Assisted Reproductive Technology and Perinatal Outcomes: Conventional versus discordant sibling design

Running title: ART & perinatal outcomes

Nafeesa N Dhalwani PhD [a,b,c*], Sheree L Boulet DrPH [a], Dmitry M Kissin MD [a], Yujia Zhang PhD [a], Patricia McKane MPH [d], Marie A Bailey MSW [e], Maria-Elena (Malena) Hood MPH [f], Laila J Tata PhD [b]

[a] Division of Reproductive Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA 30341

[b] Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK NG5 1PB

[c] Diabetes Research Centre, University of Leicester, Leicester General Hospital, Gwendolen Road, Leicester, UK LE5 4PW

[d] Michigan Department of Health and Human Services, Lansing, USA 48913

[e] Florida Department of Health, Tallahassee, USA 32399

[f] Massachusetts Department of Public Health, Boston, USA 02108

*Email for correspondence: nnd2@le.ac.uk

Phone: +(44)1162584974

Word count: 3,210

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

ABSTRACT CAPSULE

The sibling-discordant study found an increased risk of low birthweight and preterm birth associated with ART, but with a lower magnitude of effect, compared to the general population

ABSTRACT

Objective: To compare risks of adverse perinatal outcomes between Assisted Reproductive Technology (ART) and naturally conceived singleton births using a dual design approach.

Study design: Discordant sibling and conventional cross-sectional general population comparison.

Setting: National ART Surveillance System from Michigan, Massachusetts and Florida (2000-2010) linked to birth records.

Patients: all singleton live births, conceived naturally or via ART

Interventions: None

Main outcome measures: Birthweight, gestational age, low birthweight, preterm delivery, small-for-gestational age (SGA), low Apgar score.

Results: 32,762(0.8%) of 3,896,242 singleton live births in the three states were conceived via ART. In 6,458 sibling pairs, ART conceived singletons were 33g lighter (Adjusted β=-33.40, 95% Confidence Interval (CI) -48.60,-18.21) and born half a day sooner (β=-0.58, 95% CI -1.02,-0.14) than singletons conceived naturally. The absolute risk of low birth weight and preterm birth was 6.8% and 9.7% respectively in the ART group and 4.9% and 7.9% in the non-ART group respectively. The odds of low birthweight were 33% higher (Adjusted Odds Ratio (aOR) =1.33, 95% CI 1.13, 1.56) and 20% higher for preterm birth (aOR=1.20, 95% CI 1.07, 1.34). The odds of SGA and low Apgar score were not significantly different in both the groups (aOR=1.22, 95% CI 0.88, 1.68 and aOR=0.75, 95% CI 0.54, 1.05 respectively). Results of conventional analyses were similar, although the magnitude of risk was higher for pre-term birth (aOR 1.51, 95% CI 1.46, 1.56).

Conclusion: Despite some inflated risks in the general population comparison, ART remained associated with increased likelihood of low birthweight and preterm birth when underlying maternal factors were kept constant using discordant-sibling comparison.

Keywords: Assisted reproductive technology, low birth weight, preterm birth, small-for-gestational-age, sibling-discordant design

INTRODUCTION

Reporting of fertility problems has increased substantially over the last decades, with more couples seeking medical consultation as a result of delaying pregnancy, and increased awareness and wider availability of fertility treatments.(1) Consequently, the use of assisted reproductive technology (ART) has increased such that ART now accounts for approximately 1.5% of live births in the United States (US) and 2% in the United Kingdom (UK).(2, 3) The association between ART and adverse perinatal outcomes has previously been assessed for different types of ART, causes of infertility and for singleton and twin or higher order deliveries. (4-8) A population-based study from the US linking the ART surveillance system data with Massachusetts live birth-infant death records data for 1997-1998 found the risk of preterm delivery (<37 weeks of gestation) and low birthweight (birthweight <2500g) to be over twice as high in singletons conceived through ART compared to natural births.(5) A more recent study of 306,995 births in South Australia showed that compared with non-ART singletons, ART singletons had compromised perinatal outcomes varying by the type of ART used.(9) One potential explanation for ART-associated adverse perinatal outcomes is underlying maternal or paternal factors that may result in infertility, and also cause adverse perinatal outcomes. However, studies often compare couples undergoing ART with couples conceiving naturally and most have not been able to adjust for underlying parental factors which may result in infertility and consequently adverse perinatal outcomes. One alternative approach is to compare pregnancies in the same women using discordant-sibling analysis which keeps many maternal factors fairly constant between comparisons, minimizing related residual confounding. Only two studies to date have used this approach and found conflicting results on the association between ART and perinatal outcomes.(10, 11)

Using the recent linkage of the National ART Surveillance System (NASS) with birth records from three US states, we compared the risk of adverse perinatal outcomes in singleton infants conceived through ART with those conceived naturally using a sibling analysis approach and compared results to those

obtained with a conventional cross-sectional approach using all non-ART singletons from the general population as the comparison group.

METHODS

Data source

We used data from the States Monitoring Assisted Reproductive Technology (SMART) Collaborative, a collaboration between the Centers for Disease Control and Prevention's (CDC) Division of Reproductive Health, Florida Department of Health, the Massachusetts Department of Public Health and the Michigan Department of Community Health. As part of the SMART Collaborative, data from NASS, collected by CDC on all ART cycles performed in the US, have been linked to birth certificate data from the three participating states. Data are linked using the Link Plus software (CDC, Atlanta, Georgia) and a probabilistic linkage algorithm using maternal date of birth, infant date of birth, plurality, maternal residence ZIP code and gravidity as the primary linkage variables, with a linkage rate of 90.2%. A detailed description of data collection and linkages for the SMART collaborative has been reported elsewhere.(12, 13)

Study population

The study population included all singleton live-births in the three participating states between 2000 and 2010. For Massachusetts and Florida, linked data were available for all study years while for Michigan linked data were only available until 2009. Infants born to mothers aged less than 20 years at delivery and those having gestational age of less than 20 weeks (estimated clinically) or greater than 46 weeks and birthweight above or below the national reference range for gestational age(14) were excluded. ART births were identified using the NASS-linked data which provided detailed information on ART procedures and reasons for infertility treatment.

To create discordant sibling pairs we restricted the study population to singleton live births where one sibling was conceived through ART and the other was conceived without ART, regardless of the order of

conception. In the instance of a woman having more than one ART or non-ART singleton live birth during the study period, one birth of each type was randomly selected to create the sibling pair.

Main outcome measures

We examined birthweight as a continuous variable and also assessed whether the infant had low birthweight, defined as less than 2,500g according to the World Health Organization (WHO) definition.(15) Similarly, gestational age at delivery was used as a continuous variable and we also assessed preterm birth, defined as birth before 37 weeks of gestation in line with the WHO definition.(16) Small-for-gestational age (SGA) was defined as a birthweight below 2 standard deviations of mean birthweight for gestational age according to the infant's sex in line with the International Societies of Pediatric Endocrinology and the Growth Hormone Research Society definition.(17) We used this definition as opposed to using the more common definition (weight below 10^{th} percentile for gestational age) as it identifies most infants at risk of developing childhood or adulthood morbidity(17) and also facilitates comparisons with the previous study.(10) Apgar score is a scoring system to assess the clinical status of the newborn at one and five minutes and is comprised of five components: heart rate, respiratory effort, muscle tone, reflex irritability, and color, each of which is given a score of 0, 1, or 2. A low Apgar score at 5 minutes has been shown to be associated with neonatal mortality and adverse neurological outcomes in infants.(18) We categorized the 5 minute Apgar into normal (≥7) and low (<7) in line with previous studies.(5)

Statistical analysis

We compared baseline birth characteristics in the ART and non-ART groups using chi-squared tests and assessed types of ART and reasons for treatment in the population using proportions. We used two approaches to multivariable analysis; one using births from the three states in a conventional cross-sectional analysis where all ART births were compared with all non-ART births, and one comparing ART

births to non-ART births identified from only the discordant-sibling pairs. For both approaches, we used random effects generalized estimation equation (GEE) models with exchangeable correlation structure to allow for potential clustering of factors between pregnancies occurring in the same woman. We calculated the mean birthweight and gestational age for ART and non-ART groups and used linear regression to estimate mean differences (β co-efficient) and 95% confidence intervals (CI) between the two groups adjusting the gestational age model for maternal age at delivery, year of birth, parity, infant's sex and time since last recorded delivery and the birthweight model for all of the above covariates and gestational age. We then estimated the absolute risks of low birthweight, pre-term birth, SGA and low Apgar score for ART and non-ART groups. Odds ratios (OR) and 95% CIs were calculated for each outcome using unconditional logistic regression. Preterm birth and SGA estimates were adjusted for maternal age, year of birth, parity, infant's sex and time since last recorded delivery; low birthweight was adjusted for all aforementioned covariates and gestational age; low Apgar was adjusted for all of the above and mode of delivery. It is argued that gestational age may lie on the causal pathway for low birth weight and may therefore introduce bias in the estimation if used as a confounder in the risk estimation.(19) Therefore, we recalculated the difference in birth weight in ART and non-ART groups and the relative risk of low birth weight without adjustment for gestational age. Lastly, we stratified the ART group by three major indications for fertility treatment: male factor infertility only, female factor infertility only and unexplained infertility, and repeated analyses comparing the ART with the non-ART group for all perinatal outcomes. Pre-pregnancy body mass index (BMI), only available for Florida from 2004 and Michigan from late 2007, did not result in a significant change in model effect estimates in these subgroups and was therefore dropped from all models. Smoking, alcohol and caesarean section were also dropped from the models for similar reasons. Ethical approval was obtained from the institutional review boards of Michigan, Massachusetts, Florida and the CDC. All statistical analyses were

96 conducted in STATA 13 MP (StataCorp LP, College Station, Texas) and SUDAAN Release 11 (RTI

97 International, Research Triangle Park, North Carolina).

RESULTS

Baseline characteristics

Of the 4,344,247 singleton live births between 2000 and 2010 a total of 3,896,242 (89.7%) met the inclusion criteria and were included in the study. 2,025,502 (52.0%) births were from Florida, 782,522 (20.1%) from Massachusetts and 1,088,218 (27.9%) were from Michigan (Table 1). Of these, 32,762 (0.8%) were conceived through ART. Mothers in the ART group were older and had fewer previous live births compared with mothers in the non-ART group (p-value <0.001). Infants' sex was not significantly different between the ART and the non-ART group (p-value=0.789). The prevalence of reported smoking during pregnancy was lower in the ART group compared to the non-ART group (1.2% vs 10.0 % respectively, p-value <0.001). Cesarean section deliveries were more common in the ART group than the non-ART group (46.6% vs. 30.9% respectively, p-value <0.001). The average time between recorded deliveries was 2.9 years for non-ART group and 3.2 years for the ART group (p<0.0001).

Of the 32,762 ART births, 32,383 (98.8%) were conceived through *in vitro* fertilization (IVF) and 1.2% were conceived through other ART procedures including gamete intra-fallopian transfer, zygote intra-fallopian transfer or a combination of these. Reasons for ART treatment included female factor infertility in 63.8% of cases, male factor infertility in 38.3% of cases and unexplained infertility in 13.7% of cases. The most common reason for ART treatment for female infertility was tubal factor (Table 2).

When the study population was restricted to discordant sibling pairs the sample size reduced to 12,916 with 6,458 live births in the ART group and 6,458 in the non-ART group and the birth characteristics were highly comparable in both the groups (Supplementary Table 1).

ART and birth outcomes

In adjusted models from the conventional analysis, infants in the ART group weighed 64 grams less than infants in the non-ART group (β coefficient= -64.11g, 95% CI -69.74,-58.48). In comparison the

mean reduction in birthweight in the ART group compared to non-ART group in the sibling pair analysis was 33 grams (β coefficient= -33.40 g, 95% CI -48.60,-18.21). Gestational age at delivery was lower in the ART group compared with the non-ART group, the conventional analysis again showing a more marked difference between groups (-1.66 days) than the sibling analysis (-0.58 days) (Table 3).

In the conventional analysis, the absolute risks of low birthweight and preterm birth were 1.4% higher in the ART group compared to non-ART group. After adjusting for confounders, infants in the ART group still had a 38% increase in the odds of low birthweight (OR=1.38, 95% CI 1.32, 1.43) and a 51% increase in the odds of preterm birth (OR=1.51, 95% CI 1.46, 1.56) compared with the non-ART group. The relative risk of low birth weight was similar when gestational age was removed from the birth weight model (OR 1.48, 95% CI 1.42, 1.54). The absolute risk for SGA was very similar in the ART and non-ART groups (1.8% and 1.7% respectively). However, after adjusting for confounders there was an 11% increase in the odds of SGA in the ART group compared to the non-ART group (OR=1.11, 95% CI 1.03, 1.21). The odds of low Apgar score was not different between groups (OR=0.99, 95% CI 0.90, 1.09) (Table 4).

In the discordant sibling pair analyses, absolute risk differences of low birthweight between the ART and non-ART groups were smaller such that the absolute risk was 6.8% in the ART group compared to 4.9% in the non-ART group with a similar adjusted OR as the conventional analysis (OR=1.33, 95% CI 1.13, 1.56 and OR=1.45, 95% CI 1.26, 1.67 after excluding gestational age from the model). After adjusting for confounders there was still a statistically significant increased risk of preterm birth in the ART group compared to the non-ART group however the effect size was smaller (OR=1.20, 95% CI 1.07,1.34) than the conventional analysis estimate. There was no increased risk of low Apgar score in relation to ART (OR 0.75, 95% CI 0.54, 1.05). The odds ratio for SGA in the ART group relative to the non-ART group was higher in the sibling analysis than the conventional analysis (OR 1.22, 95% CI

147 0.88, 1.68), however the confidence intervals for the estimates from these two approaches
148 overlapped (Table 4).

149 When analyses were stratified by three indications for ART (Supplementary Table 2), the magnitudes
150 of association between most birth outcomes and ART were similar or smaller in the sibling pair
151 analysis compared with the conventional analysis. When comparing the overall ART sibling pair
152 analysis with the sibling pair analysis by ART indication, the magnitude of effect for female infertility
153 were very similar to the overall estimate (OR for low birthweight 1.45(95% CI 1.17, 1.78) , however;
154 the relative risks of adverse perinatal outcomes in the male infertility group were lower than the
155 overall risks and not significantly different from the non-ART group (OR for low birthweight 0.91
156 (95% CI 0.67-1.23)).

DISCUSSION

Principal findings

Using a large population-based dataset and both a conventional data analysis approach and a discordant sibling pair design, we found that ART births were associated with an increased risk of low birthweight, preterm birth and SGA compared with non-ART births. Conventional analyses provided very similar risk estimates except for pre-term birth where the adjusted risk was found to be higher in conventional analysis compared with sibling analysis.

Strengths and limitations

To our knowledge, this is the first study to date in the US to assess the association between ART and adverse birth outcomes using a discordant sibling pair design. The number of sibling pairs in our study was about three times higher than the two previous studies using this approach (6,458 pairs in the current study vs. 2,204 pairs in Romundstad et al.(10) and 3,879 in Henningsen et al.(11)) making our estimates more precise. Order of birth has been shown to affect birth weight and birth parameters however, to optimize the sample size for the study sibling pairs were randomly selected rather than selecting consecutive births, and adjustments were made for parity in the statistical analysis. The data included in this study only covers ART deliveries and natural births between 2000 and 2010 as the linkage of NASS and birth records data was only complete up until 2010 for two states and up until 2009 for one at the time of the study. Our ascertainment of the outcomes (gestational age, birthweight and Apgar score) is based on birth record data which have been shown to be reliable for these outcomes. A study comparing data from birth certificates to medical records in Indiana found high comparability between the two sources for low birthweight, preterm birth and Apgar score (Kappa=0.88,0.79 and 0.91 respectively).(20) We only compared outcomes in pregnancies ending in live births as NASS data are not linked for stillbirths. This may potentially bias our findings towards null as pregnancies conceived through ART are more likely to end in stillbirths compared to naturally conceived pregnancies(9) and fetal growth restriction is one of the most

important underlying factors for stillbirth,(21) which also results in low birthweight. Additionally, despite the discordant sibling design and adjustment of potential confounders, some maternal factors may still be different between the two pregnancies under comparison (e.g. complications related to prior pregnancies, use of medications etc.) which we were not able to control for in our analyses. Also, to be included in the sibling discordant analysis mothers had to have a non-ART conception and an ART conception, which may not be the case for all women undergoing ART. Furthermore, we used unconditional logistic regression with the maternal strata as a random effect (rather than using conditional logistic regression) which may have resulted in an underestimation of the effect size. We were able to keep maternal factors as constant as possible within the sibling analysis; however we did not have any information for additional adjustments by paternal factors, which could potentially explain some difference in the effect estimates in the sibling design, especially if both siblings had different fathers. Nevertheless, the impact of paternal factors on the association between ART and perinatal outcomes is unclear. Additionally, we were unable to identify whether non-ART births were conceived using other fertility treatments such as ovulation induction etc., which could potentially bias our results towards null. Lastly, we did not take any treatment factors (e.g. fresh vs. frozen embryo transfer, autologous vs. donor eggs, number of eggs transferred etc.) into account, which also have an impact on the perinatal outcomes. However, one of our recent studies found that between 70-80% of all ART deliveries are fresh non-donor,(22) hence stratification by such treatment factors in the sibling analysis would yield very small subgroups with inadequate numbers to assess the effect of these treatment factors, especially considering the absolute rates of low birth weight and preterm birth are less than 10%.

Comparison with current literature

We found the odds of low birthweight to be 38% higher and the odds of preterm birth to be 51% higher when all singleton ART births were compared to non-ART births. This is slightly lower than estimates from a previous study from Massachusetts (1997-1998) which found that the odds of low birthweight were 80% higher (OR 1.8, 95% CI 1.6-2.0) and the odds of pre-term delivery were 50%

higher (OR 1.5, 95% CI 1.2-1.7) in singleton ART births compared to singleton non-ART births. The study from South Australia also found similar estimates with 98% higher odds of low birthweight, 64% higher odds of pre-term birth, 22% higher odds of SGA and no association between ART and low Apgar scores at 5 minutes (OR 1.70, 95% CI 0.92-1.49).(9) A potential explanation for lower odds compared to previous studies could be an increase in single embryo transfers over time consequently improving perinatal outcomes of ART.(23, 24)

In the sibling analysis, we found 33g reduction in the birth weight in the ART group compared to the non-ART group (β=-33.40, 95% CI -48.60,0-18.21). The sibling discordant study from Denmark, including 3,879 mothers with both IVF-ICSI and natural conception found similar results with a 65g reduction in birth weight in babies conceived through IVF-ICSI compared to natural conception(11), as opposed to no statistically significant difference in birth weight reported in the Norwegian study with 2,204 sibling pairs.(10) The odds of low birthweight were still 33% higher in the ART group compared to non-ART group. The odds of pre-term birth among ART as compared with non-ART infants was reduced by over half, i.e. 20% increased odds compared to a 51% increased odds in the conventional analysis, however the increased risk remained statistically significant for ART infants. This is in line with the results of the meta-analysis combining results from the previous sibling discordant analyses, which found the odds of pre-term birth to be 27% higher in the ART group compared to non-ART conception (pooled OR 1.27, 95% CI 1.08-1.49).(25) We did not find a significant association between ART and mean Apgar score at 5 minutes in either the conventional or sibling analysis, which is consistent with previous studies.(5, 9). We noted that the risk of all adverse perinatal outcomes was smaller in the ART group with only male factor infertility compared to female factor such that there was a 35% increased risk of preterm birth associated with ART with underlying female infertility and no significant increase in the risk associated with underlying male infertility. A previous study using data from the Society for Assisted Reproductive Technology Clinical Outcome Reporting System between 2004 and 2006 found the odds of low birthweight associated with ART with underlying tubal factor infertility and uterine factor infertility to be 78%

and 200% higher respectively compared to male factor infertility.(26) The exact mechanisms to explain this difference are not known.

Our study, with the largest number of sibling pairs represented in the published literature so far, found higher odds of low birthweight, pre-term deliveries and SGA in singleton infants born after ART compared to non-ART births, reiterating the increased risks of adverse perinatal outcomes associated with ART. Although the magnitude of risk associated with ART found in our study is lower than the previous studies, there was still an absolute risk difference of about 2% for low birthweight and preterm birth between the ART and non-ART groups. Given the long term effects of preterm births and low birthweight and the increasing number of ART births, couples undergoing any kind of ART should be counselled on these risks. Further robust research especially assessing the effects of different ART procedures like fresh embryos versus frozen embryos etc. on maternal and perinatal outcomes is also needed to improve ART treatments to minimize the risk of these adverse outcomes.

AUTHOR ROLES

NND and LJT conceived the idea of the study, which was designed with input from SB and DK. YZ created the dataset for analysis. NND conducted the analysis under supervision from SB, DK and LJT and with assistance from YZ. NND prepared the first draft of the manuscript. SB, DK, LJT, MB, PM and MH helped in interpretation of the findings and critical revision of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

Dhalwani was funded through Building Experience and Skills Travel Scholarship (BESTS) awarded by the Graduate School, University of Nottingham for the duration of the research project. The authors would like to thank Dr Lisa Szatkowski (PhD), University of Nottingham, Professor Tim Coleman (MD), University of Nottingham and Allison Mneimneh, CDC for their academic and administrative support during the project. Dhalwani acknowledges support from the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care – East Midlands (NIHR CLAHRC – EM), the Leicester Clinical Trials Unit and the NIHR Leicester-Loughborough Diet, Lifestyle and Physical Activity Biomedical Research Unit, which is a partnership between University Hospitals of Leicester NHS Trust, Loughborough University and the University of Leicester. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

CONFLICT OF INTEREST

None

REFERENCES

1. Dhalwani N, Fiaschi L, West J, Tata L. Occurrence of fertility problems presenting to primary care: population-level estimates of clinical burden and socioeconomic inequalities across the UK. Human Reproduction. 2013;28(4):960-8.
2. Sunderam S, Kissin DM, Crawford SB, Folger SG, Jamieson DJ, Barfield WD, et al. Assisted reproductive technology surveillance--United States, 2011. MMWR Surveill Summ. 2014;63(10):1-28.
3. Human Fertilisation and Embryology Authority. Fertility Treatment in 2013- trends and figures. London: Human Fertilisation and Embryology Authority, 2014.
4. Boulet SL, Schieve LA, Nannini A, Ferre C, Devine O, Cohen B, et al. Perinatal outcomes of twin births conceived using assisted reproduction technology: a population-based study†. Human reproduction. 2008;23(8):1941-8.
5. Schieve LA, Cohen B, Nannini A, Ferre C, Reynolds MA, Zhang Z, et al. A population-based study of maternal and perinatal outcomes associated with assisted reproductive technology in Massachusetts. Maternal and child health journal. 2007;11(6):517-25.
6. Tepper NK, Farr SL, Cohen BB, Nannini A, Zhang Z, Anderson JE, et al. Singleton preterm birth: Risk factors and association with assisted reproductive technology. Maternal and child health journal. 2012;16(4):807-13.
7. Grigorescu V, Zhang Y, Kissin DM, Sauber-Schatz E, Sunderam M, Kirby RS, et al. Maternal characteristics and pregnancy outcomes after assisted reproductive technology by infertility diagnosis: ovulatory dysfunction versus tubal obstruction. Fertility and Sterility. 2014.
8. Bower C, Hansen M. Assisted reproductive technologies and birth outcomes: overview of recent systematic reviews. Reproduction, Fertility and Development. 2005;17(3):329-33.
9. Marino JL, Moore VM, Willson KJ, Rumbold A, Whitrow MJ, Giles LC, et al. Perinatal Outcomes by Mode of Assisted Conception and Sub-Fertility in an Australian Data Linkage Cohort. PloS one. 2014;9(1):e80398.
10. Romundstad LB, Romundstad PR, Sunde A, von Düring V, Skjærven R, Gunnell D, et al. Effects of technology or maternal factors on perinatal outcome after assisted fertilisation: a population-based cohort study. The Lancet. 2008;372(9640):737-43.
11. 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. 2011;95(3):959-63.
12. Mneimneh A, Boulet S, Sunderam S, Zhang Y, Jamieson D, Crawford S, et al. States Monitoring Assisted Reproductive Technology (SMART) Collaborative: data collection, linkage, dissemination, and use. Journal of Womens Health. 2013;22(7).
13. Zhang Y, Cohen B, Macaluso M, Zhang Z, Durant T, Nannini A. Probabilistic linkage of assisted reproductive technology information with vital records, Massachusetts 1997-2000. Matern Child Health J. 2012;16(8):1703-8.
14. Alexander G, Himes J, Kaufman R, Mor J, Kogan M. A United States National Reference for Fetal Growth. Obstetrics & Gynecology. 1996;87(2):163-8.
15. World Health Organization. Health statistics and health information systems [cited 2013 29 August]. Available from: http://www.who.int/healthinfo/statistics/indlowbirthweight/en/.
16. World Health Organization. Preterm birth 2012 [cited 2013 5 November]. Available from: http://www.who.int/mediacentre/factsheets/fs363/en/.
17. Clayton P, Cianfarani S, Czernichow P, Johannsson G, Rapaport R, Rogol A. Management of the child born small for gestational age through to adulthood: a consensus statement of the International Societies of Pediatric Endocrinology and the Growth Hormone Research Society. The Journal of Clinical Endocrinology & Metabolism. 2007;92(3):804-10.
18. American Academy of Pediatrics, American College of Obstetricians & Gynecologists. The Apgar Score. Pediatrics. 2006;117(4).

19. Delbaere I, Vansteelandt S, De Bacquer D, Verstraelen H, Gerris J, De Sutter P, et al. Should we adjust for gestational age when analysing birth weights? The use of z-scores revisited. Human Reproduction. 2007;22(8):2080-3.
20. Zollinger TW, Przybylski MJ, Gamache RE. Reliability of Indiana birth certificate data compared to medical records. Ann Epidemiol. 2006;16(1):1-10.
21. Gardosi J, Madurasinghe V, Williams M, Malik A, Francis A. Maternal and fetal risk factors for stillbirth: population based study. BMJ. 2013;346:f108.
22. Boulet SL, Crawford S, Zhang Y, Sunderam S, Cohen B, Bernson D, et al. Embryo transfer practices and perinatal outcomes by insurance mandate status. Fertil Steril. 2015;104(2):403-9 e1.
23. Joshi N, Kissin D, Anderson JE, Session D, Macaluso M, Jamieson DJ. Trends and correlates of good perinatal outcomes in assisted reproductive technology. Obstet Gynecol. 2012;120(4):843-51.
24. Steinberg ML, Boulet S, Kissin D, Warner L, Jamieson DJ. Elective single embryo transfer trends and predictors of a good perinatal outcome--United States, 1999 to 2010. Fertil Steril. 2013;99(7):1937-43.
25. Pinborg A, Wennerholm UB, Romundstad LB, Loft A, Aittomaki K, Soderstrom-Anttila V, et al. Why do singletons conceived after assisted reproduction technology have adverse perinatal outcome? Systematic review and meta-analysis. Human Reproduction Update. 2013;19(2):87-104.
26. Gibbons WE, Cedars M, Ness RB, Society for Assisted Reproductive Technologies Writing G. Toward understanding obstetrical outcome in advanced assisted reproduction: varying sperm, oocyte, and uterine source and diagnosis. Fertil Steril. 2011;95(5):1645-9 e1.

Table 1 - Birth characteristics of the study population – (Florida & Massachusetts (2000-2010), Michigan (2000-2009))

	ART group (N=32,762)		Non-ART group (N= 3,863,480)		p-value
	n	%	n	%	
Maternal age at delivery in years					
20-24	199	0.6%	1,011,051	26.2%	
25-29	3,033	9.3%	1,156,781	29.9%	<0.001
30-34	10,781	32.9%	1,049,461	27.2%	
35-39	12,144	37.1%	530,357	13.7%	
>=40	6,605	20.2%	115,830	3.0%	
Parity					
0	21,298	65.0%	1,458,775	37.8%	
1	8,428	25.7%	1,326,951	34.3%	<0.001
2-9	2,997	9.1%	1,071,785	27.7%	
≥10	0	0.0%	1,236	0.0%	
Unknown	39	0.1%	4,731	0.1%	
Caesarean deliveries					
Yes	15,259	46.6%	1,192,890	30.9%	<0.001
No	17,503	53.4%	2,670,590	69.1%	
Infant's Sex					
Male	16,805	51.3%	1,979,007	51.2%	0.798
Female	15,957	48.7%	1,884,473	48.8%	
Smoking status during pregnancy					
Smoked during pregnancy	390	1.2%	383,652	10.0%	<0.001
Did not smoke during pregnancy	32,267	98.5%	3,464,240	89.7%	
Unknown	105	0.3%	15,588	0.4%	
Alcohol Use During Pregnancy					
No	30,855	94.2%	3,600,520	93.2%	<0.001
Yes	250	0.8%	26,023	0.7%	
Not collected	1,568	4.8%	222,058	5.7%	
Unknown	89	0.3%	14,879	0.4%	
Pre-Pregnancy BMI (kg/m^2)†					
Underweight	310	1.9%	65,495	2.1%	
Normal	5,098	31.1%	708,899	22.9%	
Overweight	1,961	11.9%	358,415	11.6%	<0.001
Obese	1,142	7.0%	299,997	9.7%	
Not yet collected	7,355	44.8%	1,563,338	50.5%	
Unknown	552	3.4%	101,158	3.3%	
Year of delivery					
2000-2005	16,027	48.9%	2,141,095	55.4%	<0.001
2006-2010	16,735	51.1%	1,722,385	44.6%	
State					
Florida	10,543	32.2%	2,014,959	52.2%	<0.001
Massachusetts	16,344	49.9%	766,178	19.8%	
Michigan	5,875	17.9%	1,082,343	28.0%	

† BMI not collected for Massachusetts, available for Florida 2004 onwards and for Michigan from late 2007.

Table 2 - Details of the ART procedures and causes (Florida & Massachusetts (2000-2010), Michigan (2000-2009))

	N =32,762	%
Reasons for ART*		
Diminished Ovarian Reserve	4,676	14.2
History of Endometriosis	4,307	13.5
Male Infertility	12,534	38.3
Other factors	4,778	14.6
Ovulation disorder/ Polycystic Ovaries	4,433	13.5
Tubal factor	6,234	19.0
Unexplained infertility	4,482	13.7
Uterine factor	1,172	3.6
Type of ART procedure		
IVF	32,383	98.8
Other(GIFT,ZIFT, combination, unknown)	379	1.2

GIFT=Gamma Intrafallopian Transfer, ZIFT=Zygote Intrafallopian Transfer

*The sum of all columns may not be equal to 32,764 or 100% as one woman may have more than one reason for ART

Table 3 - Mean birthweight and gestational age and mean differences between ART and non-ART group

	ART group	Non-ART group	Unadjusted β coefficient (95% CI)	Adjusted β coefficient (95% CI)
	Mean (sd)	Mean (sd)		
		Conventional Analysis		
	n=32,762	**n=3,863,480**		
Birthweight (grams)	3296.0 (614.8)	3334.3 (563.7)	-44.61 (-51.36,-37.87)	-64.11 (-69.74, -58.48)*
Gestational Age (days)	269.0 (15.5)	270.7 (13.7)	-1.69 (-1.86,-1.52)	-1.66 (-1.83,-1.49)**
		Discordant sibling pair analysis†		
	n=6,458	**n=6,458**		
Birthweight (grams)	3340.6 (584.2)	3397.9 (569.6)	-57.27 (-72.86,-41.68)	-33.40 (-48.60,-18.21)*
Gestational Age (days)	269.6 (14.0)	270.3 (13.9)	-0.58 (-0.99,-0.17)	-0.58 (-1.02,-0.14)**

† One sibling conceived naturally and the other one conceived through ART

*Adjusted for maternal age, year of birth, parity, infant's sex, gestational age, time between the current recorded delivery and the last recorded delivery

**Adjusted for maternal age, year of birth, parity, infant's sex, time between the current recorded delivery and the last recorded delivery

Table 4- Association between ART and low birthweight, pre-term birth, low Apgar score and SGA

	ART group	Non-ART group	Unadjusted OR (95% CI)	p-val	Adjusted OR (95% CI)	p-val
	n(%)	n(%)				
Conventional analysis						
	n=32,762	**n=3,863,480**				
Low birthweight	2,762 (8.4)	230,048 (6.0)	1.46 (1.40,1.51)	<0.001	1.38 (1.32,1.43)	<0.001**
Pre-term birth	3,813 (11.6)	307,327 (8.0)	1.52 (1.47,1.58)	<0.001	1.51 (1.46,1.56)	<0.001*
Low Apgar (<7)	424 (1.3)	45,599 (1.2)	1.09 (0.99,1.21)	0.059	0.99 (0.90,1.09)	0.888***
SGA‡	593 (1.8)	67,350 (1.7)	1.04 (0.96-1.13)	0.316	1.11 (1.03,1.21)	0.01*
Discordant sibling pair analysis						
	n=6,458	**n=6,458**				
Low birthweight	436 (6.8)	314 (4.9)	1.41 (1.24,1.62)	<0.001	1.33 (1.13,1.56)	<0.001**
Pre-term birth	627 (9.7)	516 (7.9)	1.24 (1.11,1.38)	0.001	1.20 (1.07,1.34)	0.002*
Low Apgar (<7)	64 (1.0)	84 (1.3)	0.76 (0.55,1.06)	0.101	0.75 (0.54,1.05)	0.096***
SGA‡	94 (1.4)	75 (1.2)	1.25 (0.93,1.69)	0.132	1.22 (0.88,1.68)	0.237*

† One sibling conceived naturally and the other one conceived through ART

*Adjusted for maternal age, year of birth, parity, infant's sex, time since last recorded delivery

** Adjusted for maternal age, year of birth, parity, infant's sex, gestational age, time since last recorded delivery

***Adjusted for maternal age, year of birth, parity, infant's sex, gestational age, delivery type, time since last recorded delivery

‡2 SD lower than the mean birthweight for gestational age and sex

Supplementary Table 1 - Birth characteristics of the study population for discordant sibling pairs – (Florida & Massachusetts (2000-2010), Michigan (2000-2009))

	ART group (N=6,458)		Non-ART group (N= 6,458)		p-value
	n	%	n	%	
Maternal age at delivery in years					
20-24	46	0.7%	113	1.7%	
25-29	743	11.5%	773	12.0%	<0.001
30-34	2,569	39.8%	2,367	36.7%	
35-39	2,442	37.8%	2,442	37.8%	
>=40	763	11.8%	763	11.8%	
Parity					
0	3,300	51.1%	2164	33.5%	
1	2,532	39.2%	3082	47.7%	<0.001
2-9	618	9.6%	1207	18.9%	
Unknown	8	0.1%	5	0.1%	
Caesarean deliveries					
Yes	2,779	43.0%	2,804	43.4%	0.026
No	3,679	57.0%	3,654	56.6%	
Infant's Sex					
Male	3,347	51.8%	3,388	52.5%	0.473
Female	3,111	48.2%	3,070	47.5%	
Smoking status during pregnancy					
Smoked during pregnancy	53	0.8%	75	1.2%	<0.268
Did not smoke during pregnancy	6,384	98.9%	6,366	98.6%	
Unknown	21	0.3%	17	0.3%	
Alcohol Use During Pregnancy					
No	6,168	95.5%	6,130	94.9%	0.574
Yes	56	0.9%	66	1.0%	
Not collected	218	3.4%	246	3.8%	
Unknown	16	0.2%	16	0.2%	
Pre-Pregnancy BMI (kg/m^2)†					
Underweight	52	0.8%	68	1.1%	
Normal	964	14.9%	948	14.7%	
Overweight	303	4.7%	313	4.8%	0.476
Obese	174	2.7%	177	2.7%	
Not yet collected	1443	22.3%	1,442	22.3%	
Unknown	100	1.5%	88	1.4%	
Year of delivery					
2000-2005	3,294	51.0%	3,499	54.2%	<0.001
2006-2010	3,164	49.0%	2,959	45.8%	
State					
Florida	1,913	29.6%	1,913	29.6%	
Massachusetts	3,422	53.0%	3,422	53.0%	*
Michigan	1,123	17.4%	1,123	17.4%	

† BMI not collected for Massachusetts, available for Florida 2004 onwards and for Michigan from late 2007

*no p-values reported as the states for both the birth were the same

Supplementary Table 2 – Association between ART and birth outcomes by main indication for fertility treatment

Indication for fertility treatment	Mean difference in infant birthweight (β (95% CI))**	OR for Low Birthweight (95% CI) **	Mean difference in gestational age (β (95% CI))*	OR for Preterm birth (95% CI)*	OR for Small-for-gestational age (95% CI) *	OR for low Apgar Score (95% CI)***
Conventional analysis						
Male factor only	-22.06 (-34.34,-9.77)	1.15 (1.04,1.27)	-0.45 (-0.78,-0.12)	1.19 (1.10,1.29)	1.08 (0.91,1.29)	0.81 (0.65,1.03)
Female factor only	-56.38 (-66.45,-46.31)	1.42 (1.33,1.52)	-2.64 (-2.94,-2.35)	1.72 (1.64,1.82)	1.05 (0.91,1.21)	1.05 (0.90,1.22)
Unexplained infertility only	-50.41 (-65.43,-35.38)	1.45 (1.29,1.63)	-0.38 (-0.81,0.04)	1.30 (1.18,1.43)	1.22 (0.98,1.51)	0.85 (0.62,1.13)
Discordant sibling pair analysis†						
Male factor only	-17.54 (-44.17,9.07)	0.91 (0.67,1.23)	0.15 (-0.61,0.90)	0.96 (0.77,1.20)	0.78 (0.42,1.46)	0.49 (0.24,0.98)
Female factor only	-45.16 (-68.19,-22.12)	1.45 (1.17,1.78)	-1.22 (-1.87,-0.57)	1.35 (1.16,1.58)	1.39 (0.92,2.09)	0.55 (0.33,0.90)
Unexplained infertility only	-17.44 (-44.95, 10.07)	1.27 (0.95,1.68)	0.21 (-0.54,0.96)	1.06 (0.85,1.31)	1.23 (0.70,2.14)	0.79 (0.43,1.44)

β = mean difference in ART-conceived compared with naturally conceived births as the baseline

† One sibling conceived naturally and the other one conceived through ART

OR = odds in ART-conceived compared with odds naturally conceived births as the baseline

*Adjusted for mother's age, parity, infant's sex and year of birth, time since last recorded delivery

** Adjusted for mother's age, parity, infant's sex and year of birth, gestational age, time since last recorded delivery

*** Adjusted for mother's age, parity, infant's sex and year of birth, gestational age, mode of delivery, time since last recorded delivery