this equates to an absolute risk of 6.5–7.0%. Alongside this information it would certainly be worth mentioning that the risk of birth defects also appears increased in subfertile couples who do not use ART but conceive naturally. We would, however, strongly caution against inferring that the entire risk of birth defects following ART can be attributed to 'subfertility' (Ludwig, 2012; Rimm and Katayama, 2012b). ART is a rapidly changing field and the data upon which we are basing our estimates of risk may not be representative of the techniques and clinical practice in most frequent use today, or the reasons for undergoing treatment.

Limitations

This meta-analysis groups data that cover a long time period in a field that is moving extremely quickly. New techniques or adjustments to laboratory conditions are frequently being introduced and essentially changing the characteristics of the exposure as well as the case-mix of people who access treatment. Few studies have examined whether the prevalence of birth defects in a cohort of ART infants has changed over time. Our own work in Western Australia (Hansen et al., 2012) and data from the large Swedish series (Kallen et al., 2010a) suggest a drop in birth defect prevalence although this is not reflected in the results of our meta-analysis which are similar to those of our earlier meta-analysis (Hansen et al., 2005). A drop in birth defect prevalence is plausible given data from the large Swedish registers suggest important improvements over time in other perinatal outcomes including low-birthweight, preterm birth and small for gestational age in ART infants (Kallen et al., 2010b). Whilst these changes can for the most part be attributed to a large decline in the multiple pregnancy rate which has occurred following the shift towards double and then single embryo transfer (SET) in Sweden, improvements in perinatal outcomes have also been reported for singletons only (Kallen et al., 2010b; Finnstrom et al., 2011; Sazonova et al., 2011). Potential causes for improved perinatal

outcomes and declining birth defect rates include changes to patient-mix over time with couples seeking treatment earlier and for a broader range of infertility diagnoses in more recent years (Kallen et al., 2010b). Other factors may include changes to clinical practice and culture media, more optimal culture conditions and milder stimulation protocols potentially contributing to improvements in embryo quality and/or uterine receptivity (Hansen et al., 2012). Although the results of the two studies (Kallen et al., 2010a; Hansen et al., 2012) suggesting a decline in birth defect prevalence over time are encouraging, we caution that insufficient birth defect data exist to ascertain risk for many of the newer techniques in common use in recent years such as blastocyst culture and vitrification.

In order to provide better information for patient counselling and to determine whether some ART exposures carry higher risk than others we have suggested that subgroup analyses restricting comparisons to include more homogenous exposure groups will become increasingly important. We have shown that grouping data by plurality highlights important differences in the magnitude of birth defect risk for singleton and multiple births. However, we cannot yet answer important questions about birth defect risk following frozen embryo versus fresh embryo transfer, blastocyst versus cleavage stage transfer, singletons born following SET versus DET, vitrification versus slow-freezing, ICSI with non-ejaculated versus ejaculated sperm, PGD/preimplantation genetic screening and in vitro maturation of oocytes, for example. To our knowledge there has not been a single large population-based study examining birth defect prevalence following the use of different types of culture media.

In addition to refining our subgroups according to specific types of ART treatment and underlying infertility, it may become increasingly important to examine data pooled according to region. There are important regional differences in a range of ART treatment exposures. ICSI, for example, is used in 85–96% of cycles in Latin America and the Middle East, but less so in Australia, Europe and the USA (66–73%). SET varies dramatically from 13 to 14% of cycles in Latin America and the USA to 70% in Australia. In Europe, SET is used in 22.4% of cycles although there is considerable variation between individual countries; in Sweden, for example, SET is used in 70–80% of cycles. There is also much variation in the use of blastocyst transfer, representing only 10% of cycles in Latin America, 32% in the UK, 36% in the USA and 50% in Australia (Zegers-Hochschild et al., 2009; Centers for Disease Control and Prevention, 2011; Nygren et al., 2011; Wang et al., 2011; Ferraretti et al., 2012). We grouped studies by region in a subgroup analysis but there were probably too few studies in all regions except Europe to allow for patient counselling based on regional data.

IVF versus ICSI

Although we have shown comparisons of IVF infants to non-ART infants and ICSI infants to non-ART infants in subgroup analyses, our meta-analysis does not specifically address the question of whether birth defects are more common in ICSI compared with IVF infants. There are two meta-analyses that do address this question (Lie et al., 2005; Wen et al., 2012) and both suggest no significant difference in risk between the two techniques. Lie et al. combined the results of four studies to obtain pooled estimates of birth defect risk in ICSI compared with IVF infants of 1.12 (95% CI 0.97–1.28). Wen

et al. combined the results of 24 studies to report a pooled estimate of 1.05 for IVF births compared with ICSI births which converts to an estimate of 0.95 (95% CI 0.83–1.10) for ICSI versus IVF infants.

We have not systematically reviewed these studies to determine whether all the relevant literature is included or to judge the methodological quality of the included studies so we cannot make detailed comment on the results of these meta-analyses. The results are in contrast to a recently published population-based study from Australia which found a significantly increased risk of birth defects in ICSI infants compared with IVF (OR 1.47, 95% CI 1.15–1.89; Davies et al., 2012a). We stress, however, that the latter was the only study included in our meta-analysis which reported a significantly increased birth defect risk in ICSI compared with IVF infants and the study included information on births 1986–2002. Pinborg et al. (2012b) have reviewed potential reasons for differences between the Davies et al. results and the rest of the literature. They suggest that more recent data (available in the Nordic countries) may show a lower birth defect risk for ICSI infants because the technique is now used to treat a broader range of infertile couples rather than being restricted to those with severe male infertility (Pinborg et al., 2012b).

Why does risk differ for ART multiples?

There are a number of potential reasons why the association between the use of ART and birth defects appears to be weaker for studies comparing ART and non-ART multiples. ART multiples may be less frequently exposed to the early loss of one fetus than ART singletons (the majority of ART twins being born following the transfer of two embryos). Estimates in the literature suggest that up to 10.4% of ART singletons originated from a twin gestation in early pregnancy (Pinborg et al., 2005; Shebl et al., 2008a, b). Where early pregnancy loss of one fetus does occur it has been shown to have negative consequences for the surviving fetus in terms of birthweight, growth and neurological findings (Pinborg et al., 2005, 2007; Pinborg, 2010b) and an impact on birth defect prevalence has also been shown (Pharoah et al., 2009). Comparisons of ART and non-ART twins may be influenced to a greater extent by contamination of the 'non-ART' comparison group with babies born following IUI or OI which may themselves lead to small increases in birth defect risk (Reefhuis et al., 2011). Finally, many studies have failed to consider differences in zygosity distribution between ART and non-ART twins. ART twins are much more likely to be dizygous compared with non-ART twins and dizygous twins have a reduced risk of birth defects compared with monozygous twins. Our subgroup analysis suggests that making some adjustment for differences in zygosity may increase estimates of birth defect risk, however, these results need confirmation in larger studies of high quality.

Conclusions

This meta-analysis includes birth defect information on 92 671 ART infants, which represent <2% of the estimated 5 million babies born from these techniques worldwide. It took 25 years to reach agreement that an increased birth defect risk exists in ART infants (Rimm et al., 2004; Hansen et al., 2005), partly because the magnitude of the risk is small, but also because of the small size of many of the earlier studies and the reluctance to conclude that an increased risk existed in the absence of statistical significance. We need to speed

up our assessment of newer techniques and meta-analyses provide an important method of doing this, provided study results are published in such a way that they can be pooled. If authors in this area publish detailed information about major birth defects for different subgroups of ART exposure (ICSI, IVF, singletons, multiples, unlike-sex twins, fresh versus frozen transfer, slow-freezing versus vitrification, blastocyst versus cleavage stage transfer) and (where available) according to different types of underlying infertility, meta-analyses of these results may help to highlight the safest treatment options to minimize birth defect risk. When enough information is available the results may also allow for region- and treatment-specific counselling on birth defect risk following different forms of ART.

Supplementary data

Supplementary data are available at http://humupd.oxfordjournals.org/.

Authors' roles

M.H. performed the literature search and extracted the data; M.H. and C.B. performed the quality assessment of included studies; M.H. performed the analyses and wrote the initial draft of the manuscript. All authors contributed to the design, interpretation of the results and critical revision of the manuscript.

Funding

Conflict of interest

None declared.

References

Addor V, Santos-Eggimann B, Fawer CL, Paccaud F, Calame A. Impact of infertility treatments on the health of newborns. *Fertil Steril* 1998;**69**:210–215.

Adler-Levy Y, Lunenfeld E, Levy A. Obstetric outcome of twin pregnancies conceived by in vitro fertilization and ovulation induction compared with those conceived spontaneously. *Eur J Obstet Gynecol Reprod Biol* 2007;**133**:173–178.

Agarwal P, Loh SK, Lim SB, Sriram B, Daniel ML, Yeo SH, Heng D. Two-year neurodevelopmental outcome in children conceived by intracytoplasmic sperm injection: prospective cohort study. *Br J Obstet Gynecol* 2005;**112**:1376–1383.

Al-Fifi S, Al-Binali A, Al-Shahrani M, Shafiq H, Bahar M, Almushait M, Sobandi L, Eskandar M. Congenital anomalies and other perinatal outcomes in ICSI vs. naturally conceived pregnancies: a comparative study. *J Assist Reprod Genet* 2009;**26**:377–381.

Allen C, Bowdin S, Harrison RF, Sutcliffe AG, Brueton L, Kirby G, Kirkman-Brown J, Barrett C, Reardon W, Maher E. Pregnancy and perinatal outcomes after assisted reproduction: a comparative study. *Ir J Med Sci* 2008;**177**:233–241.

Anthony S, Buitendijk SE, Dorrepaal CA, Lindner K, Braat DD, den Ouden AL. Congenital malformations in 4224 children conceived after IVF. *Hum Reprod* 2002;**17**:2089–2095.

Apantaku O, Chandrasekaran I, Bentick B. Obstetric outcome of singleton pregnancies achieved with in vitro fertilisation and intracytoplasmic sperm injection: experience from a district general hospital. *J Obstet Gynaecol* 2008;**28**:398–402.

Aslan H, Cebeci A, Yildirim G, Ceylan Y. Neonatal outcomes of twin pregnancies after assisted reproductive techniques. *J Turkish German Gynecol Assoc* 2005;**6**:111–113.

Banerjee I, Shevlin M, Taranissi M, Thornhill A, Abdalla H, Ozturk O, Barnes J, Sutcliffe A. Health of children conceived after preimplantation genetic diagnosis: a preliminary outcome study. *Reprod BioMed Online* 2008;**16**:376–381.

Baxi A, Kaushal M. Outcome of twin pregnancies conceived after assisted reproductive techniques. *J Hum Reprod Sci* 2008;**1**:25–28.

Belva F, Henriet S, Liebaers I, Van Steirteghem A, Celestin-Westreich S, Bonduelle M. Medical outcome of 8-year-old singleton ICSI children (born > or =32 weeks' gestation) and a spontaneously conceived comparison group. *Hum Reprod* 2007;**22**:506–515.

Beral V, Doyle P. Births in Great Britain resulting from assisted conception, 1978–87 (MRC Working Party on Children Conceived by In Vitro Fertilisation). *Br Med J* 1990;**300**:1229–1233.

Berg I, Finnstrom O, Nygren KG. [Children born after in vitro fertilization in Sweden 1982–1997. Small, but significant increase of risks among test-tube children]. *Lakartidningen* 2001;**98**:4020–4021, 4024–4025.

Bergh T, Ericson A, Hillensjo T, Nygren KG, Wennerholm UB. Deliveries and children born after in-vitro fertilisation in Sweden 1982–95: a retrospective cohort study. *Lancet* 1999;**354**:1579–1585.

Bonduelle M, Wilikens A, Buysse A, Van Assche E, Wisanto A, Devroey P, Van Steirteghem AC, Liebaers I. Prospective follow-up study of 877 children born after intracytoplasmic sperm injection (ICSI), with ejaculated epididymal and testicular spermatozoa and after replacement of cryopreserved embryos obtained after ICSI. *Hum Reprod* 1996;**11**:131–155.

Bonduelle M, Liebaers I, Deketelaere V, Derde MP, Camus M, Devroey P, Van Steirteghem A. Neonatal data on a cohort of 2889 infants born after ICSI (1991–1999) and of 2995 infants born after IVF (1983–1999). *Hum Reprod* 2002;**17**:671–694.

Bonduelle M, Bergh C, Niklasson A, Palermo GD, Wennerholm UB. Medical follow-up study of 5-year-old ICSI children. *Reprod BioMed Online* 2004;**9**:91–101.

Bonduelle M, Wennerholm UB, Loft A, Tarlatzis BC, Peters C, Henriet S, Mau C, Victorin-Cederquist A, Van Steirteghem A, Balaska A *et al.* A multi-centre cohort study of the physical health of 5-year-old children conceived after intracytoplasmic sperm injection, in vitro fertilization and natural conception. *Hum Reprod* 2005;**20**:413–419.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. United Kingdom: John Wiley & Sons Ltd, 2009.

Bowen JR, Gibson FL, Leslie GI, Saunders DM. Medical and developmental outcome at 1 year for children conceived by intracytoplasmic sperm injection. *Lancet* 1998;**351**:1529–1534.

Brandes JM, Scher A, Itzkovits J, Thaler I, Sarid M, Gershoni-Baruch R. Growth and development of children conceived by in vitro fertilisation. *Pediatrics* 1992;**90**:424–429.

Buckett WM, Chian RC, Holzer H, Dean N, Usher R, Tan SL. Obstetric outcomes and congenital abnormalities after in vitro maturation, in vitro fertilization, and intracytoplasmic sperm injection. *Obstet Gynecol* 2007;**110**:885–891.

Cadman J, Richards B, D'Souza S, Lieberman BA, Buck P, Rivlin E. A computerised study on the results of in-vitro-fertilisation. *Stud Health Technol Inform* 1999;**68**:343–346.

Caserta D, Marci R, Tatone C, Schimberni M, Vaquero E, Lazzarin N, Fazi A, Moscarini M. IVF pregnancies: neonatal outcomes after the new Italian law on assisted reproduction technology (law 40/2004). *Acta Obstet Gynecol Scand* 2008;**87**:935–939.

Cederblad M, Friberg B, Ploman F, Sjoberg NO, Stjernqvist K, Zackrisson E. Intelligence and behaviour in children born after in-vitro fertilisation treatment. *Hum Reprod* 1996;**11**:2052–2057.

Centers for Disease Control and Prevention, American Society for Reproductive Medicine, Society for Assisted Reproductive Technology. 2009 Assisted Reproductive Technology Success Rates: National Summary and Fertility Clinic Reports. Atlanta, Centers for Disease Control and Prevention, 2011.

Chou HC, Tsao PN, Yang YS, Tang JR, Tsou KI. Neonatal outcome of infants born after in vitro fertilization at National Taiwan University Hospital. *J Formosan Med Assoc* 2002;**101**:203–205.

Daniel Y, Ochshorn Y, Fait G, Geva E, Bar-Am A, Lessing JB. Analysis of 104 twin pregnancies conceived with assisted reproductive technologies and 193 spontaneously conceived twin pregnancies. *Fertil Steril* 2000;**74**:683–689.

Davies MJ, Moore VM, Willson KJ, Van Essen P, Priest K, Scott H, Haan EA, Chan A. Reproductive technologies and the risk of birth defects. *N Engl J Med* 2012a;**366**:1803–1813.

Davies MJ, Moore VM, Haan EA. Reproductive technologies and the risk of birth defects: the authors reply. *N Engl J Med* 2012b;**367**:875–876.

Desmyttere S, De Schepper J, Nekkebroeck J, De Vos A, De Rycke M, Staessen C, Liebaers I, Bonduelle M. Two-year auxological and medical outcome of singletons born after embryo biopsy applied in preimplantation genetic diagnosis or preimplantation genetic screening. *Hum Reprod* 2009;**24**:470–476.

Dhont M, De Neubourg F, Van der Elst J, De Sutter P. Perinatal outcome of pregnancies after assisted reproduction: a case–control study. *J Assist Reprod Genet* 1997;**14**:575–580.

Dhont M, De Sutter P, Ruyssinck G, Martens G, Bekaert A. Perinatal outcome of pregnancies after assisted reproduction: a case–control study. *Am J Obstet Gynecol* 1999;**181**:688–695.

D'Souza SW, Rivlin E, Cadman J, Richards B, Buck P, Lieberman BA. Children conceived by in vitro fertilisation after fresh embryo transfer. *Arch Dis Child* 1997;**76**:F70–F74.

Dumoulin JC, Land JA, Van Montfoort AP, Nelissen EC, Coonen E, Derhaag JG, Schreurs IL, Dunselman GA, Kester AD, Geraedts JP *et al.* Effect of in vitro culture of human embryos on birthweight of newborns. *Hum Reprod* 2010;**25**:605–612.

El-Chaar D, Yang Q, Gao J, Bottomley J, Leader A, Wen SW, Walker M. Risk of birth defects increased in pregnancies conceived by assisted human reproduction. *Fertil Steril* 2009;**92**:1557–1561.

El Hage S, Ghanem I, Safi CA, Afif N, Azoury J, Kharrat K, Dagher F. The risk of neuro-orthopaedic malformations following in-vitro fertilization. *J Pediatr Orthop B* 2006;**15**:229–232.

Ericson A, Kallen B. Congenital malformations in infants born after IVF: a population-based study. *Hum Reprod* 2001;**16**:504–509.

Ericson A, Nygren KG, Olausson PO, Kallen B. Hospital care utilization of infants born after IVF. *Hum Reprod* 2002;**17**:929–932.

Ezechi OC, Ndububa VI, Loto OM, Ezeobi PM, Kalu BKE, Njokanma OF, Nwokoro CA. Pregnancy, obstetric and neonatal outcome after assisted reproduction in Nigerians. *J Matern Fetal Neonatal Med* 2008;**21**:261–266.

Ferraretti AP, Goossens V, de Mouzon J, Bhattacharya S, Castilla JA, Korsak V, Kupka M, Nygren KG, Nyboe Andersen A. Assisted reproductive technology in Europe, 2008: results generated from European registers by ESHRE. *Hum Reprod* 2012;**27**:2571–2584.

Finnstrom O, Kallen B, Lindam A, Nilsson E, Nygren KG, Olausson PO. Maternal and child outcome after in vitro fertilization—a review of 25 years of population-based data from Sweden. *Acta Obstet Gynecol Scand* 2011;**90**:494–500.

Fisch B, Harel L, Kaplan B, Pinkas H, Amit S, Ovadia J, Tadir Y, Merlob P. Neonatal assessment of babies conceived by in vitro fertilization. *J Perinatol* 1997;**17**:473–476.

FIVNAT. Pregnancies and births resulting from in vitro fertilization: French national registry, analysis of data 1986 to 1990. *Fertil Steril* 1995;**64**:746–756.

Fujii M, Matsuoka R, Bergel E, van der Poel S, Okai T. Perinatal risk in singleton pregnancies after in vitro fertilization. *Fertil Steril* 2010;**94**:2113–2117.

Ghazi HA, Spielberger C, Kallen B. Delivery outcome after infertility—a registry study. *Fertil Steril* 1991;**55**:726–732.

Gojnic M, Jeremic K, Boskovic V, Fazlagic A, Stefanovic A, Pervulov M. Perinatal outcome in multiple pregnancies—spontaneous gestation versus. *Clin Exp Obstet Gynecol* 2005;**32**:65–67.

Gonzalez L, Berzosa J, Alonso R, Ayuso F, Izquierdo M. [Spontaneous versus assisted conception multiple pregnancies]. *Clin Invest Ginecol Obstet* 2006;**33**:7–11.

Greenland S, Thomas DC. On the need for the rare disease assumption in case–control studies. *Am J Epidemiol* 1982;**116**:547–553.

Gupta P, Nayan N, Sharma M. Perinatal outcomes among children born by assisted reproductive techniques—a hospital-based case control study. *Med J Armed Forces India* 2012;**68**:132–135.

Halliday JL, Ukoumunne OC, Baker HW, Breheny S, Jaques AM, Garrett C, Healy D, Amor D. Increased risk of blastogenesis birth defects, arising in the first 4 weeks of pregnancy, after assisted reproductive technologies. *Hum Reprod* 2010;**25**:59–65.

Hansen M, Kurinczuk JJ, Bower C, Webb S. The risk of major birth defects after intracytoplasmic sperm injection and in vitro fertilization. *N Engl J Med* 2002;**346**:725–730.

Hansen M, Bower C, Milne E, de Klerk N, Kurinczuk JJ. Assisted reproductive technologies and the risk of birth defects—a systematic review. *Hum Reprod* 2005;**20**:328–338.

Hansen M, Colvin L, Petterson B, Kurinczuk JJ, de Klerk N, Bower C. Admission to hospital of singleton children born following assisted reproductive technology (ART). *Hum Reprod* 2008;**23**:1297–1305.

Hansen M, Colvin L, Petterson B, Kurinczuk JJ, de Klerk N, Bower C. Twins born following assisted reproductive technology: perinatal outcome and admission to hospital. *Hum Reprod* 2009;**24**:2321–2331.

Hansen M, Milne E, de Klerk N, Kurinczuk JJ, Jacoby P, Bower C. ART, birth defects and subfertility—what should prospective patients be told? *J Assist Reprod Genet* 2011;**28**:1229–1230.

Hansen M, Kurinczuk JJ, de Klerk N, Burton P, Bower C. Assisted reproductive technology and major birth defects in Western Australia. *Obstet Gynecol* 2012;**120**:852–863.

Harrison RF, Hennelly B, Woods T, Lowry K, Kondaveeti U, Barry-Kinsella C, Nargund G. Course and outcome of IVF pregnancies and spontaneous conceptions within an IVF setting. *Eur J Obstet Gynecol Reprod Biol* 1995;**59**:175–182.

Henningsen AK, Pinborg A, Lidegaard O, Vestergaard C, Forman JL, Andersen AN. Perinatal outcome of singleton siblings born after assisted reproductive technology and spontaneous conception: Danish national sibling-cohort study. *Fertil Steril* 2011a;**95**:959–963.

Henningsen AKA, Romundstad LB, Gissler M, Nygren KG, Lidegaard O, Skjaerven R, Tiitinen A, Andersen AN, Wennerholm UB, Pinborg A. Infant and maternal health monitoring using a combined Nordic database on ART and safety. *Acta Obstet Gynecol Scand* 2011b;**90**:683–691.

Ho CH, Peng FS, Chen HF, Lien YR, Chen SU, Yang YS. Twin pregnancies conceived by assisted reproductive technology: maternal and perinatal outcomes. *Taiwanese J Obstet Gynecol* 2005;**44**:332–337.

Isaksson R, Gissler M, Tiitinen A. Obstetric outcome among women with unexplained infertility after IVF: a matched case–control study. *Hum Reprod* 2002;**17**:1755–1761.

Jaques AM, Amor DJ, Baker HW, 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;**94**:2674–2679.

Joy J, McClure N, Cooke IE. A comparison of spontaneously conceived twins and twins conceived by artificial reproductive technologies. *J Obstet Gynaecol* 2008;**28**:580–585.

Kallen B, Finnstrom O, Nygren KG, Olausson PO. In vitro fertilization (IVF) in Sweden: risk for congenital malformations after different IVF methods. *Birth Defects Res A Clin Mol Teratol* 2005;**73**:162–169.

Kallen B, Finnstrom O, Lindam A, Nilsson E, Nygren KG, Otterblad PO. Congenital malformations in infants born after in vitro fertilization in Sweden. *Birth Defects Res A Clin Mol Teratol* 2010a;**88**:137–143.

Kallen B, Finnstrom O, Lindam A, Nilsson E, Nygren KG, Otterblad Olausson P. Trends in delivery and neonatal outcome after in vitro fertilization in Sweden: data for 25 years. *Hum Reprod* 2010b;**25**:1026–1034.

Kanat-Pektas M, Kunt C, Gungor T, Mollamahmutoglu L. Perinatal and first year outcomes of spontaneous versus assisted twins: a single center experience. *Arch Gynecol Obstet* 2008;**278**:143–147.

Kanyo K, Konc J. A follow-up study of children born after diode laser assisted hatching. *Eur J Obstet Gynecol Reprod Biol* 2003;**110**:176–180.

Katalinic A, Rosch C, Ludwig M. Pregnancy course and outcome after intracytoplasmic sperm injection: a controlled, prospective cohort study. *Fertil Steril* 2004;**81**:1604–1616.

Klemetti R, Gissler M, Sevon T, Koivurova S, Ritvanen A, Hemminki E. Children born after assisted fertilization have an increased rate of major congenital anomalies. *Fertil Steril* 2005;**84**:1300–1307.

Knoester M, Helmerhorst FM, Vandenbroucke JP, van der Westerlaken LA, Walther FJ, Veen S. Perinatal outcome, health, growth, and medical care utilization of 5- to 8-year-old intracytoplasmic sperm injection singletons. *Fertil Steril* 2008;**89**:1133–1146.

Koivurova S, Hartikainen AL, Gissler M, Hemminki E, Sovio U, Jarvelin MR. Neonatal outcome and congenital malformations in children born after in-vitro fertilization. *Hum Reprod* 2002;**17**:1391–1398.

Kor-anantakul O, Suwanrath C, Suntharasaj T, Getpook C, Leetanaporn R. Outcomes of multifetal pregnancies. *J Obstet Gynaecol Res* 2007;**33**:49–55.

Koudstaal J, Braat DDM, Bruinse HW, Naaktgeboren N, Vermeiden JPW, Visser GHA. Obstetric outcome of singleton pregnancies after IVF: a matched control study in four Dutch university hospitals. *Hum Reprod* 2000a;**15**:1819–1825.

Koudstaal J, Bruinse HW, Helmerhorst FM, Vermeiden JPW, Willemsen WNP, Visser GHA. Obstetric outcome of twin pregnancies after in-vitro fertilization: a matched control study in four Dutch University hospitals. *Hum Reprod* 2000b;**15**:935–940.

Kozinszky Z, Zadori J, Orvos H, Katona M, Pal A, Kovacs L. Obstetric and neonatal risk of pregnancies after assisted reproductive technology: a matched control study. *Acta Obstet Gynecol Scand* 2003;**82**:850–856.

Kurinczuk JJ, Bower C. Birth defects in infants conceived by intracytoplasmic sperm injection: an alternative interpretation. *Br Med J* 1997;**315**:1260–1265.

Kurinczuk JJ, Hansen M, Bower C. Methodological considerations when designing studies to examine the health of children born following ART. In: Sutcliffe AG (ed). *Health and Welfare of ART Children*. London: Informa Healthcare, 2006, 11–41.

Kuwata T, Matsubara S, Ohkuchi A, Watanabe T, Izumi A, Honma Y, Yada Y, Shibahara H, Suzuki M. The risk of birth defects in dichorionic twins conceived by assisted reproductive technology. *Twin Res* 2004;**7**:223–227.

Lahat E, Raziel A, Friedler S, Schieber-Kazir M, Ron-El R. Long-term follow-up of children born after inadvertent administration of a gonadotrophin-releasing hormone agonist in early pregnancy. *Hum Reprod* 1999;**14**:2656–2660.

Lambalk CB, van Hooff M. Natural versus induced twinning and pregnancy outcome: a Dutch nationwide survey of primiparous dizygotic twin deliveries. *Fertil Steril* 2001;**75**:731–736.

Lancaster PAL. Registers of in-vitro fertilisation and assisted conception. *Hum Reprod* 1996;**11**:89–104.

Lehnen H, Schafer S, Reineke T, Puchooa A, Maiwald R, Zechner U. Twin pregnancies conceived spontaneously and by ART (Assisted Reproductive Technologies)—a retrospective analysis and review. *Geburtsh Frauenheilk* 2011;**71**:669–676.

Leslie GI, Gibson FL, McMahon C, Tennant C, Saunders DM. Infants conceived using in-vitro fertilization do not over-utilize health care resources after the neonatal period. *Hum Reprod* 1998;**13**:2055–2059.

Lie RT, Lyngstadaas A, Orstavik KH, Bakketeig LS, Jacobsen G, Tanbo T. Birth defects in children conceived by ICSI compared with children conceived by other IVF-methods; a meta-analysis. *Int J Epidemiol* 2005;**34**:696–701.

Liu F, He L. Evaluate the security of assisted reproductive technology in the near future. [Chinese]. *Prog Obstet Gynecol* 2010;**19**:179–184.

Ludwig M. Malformation risk in subfertile couples. *Reprod Biomed Online* 2012;**25**:225–226.

Ludwig M, Diedrich K. Follow-up of children born after assisted reproductive technologies. *Reprod Biomed Online* 2002;**5**:317–322.

Ludwig M, Katalinic A. Malformation rate in fetuses and children conceived after ICSI: results of a prospective cohort study. *Reprod Biomed Online* 2002;**5**:171–178.

Manoura A, Korakaki E, Hatzidaki E, Bikouvarakis S, Papageorgiou M, Giannakopoulou C. Perinatal outcome of twin pregnancies after in vitro fertilization. *Acta Obstet Gynecol Scand* 2004;**83**:1079–1084.

Marino JL, Moore VM, Rumbold AR, Davies MJ. Fertility treatments and the young women who use them: an Australian cohort study. *Hum Reprod* 2011;**26**:473–479.

McDonald S, Murphy K, Beyene J, Ohlsson A. Perinatal outcomes of in vitro fertilization twins: a systematic review and meta-analyses. *Am J Obstet Gynecol* 2005a;**193**:141–152.

McDonald SD, Murphy K, Beyene J, Ohlsson A. Perinatal outcomes of singleton pregnancies achieved by in vitro fertilization: a systematic review and meta-analysis. *J Obstet Gynaecol Can* 2005b;**27**:449–459.

McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003;**157**:940–943.

Merlob P, Fisch B. Neonatal outcome and congenital malformations in children born after IVF. *Hum Reprod* 2002;**17**:3004–3005; author reply 3005.

Merlob P, Sapir O, Sulkes J, Fisch B. The prevalence of major congenital malformations during two periods of time, 1986–1994 and 1995–2002 in newborns conceived by assisted reproduction technology. *Eur J Med Genet* 2005;**48**:5–11.

Minakami H, Sayama M, Honma Y, Matsubara S, Koike T, Sato I, Uchida A, Eguchi Y, Momoi M, Araki S. Lower risks of adverse outcome in twins conceived by artificial reproductive techniques compared with spontaneously conceived twins. *Hum Reprod* 1998;**13**:2005–2008.

Morcel K, Lavoue V, Beuchee A, Le Lannou D, Poulain P, Pladys P. Perinatal morbidity and mortality in twin pregnancies with dichorionic placentas following assisted reproductive techniques or ovarian induction alone: a comparative study. *Eur J Obstet Gynecol Reprod Biol* 2010;**153**:138–142.

Morin NC, Wirth FH, Johnson DH, Frank LM, Presburg HJ, Van de Water VL, Chee EM, Mills JL. Congenital malformations and psychosocial development in children conceived by in vitro fertilization. *J Pediatr* 1989;**115**:222–227.

Nassar S, Boutros J, Aboulghar H, Mansour R, Hussein M, Aboulghar M. Perinatal outcome after in vitro fertilization and spontaneous pregnancy: a comparative study. *Middle East Fertil Soc J* 1996;**1**:151–158.

Nassar AH, Usta IM, Rechdan JB, Harb TS, Adra AM, Abu-Musa AA. Pregnancy outcome in spontaneous twins versus twins who were conceived through in vitro fertilization. *Am J Obstet Gynecol* 2003;**189**:513–518.

Nuojua-Huttunen S, Gissler M, Martikainen H, Tuomivaara L. Obstetric and perinatal outcome of pregnancies after intrauterine insemination. *Hum Reprod* 1999;**14**:2110–2115.

Nygren KG, Sullivan E, Zegers-Hochschild F, Mansour R, Ishihara O, Adamson GD, de Mouzon J. International Committee for Monitoring Assisted Reproductive Technology (ICMART) world report: assisted reproductive technology 2003. *Fertil Steril* 2011;**95**:2209–2222.

Olson CK, Keppler-Noreuil KM, Romitti PA, Budelier WT, Ryan G, Sparks AE, Van Voorhis BJ. In vitro fertilization is associated with an increase in major birth defects. *Fertil Steril* 2005;**84**:1308–1315.

Ombelet W, Peeraer K, De Sutter P, Gerris J, Bosmans E, Martens G, Ruyssinck G, Defoort P, Molenberghs G, Gyselaers W. Perinatal outcome of ICSI pregnancies compared with a matched group of natural conception pregnancies in Flanders (Belgium): a cohort study. *Reprod Biomed Online* 2005;**11**:244–253.

Ozer EA, Turkoglu E, Balli T, Sutcuoglu S, Erdemir A, Cosar H, Kahramaner Z, Kanik A, Men G, Yaprak I. Neonatal mortality and short term prognosis in newborns born after assisted reproduction techniques. *Turk Arch Ped* 2011;**46**:37–41.

Palermo GD, Neri QV, Hariprashad JJ, Davis OK, Veeck LL, Rosenwaks Z. ICSI and its outcome. *Semin Reprod Med* 2000;**18**:161–169.

Palermo GD, Neri QV, Takeuchi T, Squires J, Moy F, Rosenwaks Z. Genetic and epigenetic characteristics of ICSI children. *Reprod Biomed Online* 2008;**17**:820–833.

Pandey S, Shetty A, Hamilton M, Bhattacharya S, Maheshwari A. Obstetric and perinatal outcomes in singleton pregnancies resulting from IVF/ICSI: a systematic review and meta-analysis. *Hum Reprod Update* 2012;**18**:485–503.

Petersen K, Hornnes PJ, Ellingsen S, Jensen F, Brocks V, Starup J, Jacobsen JR, Andersen AN. Perinatal outcome after in vitro fertilisation. *Acta Obstet Gynecol Scand* 1995;**74**:129–131.

Pharoah PO, Glinianaia SV, Rankin J. Congenital anomalies in multiple births after early loss of a conceptus. *Hum Reprod* 2009;**24**:726–731.

Pinborg A. Vanishing twins in in-vitro fertilization (IVF). *Twin Res Hum Genet* 2010b;**13**:283.

Pinborg A, Loft A, Rasmussen S, Schmidt L, Langhoff-Roos J, Greisen G, Andersen AN. Neonatal outcome in a Danish national cohort of 3438 IVF/ICSI and 10,362 non-IVF/ICSI twins born between 1995 and 2000. *Hum Reprod* 2004;**19**:435–441.

Pinborg A, Lidegaard O, la Cour Freiesleben N, Andersen AN. Consequences of vanishing twins in IVF/ICSI pregnancies. *Hum Reprod* 2005;**20**:2821–2829.

Pinborg A, Lidegaard O, Freiesleben NC, Andersen AN. Vanishing twins: a predictor of small-for-gestational age in IVF singletons. *Hum Reprod* 2007;**22**:2707–2714.

Pinborg A, Loft A, Aaris Henningsen AK, Rasmussen S, Andersen AN. Infant outcome of 957 singletons born after frozen embryo replacement: the Danish National Cohort Study 1995–2006. *Fertil Steril* 2010a;**94**:1320–1327.

Pinborg A, Wennerholm UB, Romundstad LB, Loft A, Aittomaki K, Söderström-Anttila V, Nygren KG, Hazekamp J, Bergh C. Why do singletons conceived after assisted reproduction technology have adverse perinatal

outcome? Systematic review and meta-analysis. *Hum Reprod Update* 2012a; doi:10.1093/humupd/dms044.

Pinborg A, Loft A, Henningsen AK, Ziebe S. Does assisted reproductive treatment increase the risk of birth defects in the offspring? *Acta Obstet Gynecol Scand* 2012b; **91**:1245–1246.

Place I, Englert Y. 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* 2003;**80**:1388–1397.

Raatikainen K, Kuivasaari-Pirinen P, Hippelainen M, Heinonen S. Comparison of the pregnancy outcomes of subfertile women after infertility treatment and in naturally conceived pregnancies. *Hum Reprod* 2012;**27**:1162–1169.

Reefhuis J, Honein MA, Schieve LA, Correa A, Hobbs CA, Rasmussen SA. Assisted reproductive technology and major structural birth defects in the United States. *Hum Reprod* 2009;**24**:360–366.

Reefhuis J, Honein MA, Schieve LA, Rasmussen SA. Use of clomiphene citrate and birth defects, National Birth Defects Prevention Study, 1997–2005. *Hum Reprod* 2011;**26**:451–457.

Rimm AA, Katayama AC. Reproductive technologies and the risk of birth defects. *N Engl J Med* 2012a;**367**:874–875; author reply 875–876.

Rimm AA, Katayama AC. The impact of ART on major malformations is not known at this time. *J Assist Reprod Genet* 2012b;**29**:699–700.

Rimm AA, Katayama AC, Diaz M, Katayama KP. A meta-analysis of controlled studies comparing major malformation rates in IVF and ICSI infants with naturally conceived children. *J Assist Reprod Genet* 2004;**21**:437–443.

Rimm AA, Katayama AC, Katayama KP. A meta-analysis of the impact of IVF and ICSI on major malformations after adjusting for the effect of subfertility. *J Assist Reprod Genet* 2011;**28**:699–705.

Rizk B, Doyle P, Tan SL, Rainsbury P, Betts J, Brinsden P, Edwards R. Perinatal outcome and congenital malformations in in-vitro fertilisation babies from the Bourn-Hallam group. *Hum Reprod* 1991;**6**:1259–1264.

Rodrigues M, Pinelo S, Carvalho I, Figueiredo H, Felgueira E, Joao A, Pinho M. Comparative study of the obstetric and neonatal complications of in vitro fertilization/intracitoplasmatic sperm injection twin pregnancies vs naturally conceived twin gestations [Spanish]. *Rev Iberoam Fertil Reprod Hum* 2005;**22**:395–401.

Romundstad LB, Romundstad PR, Sunde A, von During 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;**372**:737–743.

Ron-El R, Lahat E, Golan A, Lerman M, Bukovsky I, Herman A. Development of children born after ovarian superovulation induced by long-acting gonadotropin-releasing hormone agonist and menotropins, and by in vitro fertilization. *J Pediatr* 1994;**125**:734–737.

Rossi AC, D'Addario V. Neonatal outcomes of assisted and naturally conceived twins: systematic review and meta-analysis. *J Perinat Med* 2011;**39**:489–493.

Sagot P, Bechoua S, Ferdynus C, Facy A, Flamm X, Gouyon JB, Jimenez C. Similarly increased congenital anomaly rates after intrauterine insemination and IVF technologies: a retrospective cohort study. *Hum Reprod* 2012;**27**:902–909.

Sala P, Ferrero S, Buffi D, Pastorino D, Bertoldi S, Vaccari L, Bentivoglio G, Venturini PL, De Biasio P. Congenital defects in assisted reproductive technology pregnancies. *Minerva Ginecol* 2011;**63**:227–235.

Sanchez-Albisua I, Borell-Kost S, Mau-Holzmann UA, Licht P, Krageloh-Mann I. Increased frequency of severe major anomalies in children conceived by intracytoplasmic sperm injection. *Dev Med Child Neurol* 2007;**49**:129–134.

Sanchis Calvo A, Marcos Puig B, Juan Garcia L, Morales Suarez-Varela MM, Abeledo Gomez A, Balanza MacHancosa R, Pineda Caplliure A, Tamarit Bordes G, Gimeno Clemente N, Cervero Marti L. [Birth characteristics due to in vitro fertilization (IVF) techniques]. *An Pediatr* 2009;**70**:333–339.

Saunders K, Spensley J, Munro J, Halasz G. Growth and physical outcome of children conceived by in vitro fertilisation. *Pediatrics* 1996;**97**:688–692.

Saygan-Karamursel B, Teksam O, Aksu T, Yurdakok M, Onderoglu L. Perinatal outcomes of spontaneous twins compared with twins conceived through intracytoplasmic sperm injection. *J Perinat Med* 2006;**34**:132–138.

Sazonova A, Kallen K, Thurin-Kjellberg A, Wennerholm UB, Bergh C. Obstetric outcome after in vitro fertilization with single or double embryo transfer. *Hum Reprod* 2011;**26**:442–450.

Sazonova A, Kallen K, Thurin-Kjellberg A, Wennerholm UB, Bergh C. Obstetric outcome in singletons after in vitro fertilization with cryopreserved/thawed embryos. *Hum Reprod* 2012;**27**:1343–1350.

Schimmel MS, Hammerman C, Lusky A, Reichman B. Very low-birth-weight-infants conceived by in vitro fertilization are not at higher risk for mortality and morbidity: a population-based study. *Fertil Steril* 2006;**85**:907–912.

Sebastiani G, Pertierra Cortada A, Vidal Sorde E, Figueras Aloy J, Balasch Cortina J. [Factors associated with assisted reproduction technologies and neonatal outcomes]. *An Pediat* 2009;**70**:323–332.

Shebl O, Ebner T, Sommergruber M, Sir A, Tews G. Risk in twin pregnancies after the use of assisted reproductive techniques. *J Reprod Med* 2008a;**53**:798–802.

Shebl O, Ebner T, Sommergruber M, Sir A, Tews G. Birth weight is lower for survivors of the vanishing twin syndrome: a case–control study. *Fertil Steril* 2008b;**90**:310–314.

Shevell T, Malone FD, Vidaver J, Porter TF, Luthy DA, Comstock CH, Hankins GD, Eddleman K, Dolan S, Dugoff L et al. Assisted reproductive technology and pregnancy outcome. *Obstet Gynecol* 2005;**106**:1039–1045.

Shu-Chi M, Jiann-Loung H, Yu-Hung L, Tseng-Chen S, Ming-I L, Tsu-Fuh Y. Growth and development of children conceived by in-vitro maturation of human oocytes. *Early Hum Dev* 2006;**82**:677–682.

Silver RI, Rodriguez R, Chang TSK, Gearhart JP. In Vitro Fertilisation is associated with an increased risk of hypospadias. *J Urol* 1999;**161**:1954–1957.

Simpson JL. Registration of congenital anomalies in ART populations: pitfalls. *Hum Reprod* 1996;**11**:81–88.

Sipek A, Gregor V, Rychtarikova J, Mardesic T, Horacek J, Masatova D. [Condition of children after assisted reproduction in the Czech Republic in the period of 1995–1999]. *Ceska Gynekol* 2004;**69**:358–365.

Smithers PR, Halliday J, Hale L, Talbot JM, Breheny S, Healy D. High frequency of Cesarean section, antepartum hemorrhage, placenta previa, and preterm delivery in in-vitro fertilization twin pregnancies. *Fertil Steril* 2003;**80**:666–668.

Snajderova M, Zemkova D, Mardesic T, Sipek A, Gregor V, Krejcirova D, Sobotkova D, Kraus J, Lanska V. [Birth defects, medical outcome and somatic development in children conceived after intracytoplasmic sperm injection (ICSI)]. *Ceska Gynekol* 2008;**73**:22–29.

Sperling L, Kill C, Larsen LU, Brocks V, Wojdemann KR, Qvist I, Schwartz M, Jorgensen C, Espersen G, Skajaa K et al. Detection of chromosomal abnormalities, congenital abnormalities and transfusion syndrome in twins. *Ultrasound Obstet Gynecol* 2007;**29**:517–526.

Stromberg B, Dahlquist G, Ericson A, Finnstrom O, Koster M, Stjernqvist K. Neurological sequelae in children born after in-vitro fertilisation: a population-based study. *Lancet* 2002;**359**:461–465.

Sutcliffe AG, D'Souza SW, Cadman J, Richards B, McKinlay IA, Lieberman B. Minor congenital anomalies, major congenital malformations and development in children conceived from cryopreserved embryos. *Hum Reprod* 1995;**10**:3332–3337.

Sutcliffe AG, Taylor B, Li J, Thornton S, Grudzinskas JG, Lieberman BA. Children born after intracytoplasmic sperm injection: population control study. *Br Med J* 1999;**318**:704–705.

Sutcliffe AG, Taylor B, Saunders K, Thornton S, Lieberman BA, Grudzinskas JG. Outcome in the second year of life after in-vitro fertilisation by intracytoplasmic sperm injection: a UK case–control study. *Lancet* 2001;**357**:2080–2084.

Sutcliffe AG, Saunders K, McLachlan R, Taylor B, Edwards P, Grudzinskas G, Leiberman B, Thornton S. 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* 2003;**79**:512–516.

Szymankiewicz M, Jedrzejczak P, Rozycka J, Gadzinowski J, Pawelczyk L. Newborn outcome after assisted reproductive technology: experiences and reflections from Poland. *Int J Fertil* 2004;**49**:150–154.

Tanbo T, Dale PO, Lunde O, Moe N, Abyholm T. Obstetric outcome in singleton pregnancies after assisted reproduction. *Obstet Gynecol* 1995;**86**:188–192.

Tararbit K, Houyel L, Bonnet D, De Vigan C, Lelong N, Goffinet F, Khoshnood B. Risk of congenital heart defects associated with assisted reproductive technologies: a population-based evaluation. *Eur Heart J* 2011;**32**:500–508.

Thomson F, Shanbhag S, Templeton A, Bhattacharya S. Obstetric outcome in women with subfertility. *Br J Obstet Gynaecol* 2005;**112**:632–637.

Tomic V, Tomic J. Neonatal outcome of IVF singletons versus naturally conceived in women aged 35 years and over. *Arch Gynecol Obstet* 2011;**284**:1411–1416.

Vasario E, Borgarello V, Bossotti C, Libanori E, Biolcati M, Arduino S, Spinelli R, Delle Piane L, Revelli A, Todros T. IVF twins have similar obstetric and neonatal outcome as spontaneously conceived twins: a prospective follow-up study. *Reprod Biomed Online* 2010;**21**:422–428.

Verlaenen H, Cammu H, Derde MP, Amy JJ. Singleton pregnancy after in vitro fertilization: expectations and outcome. *Obstet Gynecol* 1995;**86**:906–910.

Wang JX, Norman RJ, Kristiansson P. The effect of various infertility treatments on the risk of preterm birth. *Hum Reprod* 2002;**17**:945–949.

Wang Q-Q, Zhu Y-M, Wu M-Y. [Health status of mothers undergoing in vitro fertilization and their offspring]. *Zhejiang Da Xue Xue Bao Yi Xue Ban* 2009; **38**:515–520.

Wang YA, Macaldowie A, Hayward I, Chambers GM, Sullivan EA. *Assisted reproductive technology in Australia and New Zealand 2009*. Assisted reproduction technology series no. 15. Cat. no. PER 51. Canberra: AIHW, 2011.

Welmerink DB, Voigt LF, Daling JR, Mueller BA. Infertility treatment use in relation to selected adverse birth outcomes. *Fertil Steril* 2010;**94**:2580–2586.

Wen SW, Leader A, White RR, Leveille MC, Wilkie V, Zhou J, Walker MC. A comprehensive assessment of outcomes in pregnancies conceived by in vitro fertilization/intracytoplasmic sperm injection. *Eur J Obstet Gynecol Reprod Biol* 2010;**150**:160–165.

Wen J, Jiang J, Ding C, Dai J, Liu Y, Xia Y, Liu J, Hu Z. Birth defects in children conceived by in vitro fertilization and intracytoplasmic sperm injection: a meta-analysis. *Fertil Steril* 2012;**97**:1331–1337 e1331–1334.

Wennerholm UB, Hamberger L, Nilsson L, Wennergren M, Wikland M, Bergh C. Obstetric and perinatal outcome of children conceived from cryopreserved embryos. *Hum Reprod* 1997;**12**:1819–1825.

Wennerholm UB, Albertsson-Wikland K, Bergh C, Hamberger L, Niklasson A, Nilsson L, Thiringer K, Wennergren M, Wikland M, Borres MP. Postnatal growth and health in children born after cryopreservation as embryos. *Lancet* 1998;**351**:1085–1090.

Wennerholm UB, Bergh C, Hamberger L, Lundin K, Nilsson L, Wikland M, Kallen B. Incidence of congenital malformations in children born after ICSI. *Hum Reprod* 2000;**15**:944–948.

Westergaard HB, Johansen AM, Erb K, Andersen AN. Danish National In-Vitro Fertilization Registry 1994 and 1995: a controlled study of births, malformations and cytogenetic findings. *Hum Reprod* 1999;**14**:1896–1902.

Yan J, Huang G, Sun Y, Zhao X, Chen S, Zou S, Hao C, Quan S, Chen ZJ. Birth defects after assisted reproductive technologies in China: analysis of 15,405 offspring in seven centers (2004 to 2008). *Fertil Steril* 2011; **95**:458–460.

Yang H, Choi YS, Nam KH, Kwon JY, Park YW, Kim YH. Obstetric and perinatal outcomes of dichorionic twin pregnancies according to methods of conception: spontaneous versus in-vitro fertilization. *Twin Res Hum Genet* 2011; **14**:98–103.

Yeh J, Leipzig S, Friedman EA, Seibel MM. Results of in vitro fertilization pregnancies: experience at Boston's Beth Israel Hospital. *Int J Fertil* 1990; **35**:116–119.

Zadori J, Kozinszky Z, Orvos H, Katona M, Kaali SG, Pal A. The incidence of major birth defects following in vitro fertilization. *J Assist Reprod Genet* 2003; **20**:131–132.

Zadori J, Kozinszky Z, Orvos H, Katona M, Kaali SG, Pal A. Birth weight discordance in spontaneous versus induced twins: impact on perinatal outcome. *J Assist Reprod Genet* 2004;**21**:85–88.

Zegers-Hochschild F, Schwarze JE, Musri C, Crosby J. *Red Latinoamericana de Reproduccion Asistida*. Santiago: Chile, 2009.

Zhu JL, Basso O, Obel C, Bille C, Olsen J. Infertility, infertility treatment, and congenital malformations: Danish national birth cohort. *Br Med J* 2006; **333**:679–683.

Zuppa AA, Maragliano G, Scapillati ME, Crescimbini B, Tortorolo G. Neonatal outcome of spontaneous and assisted twin pregnancies. *Eur J Obstet Gynecol Reprod Biol* 2001;**95**:68–72.