Fathers of children conceived using ART have higher cognitive ability scores than fathers of naturally conceived children

Bratsberg, Bernt, Ole Røgeberg and Vegard Skirbekk

This version: March 9, 2020

Fathers of children conceived using ART have higher cognitive ability scores than fathers of naturally conceived children

Bratsberg, Bernt; Rogeberg, Ole; Skirbekk, Vegard

Abstract

STUDY QUESTION: Does paternal cognitive ability differ for children conceived with and without Assisted Reproductive Technology (ART)?

SUMMARY ANSWER: Young fathers of ART conceived children tend to score cognitively below their same-age natural conception (NC) counterparts and older (above 35) fathers of ART conceived children tend to score above.

WHAT IS KNOWN ALREADY: Cognitive ability is a genetically and socially transmitted trait, and If ART and NC children have parents with different levels of this trait, then this would in itself predict systematic differences in child cognitive outcomes. Research comparing cognitive outcomes of children with different modes of conception finds conflicting results, and studies may be influenced by selection and confounding.

STUDY DESIGN, SIZE, DURATION: This is a population-based study based on Norwegian data, combining information from the Medical Birth Registry (births through 2012), military conscription tests (birth cohorts 1955–1977) and the population registry. These data allow us to compare the cognitive ability scores of men registered as the father of an ART-conceived child to the cognitive abilities of other fathers and to average scores in the paternal birth cohorts.

PARTICIPANTS/MATERIALS, SETTINGS, METHODS: The population level study included 18 566 births after ART (5810 after ICSI, 12 756 after IVF), and 1 048 138 NC births. It included all Norwegian men who received a cognitive ability score after attending military conscription between 1973 and 1995. This constituted 614 827 men (89.4% of the male birth cohorts involved). An additional 77 650 unscored males were included in sensitivity analyses.

MAIN RESULTS AND THE ROLE OF CHANCE: Paternal cognitive level was assessed using intelligence quotients (IQ) converted from stanine scores on a three-part cognitive ability test with items measuring numeracy, vocabulary and abstract thought (Raven-like matrices). ART fathers averaged 1.95 IQ points above the average of their own birth cohort (P-value < 0.0005) and 1.83 IQ points above NC fathers in their own birth cohort (P<0.0005). Comparisons of the IQ of ART fathers to those of NC fathers of similar age and whose children were born in the same year, however, found average scores to be more similar (point estimate 0.24, P.0.023). These low average differences were found to differ substantially by age of fatherhood, with young ART fathers scoring below their NC counterparts and older ART fathers scoring above their NC counterparts.

LIMITATIONS, REASONS FOR CAUTION: We do not have information on maternal cognition. We also lack information on unsuccessful infertility treatments that did not result in a live birth.

WIDER IMPLICATIONS OF THE FINDINGS: Paternal cognitive ability of ART children differs from that of NC children, and this difference varies systematically with paternal age at child birth. Selection effects into ART may help explain differences between ART and NC children and need to be adequately controlled for when assessing causal effects of ART treatment on child outcomes.

STUDY FUNDING/COMPETING INTEREST(S): This research has also been supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 262700 (Centre for Fertility and Health). It has also been supported by the Research Council of Norway's Project 236992 (Egalitarianism under pressure? New perspectives on inequality and social cohesion). There are no competing interests.

TRIAL REGISTRATION NUMBER: N/A

Key words: cognitive ability / assisted reproductive technologies / selectivity /fathers / ages at parenthood

Introduction

Assisted Reproductive Technology (ART) birth proportions are rising, particularly in richer countries (Wahlberg 2018). In Norway, the first birth following ART treatment occurred in 1984 and involved In-Vitro Fertilization (IVF), with births after Intracytoplasmic Sperm Injection (ICSI) being observed since 1996. The use of both ART technologies have been increasing, and since the mid-2010s, more than 3% of all births in Norway were the result of ART treatments (Präg, Sobotka et al. 2017, Medical Birth Registry 2019).

Extensive research has been carried out to assess whether the mode of conception may itself influence child outcomes (Steel and Sutcliffe 2009, Pottinger and Palmer 2013, Adams, Clark et al. 2017, Catford, McLachlan et al. 2017, Hansen, Greenop et al. 2018). There is concern that ART children experience adverse developmental trajectories, with studies assessing both the neurological and cognitive development of children (Leslie, Gibson et al. 2003, Carson, Kurinczuk et al. 2009) and adolescents (Spangmose, Malchau et al. 2017, Norrman, Petzold et al. 2018). As highlighted in a recent meta-review, a primary challenge is adequately to control for selection bias and/or confounding by family background (Rumbold, Moore et al. 2017).

Under a counterfactual definition of causality (Rubin 1974, Bratsberg and Rogeberg 2018) the causal effect of ART on child outcomes is the expected difference between the life outcomes of a child conceived using ART and the life outcomes of a child naturally conceived *by the same parents at the same time*. As only one of these potential outcomes will be observed, the effects of ART treatment can only be estimated by identifying a control group whose expected outcomes can credibly represent the expected counterfactual outcomes of the ART-conceived children. This is difficult, as neither interest in nor the opportunity to use ART will be statistically independent of parental traits that influence child outcomes through genetic and social factors. A growing literature considers demographic and social dimensions of the users of IVF (Barbuscia, Myrskylä et al. 2019, Gleicher and Barad 2019).

For studies assessing the effects of ART on child cognitive outcomes, parental cognitive ability will typically be an unobserved potential confounder. If the cognitive abilities of parents conceiving through ART differs from that of other parents, this would in itself predict that the cognitive abilities of their children will differ from those born to other parents. Cognitive ability has a strong heritable component (Plomin and Kosslyn 2001, Grönqvist, Öckert et al. 2017),[1] and children whose parents

[1] Conceptions involving donor sperm would imply social – but not genetic – transmission from the registered father to the child. While the use of donor sperm was not reported to the Medical Birth Registry, its use was rare until 2007, increasing from 2012 and still below 10% of ART births in 2017 (source: private communication with ART specialist Dr Liv Bente Romundstad).

score poorly on IQ tests may also have an increased risk of conduct, emotional, and attention problems (Whitley, Gale et al. 2011).

The current study uses administrative register data with cognitive ability scores for males born from 1950 to 1991 to assess how paternal cognitive ability is associated with ART use. Higher cognitive functioning relates to higher income levels (Nisbett, Aronson et al. 2012) – and thereby greater affordability of ART services (Fahlén and Oláh 2015). More speculatively, higher cognitive ability can be associated with a heightened awareness of available medical techniques and higher effectiveness in negotiating for novel treatments in a public health system.

On the other hand, an association between cognitive ability and ART might also result from an increased risk of problems conceiving, as higher cognition is associated with delayed parenthood (Shearer, Mulvihill et al. 2002). A common reason for using ART is that prospective parents have waited until an age where it is more difficult to conceive naturally (Klemetti, Raitanen et al. 2010, Schmidt, Sobotka et al. 2012, Ventimiglia, Capogrosso et al. 2015). Higher education relates to both higher cognition and older ages of family formation (Rodgers, Kohler et al. 2008, Nisén, Martikainen et al. 2017, Tan 2017).

Cognitive ability is also associated with a greater likelihood of holding career-oriented jobs, where many seek to avoid early childbearing as it can worsen career trajectories (Balbo, Billari et al. 2013, Colleran, Jasienska et al. 2015). Hence, a greater proportion of those with higher cognitive ability may postpone childbearing until ages where the likelihood of natural conception is reduced, and this could lead to a greater use of ART in this group (Gustafsson 2001, Skirbekk, Kohler et al. 2004, Monstad, Propper et al. 2008). Evidence from the Nordic countries suggests that this latter mechanism may be relevant for Norway. Nordic welfare states tend to incentivize completion of education and establishing oneself in the labor market before having children, as many childbearing and parental benefits require that one has work experience (Ellingsæter and Pedersen 2015, Lopoo and Raissian 2018). Education levels among men and women in the Nordic countries are high in comparison with most other countries (Barro and Lee 2015), and cohort fertility levels among those with tertiary education attainment tend to be slightly higher – but delayed – relative to those with less education (Thévenon 2011).

The current study investigates whether fathers of ART children differ from fathers of non-ART children in terms of cognitive ability, and whether any such differences persist after controlling for paternal and maternal age.

Methods and materials

Materials

For the 1984 to 2017 period, the Norwegian Medical Birth Registry recorded a total of 40 171 births using ART in Norway. Of these, 22 324 (56%) were registered as using IVF treatment alone, 13 472 (34%) ICSI, 180 (0.4%) both IVF and ICSI, 637 (1.6%) AIH, 230 (0.6%) as “other” methods, and 3328 (8.3%) without a specified technology (Medical birth registry 2019). Figure 1 displays the development of the proportion of births with ART treatment over time, grouping all non-ICSI births as IVF. As the figure shows, the number of ICSI births have increased since the late 1990s.

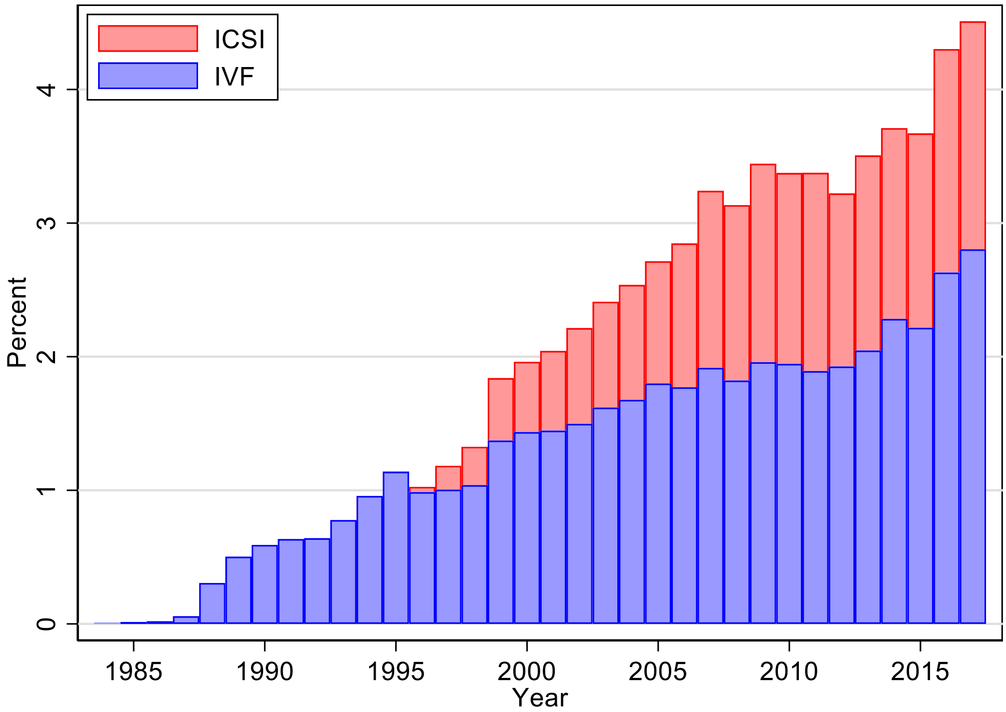

Figure 1 – Proportion of recorded births with ART (source: Medical Birth Registry)

Note: The statistics cover 2 000 827 births in the 1984-2017 period, of which 22 324 used IVF treatment alone, 13 472 ICSI, 1 047 either a combination of the two or a different method, and 3 328 a non-recorded method. For simplicity, we group the non-ICSI births with IVF treatments.

Our analysis data draw on extracts from three administrative data sets: the central population register, the military conscription registry, and the medical birth registry. A pseudonymous identifier allows us to match individual records across these data sources and to identify parents from the birth record. Individual data on ART treatment are available for this research project through 2012. We collect cognitive test scores from the conscription register, where scores are available for the vast majority of Norwegian men born between 1950 and 1991. Scores are based on three speeded tests of arithmetic (30 items), word similarities (54 items), and figures (36 items), and are aggregated into nine stanine scores. Stanine scores aim to have a mean of 5 and a standard deviation of 2. To convert this into IQ scores, which have a mean of 100 and standard deviation of 15, we thus calculate IQ = 100 + 7.5*(stanine-5)). The birth registry states whether each child was conceived with ART, but does not indicate why the treatment was required. Because ART treatment first became available in Norway in the mid-1980s, and because we want to study fertility well into the 30s, we restrict our sample to men born between 1955 and 1977 and who were present in Norway on their 18th birthday. The resulting sample covers 687 502 males, of which 614 827 (89.4%) have a valid conscription test score. Of these, 474 727 entered fatherhood by 2012, fathering 1 066 705 children of which 18 566 followed successful ART treatment (12 756 IVF and 5 810 ICSI).[2] As there are no indications of substantial differences in paternal age or IQ between the two forms of ART treatment, we pool the data across these categories in the later statistical analyses.[3]

[2] Among *fathers*, 91.0 % have a valid cognitive ability score; see supplementary materials, Table S3.

[3] Results from separate analyses for ICSI and IVF births, where the latter category includes those with combined ICSI/IVF and other ART treatments, are available in the supplementary materials.

Statistical method

To assess whether ART is associated with paternal cognitive ability we conduct five statistical comparisons. We first compare the IQ of males with at least one ART-conceived child during the observation period (ever-ART) to all other males in their own cohort. This comparison uses an ordinary least squares (OLS) regression specified as

$$IQ_{i,c} = \alpha + \eta I(ART_i) + \gamma_c + \epsilon_i \qquad (1)$$

In this equation, $I(ART_i)$ is an indicator variable taking the value 1 if individual i is ever observed as the father of an ART child, while γ_c is a cohort fixed effect and ϵ_i a classical noise term. The parameter of interest is η, which measures the difference in mean IQ for men fathering an ART child relative to other men in their own cohort.

Since there may be selection on IQ into fatherhood (irrespective of ART use), we next compare the IQ of ever-ART males to all males with naturally conceived children only (never-ART *fathers*), by running the same model on a data set conditioned on all included men being fathers.

To control for selection on IQ into fertility *timing*, we turn to a dataset with successful births as the observational unit. In this dataset, each father is observed once for each child fathered during the observation period. We first compare the paternal IQ of ART children to the paternal IQ of non-ART children without controlling for paternal age. In this analysis, a male can appear as the father of an ART child in one year and as the father of an NC child in another. We do not expect this to substantially alter the outcome of the comparison, however, and include this analysis primarily to link the estimates using fathers as the observational unit to those using births.[4]

Next, we account for fertility timing by comparing males of the same age with births in the same year. The model is specified as

$$IQ_{i,c,a} = \alpha + \eta I\big(ART_{i,a}\big) + \gamma_{c,a} + \epsilon_i \qquad (2)$$

where $\gamma_{c,a}$ represents a set of fixed effects for each specific combination of father's birth cohort and age. The flexible specification of fixed effects avoids any assumptions on how cohort and age effects interact.

This specification will capture any systematic differences in *average* IQ between ART and naturally conceiving fathers, accounting for fertility timing.

Finally, to assess whether any IQ difference between ART and NC fathers varies by father age, we estimate a model that allows for an interaction between father's age and ART treatment

$$IQ_{i,c,a} = \alpha + \eta_a I\big(ART_{i,a}\big) + \gamma_{c,a} + \epsilon_i \qquad (3)$$

In words, this model estimates whether fathers of ART and NC children have different average cognitive ability scores, and how this difference varies with father age. We apply two alternative specifications of the age-ART association: a simple linear interaction where ART treatment is interacted with father age and a flexible interaction model where the association is allowed to differ across two-year age intervals (two-year bins to avoid small cells). To assess the appropriateness of the linear age trend assumption across age, we compare the estimated linear trend across the 1st to

[4] In the supplementary materials, we present evidence showing that fathers with both NC and ART children do not differ from fathers with ART children only.

99th percentile ages of the ART fathers to estimates from the flexible specification with dummies for age categories.

As a robustness check, we also redid all the analyses with imputed cognitive ability scores for the unscored males. As previous research has shown that the probability of a missing score is strongly linked to low cognitive ability (Bratsberg and Rogeberg 2018), we assigned the lowest score to all unscored males.

Results

Table 1 gives descriptive statistics for fathers of IVF, ICSI, and NC births in the analysis sample. As the table shows, fathers of children conceived through ART treatment have higher IQ scores and are considerably older than fathers of naturally conceived children.

Table 1 – Descriptive statistics, analysis sample

	IVF births (1)	ICSI births (2)	Naturally conceived (3)
Father's IQ score	102.9	103.0	100.9
	(13.2)	(13.0)	(13.2)
Father's age	36.1	37.1	31.0
	(4.5)	(4.7)	(5.6)
Observations (unique pregnancies)	10 477	5 005	1 034 900
Births	12 756	5 810	1 048 138
Unique fathers	9 134	4 201	467 563

Note: Standard deviations are shown in parentheses. Sample consists of births through 2012 of fathers born 1955-1977 and with valid conscription test data. Omitted are 157 births with indeterminate ART treatment.

The probability that one of the cognitively highest scoring men born in the 1955-1977 cohorts will benefit from ART is twice the probability that one of the lowest scoring men will do so (2.8 vs 1.4%), and there is a clear gradient across the range of our cognitive ability measure (Figure 2). Using model specification (1), we find that ever-ART fathers average 1.9 IQ points above never-ART males from their own cohort ($p < 0.0005$), and 1.8 IQ points above never-ART *fathers* from their own cohort ($p < 0.0005$) (Figure 4). The comparison between ever-ART and never-ART fathers is similar to that between ART and non-ART fathers when we use births as observational unit (Figure 4).

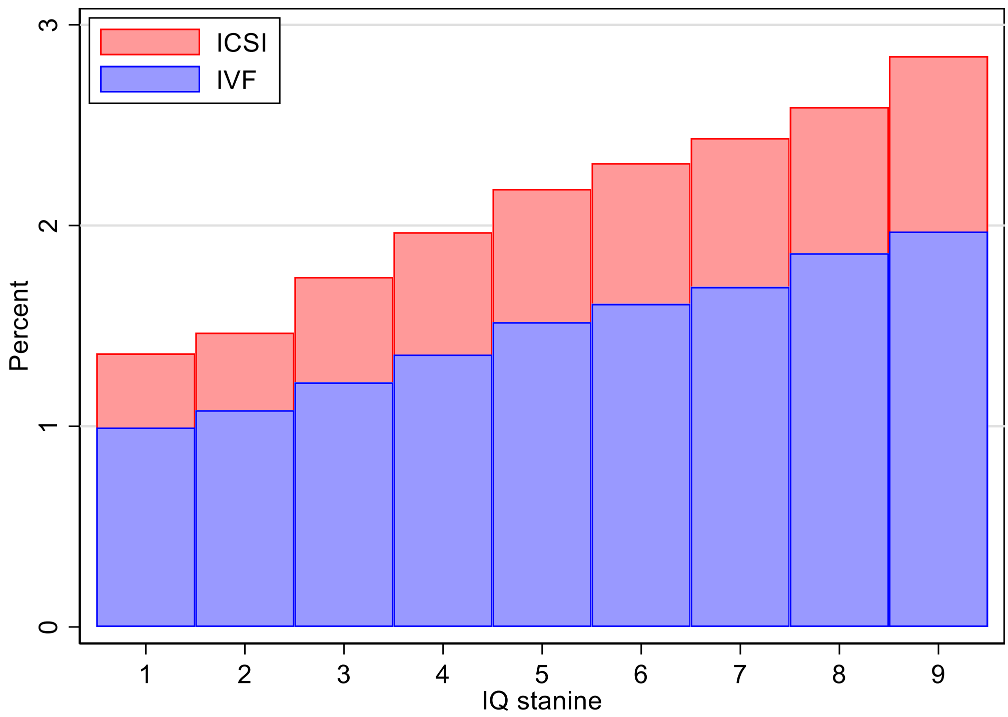

Figure 2 – Proportion of fathers with ART treatment by IQ stanine

Note: Sample consists of Norwegian men born between 1955 and 1977 and with valid conscription test score. Vertical axis shows the fraction with successful IVF or ICSI treatment by end of 2012. Linear regressions of ART treatment on IQ stanine yielded coefficients of 0.0018 (p < 0.0005) for ART, 0.0012 (p < 0.0005) for IVF, and 0.0006 ((p < 0.0005) for ICSI treatment; N = 614 824.

One possible explanation for this pattern relates to birth timing. While the median age of fathers in the sample is 31, the use of ART is concentrated in the tail of older fathers (Figure 3, panel A) who tend to have higher cognitive ability scores irrespective of their ART use (Figure 3, panel B). This is assessed using model specification (2) on the dataset where individual births are the observational unit. Controlling for age and birth cohort of the father, most of the difference in paternal average IQ scores disappears (Figure 4).

Although there is little difference on average between fathers of ART and non-ART children after accounting for age of father, we do find a significant heterogeneity across father's age using model specification (3): The IQ difference between fathers of ART and NC children increases with father's age, with young fathers of ART children scoring below their naturally conceiving peers and older fathers of ART children scoring significantly above (Figure 5). The estimates using two-year age groups to assess the specification largely follow the estimates from the linear interaction model, indicating that the linear trend is an acceptable simplification.

In supplementary analyses, we addressed whether the rising IQ difference between fathers of ART and NC children with father age might reflect associations with calendar time or mother age, as father age of ART children will be positively correlated with these variables in a setup such as ours. These analyses revealed that the coefficient of the interaction term between father age and ART is robust to model inclusion of the additional interactions: the coefficient estimate of 0.130 (p < 0.0005) displayed in Figure 5 declines slightly to 0.115 (p < 0.0005) in the extended model. Separate results for ICSI and IVF cases were highly similar, and are available in the supplementary materials.

Finally, redoing the main analyses with imputed ability scores for unscored males did not lead to any substantial changes in the estimates (see supplementary materials).

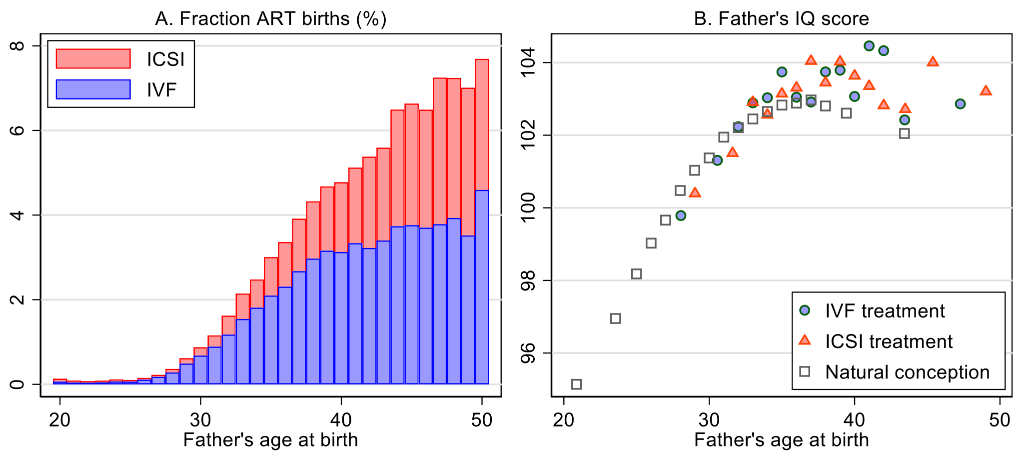

Figure 3 – Fraction ART births and father's IQ score by father's age

Note: Sample consists of Norwegian men born between 1955 and 1977 with valid IQ data and children born through 2012. Panel B shows binned scatter plots of the relation between father's age and IQ score in three samples by mode of conception, with scatter points representing equal shares of each sample.

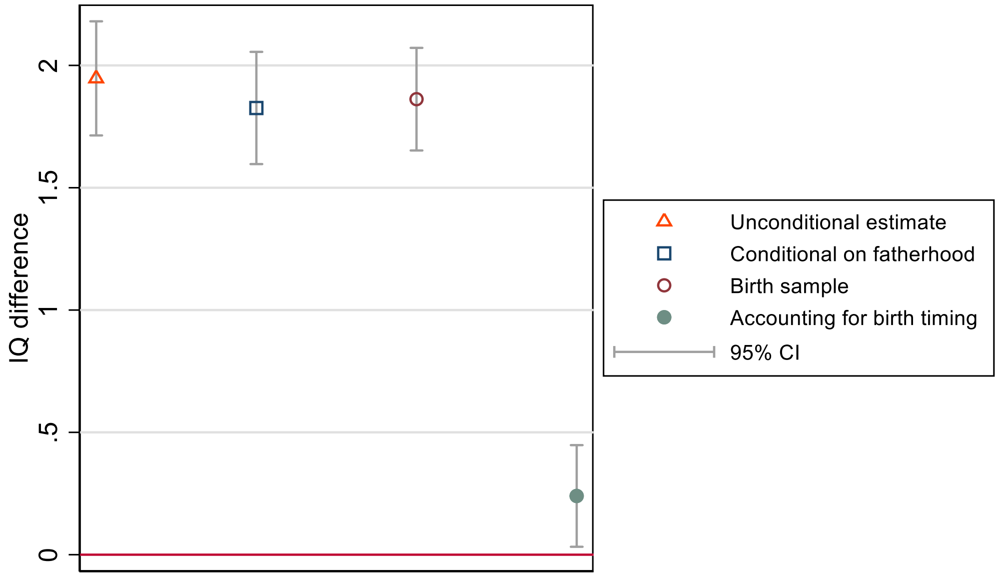

Figure 4 – Estimated IQ difference between men with and without ART treatment, accounting for age of fatherhood

Note: Vertical bars depict 95% percent confidence interval around estimate. All estimates are adjusted for father's birth cohort. Unconditional estimates are based on the full sample of men born between 1955 and 1977 with valid IQ data (N=614 827), while conditional estimates only use the subsample who entered fatherhood by 2012 (N=475 907). Estimates from the birth sample and those correcting for fertility timing are based on a data set with one observation for each unique pregnancy (N=1 050 382); regressions include fixed effects for 23 birth cohorts and 713 combinations of birth cohort and father's age, respectively.

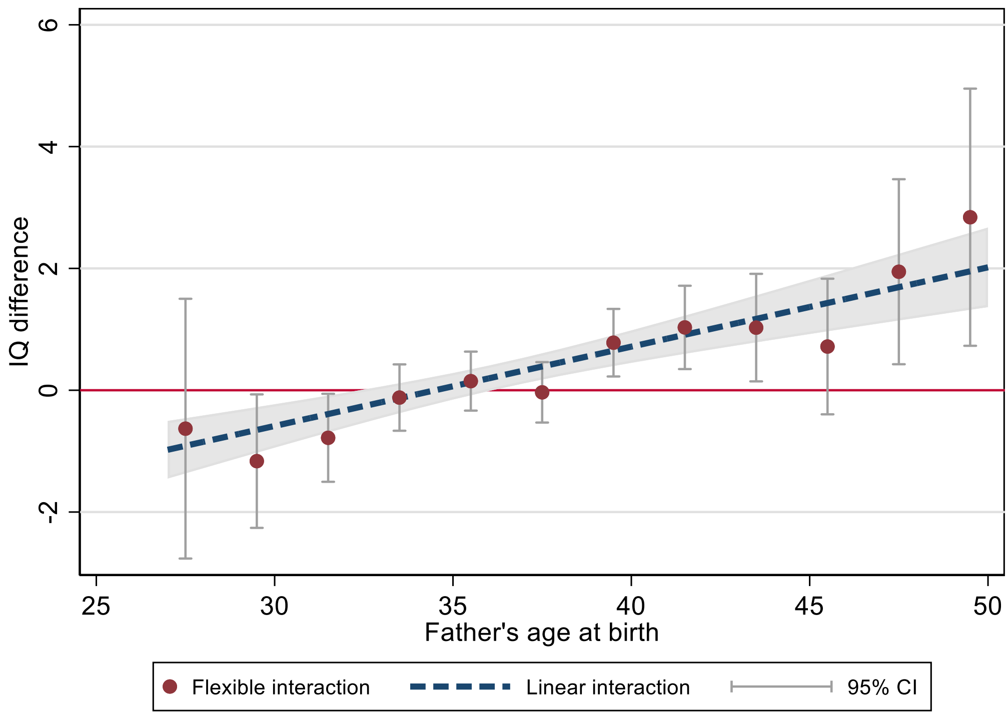

Figure 5 -- Estimated IQ difference between men with and without ART treatment (both IVF and ICSI methods) by age of fatherhood

Note: Estimates are based on a data set with one observation for each unique pregnancy (N=1 050 382); regressions include fixed effects for 713 combinations of birth cohort and father's age. Displayed age range covers the 1st and the 99th percentiles of age of fathers with ART treatment. In the linear interaction model, the coefficient estimate of the interaction term between father age and ART is 0.130 ($p < 0.0005$).

Discussion

A recent meta-review concluded that there are indications from high-quality evidence that some ART-treatments may negatively impact cognitive development in childhood and adolescence (Rumbold, Moore et al. 2017). The review also stressed, however, that results are inconsistent across studies, and that estimates may be affected by selection bias and confounding by family background.

To the extent that the traits studied are genetically and socially transmitted from the parents, our results highlight the methodological challenge involved in such child comparisons. The difficulty lies in identifying an appropriate control group that can credibly represent the counterfactual outcomes of children conceived using ART, and our results show that this would be challenging for cognitive ability.

We find clear evidence that the cognitive ability of ART fathers in Norway differs from that of naturally conceiving fathers, in a way that would be expected to have positive implications for these children and their development. The average difference between ever-ART and never-ART fathers is 1.8 IQ points. In our sample, this average difference largely reflects the raised probability that males with high cognitive ability will postpone fertility to ages where there is a raised risk of fertility problems. Despite this, we find that it would be insufficient to control for father's age (or even parental ages) in an analysis of ART-child outcomes, as there is clear evidence of substantial differences related to paternal age: While ART fathers in their early thirties tend to have slightly lower cognitive ability than the NC fathers of the same age, older ART fathers have higher cognitive ability than same-aged NC fathers. In the upper ranges of the paternal ages observed, ART fathers have about 2 IQ points higher scores than NC fathers of the same age. The results are similar when assessed for ICSI and other fertility treatments separately. This suggests that estimates of how ART treatment affects child cognitive outcomes could have a negative bias if estimated on a sample of Norwegian ART-children with young fathers, and a positive bias if estimated on a sample with older fathers. Although challenging, researchers need to find designs robust to such selection effects. This may involve directly controlling for parental cognitive ability, or indirectly controlling for it by comparing ART and NC children with the same parents. If selection into different types of ART is similar, comparing child cognitive outcomes across ART technologies would be less subject to this bias.

Our study is a descriptive analysis of successful births, which means that the results may reflect mechanisms at different parts of the causal chain. A successful ART birth is the end result of a long chain of events, typically beginning with the intention to have a child. At least one in the couple needs to have a medical issue that makes it difficult to conceive naturally, and the couple needs to realize this, initiate contact with the health care system, be diagnosed and referred to ART treatment, complete treatment, and achieve a successful conception and pregnancy outcome. Our analysis cannot identify where in this chain of events a father's cognitive ability would influence the probability of a successful ART birth.

This descriptive focus means that our results may fail to generalize outside of the Norwegian (or Nordic) context, as several mechanisms potentially influencing these associations may differ across countries. The association between cognitive ability and timing of fertility may differ across educational systems, as men and women with more education tend to have their children relatively late in life (Skirbekk, Kohler et al. 2004, Iyigun and Lafortune 2016). When a couple realizes they have problems conceiving, differences in health care systems are likely to play a role: In Norway, the tax-financed health care system pays for up to three cycles of IVF treatment, provided the mother is aged 38 or below (in the late 2010s reduced from the age of 40) and cohabiting or married (lesbian couples qualifying from 2009). Relative to a more market-based health care system, this means that selection into partnerships matters more while household income or wealth matters less. Higher cognitive ability is associated with a higher probability of being in a stable partnership (Frenette 2011). To the extent that multiple ART cycles are needed, however, selection on income/wealth would arise in Norway too: If further ART cycles are needed beyond the three provided by the public health system, these additional cycles must be privately funded (Helsenorge 2019). The cost in the Nordic countries have been estimated to be around 88 000 NOK in 2014, almost a quarter of average gross female annual income that year (Christiansen, Erb et al. 2014). ART use could also relate to cognition to the extent that higher cognitive ability made it easier to get an overview of relevant reproductive technologies, understand their respective procedures, efficiency and side-effects. Cognitive ability may also affect how difficult it is for households to find accurate information and utilize it effectively when negotiating with doctors and the bureaucracy of the health care system. Finally, the availability and affordability of ART treatments may themselves influence education and economic behavior: Israel's 1994 policy change to make in vitro fertilization free has been argued to have important social effects, leading women to marry later, achieve more education and better labor market outcomes (Gershoni and Low 2017).

The relationship we observe between paternal age and the size of the ART-NC father IQ difference can suggest that there may be more than one mechanism driving the observed associations. In particular, we found it surprising that the direction of the difference varies between young and old fathers. Speculatively, this could reflect a raised risk of fertility problems at young ages for lower ability males, combined with a raised probability of being successfully treated for higher ability males. As paternal age increases, the proportion of higher ability males with fertility problems would increase, and if they had a higher probability of being treated this would gradually erase and then reverse the negative association seen at low paternal ages.

A second hypothesis is that the gradient between paternal age and the ART-NC father IQ difference is caused by a delay in conception due to fertility problems. In our analyses, we compare ART and NC fathers of the same age, but it takes time to identify and treat fertility issues. A 30-year old male discovering that he or his wife has fertility problems may as a result see a successful child birth delayed by, e.g., two years. Since the average IQ of NC fathers aged 32 is higher than for those aged 30, this delay reduces the average IQ of ART fathers relative to NC fathers of the same age. As the slope of the IQ-age gradient shown in Figure 3 declines and flattens around 35, this effect is reduced and a more positive relationship between ART and IQ would be observed.

Importantly, this study only assesses paternal age and IQ, while fertility problems may arise in both males and females. Since our study sample only includes births where a registered father has a cognitive ability score from military conscription, our sample of ART fathers is thus a mix of “males with fertility problems” and “males married to a woman with fertility problems.” This means that there will be systematic differences between the female partners of young and old ART fathers. As fertility problems increase with age in both sexes, the older ART-fathers tend to have partners considerably younger than themselves (see supplementary materials, Figure S4). At the other end of the age range, young fathers of ART children tend to have older partners than similarly aged fathers of NC children.

Our main focus in this paper has been to document how a typically unobserved confounder – paternal cognitive ability – can bias comparisons of child outcomes across modes of conception. Beyond this methodological point, however, the results also speak to how access to and utilization of ART technology is distributed in Norway. As noted above, several plausible mechanisms could generate a cognitive gradient in the use of ART technologies, but we find no strong indications of such a gradient at the overall level in Norway. Controlling for differences in fertility timing associated with paternal cognitive ability, utilization seems similar on average, with lower ability males somewhat over-represented in treatment at young ages and higher ability males at older. While in no way conclusive, this evidence is consistent with an equitable access to assisted reproductive technologies in Norway’s publicly funded healthcare system.

Authors’ roles: BB carried out the analysis and contributed to design, manuscript drafting and critical discussion. OR worked on design, execution, manuscript drafting and critical discussion. VS focused on the idea, design, manuscript drafting and critical discussion.

Acknowledgments: We acknowledge helpful comments from Alice Goisis. Data on loan from Statistics Norway have been essential for the analyses.

References

Adams, D., R. Clark, M. Davies and S. De Lacey (2017). "A meta-analysis of sperm donation offspring health outcomes." Journal of developmental origins of health and disease **8**(1): 44-55.

Balbo, N., F. C. Billari and M. Mills (2013). "Fertility in advanced societies: a review of research." European Journal of Population/Revue Européenne de Démographie **29**(1): 1-38.
Barbuscia, A., M. Myrskylä and A. Goisis (2019). "The psychosocial health of children born after medically assisted reproduction: Evidence from the UK Millennium Cohort Study." SSM-population health **7**: 100355.
Barro, R. J. and J.-W. Lee (2015). Education matters: Global schooling gains from the 19th to the 21st century, Oxford university press.
Bratsberg, B. and O. Rogeberg (2018). "Flynn effect and its reversal are both environmentally caused." Proceedings of the National Academy of Sciences: 201718793.
Carson, C., J. Kurinczuk, A. Sacker, Y. Kelly, R. Klemetti, M. Redshaw and M. Quigley (2009). "Cognitive development following ART: effect of choice of comparison group, confounding and mediating factors." Human reproduction **25**(1): 244-252.
Catford, S., R. McLachlan, M. O'bryan and J. Halliday (2017). "Long-term follow-up of intra-cytoplasmic sperm injection-conceived offspring compared with in vitro fertilization-conceived offspring: a systematic review of health outcomes beyond the neonatal period." Andrology **5**(4): 610-621.
Christiansen, T., K. Erb, A. Rizvanovic, S. Ziebe, A. L. M. Englund, F. Hald, J. Boivin and L. Schmidt (2014). "Costs of medically assisted reproduction treatment at specialized fertility clinics in the Danish public health care system: results from a 5-year follow-up cohort study." Acta obstetricia et gynecologica Scandinavica **93**(1): 64-72.
Colleran, H., G. Jasienska, I. Nenko, A. Galbarczyk and R. Mace (2015). Fertility decline and the changing dynamics of wealth, status and inequality. Proc. R. Soc. B, The Royal Society.
Ellingsæter, A. L. and E. Pedersen (2015). "Institutional trust: Family policy and fertility in Norway." Social Politics: International Studies in Gender, State & Society **23**(1): 119-141.
Fahlén, S. and L. Oláh (2015). The impact of economic uncertainty on childbearing intentions in Europe, FamiliesAndSocieties Working Paper 36.
Frenette, M. (2011). "Why do larger families reduce parental investments in child quality, but not child quality per se?" Review of Economics of the Household **9**(4): 523-537.
Gershoni, N. and C. Low (2017). "The Impact of Extended Reproductive Time Horizons: Evidence from Israel's Expansion of Access to IVF."
Gleicher, N. and D. H. Barad (2019). "Assessing in-vitro fertilisation at age 40 years." The Lancet **393**(10177): 1181-1183.
Grönqvist, E., B. Öckert and J. Vlachos (2017). "The intergenerational transmission of cognitive and noncognitive abilities." Journal of Human Resources **52**(4): 887-918.
Gustafsson, S. (2001). "Optimal Age at Motherhood. Theoretical and Empirical Considerations on Postponement of Maternity in Europe." Journal of Population Economics **14**(2): 225-247.
Hansen, M., K. R. Greenop, J. Bourke, G. Baynam, R. J. Hart and H. Leonard (2018). "Intellectual Disability in Children Conceived Using Assisted Reproductive Technology." Pediatrics **142**(6): e20181269.
Helsenorge (2019). "https://helsenorge.no/refusjon-og-stotteordninger/ufrivillig-barnloshet-og-infertilitetsbehandling."
Iyigun, M. and J. Lafortune (2016). "Why wait? A century of education, marriage timing and gender roles."
Klemetti, R., J. Raitanen, S. Sihvo, S. Saarni and P. Koponen (2010). "Infertility, mental disorders and well-being–a nationwide survey." Acta obstetricia et gynecologica Scandinavica **89**(5): 677-682.
Leslie, G., F. Gibson, C. McMahon, J. Cohen, D. Saunders and C. Tennant (2003). "Children conceived using ICSI do not have an increased risk of delayed mental development at 5 years of age." Human Reproduction **18**(10): 2067-2072.
Lopoo, L. M. and K. M. Raissian (2018). "Fertility Policy in Developed Countries." The Oxford Handbook of Women and the Economy: 173.
Medical birth registry (2019). "https://www.fhi.no/hn/helseregistre-og-registre/mfr/."

Monstad, K., C. Propper and K. G. Salvanes (2008). "Education and Fertility: Evidence from a Natural Experiment." Scandinavian Journal of Economics **110**(4): 827-852.
Nisbett, R. E., J. Aronson, C. Blair, W. Dickens, J. Flynn, D. F. Halpern and E. Turkheimer (2012). "Intelligence: New findings and theoretical developments." American Psychologist **67**(2): 130-159.
Nisén, J., P. Martikainen, M. Myrskylä and K. Silventoinen (2017). "Education, Other Socioeconomic Characteristics Across the Life Course, and Fertility Among Finnish Men." European Journal of Population: 1-30.
Norrman, E., M. Petzold, C. Bergh and U.-B. Wennerholm (2018). "School performance in singletons born after assisted reproductive technology." Human Reproduction **33**(10): 1948-1959.
Norwegian Institute of Public Health. (2016). "Medical Birth Registry of Norway (MBRN)." from www.fhi.no/artikler/?id=94819.
Plomin, R. and S. M. Kosslyn (2001). "Genes, brain and cognition." Nature neuroscience **4**(12): 1153-1154.
Pottinger, A. M. and T. Palmer (2013). "Whither IVF assisted birth or spontaneous conception? Parenting anxiety, styles and child development in Jamaican families." Journal of Reproductive and Infant Psychology **31**(2): 148-159.
Präg, P., T. Sobotka, E. Lappalainen, A. Miettinen, A. Rotkirch, J. Takács, A. Donno, M. L. Tanturri and M. C. Mills (2017). "Childlessness and Assisted Reproduction in Europe."
Rodgers, J. L., H.-P. Kohler, M. McGue, J. R. Behrman, I. Petersen, P. Bingley and K. Christensen (2008). "Education and cognitive ability as direct, mediating, or spurious influences on female age at first birth: behavior genetic models fit to Danish twin data." American Journal of Sociology **114 Suppl**: 202-232.
Rubin, D. B. (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." Journal of educational Psychology **66**(5): 688.
Rumbold, A. R., V. M. Moore, M. J. Whitrow, T. K. Oswald, L. J. Moran, R. C. Fernandez, K. T. Barnhart and M. J. Davies (2017). "The impact of specific fertility treatments on cognitive development in childhood and adolescence: a systematic review." Human Reproduction **32**(7): 1489-1507.
Schmidt, L., T. Sobotka, J. G. Bentzen and A. N. Andersen (2012). "Demographic and medical consequences of the postponement of parenthood." Human Reproduction Update **18**(1): 29-43.
Shearer, D. L., B. A. Mulvihill, L. V. Klerman, J. L. Wallander, M. E. Hovinga and D. T. Redden (2002). "Association of early childbearing and low cognitive ability." Perspectives on sexual and reproductive health: 236-243.
Skirbekk, V., H.-P. Kohler and A. Prskawetz (2004). "Birth Month, School Graduation, and the Timing of Births and Marriages." Demography **41**(3): 547-568.
Spangmose, A. L., S. Malchau, L. Schmidt, D. Vassard, S. Rasmussen, A. Loft, J. Forman and A. Pinborg (2017). "Academic performance in adolescents born after ART—a nationwide registry-based cohort study." Human Reproduction **32**(2): 447-456.
Steel, A. J. and A. Sutcliffe (2009). "Long-term health implications for children conceived by IVF/ICSI." Human fertility **12**(1): 21-27.
Tan, P. L. (2017). "The impact of school entry laws on female education and teenage fertility." Journal of Population Economics **30**(2): 503-536.
Thévenon, O. (2011). "Family Policies in OECD Countries: A Comparative Analysis." Population and Development Review **37**(1): 57-87.
Ventimiglia, E., P. Capogrosso, L. Boeri, A. Serino, M. Colicchia, S. Ippolito, R. Scano, E. Papaleo, R. Damiano and F. Montorsi (2015). "Infertility as a proxy of general male health: results of a cross-sectional survey." Fertility and sterility **104**(1): 48-55.
Wahlberg, A. (2018). Good Quality: The Routinization of Sperm Banking in China, Univ of California Press.
Whitley, E., C. R. Gale, I. J. Deary, M. Kivimaki and G. D. Batty (2011). "Association of maternal and paternal IQ with offspring conduct, emotional, and attention problem scores: transgenerational evidence from the 1958 british birth cohort study." Archives of general psychiatry **68**(10): 1032-1038.

Results for ICSI and IVF treatments

In the main paper, we analyze all ART births jointly. To verify that selection into ART-technology does not differ in ways affecting the main conclusions of the paper, we also show the results for the 5 810 ICSI births and 12 756 IVF births separately, where the latter category also includes ART births involving both IVF and ICSI (142 observations) and ART births with missing information on ART technology (1 625 observations).

As shown in Figure S1, when ART parents are compared to males and fathers from their own cohort (model specification (1)), the average IQ difference is similar for IVF and ICSI fathers. In both cases, this average difference is strongly reduced after controlling for paternal age (model specification (2)). Allowing the IQ difference between ART and non-ART fathers to vary by paternal age, model specification (3) finds positive and similar gradients for IVF and ICSI fathers (Figure S2). Table A1 shows the estimated coefficients of the interaction terms displayed in Figures 5 and S2 (see columns (1) and (3)), as well as estimates from the extended model specification where the IQ difference between fathers of ART and naturally conceived births is allowed to vary with father's age, mother's age, and birth year (see columns (2) and (4)).

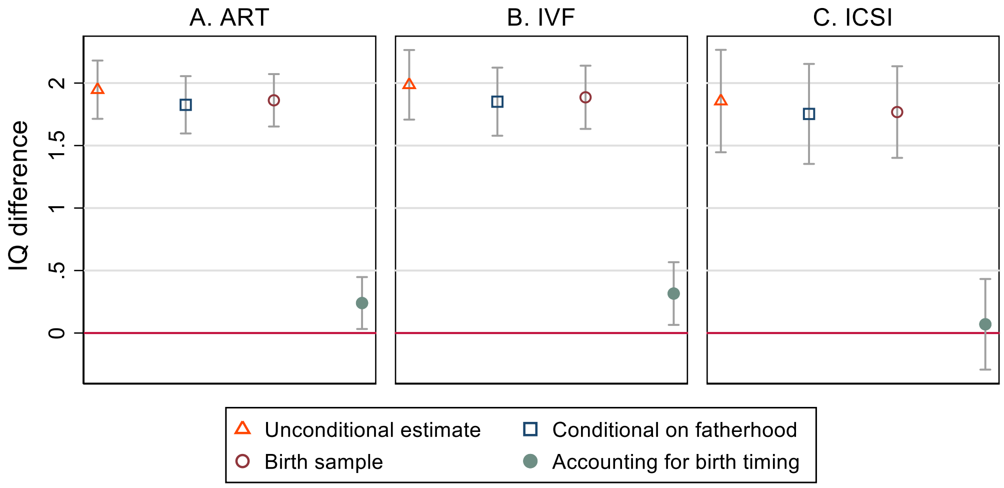

Figure S1 – Estimated IQ difference between men with and without ART treatment, by method and accounting for age of fatherhood

Note: Vertical bars depict 95% percent confidence interval around estimate. All estimates are adjusted for father's birth cohort. Unconditional estimates are based on the full sample of men born between 1955 and 1977 with valid IQ data (N=614 827), while conditional estimates only use the subsample who entered fatherhood by 2012 (N=475 907). Estimates from the birth sample and those correcting for fertility timing are based on a data set with one observation for each unique pregnancy (N=1 050 382); regressions include fixed effects for 23 birth cohorts and 713 combinations of birth cohort and father's age, respectively.

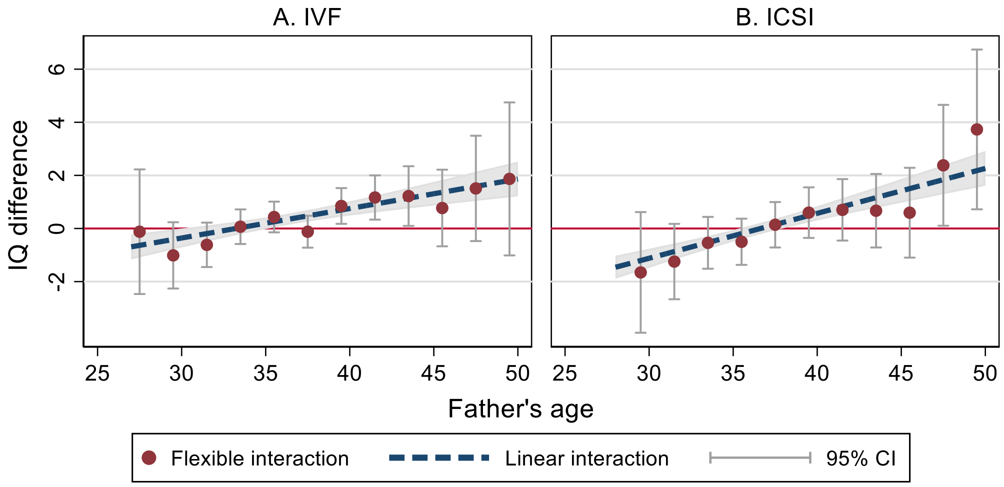

Figure S2 -- Estimated IQ difference between men with and without IVF and ICSI treatment by age of fatherhood

Note: Displayed age range is 1st – 99th sample percentiles of father's age; 1st percentile is 27 for IVF treatment and 28 for ICSI treatment. In the linear interaction models, the coefficient estimate of the interaction term between father age and IVF is 0.113 ($p < 0.0005$) and that between father age and ICSI 0.173 (($p < 0.0005$).

Table S1 – IQ difference between fathers of ART and naturally conceived births, linear interaction models

	(1)	(2)	(3)	(4)
ART*father's age	0.130***	0.115***		
	(0.023)	(0.029)		
ART*mother's age		0.049		
		(0.033)		
ART*birth year		0.032		
		(0.020)		
IVF*father's age			0.113***	0.069*
			(0.028)	(0.036)
IVF*mother's age				0.095**
				(0.041)
IVF*birth year				0.038
				(0.023)
ICSI*father's age			0.173***	0.194***
			(0.039)	(0.047)
ICSI*mother's age				-0.038
				(0.057)
ICSI*birth year				0.024
				(0.048)
Observations (unique pregnancies)	1 050 382	1 050 382	1 050 382	1 050 382

/*/**** *Statistically significant at 10/5/1 percent level.*

Note: Standard errors are shown in parentheses. Sample consists of births through 2012 of fathers born 1955-1977 and with valid conscription test data. Omitted are 157 births with indeterminate ART treatment. Regressions include fixed effects for 713 combinations of birth cohort and father's age.

Men with both ART and NC children

Our main analysis draws on the birth sample, where the observation unit is a unique birth and where there is no ambiguity whether conception is from ART or NC. Certain results build, however, on samples of all men or all fathers from the 1955-1977 cohorts, where we define "ever-ART" males as those with at least one ART-conceived child. Some of these fathers will also have NC children, and the question arises whether fathers with both ART and NC children differ from men with only ART-conceived children. In Table S2, we investigate this question. The table reproduces two coefficient estimates displayed in Figure 4 (columns 2 and 5) as well as estimates from an extended regression specification where we add an indicator variable set to unity for men with both ART and NC children (columns 3 and 6). In the extended model, the coefficient of "ART child" gives the IQ difference between men with only ART-conceived children and men without ART children, while the coefficient of "Both ART and NC children" gives the IQ difference between fathers with both modes of conception and those with ART children only. The table shows that, while 46 percent of ever-ART fathers also conceived naturally (e.g., 0.013/0.028; see column 4), there is no indication of any statistically significant difference between the IQ scores of the two groups.

Table S2 – IQ difference between men with only ART and men with both ART and NC children

	All men			Fathers only		
	Fraction of sample (1)	Model used in Figure 4 (2)	Extended model (3)	Fraction of sample (4)	Model used in Figure 4 (5)	Extended model (6)
Reference: no ART children						
ART child ("ever-ART")	0.021	1.947***	1.861***	0.028	1.826***	1.752***
		(0.119)	(0.161)		(0.117)	(0.158)
Both ART and NC children	0.010		0.186	0.013		0.162
			(0.236)			(0.231)
Observations		614 827			475 907	

**** Statistically significant at 1 percent level.*

Note: Standard errors are shown in parentheses. Sample consists of men born between 1955 and 1977 with valid IQ data. Regressions include fixed effects for 23 birth cohorts.

Imputing IQ for those with missing data

The results in the paper are based on data for men with a valid IQ score in the military conscription registry, but some are exempted from conscription and others are exempted from cognitive ability testing during the conscription process. Our understanding, based on conversations with researchers in the military services, is that the cognitive tests were often given towards the end of a series of medical and physical tests – and that males screened out for other reasons would sometimes be excused from the rest of the day and thus miss the cognitive tests. Irrespective of the reasons for not being scored, however, evidence presented in Bratsberg and Rogeberg (2018) shows that unscored individuals are disproportionally drawn from the lower tail of the underlying ability distribution.

Table S3 shows that, depending on sample, between 9.0 and 10.6 percent of men are omitted from our main analysis because of missing IQ data.

Table S3 – Fraction of men with missing IQ data in three samples used in the main analysis

	All men (1)	Fathers (2)	Birth sample (3)
Fraction with missing score	0.106	0.090	0.092
Observations	687 502	523 261	1 156 948

Note: Samples in columns (1) and (2) consist of men born in Norway 1955-1977 and present on their 18th birthday. Sample in column (3) consists of births (unique pregnancies) through 2012.

In Table S4, we examine whether ART use in the subsamples of men with missing IQ data differs from that in the samples used in the analysis, and in particular whether ART use among unscored men differs from that in the lowest IQ score stanine. The table shows that, in terms of ART treatment, men with missing data differ significantly from the overall sample but are practically indistinguishable from men scoring in the lowest IQ bracket.

Table S4 – Fraction with ART treatment among men with and without IQ data in three samples

	Men with missing IQ score (1)	Lowest IQ stanine (2)	All scored men (3)
A. All men, observations	77 650	14 893	614 827
Fraction ever ART	0.0134	0.0136	0.0214
Difference from col (1)		0.0002	0.0080
p-value 2-sided t-test		0.808	0.000
B. Fathers, observations	47 643	9 360	475 907
Fraction ever ART	0.0218	0.0217	0.0276
Difference from col (1)		-0.0001	0.0058
p-value 2-sided t-test		0.942	0.000
C. Births, observations	106 563	20 193	1 050 385
Fraction ART	0.0114	0.0110	0.0147
Difference from col (1)		-0.0004	0.0033
p-value 2-sided t-test		0.656	0.000

Note: For description of sample, see note to Table S2.

Next, we turn to the concern that our main results may be hampered by selection bias caused by omitting about 10 percent of the samples because of missing IQ scores. The finding that men with missing data are most similar to men in the lowest IQ bracket suggests a robustness check, as a worst-case scenario, where we redo our analyses with imputed IQ scores for those with missing data, setting their score equal to that of the lowest bracket. Figure S3 and Table S5 show the results from this exercise, replicating Figure S1 and Table S1 but drawing on the full sample with imputed IQ for men with missing data. Importantly, the exercise replicates both our key results: accounting for birth timing severely attenuates the average IQ difference between fathers with and without ART treatment, but the IQ difference varies significantly with the age of the father.

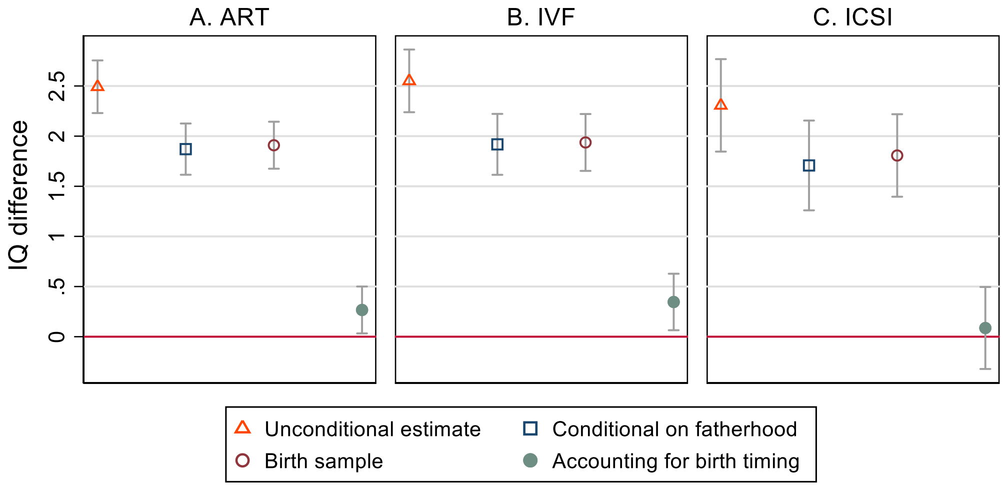

Figure S3 – Estimated IQ difference between men with and without ART treatment, by method and accounting for age of fatherhood and with imputed scores for men with missing IQ data

Note: Vertical bars depict 95% percent confidence interval around estimate. All estimates are adjusted for father's birth cohort. Unconditional estimates are based on the full sample of men born between 1955 and 1977 with valid IQ data (N=687 502), while conditional estimates only use the subsample who entered fatherhood by 2012 (N=523 261). Estimates from the birth sample and those correcting for fertility timing are based on a data set with one observation for each unique pregnancy (N=1 156 948); regressions include fixed effects for 23 birth cohorts and 713 combinations of birth cohort and father's age, respectively.

Table S5 – IQ difference between fathers of ART and naturally conceived births in samples with imputed scores for men with missing IQ data, linear interaction models

	(1)	(2)	(3)	(4)
ART*father's age	0.152***	0.153***		
	(0.026)	(0.032)		
ART*mother's age		0.033		
		(0.037)		
ART*birth year		0.022		
		(0.023)		
IVF*father's age			0.153***	0.133***
			(0.032)	(0.040)
IVF*mother's age				0.069
				(0.045)
IVF*birth year				0.027
				(0.026)
ICSI*father's age			0.163***	0.191***
			(0.043)	(0.052)
ICSI*mother's age				-0.042
				(0.064)
ICSI*birth year				0.008
				(0.054)

/**/ Statistically significant at 10/5/1 percent level.*

Note: Standard errors are shown in parentheses. Sample consists of births through 2012 of fathers born 1955-1977 and with valid conscription test data. Omitted are 157 births with indeterminate ART treatment. Regressions include fixed effects for 713 combinations of birth cohort and father's age. Regressions have 1 156 948 observations.

The relationship between father and mother's age

Results show that the IQ difference between fathers of ART and NC children varies with the age of fatherhood and is negative among young fathers and positive among old fathers. Although the pattern is robust to accounting for mother's age, the question remains whether there are systematic age differences between parents in the two groups, and whether the age difference might vary with the age of fatherhood. Plotting the binned scatter points of the relationship between the age of the father and mother (Figure S4), we see, as expected, that both paternal and maternal ages are substantially higher for ART births than for NC births (although not readily apparent from the figure, the average ages of father and mother are 36.3 and 33.6 for ART births and 31.0 and 28.5 for NC births). We also see the typical pattern that mothers in a couple tend to be younger than the father, with the plotted age relationships largely falling below the 45-degree line where maternal and paternal ages are equal. This within-couple difference increases strongly with father age – with both curves diverging from the 45-degree line. Given father's age, however, mothers of ART children are systematically older than those of NC children, with the plotted curve lying higher in the diagram and even crossing the 45-degree line for ART births with fathers around the age of 30. These differences in maternal age are statistically significant and around 1-2 years throughout the relevant age range: among 29-year old fathers, the average age of mother is 29.6 in the ART sample and 27.3 in the NC sample. This difference is highly significant at conventional levels of significance ($p<0.0005$). Among 45-year old fathers, average maternal age is 36.6 in the ART sample and 35.8 in the NC sample, with the difference significant at the 1 percent level ($p=0.009$).

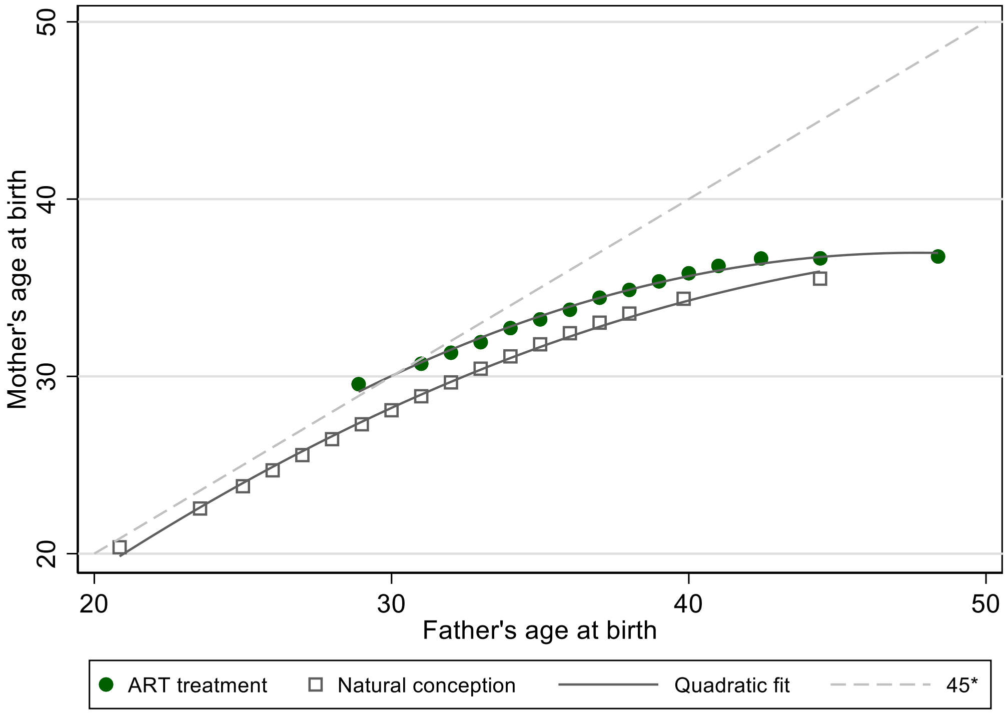

Figure S4 – Binned scatter plots of the relationship between father and mother's age by mode of conception

Note: Samples consist of unique pregnancies in the birth sample; observation counts are 16 692 ART births and 1 140 256 NC births. Scatter points represent equal shares of each sample.