pagepress

Pediatric Reports 2014; volume 6:5118

Health and development of children born after assisted reproductive technology and sub-fertility compared to naturally conceived children: data from a national study

Alastair G. Sutcliffe,[1] Edward Melhuish,[2,3] Jacqueline Barnes,[2,4] Julian Gardiner[2]

[1]Institute of Child Health, University College London; [2]Institute for the Study of Children, Families and Social Issues, Birkbeck, University of London; [3]University of Oxford; [4]Tavistock and Portman NHS Foundation Trust, London, UK

Correspondence: Julian Gardiner, Institute for the Study of Children, Families and Social Issues, Birkbeck, University of London, Malet Street, Bloomsbury, London WC1E 7HX, UK.
Tel.: +44.020.707.90853 - Fax: +44.020.707.90827.
E-mail: julian.gardiner@bbk.ac.uk

Key words: assisted reproductive technology, child health, sub-fertility.

Acknowledgments: we thank the families and their children for their cooperation throughout both projects.

Contributions: AS had the idea for the analysis and designed the questions for the 7 year sweep pertaining to fertility status in the mother interviews; JG performed statistical analyses; EM and JB were involved with colleagues for the design, collection of data and statistical planning of the cohort studies.

Conflicts of interests: the authors declare no potential conflict of interests.

Funding: Dr. Julian Gardiner is funded by Wellcome Trust grant number 089744/z/09/z.

Received for publication: 8 October 2013.
Revision received: 27 November 2013.
Accepted for publication: 27 November 2013.

Pediatric Reports 2014; 6:5118
doi:10.4081/pr.2014.5118

Abstract

In a non-matched case-control study using data from two large national cohort studies, we investigated whether indicators of child health and development up to 7 years of age differ between children conceived using assisted reproductive technology (ART), children born after sub-fertility (more than 24 months of trying for conception) and other children. Information on ART use/sub-fertility was available for 23,649 children. There were 227 cases (children conceived through ART) and two control groups: 783 children born to sub-fertile couples, and 22,639 children born to couples with no fertility issues. In models adjusted for social and demographic factors there were significant differences between groups in rate of hospital admissions before the children were 9 months old (P=0.029), with the ART group showing higher rates of hospital admission than the *no fertility issues* control group, the sub-fertile control group being intermediate between the two. Children born after ART had comparable health and development beyond 9 months of age to their naturally conceived peers. This applied to the whole sample and to a sub-sample of children from deprived neighborhoods.

Introduction

A range of studies has been performed looking at the health of children conceived using assisted reproductive technology (ART).[1-3] Existing studies have a number of weaknesses including a lack of a sub-fertile comparison group,[4] variable degrees of follow-up, low-power, lack of direct face-to-face assessments (*i.e*. health and developmental status inferred by proxy measures)[5] and other sources of bias such as failure of blinding,[6] and using non-standardized measures. Most studies have not been population based, and there has also been no study of ART specifically in families from deprived neighborhoods.

We describe a study that overcomes some of these difficulties. The study combines data from two national birth cohorts: the Millennium Cohort Study (MCS)[7] and The National Evaluation of Sure Start (NESS),[8] which use common measures. Our objective was to determine if ART conceived children assessed in some detail in national studies established for other purposes had similar outcomes to controls who were either born after parental sub-fertility not needing ART or naturally conceived.

In addition to the main analysis of a combined MCS/NESS sample, we analyze the NESS data separately; the NESS study sample consists of families from deprived neighborhoods and we believe this is the first analysis of the health of children born after ART in a deprived sample.

Materials and Methods

Sample

The study sample consisted of children from the Millennium Cohort Study and from the National Evaluation of Sure Start study.[7,8]

Eligible children for the Millennium Cohort Study were all children in England born over a period of 16 months from September 2000 and living in the 398 wards. The random sample was clustered geographically by electoral ward with some oversampling to ensure adequate representation of wards with a high ethnic minority population (≥30% population black or Asian in 1991 census) and disadvantaged areas from the poorest 25% of wards using the child poverty index,[9] which is based on the proportion of children in families receiving means tested benefits. Overall there were 188 advantaged wards (not in poorest 25%), 191 disadvantaged wards, and 19 wards with a high proportion of ethnic minority families. Children were sampled from the government's child benefit records. Child benefit is a universal provision, payable to mothers from the birth of their children. The take-up of child benefit exceeds 97%. Apart from the possibility of eligible families being too rich or too ill-informed to claim, most of the children not claimed for were ineligible as the children of non-nationals with temporary or unconfirmed residence status, such as foreign armed forces, overseas students, and recent immigrants, including asylum seekers. The attained sample at nine months was 18,552 children and families (response rate 70%). Of these, 14,898 were seen again when the children were aged 3 (80.3% retention rate), 14,678 were seen when the children were aged 5 (79.1% retention rate) and 13,363 when the children were aged 7 (72.0% retention rate). Additional children and families were recruited at three years to give a sample of 15,590 at age 3, 15,246 at age 5 and 13,857 at age 7.[10]

The sample from the National Evaluation of Sure Start study was selected from areas in England chosen to receive a Sure Start local program, all in the 20% most disadvantaged areas defined by the child poverty index.[9] From children born in 200 randomly chosen Sure Start areas during 29 months from January 2002, a random sample of 12,705 infants aged 9 months was chosen (response rate 84.0%), again using the child benefit records as a sampling frame. Of those seen at this age, 11,118 children and families were randomly selected to be followed-up when the child was aged 3 years, 9191 (82.7%) of whom participated in data collection at the 3 years age point. When the children were aged 5 years 8000 of the children and families seen when the children were aged 3 were randomly selected for follow-up; data were collected from 7258 (response rate 90.7%). When the children were aged 7 years 5940 families were selected for follow-up; data were collected from 5393 (response rate 90.8%).

We applied no exclusion criteria except to include in analyses only the first born child when multiple births occurred.

Data on time to conception and the use of ART were collected when the children were 9 months old in the MCS study and when the children were 7 years old in the NESS study. Of those present in the studies at these times, data were available for 98.7% of children in the MCS and 98.8% of children in NESS. The sample analyzed consisted of the 23,649 children for which this information was available (18,319 from the MCS and 5330 from NESS). The number of children in the analysis group at each time point is summarized in Supplementary Table 1.

Children were classified into three groups: those who were conceived using assisted reproduction techniques (*ART group*; N=227: 182 from MCS, 45 from NESS), those who were conceived naturally after more than 24 months of trying for conception (*sub-fertile group*; N=783: 612 from MCS, 171 from NESS) and the remaining children (*control group*; N=22,639: 17,525 from MCS, 5114 from NESS).

Outcomes

We chose the child outcomes because they have been used as indicators of child wellbeing in reports from the World Health Organization,[11] applied to the total population, were likely to predict later health and development, and were reliably measured by parental report or researcher. Outcomes collected by parental report were socio-emotional difficulties experienced by the children, using the strengths and difficulties questionnaire (SDQ);[12] unintentional injuries requiring medical treatment (from nurse, general practitioner, hospital, or medical clinic) in the past year and admissions to hospital in the past year. To calculate the children's body mass index a researcher measured their weight and height during home visits. We determined whether the child was overweight using BMI reference data from the Centers for Disease Control and Prevention,[13] defining *overweight* as *above the 85th percentile for child's sex and age*). In addition, researchers assessed language development using the British Ability Scales naming vocabulary subscale.[14] Not all measurements were available for all age groups (Table 1).

Analysis

The outcome variables were modeled using linear mixed-effects models for the continuous outcomes, and logistic mixed-effects models for the binary outcomes. A random effect was fitted for cluster in all cases. Analysis was carried out using R 2.11.1,[15] using the packages NLME and MASS.[16,17]

Unadjusted models

Whether the values of the outcome variables differed significantly between groups was assessed using an ANOVA comparison of a model in terms of a three-level group factor (*control/sub-fertile/ART*) and a model of the outcome in terms of an intercept only. Comparisons between each pair of groups were carried out using a Bonferroni correction factor of 3 in order to maintain the overall Type I error rate of 5%. The binary and continuous covariates were tested for group differences in the same way.

We tested for group differences in the categorical covariates using chi-square tests on the group by covariate-value contingency tables, with merging of low-count cells where necessary. Further chi-square tests were carried out to compare each pair of groups, using

Table 1. Summary of unadjusted comparisons of outcome variables by group. For the binary variables, numbers of cases are given, with the percentage of the group this represents in brackets. For the continuous variables, group means are given with their standard error in brackets. The P for an ANOVA test for group differences is shown, along with the P for tests for differences between each pair of groups, using a Bonferroni correction to maintain the Type I error rate of 5%. Higher SDQ Difficulties Score values indicate greater child difficulties. Higher BAS Naming Vocabulary Scores indicate larger child vocabulary.

Outcome variable	Age group	Group 1: controls	Group 2: sub-fertile	Group 3: ART	P for test for any group differences	P for test differences between pairs of groups 1 *vs.* 2	1 *vs.* 3	2 *vs.* 3
Binary outcome variables (%)								
Child has had accident	9 m	1865 (8.3)	74 (9.6)	16 (7.1)	0.37	0.68	1	0.72
	3 y	6350 (33.3)	234 (33.9)	57 (29.8)	0.47	1	0.74	0.68
	5 y	5278 (28.0)	162 (24.0)	57 (28.8)	0.081	0.085	1	0.57
	7 y	4257 (28.0)	145 (22.9)	46 (25.3)	0.72	1	1	1
Child has had hospital admission	9 m	3250 (14.5)	116 (15.0)	39 (17.3)	0.39	1	0.52	0.97
	3 y	4439 (24.1)	184 (27.4)	43 (22.8)	0.099	0.089	1	0.66
	5 y	2742 (14.5)	115 (17.0)	25 (12.6)	0.13	0.17	1	0.41
	7 y	1944 (11.0)	69 (10.9)	12 (6.6)	0.14	2.6	0.20	0.32
Child is overweight	3 y	5495 (33.2)	196 (31.9)	45 (25.6)	0.069	1	0.092	0.31
	5 y	5550 (30.1)	188 (28.6)	43 (22.2)	0.032*	1	0.047*	0.22
	7 y	4313 (25.0)	162 (25.9)	34 (19.4)	0.21	1	0.32	0.27
Continuous outcome variables (SE)								
Child's BMI	3 y	16.8 (0.017)	16.8 (0.083)	16.6 (0.18)	0.59	1	1	1
	5 y	16.3 (0.014)	16.3 (0.076)	16.1 (0.13)	0.29	1	0.35	0.47
	7 y	16.7 (0.018)	16.8 (0.10)	16.2 (0.17)	0.035*	1	0.049*	0.030*
SDQ difficulties score	3 y	10.9 (0.040)	10.6 (0.20)	10.3 (0.35)	0.85	1	1	1
	5 y	8.98 (0.035)	8.53 (0.18)	8.16 (0.30)	0.047*	0.12	0.48	1
	7 y	9.11 (0.040)	8.59 (0.19)	8.31 (0.34)	0.043*	0.085	0.63	1
BAS naming vocabulary	3 y	48.7 (0.087)	50.5 (0.48)	53.2 (0.82)	$<10^{-4}$***	0.0016**	0.0016**	0.47
	5 y	52.4 (0.084)	54.4 (0.46)	57.1 (0.79)	$<10^{-4}$***	0.0004***	0.0004***	0.39
Number in group		22 639	783	227				

ART, Assisted Reproductive Technology; BMI, Body Mass Index; SDQ, Strengths and Difficulties Questionnaire; BAS, British Ability Scales; m, months; y, years. Where the Bonferroni correction produced P>1 these have been reported as 1. Significant P are indicated: *P<0.05, **P< 0.01, ***P< 0.001.

a Bonferroni correction factor of 3 for multiple comparisons.

Adjusted models

Whether there were any group differences once demographic and socio-economic factors had been controlled for was determined using adjusted models. These controlled for the following covariates: child's sex, child's age (this differed slightly from the nominal age for a given sweep of the survey), mother's age at child's birth, father's age at child's birth, child's birth weight, whether child was breast fed for at least 6 weeks, mother's parity at birth of child (as a proxy for birth order), the number of siblings child has, child's ethnic group, child being raised in a workless household, child being raised by a lone parent, mother's educational attainment, household income and mother's social class (as assessed from habitual employment). Models of *child's BMI* and *child is overweight* were also controlled for mother's BMI. When the children were 9 months old, *parity* was virtually identical to *number of siblings* so *parity* was omitted from these models.

Multiple imputation

There were missing values in the data for two reasons: i) non-response on a particular item or items at a given survey, ii) attrition of the sample, where a subject has dropped out of the study completely. Multiple imputation was used to handle both sorts of missingness.[18] Missing data were imputed using the Amelia II package.[19] The imputation model assumes a multivariate normal distribution for the complete data (missing and observed). Binary, categorical and ordinal variables are incorporated into this distribution using appropriate transformations.[20] Whilst the use of the multivariate normal distribution is inevitably an approximation, its effectiveness in missing data problems is well established.[21] All the outcome variables and covariates from all time points were included in the imputation model.

Five imputations were generated, and models fitted to each imputed data set. Model results were consolidated using Rubin's Rules,[22] following Hesterberg.[23]

Power

The power of this study is limited by the relatively small sizes of the ART and sub-fertile groups. For comparisons of group means (continuous variables) or proportions (binary variables) the effect sizes required for a difference to be detected between each pair of groups with a 5% Type I error rate and power of 80% were calculated using Cohen's methods for comparing groups of different sizes as implemented by Champely.[24,25] These effect sizes were 0.12 (Control *vs*. Sub-fertile group), 0.22 (Control *vs*. ART group) and 0.24 (Sub-fertile *vs*. ART group). For the continuous outcomes, these effect sizes correspond to differences in group means in units of the standard deviation of the outcome variable. For the binary outcomes, the effect size depends on the difference between the proportions of positive outcomes in the two groups, p_1 and p_2:

$$\text{Effect size}=\frac{|p_1-p_2|}{\sqrt{p_1(1-p_1)+p_2(1-p_2)}}$$

To give a feel for the difference in proportions which this study can detect, if p_1=20%, then to give 80% power with Type I error rate of 5%, it is necessary for p_2 to be less than 13.8% or greater than 27.0% (Control *vs*. Sub-fertile group), less than 9.3% or greater than 33.3% (Control *vs*. ART group) or less than 8.2% or greater than 35.2% (Sub-fertile *vs*. ART group).

Analysis of a sub-sample from deprived neighborhoods

Analyses were repeated using data from the NESS study only, which is a sample of children from deprived neighborhoods. The power of these analyses was lower than those using the whole sample; the effect sizes which could be detected with 80% power and Type I error rate of 5% were 0.25 (Control *vs*. Sub-fertile group), 0.48 (Control *vs*. ART group) and 0.55 (Sub-fertile *vs*. ART group).

Differential drop out

In order to investigate the possible effect of differential dropout in the NESS survey, we fitted a logistic mixed-effects regression model to the 9 month survey data (N=12,705) with binary outcome *child not in study for NESS 7 year survey* and covariates as used in the adjusted models described above, with time-varying covariates evaluated at the 9 month survey.

Results

A breakdown of ART status by social class is given in Supplementary Table 2. Summary statistics for the outcome variables by group are given in Table 2 along with the results of the unadjusted models. Summary statistics and tests for group differences for the covariates are given in Tables 2-4. The results of the adjusted models are given in Table 5.

There are significant differences in the outcomes for the three groups in a number of cases (Table 1). Children in the ART group are less likely to be overweight at 5 years old than the controls, and have significantly lower BMIs at 7 years old than both controls and sub-fertile group children. There are significant group differences in socio-emotional difficulties (SDQ) score at both 5 and 7 years, with the ART group having lower mean scores than the sub-fertile group, which in turn have lower mean scores than the controls (although no pair of groups exhibit significant differences). There are highly significant ($P<10^{-4}$) differences in BAS Naming Vocabulary Score between groups, with both sub-fertile and ART groups having higher mean scores than the controls.

There are significant differences between groups on the covariates in a number of cases (Tables 2-4). ART group children are more likely to have been breast-fed than the sub-fertile group children, who in turn are more likely to have been breast fed than controls (Table 3); both ART group and sub-fertile group children are less likely to be raised in workless households than controls; the mean age of both mothers and fathers are higher in the ART group than in the controls, with the sub-fertile group intermediate between the two; birth weight in the ART group is significantly lower than in either of the other groups. As would be expected, there are significant differences between groups on number of sibs and parity (Table 2). There are significant differences in household income between groups (Table 4), with a higher proportion of high-earners in the ART and sub-fertile groups than among controls; there are also significant differences between groups in mother's socio-economic status.

The adjusted models show group differences in child outcomes in one case only (Table 5), namely *child has had hospital admission* up to 9 month old, but not at any subsequent age. In all other cases there are no significant differences in child outcomes between the three groups once the covariates have been controlled for.

In the analysis of the NESS sample of children from deprived neighborhoods there were no significant differences between groups in the adjusted models (details omitted).

A number of covariates were significantly associated with increased risk of dropout between the 9 months and 7 year sweeps of the NESS study: *child being raised in workless household* (OR=1.12, P=0.038), child's age (OR=1.08 per month of child's age, P=0.0011), child has three or more sibs (OR=1.40, $P<10^{-4}$) and *child's ethnic group Afro-Caribbean* (OR=1.76, $P<10^{-4}$). Other factors were associated with lower risk of dropout by the 7 year survey: mother's age (OR=0.83 per 5 years of maternal age, $P<10^{-4}$), child's birth weight (OR=0.93 per Kg of birth weight, P=0.032), *mother has some formal qualifications* (OR=0.50, $P<10^{-4}$) and mother's BMI (OR=0.92 per 5 units of maternal BMI, P=0.0002).

Discussion

Overview

Although there are no significant differences between groups in the children's probability of having been admitted to hospital by

Table 2. Summary of comparisons of categorical covariates by group (I). Numbers of cases are given, with the percentage of the group this represents in brackets. The P for a chi-square test for group differences is given, along with the P for tests for differences between each pair of groups, using a Bonferroni correction to maintain the Type I error rate of 5%.

Variable	Levels	Group 1: controls	Group 2: sub-fertile	Group 3: ART	Any group differences?	Differences between pairs of groups 1 *vs.* 2	1 *vs.* 3	2 *vs.* 3
Number of sibs (9 months)	0	9274 (41.0)	430 (54.9)	141 (62.1)	<10–4***	<10–4***	<10–4***	0.82
	1	7734 (34.2)	255 (32.6)	64 (28.2)				
	2	3610 (15.9)	69 (8.8)	16 (7.0)				
	3+	2021 (8.9)	29 (3.7)	6 (2.6)				
Number of sibs (3 years)	0	6201 (27.4)	270 (34.5)	113 (49.8)	<10–4***	<10–4***	<10–4***	0.0003***
	1	9574 (42.3)	369 (47.1)	84 (37.0)				
	2	4293 (19.0)	99 (12.6)	26 (11.5)				
	3+	2571 (11.4)	45 (5.7)	4 (1.8)				
Number of sibs (5 years)	0	3017 (16.0)	153 (22.6)	80 (40.4)	<10–4***	<10–4***	<10–4***	<10–4***
	1	8626 (45.6)	341 (50.4)	78 (39.4)				
	2	4593 (24.3)	129 (19.1)	33 (16.7)				
	3+	2675 (14.1)	53 (7.8)	7 (3.5)				
Number of sibs (7 years)	0	2209 (12.5)	125 (19.6)	71 (39.0)	<10–4***	<10–4***	<10–4***	<10–4***
	1	7795 (44.0)	310 (48.7)	75 (41.2)				
	2	4856 (27.4)	143 (22.4)	30 (16.5)				
	3+	2856 (16.1)	59 (9.3)	6 (3.3)				
Parity	1	9687 (47.1)	443 (60.4)	165 (80.9)	<10–4***	<10–4***	<10–4***	<10–4***
	2	6748 (32.8)	221 (30.1)	35 (17.2)				
	3+	4120 (20.0)	70 (9.5)	4 (2.0)				
Child's ethnic group	White	18323 (81.7)	642 (83.1)	199 (88.1)	0.12	1	0.13	0.57
	Mixed	776 (3.5)	18 (2.3)	7 (3.1)				
	Indian	495 (2.2)	23 (3.0)	3 (1.3)				
	Pakistani/Bangladeshi	1575 (7.0)	53 (6.9)	10 (4.4)				
	Black	834 (3.7)	25 (3.2)	6 (2.7)				
	Other	425 (1.9)	12 (1.6)	1 (0.4)				

ART, Assisted Reproductive Technology. Where the Bonferroni correction produced P>1 these have been reported as 1. Significant P are indicated: ***P<0.001.

Table 3. Summary of comparisons of binary and continuous covariates by group. For the binary variables, numbers of cases are given, with the percentage of the group this represents in brackets. For the continuous variables, group means are given with their standard error in brackets. The P for an ANOVA test for group differences is shown, along with the P for tests for differences between each pair of groups, using a Bonferroni correction to maintain the Type I error rate of 5%.

Covariate		Group 1: controls	Group 2: sub-fertile	Group 3: ART	Any group differences	Differences between pairs of groups 1 *vs.* 2	1 *vs.* 3	2 *vs.* 3
Binary covariates (%)								
Child's sex (female)		11 050 (48.8)	393 (50.2)	115 (50.7)	0.65	1	1	1
Breast fed for 6 weeks or more		9 302 (41.1)	363 (46.4)	130 (57.3)	<10–4***	0.0098**	<10–4***	0.012*
Child in workless household	9 m	5 227 (23.3)	79 (10.2)	12 (5.3)	<10–4***	<10–4***	<10–4***	0.080
	3 y	4 083 (21.3)	80 (11.5)	9 (4.7)	<10–4***	<10–4***	<10–4***	0.021*
	5 y	3 761 (19.8)	66 (9.7)	9 (4.5)	<10–4***	<10–4***	<10–4***	0.074
	7 y	3 255 (18.4)	64 (10.0)	9 (4.9)	<10–4***	<10–4***	<10–4***	0.11
Child raised by lone parent	9 m	4 592 (20.5)	54 (7.0)	10 (4.4)	<10–4***	<10–4***	<10–4***	0.51
	3 y	3 922 (20.5)	59 (8.5)	14 (7.3)	<10–4***	<10–4***	<10–4***	1
	5 y	3 984 (21.0)	71 (10.5)	17 (8.6)	<10–4***	<10–4***	0.0001***	1
	7 y	4 257 (24.0)	76 (11.9)	22 (12.1)	<10–4***	<10–4***	0.0008***	1
Continuous covariates (SE)								
Child's age	9 m	0.801 (0.00034)	0.805 (0.0018)	0.807 (0.0036)	0.017*	0.10	0.16	1
	3 y	3.15 (0.0014)	3.14 (0.0070)	3.13 (0.014)	0.073	0.38	0.25	1
	5 y	5.21 (0.0018)	5.21 (0.0098)	5.20 (0.018)	0.79	1	1	1
	7 y	7.22 (0.0019)	7.22 (0.010)	7.22 (0.020)	0.92	1	1	1
Mother's age		28.4 (0.040)	31.1 (0.18)	34.2 (0.33)	<10–4***	<10–4***	<10–4***	<10–4***
Father's age		31.5 (0.047)	34.1 (0.21)	36.0 (0.39)	<10–4***	<10–4***	<10–4***	0.0006***
Child's birth weight (Kg)		3.34 (0.0039)	3.3 (0.022)	3.02 (0.052)	<10–4***	0.35	<10–4***	<10–4***
Mother's BMI		25.7 (0.087)	27.6 (0.52)	26.9 (0.99)	0.0002***	0.0002***	0.47	1
Number in group		22 639	783	227				

ART, Assisted Reproductive Technology; BMI, Body Mass Index; m, months; y, years.
Where the Bonferroni correction produced P>1 these have been reported as 1. Significant P are indicated: *P<0.05, **P<0.01, ***P<0.001.

the age of 9 months (Table 1), significant group differences were found once the covariates were adjusted for (Table 5). This superficially puzzling result can be explained by the fact that there is a tendency to increased socio-economic status and maternal age as one moves from the controls to the sub-fertile group and on to the ART group. Both higher SES,[26,27] and higher maternal age,[28] are associated with lower rates of hospital admission. The significant differences found in hospital admission rates up to 9 months of age in the adjusted models reflect the fact that the sub-fertile and ART groups fail to show the lower admission rates that would be expected from groups with the observed levels of SES and maternal age. It is possible that the higher than expected rates of hospital admission during the first 9 months of life in the ART and sub-fertile groups could be at least partly explained by parents who have undergone ART/sub-fertility feeling greater anxiety about their babies, rather than these children actually having poorer health.

For children older than 9 months, there are no significant group differences in hospital admission rates, showing that this 9-month difference does not indicate comprised health outcomes. There were also no significant differences on any of the other outcomes. These results suggest that artificially conceived children have as good a prognosis as other children with respect to the health and well-being related outcomes analyzed in this study.

Generalizability

The study sample was drawn disproportionately from areas with higher than average levels of deprivation. Nevertheless, the study includes families from all socio-economic strata (Supplementary Table 2). This table also confirms that all social classes have access to ART – as would be expected, since at least one treatment cycle is available to all infertile couples via the NHS – although the higher rates of ART usage in the managerial and self-employed/small employer classes show that socio-economic status affects families' access to ART. Since there are data from all social classes, including substantial numbers of deprived families, all of whom could potentially access ART, and the final models control for socio-economic status it is reasonable to conclude that the results are applicable to the whole population.

Strengths and limitations of the study

The MCS and NESS studies did not originally set out to explicitly investigate the health of children born after assisted conception. It is a strength of this analysis that it is based on a cohort of children rather than on case/control data collected for the specific purpose of assessing the effects of ART or sub-fertility since the risks of participation bias and confounding are reduced. However, the relativity low rate of ART use – due in part to the limited availability of publically funded treatment – limits the power of the study, despite the fairly large sample size (N=23,649). So, as with all negative findings, the conclusions are provisional rather than definitive.[29]

Table 4. Summary of comparisons of categorical covariates by group (II). Numbers of cases are given, with the percentage of the group this represents in brackets. The P for a chi-square test for group differences is given, along with the P for post hoc tests for differences between each pair of groups, using a Bonferroni correction to maintain the Type I error rate of 5%.

Variable	Levels	Group 1: controls	Group 2: sub-fertile	Group 3: ART	Any group differences?	Differences between pairs of groups		
						1 *vs.* 2	1 *vs.* 3	2 *vs.* 3
Mother's educational attainment	None	2753 (12.2)	78 (10.0)	13 (5.7)	0.056	0.21	0.46	0.87
	GCSE or equivalent	10393 (45.9)	315 (40.3)	94 (41.4)				
	A levels or equivalent	5305 (23.4)	194 (24.8)	48 (21.1)				
	Degree/higher degree	3662 (16.2)	173 (22.1)	67 (29.5)				
	Other	515 (2.3)	22 (2.8)	5 (2.2)				
Household income (9 months)	<£11,000	7001 (33.4)	123 (17.2)	18 (8.7)	$<10^{-4}$***	$<10^{-4}$***	$<10^{-4}$***	0.013*
	£11,000 to £22,000	7806 (37.3)	294 (41.0)	82 (39.8)				
	>£22,000	6143 (29.3)	300 (41.8)	106 (51.5)				
Household income (3 years)	<£11,000	5772 (26.4)	108 (14.4)	21 (9.7)	$<10^{-4}$***	$<10^{-4}$***	$<10^{-4}$***	0.0060*
	£11,000 to £22,000	7497 (34.3)	243 (32.4)	51 (23.6)				
	>£22,000	8596 (39.3)	399 (53.2)	144 (66.7)				
Household income (5 years)	<£11,000	4950 (22.3)	91 (11.9)	20 (9.0)	$<10^{-4}$***	$<10^{-4}$***	$<10^{-4}$***	0.10
	£11,000 to £22,000	7158 (32.3)	227 (29.6)	51 (22.9)				
	>£22,000	10087 (45.4)	448 (58.5)	152 (68.2)				
Household income (7 years)	<£11,000	3846 (17.7)	65 (8.7)	16 (7.2)	$<10^{-4}$***	$<10^{-4}$***	$<10^{-4}$***	0.55
	£11,000 to £22,000	6991 (32.1)	204 (27.3)	49 (22.1)				
	>£22,000	10932 (50.2)	479 (64.0)	157 (70.7)				
Mother's social class (9 months)	Management	5379 (25.2)	265 (35.8)	104 (46.4)	$<10^{-4}$***	$<10^{-4}$***	$<10^{-4}$***	0.0043*
	Intermediate	3755 (17.6)	131 (17.7)	48 (21.4)				
	Small employer/self-employed	879 (4.1)	35 (4.7)	15 (6.7)				
	Technical	1356 (6.4)	47 (6.4)	9 (4.0)				
	Routine	9291 (43.6)	239 (32.3)	46 (20.5)				
	Unemployed	656 (3.1)	23 (3.1)	2 (0.9)				
Mother's social class (5 years)	Management	5054 (23.2)	245 (32.2)	101 (44.9)	$<10^{-4}$***	$<10^{-4}$***	$<10^{-4}$***	0.010*
	Intermediate	4581 (21.0)	157 (20.6)	43 (19.1)				
	Small employer/self-employed	1863 (8.6)	72 (9.4)	22 (9.8)				
	Technical	2069 (9.5)	65 (8.5)	16 (7.1)				
	Routine	7685 (35.3)	203 (26.6)	41 (18.2)				
	Unemployed	527 (2.4)	20 (2.6)	2 (0.9)				

ART, Assisted Reproductive Technology; GCSE, General Certificate of Secondary Education. Where the Bonferroni correction produced P>1 these have been reported as 1. Significant P are indicated: *P<0.05, **P<0.01, ***P<0.001.

Table 5. Results of adjusted models. Group differences are given as odds ratios (binary variables) and differences in units of the standard deviation (continuous variables) with 95% confidence intervals in brackets. The p-value for a test for any group differences is given. Higher SDQ Difficulties Score values indicate greater child difficulties. Higher BAS Naming Vocabulary Scores indicate larger child vocabulary.

Outcome variable		Group 2 (Sub-fertile) *vs.* Group 1 (Control)	Group 3 (ART) *vs.* Group 1 (Control)	Group 3 (ART) Group 2 *vs.* (Sub-fertile)	Any group differences? (P)
Binary covariates					
Child has had accident	9 m	1.26 (0.93 to 1.71)	0.94 (0.5 to 1.77)	0.74 (0.37 to 1.48)	0.19
	3 y	1.12 (0.91 to 1.38)	0.97 (0.64 to 1.47)	0.88 (0.55 to 1.4)	0.47
	5 y	0.91 (0.73 to 1.14)	1.11 (0.74 to 1.67)	1.22 (0.79 to 1.88)	0.44
	7 y	0.99 (0.74 to 1.33)	1.09 (0.73 to 1.63)	1.11 (0.65 to 1.88)	1.0
Child has had hospital admission	9 m	1.22 (0.95 to 1.58)	1.45 (0.94 to 2.26)	1.17 (0.71 to 1.92)	0.029*
	3 y	1.27 (0.98 to 1.65)	1.08 (0.7 to 1.65)	0.86 (0.53 to 1.39)	0.12
	5 y	1.24 (0.93 to 1.64)	0.93 (0.56 to 1.55)	0.76 (0.41 to 1.41)	0.28
	7 y	1.04 (0.76 to 1.41)	0.74 (0.35 to 1.53)	0.71 (0.35 to 1.44)	0.53
Child is overweight	3 y	0.94 (0.77 to 1.15)	0.91 (0.57 to 1.44)	0.96 (0.59 to 1.57)	0.72
	5 y	0.93 (0.75 to 1.15)	0.77 (0.49 to 1.23)	0.83 (0.52 to 1.33)	0.45
	7 y	1.02 (0.83 to 1.26)	0.9 (0.55 to 1.49)	0.88 (0.51 to 1.51)	1.0
Continuous covariates					
Child's BMI	3 y	-0.018 (-0.11 to 0.071)	0.022 (-0.15 to 0.19)	0.032 (-0.15 to 0.22)	0.84
	5 y	0.003 (-0.09 to 0.096)	-0.03 (-0.24 to 0.18)	-0.036 (-0.25 to 0.18)	0.99
	7 y	0.03 (-0.064 to 0.12)	-0.078 (-0.26 to 0.11)	-0.11 (-0.31 to 0.087)	0.44
SDQ difficulties score	3 y	0.067 (-0.018 to 0.15)	0.084 (-0.097 to 0.26)	0.011 (-0.19 to 0.21)	0.096
	5 y	0.013 (-0.08 to 0.11)	0.028 (-0.14 to 0.2)	0.012 (-0.2 to 0.22)	0.85
	7 y	0.00012 (-0.089 to 0.09)	0.012 (-0.14 to 0.17)	0.0089 (-0.17 to 0.19)	1.0
BAS naming vocabulary	3 y	-0.005 (-0.087 to 0.077)	0.0095 (-0.14 to 0.16)	0.027 (-0.14 to 0.19)	0.96
	5 y	0.0089 (-0.077 to 0.095)	0.025 (-0.12 to 0.17)	0.017 (-0.15 to 0.18)	0.87

ART, Assisted Reproductive Technology; BMI, Body Mass Index; SDQ, Strengths and Difficulties Questionnaire; BAS, British Ability Scales; m, months; y, years. *P<0.05.

There is a possibility of bias due to differential drop out. The use of multiple imputation largely forestalls this difficulty as far as the MCS data is concerned, where data on ART use/sub-fertility were collected when the children were 9 months old. In the NESS study, the data on ART/sub-fertility status was not collected until 7 years of age, by which time the sample had been reduced both by random re-sampling and potentially non-random drop-out.

Reduction in the sample between the 9 month and 7 year surveys was higher for children with lower birth weight, children from workless families, larger families and Afro-Caribbean families, and families where the mother was younger or had no qualifications. These factors are broadly associated with lower socio-economic status. This makes it probable that there was lower drop out between the 9 month and 7 year surveys among children conceived through ART than among controls. Among the controls, the children who drop out will come disproportionately from the lower SES families, which are also the children most likely to have poorer health outcomes. This makes it unlikely that the differential dropout between control and ART groups has masked poorer health or development outcomes among the ART children.

References

1. Ludwig AK, Sutcliffe AG, Diedrich K, Ludwig M. Post-neonatal health and development of children born after assisted reproduction: a systematic review of controlled studies. Eur J Obstet Gynecol Reprod Biol 2006;127:3-25.
2. Hansen M, Bower C, Milne E, et al. Assisted reproductive technologies and the risk of birth defects - a systematic review. Hum Reprod 2005;20:328-38.
3. Helmerhorst FM, Perquin DA, Donker D, Keirse MJ. Perinatal outcome of singletons and twins after assisted conception: a systematic review of controlled studies. BMJ 2004;328:261.
4. Sutcliffe AG, Ludwig M. Outcome of assisted reproduction. Lancet 2007;370:351-9.
5. Stromberg B, Dahlquist G, Ericson A, et al. Neurological sequelae in children born after in-vitro fertilisation: a population-based study. Lancet 2002;359:461-5.
6. Ludwig AK, Katalinic A, Entenmann A, et al. Can we sense ART? The blinded examiner is not blind-a problem with follow-up studies on children born after assisted reproduction. Fertil Steril 2009;92:950-2.
7. Dex S, Joshi H, eds. Millennium cohort study, first survey: a user's guide to initial findings. London: Centre for Longitudinal Studies; 2004.
8. Melhuish E, Belsky J, Leyland AH, et al. Effects of fully-established Sure Start Local Programmes on 3-year-old children and their families living in England: a quasi-experimental observational study. Lancet 2008;372:1641-7.
9. Noble M, Smith G, Penhale B, et al. Measuring multiple deprivation at the small area level: the indices of deprivation 2000. Regeneration Research Summary No. 37. London: Department of the Environment, Transport and the Regions; 2000.
10. Plewis I. Millennium cohort study first survey: technical report on sampling (4th ed). London: Centre for Longitudinal Studies, Institute of Education; 2007.
11. Irwin L, Siddiqi A, Hertzman C. Early child development: a powerful equalizer. Final report for the World Health Organization's Commission on the social determinants of health. 2007. Available from: http://www.who.int/social_determinants/resources/ecd_kn_report_07_2007.pdf
12. Goodman R. The strengths and difficulties

questionnaire: a research note. J Child Psychol Psychiatry 1997;38:581-6.
13. Centers for Disease Control and Prevention. Percentile data files with LMS values. Available from: http://www.cdc.gov/growthcharts/percentile_data_files.htm
14. Elliott C, Smith P, McCullock K. BAS II British Ability scales technical manual. Windsor: NFER Nelson; 1997.
15. R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2005.
16. Pinheiro J, Bates D. nlme (R Package). 2012. Available from: http://cran.r-project.org/web/packages/nlme/index.html.
17. Venables WN, Ripley BD. MASS (R Package). 2010. Available from: http://cran.r-project.org/web/packages/MASS/ index.html
18. Goldstein H. Handling attrition and non-response in longitudinal data. Longit Life Course Stud 2009;1:63-72.
19. Honaker J, Joseph A, King G, et al. Amelia II (R Package). 2012. Available from: http://cran.r-project.org/web/packages/Amelia/index.html
20. Honaker J, King G, Blackwell M. AMELIA II: a program for missing data. 2013. Available from: http://cran.r-project.org/web/packages/Amelia/vignettes/amelia.pdf
21. Schafer JL. Analysis of Incomplete multivariate data. Boca Raton: CRC Press; 1997.
22. Rubin DB. Multiple imputation for nonresponse in surveys. Hoboken: Wiley and Sons; 1987.
23. Hesterberg T. Combining multiple imputation *t*, chi-square, and *F* inferences. Seattle: MathSoft; 1998.
24. Cohen J. Statistical power analysis for the behavioral sciences. Routledge: Psychology Press; 1988.
25. Champely S. pwr (R Package for power analysis). 2012. Available from: http://cran.r-project.org/web/packages/pwr/index.html
26. Seagrott V, Mason A, Goldacre M. Hospital Admission rates: effect of social deprivation. Oxford: Unit Of Health-Care Epidemiology; 2004.
27. Walker A, Pearse J, Thurecht L, Harding A. Hospital admissions by socio-economic status: does the inverse care law apply to older Australians? Aust N Z J Public Health 2006;30:467-73.
28. Sutcliffe AG, Barnes J, Belsky J, et al. The health and development of children born to older mothers in the United Kingdom: observational study using longitudinal cohort data. BMJ 2012;345:e5116.
29. National Collaborating Centre for Women's and Children's Health. Fertility: assessment and treatment for people with fertility problems. 2004. Available from: http://www.rcog.org.uk/files/rcog-corp/uploaded-files/NEBFertilityFull.pdf