Human Reproduction, Vol.37, No.12, pp. 2831–2844, 2022
Advance Access Publication on September 27, 2022 https://doi.org/10.1093/humrep/deac214

human reproduction

ORIGINAL ARTICLE ***Psychology and counselling***

Mental health and behavioural problems in adolescents conceived after ART

L.A. Wijs [1,*], D.A. Doherty [1,2], J.A. Keelan [1,2,3], P. Burton [4,5], J.L. Yovich [6,7], M. Robinson [8], and R.J. Hart [1,9]

[1]Division of Obstetrics and Gynaecology, University of Western Australia, Perth, WA, Australia [2]Women and Infants Research Foundation, Perth, WA, Australia [3]School of Biomedical Sciences, University of Western Australia, Perth, WA, Australia [4]School of Medical and Health Sciences, Edith Cowan University, Perth, WA, Australia [5]Concept Fertility Centre, Perth, WA, Australia [6]School of Pharmacy and Biomedical Sciences, Curtin University, Perth, WA, Australia [7]PIVET Medical Centre, Perth, WA, Australia [8]Telethon Kids Institute, University of Western Australia, Perth, WA, Australia [9]Fertility Specialists of Western Australia, Perth, WA, Australia

*Correspondence address: Division of Obstetrics and Gynaecology, King Edward Memorial Hospital, Second Floor, A Block, 374 Bagot Road, Subiaco WA 6008, Australia. Tel: +61-8-6458-1888; Fax: +61-8-6458-1045; E-mail: lauraanna.wijs@research.uwa.edu.au
https://orcid.org/0000-0002-1486-4032

Submitted on June 1, 2022; resubmitted on August 24, 2022; editorial decision on August 31, 2022

STUDY QUESTION: Does mental health and behaviour differ between those conceived with and those conceived without ART?

SUMMARY ANSWER: Our study observed less externalizing behaviour (delinquent/aggressive), and more parent-reported internalizing behaviour, as well as more (clinical) depression at age 14 years, in adolescents conceived after ART compared to their non-ART counterparts.

WHAT IS KNOWN ALREADY: Health outcomes of ART-conceived offspring may differ from those conceived without ART, and previous studies have reported differences in behaviour and mental health, particularly in childhood.

STUDY DESIGN, SIZE, DURATION: The Growing Up Healthy Study (GUHS) is a prospective cohort study, investigating the long-term health of offspring conceived after ART (aged 14, 17 and 20 years), in the two operational fertility clinics in Western Australia 1991–2001 (n = 303). Their long-term health outcomes were compared to those of offspring conceived without ART from the Raine Study Generation 2 (Gen2) born 1989–1991 (n = 2868). Both cohorts are representative of the local adolescent population.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Mental health parameters and behaviour were assessed at ages 14 and 17 years, through the parent completed 'Child Behaviour Checklist' (CBCL; ART versus non-ART: age 14 years: N = 150 versus N = 1781, age 17 years: N = 160 versus N = 1351), and the adolescent completed equivalent 'Youth Self-Report' (YSR; age 14 years: by N = 151 versus N = 1557, age 17 years: N = 161 and N = 1232). Both tools generate a T-score (standardized for age and sex) for internalizing (withdrawn, somatic complaints, anxious/depressed), externalizing (delinquent/aggressive behaviour) and total behaviour. Adolescents also completed the 'Beck Depression Inventory for Youth' (BDI-Y; age 14 years: N = 151 versus N = 1563, age 17 years: N = 161 versus N = 1219). Higher scores indicate poorer mental health and behaviour on all the above tools. Parent-reported doctor-diagnosed conditions (anxiety, behavioural problems, attention problems and depression) were also univariately compared between the cohorts. In addition, univariate comparisons were conducted between the GUHS adolescents and Gen2 adolescents born to subfertile parents (time to pregnancy >12 months), as well as between offspring born to subfertile versus fertile parents within the Gen2 cohort. A subgroup analysis excluding offspring born preterm (<37 weeks' gestation) or at low birthweight (<2500 g) was also performed. Generalized estimating equations that account for correlated familial data were adjusted for the following covariates: non-singleton, primiparity, primary caregiver smoking, family financial problems, socio-economic status and both maternal and paternal ages at conception.

MAIN RESULTS AND THE ROLE OF CHANCE: At both 14 and 17 years of age, ART versus non-ART-conceived adolescents reported lower mean T-scores for externalizing problems (age 14 years: 49 versus 51, $P = 0.045$, age 17 years: 49 versus 52, $P < 0.001$). A similar effect was reported by parents, although not significant (age 14 years: $P = 0.293$, age 17 years: $P = 0.148$). Fewer ART-conceived adolescents reported a T-score above the clinical cut-off for externalizing behaviour (≥60; age 14 years: 7.3% versus 16.3%, $P = 0.003$, age 17 years: 8.1% versus 19.7%, $P < 0.001$). At both ages, no differences in internalizing behaviour were reported by adolescents (age 14 years: $P = 0.218$, age 17 years: $P = 0.717$); however, higher mean scores were reported by parents of the ART-conceived adolescents than by parents of the non-ART conceived adolescents (age 14 years: 51 versus 48, $P = 0.027$, age 17 years: 50 versus 46, $P < 0.001$). No differences in internalizing behaviour above the clinical cut-off (T-score ≥ 60) were observed. At age 17 years, parents who conceived through ART reported higher total behaviour scores than those parents who conceived without ART (48 versus 45, $P = 0.002$). At age 14 years, ART versus non-ART-conceived adolescents reported significantly higher mean scores on the BDI-Y (9 versus 6, $P = 0.005$); a higher percentage of adolescents with a score indicating clinical depression (≥17; 12.6% versus 8.5%, aOR 2.37 (1.18–4.77), $P = 0.016$), as well as more moderate/severe depression (≥21; 9.3% versus 4.0%, $P = 0.009$). At age 17 years, no differences were reported on the BDI-Y. There was also a higher percentage of parent-reported doctor-diagnosed anxiety in the ART cohort (age 14 years: 8.6% versus 3.5%, $P = 0.002$, at age 17 years: 12.0% versus 4.5%, $P < 0.001$). Removing adolescents born preterm or at low birthweight did not alter the above results. Comparing outcomes between GUHS adolescents and Gen2 adolescents born to subfertile parents, as well as between those born to subfertile versus fertile parents within Gen2, did not alter results for CBCL and YSR outcomes. Those born to subfertile parents showed higher rates of clinical depression than those born to fertile parents at age 14 years (13.7% versus 6.9%, $P = 0.035$).

LIMITATIONS, REASONS FOR CAUTION: The main limitation of the study is the time difference between the GUHS and Gen2 assessments. Even though we have adjusted for covariates, additional socio-economic and lifestyle factors affecting behaviour and mental well-being could have changed. We were unable to differentiate between different types of ART (e.g. IVF versus ICSI), owing to the low number of ICSI cycles at the time of study. Fertility sub-analyses need to be replicated in larger cohorts to increase power, potentially using siblingship designs. Lastly, selection bias may be present.

WIDER IMPLICATIONS OF THE FINDINGS: The reported lower prevalence of externalizing behaviour (delinquent/aggressive), and higher prevalence of internalizing behaviour, as well as more (clinical) depression at age 14 years, in ART versus non-ART-conceived adolescents, is in line with some previous studies, mostly conducted in childhood. It is reassuring that differences in the rates of depression were not observed at age 17 years, however, these findings require replication. As the use of ART is common, and mental health disorders are increasing, knowledge about a potential association is important for parents and healthcare providers alike.

STUDY FUNDING/COMPETING INTEREST(S): This project was funded by an NHMRC Grant (Hart *et al.*, ID 1042269). R.J.H. is the Medical Director of Fertility Specialists of Western Australia and a shareholder in Western IVF. He has received educational sponsorship from MSD, Merck-Serono and Ferring Pharmaceuticals. P.B. is the Scientific Director of Concept Fertility Centre, Subiaco, Western Australia. J.L.Y. is the Medical Director of PIVET Medical Centre, Perth, Western Australia.

TRIAL REGISTRATION NUMBER: N/A.

Key words: ART / IVF/ICSI outcome / long-term outcomes / mental health / behaviour / depression / Youth Self-Report / Child Behaviour Checklist / Beck Depression Inventory for Youth

Introduction

The use of ART is commonplace in many Western countries, with one in six couples experiencing fertility problems at some point during their reproductive life (Farquhar and Marjoribanks, 2018). Worldwide, over 8 million children have been born after conception with ART and currently 1 in 20 children born in Australia are conceived using such techniques (Fauser, 2019; Newman *et al.*, 2020). With a greater risk of adverse short-term health outcomes in these offspring, such as congenital malformations, imprinting disorders, preterm birth (PTB) and low birthweight (LBW), increasingly studies are investigating their longer-term health beyond early childhood (Schieve *et al.*, 2002; Pandey *et al.*, 2012; Vermeiden and Bernardus, 2013; Declercq *et al.*, 2015; Qin *et al.*, 2015). Because of a greater risk of such adverse perinatal outcomes, particularly congenital malformations and cerebral palsy, it makes sense to investigate the longer term mental and psychological health of these offspring (Ludwig *et al.*, 2006; Hvidtjorn *et al.*, 2009; Hansen *et al.*, 2013; Pinborg *et al.*, 2013). LBW, being small for gestational age and PTB are all associated with neurodevelopmental disorders (Largo *et al.*, 1989; Hutton *et al.*, 1997; Arcangeli *et al.*, 2012), which could explain a potential association between conception after ART and these outcomes. Another potential explanation could be the epigenetic alterations that have been shown to occur in the gametes and embryos during the vulnerable window of conception in ART pregnancies, which can in turn have long-term health effects in the child (Jiang *et al.*, 2017; Fleming *et al.*, 2018; Roseboom, 2018).

Studies to date investigating mental health show conflicting results and have mainly focussed on the earlier years of life. Wilson *et al.* reviewed the psychosocial health outcomes in adolescents conceived by ART in a review of 17 studies, of which 10 studies investigated psychosocial health, and 5 of these 10 studies investigated behaviour. Although results were conflicting, most of the studies concluded that ART offspring do well cognitively and psychologically (Wilson *et al.*, 2011). Studies investigating the prevalence of attention deficit hyperactivity disorder (ADHD) in ART-conceived offspring show contradicting results, however there is some evidence of a higher prevalence of ADHD in such offspring (Wagenaar *et al.*, 2009b; Beydoun *et al.*, 2010). It is unclear whether this is due to the ART treatment itself, or the underlying infertility and intrinsic factors of couples needing ART (Kallen *et al.*, 2011; Bay *et al.*, 2013; Svahn *et al.*, 2015). Most studies investigating mental and emotional well-being of offspring conceived after ART report reassuring findings, but some studies report an increase in adverse outcomes, specifically in depression, which requires further research (Hart and Norman, 2013). A large Swedish study reported an increase in behavioural problems as hospital discharge diagnosis in ART-conceived offspring (Kallen *et al.*, 2005), whereas other studies report no differences in behaviour between ART and non-ART conceived offspring (Barbuscia *et al.*, 2019; Heineman *et al.*, 2019). A recent study reported reassuring findings considering social and mental health in ART-conceived offspring; however, the risk for mental disorders increased when adjusting for the favourable family sociodemographic characteristics of the ART cohort (Remes *et al.*, 2022). The majority of studies to date were carried out in younger children, and there is a paucity of longer term follow-up. Yet many mental health problems do not fully emerge until adolescence (Paus *et al.*, 2008).

Considering the evidence to date, and for the purpose of this study, we studied adolescents and focussed particularly on behaviour that can be assessed using the Youth Self-Report (YSR) as well as the Child Behaviour Checklist (CBCL), such as depressed and anxious behaviour, as well as attention problems. We further assessed (severity of) depression using the Beck Depression Inventory for Youth (BDI-Y). In addition, the prevalence of parent-reported diagnosed mental health conditions was compared between the cohorts. This study was achieved by comparing two population-based Australian studies, where a cohort of ART-conceived adolescents replicated the same assessments undertaken by their non-ART counterparts, from within the same geographical region. Compared to the limited other studies investigating mental health and behaviour in ART-conceived adolescents, this study uses both parent and adolescent data as well as different (validated) measurement tools, and therefore adds to the existing knowledge.

Materials and methods

Mental health outcomes were compared between adolescents conceived after ART from the Growing Up Healthy Study (GUHS) and adolescents conceived without ART from the Raine Study Generation 2 (Gen2).

ART cohort

The GUHS is a prospective cohort study, which aimed to recruit all adolescents and young adults conceived through ART and born 1991–2001 in the two operational fertility clinics in Perth, Western Australia, Australia (PIVET Medical Centre and CONCEPT Fertility Centre). The n = 303 included adolescents who completed various health assessments between 2013 and 2017 at ages 14, 17 and 20 years. Depending on age at enrolment, participants completed assessments at one or two of the follow-ups. Mental health assessments were conducted through questionnaires at ages 14 and 17 years. The full recruitment process can be viewed in Fig. 1. The GUHS participants replicated the assessments previously completed by adolescents conceived without ART from the Raine Study. Because only n = 64 of the GUHS adolescents completed assessments at both the 14- and 17-year follow-up, outcomes have been analysed separately per age group. To account for those 64 who undertook both follow-ups, a sensitivity analysis of those with repeated measures has been performed. Mental health assessments were completed by n = 151 and n = 164 age-eligible participants, and n = 150 and n = 163 primary caregivers, at the 14- and 17-year follow-ups, respectively. After the exclusion of n = 3 participants conceived through gamete intrafallopian transfer (GIFT) from the 17-year follow-up, results from n = 161 participants and n = 160 primary caregivers at age 17 were used in this study. These participants were conceived through IVF (n = 174) and ICSI (n = 60), of which n = 141 were fresh embryo transfers (ETs) and n = 93 were frozen ETs. ART status could not be confirmed for n = 15 participants.

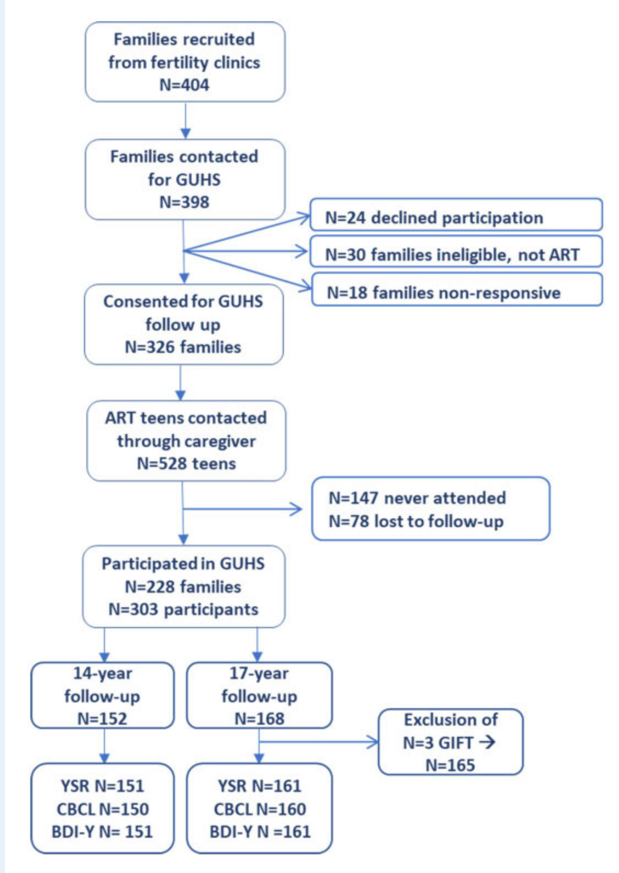

Figure 1. Flowchart of study recruitment process of the Growing Up Healthy Study. GUHS, Growing Up Healthy Study; GIFT, Gamete intrafallopian transfer; YSR, Youth Self-Report; CBCL, Child Behaviour Checklist; BDI-Y, Beck Depression Inventory for Youth.

Non-ART cohort

Adolescents from the Raine Study were used for the cohort of participants who were not conceived after ART (i.e. naturally conceived). The Raine Study was established to investigate the safety and effects of ultrasound on the foetus, and recruited pregnant women in Western Australia between 1989 and 1991 (https://www.rainestudy.org.au; Newnham *et al.*, 1991). A total n = 2868 children born to n = 2804 mothers formed Generation 2 (Gen2) of the Raine Study. The Gen2 participants have completed health assessments at regular intervals from birth to early adulthood, to investigate the effect of perinatal health on subsequent childhood and adult health (Straker *et al.*, 2017). The cohort has a current follow-up rate of 70%, including 1800 participants just over the age of 30. The Raine Study adolescents have been recognized to be representative of the Western Australian adolescent population at the time of assessments (Robinson *et al.*, 2010).

The 14- and 17-year follow-up including mental health assessments were completed by n = 1557 and n = 1236 Gen2 participants, and n = 1781 and n = 1355 primary caregivers, respectively (2003–2005 and 2006–2009). After the exclusion of n = 3 participants conceived after IVF and n = 1 participant conceived after GIFT from the 17-year follow-up, results from n = 1232 Gen2 participants and n = 1351 primary caregivers at age 17 years were used in this study. The BDI-Y was completed by n = 1563 participants at age 14 and n = 1219 participants at age 17 years.

Assessments

Mental health assessments

At the 14- and the 17-year follow-up, primary caregivers of both cohorts completed the CBCL, and adolescents completed the equivalent YSR. The CBCL and YSR are frequently used, validated and standardized questionnaires, recognized to reliably measure behaviour of the adolescents during the preceding 6 months (Achenbach, 1991; Zubrick *et al.*, 1997). They consist of 118 items on behaviour and emotional problems, which are all scored on a three-point scale: not true (=0), somewhat or sometimes true (=1) and very true or very often true (=2). By adding up item scores, both questionnaires produce raw scores for eight problem scales, which are then translated to internalizing, externalizing and total (problem) behaviour scores, and were transformed into T-scores (standardized for age and sex), with higher scores indicating more problems. The eight problem scales are: withdrawn, somatic complaints, anxious/depressed, social problems, thought problems, attention problems, delinquent behaviour and aggressive behaviour. Withdrawn, somatic and depressed/anxious behaviour were scored into internalizing problems, and delinquent and aggressive behaviour into externalizing problems (Achenbach, 1991). Continuous T-scores for internalizing, externalizing and total behaviour, as well as raw problem scale scores were compared between the cohorts. The percentage of participants with a score above the recommended clinical cut-off of T-score ≥60 was also compared between the cohorts (Achenbach, 1991).

Participants also completed the BDI-Y at ages 14 and 17 years. The BDI-Y is a well-established and validated 20-item questionnaire assessing the feelings experienced over the 2 weeks preceding completion of the questionnaire. The total raw score is calculated by adding up all 20 item scores (range 0–60), with higher scores reflecting the presence of more depressive symptoms (Beck *et al.*, 2001). We further compared the cohort percentages with scores in the normal range (≤16) and with scores suggesting mild depressive symptoms (17–20) and moderate/severe depressive symptoms (≥21), following the recommended clinical cut-offs (Beck *et al.*, 2001).

Questionnaire data

Further data collected from primary caregiver completed questionnaires included information regarding doctor-diagnosed conditions (specifically anxiety, attention problems, behavioural problems and depression), parental mental health, family financial problems in preceding year, maternal and paternal age at conception, and primary caregiver smoking. Socio-economic status (SES) at date of assessment is based on postcodes and reported as advantage–disadvantage deciles based on the Socio-Economic Indexed for Areas scores for Western Australia, from the Australian Bureau of Statistics (from lowest score of 1 to the highest score of 10; Australian Bureau of Statistics, 2006, 2011, 2016). For presentation, we transformed deciles into quintiles (from lowest score of 1 to highest score of 5). For the Gen2 participants, information regarding type of conception and pregnancy outcomes (i.e. time to pregnancy (TTP), birthweight (BW), gestational age (GA), parity and plurality) was collected via midwives and clinical records of the birthing hospital.

Primary outcomes of this study were the T-scores (continuous and categorical) of internalizing, externalizing and total behaviour, from the CBCL, YSR and BDI-Y scales, and secondary outcomes were problem scales derived from the CBCL and YSR, as well as parent-reported doctor-diagnosed conditions.

Fertility clinic records

For GUHS participants, information regarding IVF cycle of interest and pregnancy outcomes (GA, BW, birth length and multiplicity) was collected from clinical records from Concept Fertility Centre and PIVET Medical Centre. The Midwives Notification System from the Western Australian Department of Health provided additional pregnancy outcomes for the GUHS participants, as fertility clinics do not always hold information regarding pregnancy outcomes.

Ethical approvals

The GUHS received ethical approval from the Human Research Ethics Office (HREC) of University of Western Australia (RA/4/1/5860) and from the Human Research Ethics Committee of the Department of Health in Western Australia (project number 2013/25). Informed and written consent was obtained from participating families at each follow-up. Each Raine Study follow-up was approved by the HREC of the University of Western Australia.

Statistical analyses

Univariate comparisons between the cohorts were evaluated by Student's *t*-test and Mann–Whitney *U*-test for continuous outcomes, and Chi^2 or Fisher's exact test for categorical outcomes. Adjusted analyses were performed using generalized estimating equations with individual families modelled as random effects and adjusted for: non-singleton, primiparity, sex, primary caregiver smoking, family financial problems in preceding year, SES and maternal and paternal age at conception. Subgroup analyses among adolescents born at term (>37 weeks of gestation) and/or with a BW above 2500 g were performed. To attempt controlling for subfertility, additional subgroup analyses compared outcomes between the GUHS participants and Gen2 participants born to subfertile couples (TTP > 12 months), as well as between Gen2 participants born to subfertile versus fertile parents. Means or odds ratios (ORs) together with their 95% CI were reported to summarize the effects of continuous and categorical outcomes, respectively. A sensitivity analysis was conducted on those participants with repeated measures at both ages (14 and 17 years). Agreement between parents and adolescents was assessed using the Pearson Correlation Coefficient for continuous outcomes, and McNemar's Tests for binary outcomes, in each cohort separately. No adjustments for multiple testing were applied in accordance with the recommendations of the American Statistical Association (Lew, 2016; Wasserstein and Lazar, 2016). All analyses were performed using SPSS version 28.0. (IBM Corp., Armonk, NY, USA) $P < 0.05$ was considered significant.

Results

Demographic characteristics

Demographic differences are shown in Table I. The following differences were detected at both ages in the ART versus non-ART cohort: lower BW and GA, more multiples and primiparity, higher maternal and paternal age, and fewer smokers among primary caregivers. No differences were detected in sex distribution and previous maternal mental health problems. At the 14-year follow-up, the ART participants were older and no significant differences in SES are reported. At the 17-year follow-up, the ART participants were younger, had a more disadvantaged SES, with fewer smokers and fewer reporting binge drinking.

YSR reports

YSR results at ages 14 and 17 years are shown in Table II. At age 14 years, lower externalizing behaviour T-scores were found among the ART-conceived adolescents (49 versus 51, $P = 0.045$) in adjusted analyses. No differences between the cohorts were reported in adjusted analyses in continuous T-scores and T-scores ≥60 for internalizing, externalizing and total behaviour.

At age 17 years, lower externalizing behaviour T-scores were reported in the ART cohort (49 versus 52, $P < 0.001$) in adjusted analyses. Fewer ART-conceived adolescents had a T-score ≥60 for externalizing behaviour in univariate analysis (8.1% versus 19.7%, $P < 0.001$), and after adjustment (aOR 0.44 (0.23–0.82), $P = 0.010$). No differences in internalizing or total behaviour were reported in continuous T-scores or T-score ≥60 in adjusted analyses.

Univariate comparisons of problem scales at age 14 years demonstrated lower raw scores among ART-conceived adolescents for delinquent and aggressive behaviour (2.2 versus 2.9 (out of a total of 26), $P < 0.001$; and 6.5 versus 8.2 (out of a total of 40), $P < 0.001$, respectively). At age 17, lower raw scores amongst GUHS adolescents were found for delinquent and aggressive behaviour (2.7 versus 4.9 (out of a total of 26), $P < 0.001$; and 6.0 versus 9.2 (out of a total of 40), $P < 0.001$, respectively), as well as for somatic complaints (2.9 versus 4.3 (out of a total of 18), $P = 0.023$) and thought problems (2.3 versus 3.4 (out of a total of 14), $P < 0.001$). These univariate comparisons are not shown in tables.

CBCL reports

CBCL results at ages 14 and 17 years are shown in Table III. At age 14, parents reported higher mean T-scores for internalizing behaviour in adjusted analyses (51 versus 48, $P = 0.027$). No differences were reported in adjusted analyses for externalizing and total behaviour continuous T-scores, or T ≥ 60 for internalizing, externalizing or total behaviour.

At age 17 years, a higher mean T-score for internalizing behaviour (50 versus 46, $P < 0.001$), and for total behaviour (48 versus 45, $P = 0.002$) was reported in adjusted analyses. No differences were seen in adjusted analyses in continuous T-scores for externalizing behaviour, or in T ≥ 60 for internalizing, externalizing or total behaviour.

Univariate comparisons of problem scales at age 14 years demonstrated lower raw scores among ART-conceived adolescents reported by parents for somatic complaints (2.3 versus 1.4 (out of a total of 18), $P < 0.001$) and delinquent and aggressive behaviour (0.9 versus 1.5 (out of a total of 26), $P = 0.003$; and 3.9 versus 5.5, $P < 0.001$ (out of a total of 40), respectively), as well as for withdrawn and anxious/depressed behaviour (1.6 versus 1.7 (out of a total of 18), $P = 0.042$; and 2.0 versus 2.5 (out of a total of 28), $P = 0.025$, respectively). At age 17 years, lower raw scores for ART-conceived adolescents were reported by parents for delinquent behaviour (0.9 versus 1.4 (out of a total of 26), $P < 0.001$), and higher raw scores were reported by parents of ART-conceived adolescents for somatic complaints (1.9 versus 1.1 (out of a total of 18), $P < 0.001$) and attention problems (1.8 versus 1.3 (out of a total of 22), $P < 0.001$). These univariate comparisons are not shown in tables.

Beck Depression Inventory for Youth

BDI-Y results are shown in Table IV. Raw BDI-Y scores at age 14 years were higher in the ART cohort in both univariate analysis ($P = 0.011$), and after adjustment (9 versus 6, $P = 0.005$). More ART offspring had BDI-Y scores above the clinical cut-point of 17 after covariate adjustment (12.6% versus 8.5%, $P = 0.093$, aOR 2.37 (1.18–4.77), $P = 0.016$). Further categorization of BDI-Y > 17 showed a higher percentage of ART offspring scored above 21, indicating moderate/severe depression (9.3% versus 4.0%, $P = 0.009$).

These differences were not detected at age 17 years, where raw scores were similar between the cohorts ($P = 0.950$) and prevalence of clinical depression symptoms did not differ ($P = 0.671$).

Other diagnosed mental health conditions

Univariate comparisons of parent-reported diagnosed conditions at both ages are shown in Table V and revealed anxiety to be more often diagnosed in the ART versus the non-ART cohort (age 14: 8.6% versus 3.5%, $P = 0.002$, at age 17 years: 12.0% versus 4.5%, $P < 0.001$). No differences in attention problems, behavioural problems and depression were evident at both ages.

The sensitivity analyses of those individuals with repeated measures (i.e. partaking in assessments both follow-ups; GUHS: n = 64, Gen2: YSR: n = 1112, CBCL: n = 1245 and BDI-Y: n = 1105) did not alter the results (data not shown).

Correlation coefficients between parent (CBCL) and adolescent (YSR) answers in the GUHS ranged from 0.4 to 0.6, and in the Raine Study Gen2 from 0.3 to 0.4. The GUHS appeared to have stronger correlations on the McNemar's Test for YSR and CBCL T-scores ≥ 60, than the Raine Study (data not shown).

ART versus non-ART among term offspring with BW **>** *2500 g*

All outcomes were also analysed with participants born preterm (<37 weeks; GUHS n = 58, Gen2 n = 203) and/or with LBW (<2500 g; GUHS n = 33, Gen2 n = 159) removed from analyses. This did not alter the results (data not shown).

GUHS versus Gen2 born to subfertile parents

When comparing outcomes between the GUHS and Gen2 participants born to subfertile parents (TTP >12 months; age 14 years: n = 82, age 17 years: n = 73) to minimize the effect of the underlying subfertility, the YSR and CBCL outcomes at both ages were the same as when comparing GUHS to the full Gen2 cohort, although some lost significance because of a lack of power, while the magnitude of others

Table 1 Demographic characteristics of the ART (GUHS) and non-ART (Gen2) cohorts, presented separately for age 14 and 17 years.

Demographics age 14 years	ART (152)	Non-ART (1820)	P-value
Sex (male)	70 (46.1%)	932 (51.2%)	0.222
Age (years)	14.8 (14.0–15.5) [13.1–16.1]	14.1 (14.0–14.2) [13.0–15.8]	**<0.001**
Birthweight (g)	3165 (2775–3546) [970–4470]	3375 (3005–3690) [750–5550]	**<0.001**
Birthweight (z-score)	0.1 (−0.6 to 0.6) [−3.1 to 8.4]	0.1 (−0.5 to 0.6) [−2.9 to 3.7]	0.968
Birth length (cm)	49.0 (47.0–51.0) [34.0–56.0]	49.0 (48.0–50.5) [30.5–57.0]	0.267
Gestational age (weeks)	38.4 (36.5–39.4) [26.9–42.0]	39.6 (38.4–40.6) [23.6–42.9]	**<0.001**
Parity			
Primiparous	92 (60.5%)	843 (46.3%)	**<0.001**
Plurality			
Multiples	42 (27.6%)	82 (4.5%)	**<0.001**
Height (m)	1.71 (1.63–1.76) [1.48–1.91]	1.64 (1.60–1.70) [1.19–1.90]	**<0.001**
Weight (kg)	58.4 (51.8–68.2) [33.7–113.7]	55.4 (48.8–63.8) [27.9–135.4]	**<0.001**
BMI (m^2/kg)	20.2 (18.5–22.7) [14.7–37.3]	20.4 (18.5–23.0) [14.0–43.8]	0.769
Maternal age at conception	33.9 (31.1–36.3) [23.3–45.0]	28.2 (24.0–32.1) [14.5–42.9]	**<0.001**
Paternal age at conception	37.0 (34.0–41.0) [24.0–51.0]	30.0 (26.0–35.0) [15.0–62.0]	**<0.001**
Mother ever mental health problems	38 (25.2%)	472 (27.1%)	0.608
Primary caregiver smoking at time of assessment	7 (4.6%)	410 (22.9%)	**<0.001**
SES based on SEIFA			
1st Quintile	6 (4.0%)	112 (7.2%)	
2nd Quintile	22 (14.8%)	169 (10.8%)	
3rd Quintile	23 (15.4%)	216 (13.8%)	0.141
4th Quintile	25 (16.8%)	360 (23.0%)	
5th Quintile	73 (49.0%)	706 (45.2%)	
Family financial problems in last year	39 (27.1%)	698 (39.3%)	**0.004**
Demographics age 17 years	**ART (163)**	**Non-ART (1642)**	**P-value**
Sex (male)	84 (51.5%)	818 (49.8%)	0.676
Age (years)	16.8 (16.4–17.6) [15.6–19.3]	17.0 (16.9–17.2) [14.2–18.9]	**<0.001**
Birthweight (g)	3200 (2805–3545) [970–4620]	3375 (3015–3690) [750–5550]	**<0.001**
Birthweight (z-score)	0.1 (−0.5 to 0.6) [−2.4 to 8.4]	0.1 (−0.5 to 0.6) [−2.6 to 3.7]	0.762
Birth length (cm)	49.0 (47.0–51.0) [34.0–55.0]	49.0 (48.0–50.5) [30.5–57.0]	0.107
Gestational age (weeks)	38.4 (36.9–39.9) [26.9–42.3]	39.6 (38.4–40.6) [23.6–42.9]	**<0.001**
Parity			
Primiparous	112 (68.7%)	780 (47.5%)	**<0.001**

(continued)

Table I Continued

Demographics age 17 years	ART (163)	Non-ART (1642)	*P*-value
Plurality			
Multiples	32 (19.6%)	70 (4.3%)	**<0.001**
Height (m)	1.74 (1.68–1.80) [1.53–1.99]	1.72 (1.65–1.79) [1.36–1.99]	**0.042**
Weight (kg)	64.3 (56.9–70.8) [45.2–104.5]	65.4 (58.2–74.6) [41.1–137.9]	0.133
BMI (m^2/kg)	21.2 (19.6–23.2) [16.0–34.6]	22.1 (20.0–24.3) [14.3–50.2]	**0.004**
Maternal age at conception	34.0 (31.8–36.6) [25.2–45.0]	28.3 (24.1–32.1) [14.5–42.9]	**<0.001**
Paternal age at conception	36.5 (34.0–41.0) [27.0–51.0]	30.0 (27.0–35.0) [15.0–62.0]	**<0.001**
Mother ever mental health problems	44 (28.4%)	405 (29.3%)	0.803
Primary caregiver smoking at time of assessment	8 (5.3%)	283 (20.1%)	**<0.001**
SES based on SEIFA			
1st Quintile	3 (1.9%)	35 (3.2%)	
2nd Quintile	**26 (16.0%)**	**82 (7.5%)**	
3rd Quintile	26 (16.0%)	177 (16.3%)	**0.015**
4th Quintile	23 (14.2%)	194 (17.8%)	
5th Quintile	84 (51.9%)	600 (55.1%)	
Family financial problems in last year	32 (22.5%)	406 (29.0%)	0.106
Current smoking (participant)	7 (4.3%)	192 (15.6%)	**<0.001**
Binge drinking*			
No	**123 (79.4%)**	**475 (39.0%)**	
Yes once	**15 (9.7%)**	**279 (22.9%)**	**<0.001**
Yes > once	**17 (11.0%)**	**465 (38.1%)**	

Data are shown as median (25–75th percentile) [Range] and as N (%). Comparisons between the cohorts were evaluated using Mann–Whitney *U*-tests for continuous outcomes, and Chi^2 tests for categorical outcomes. SES, socio-economic status; SEIFA, Socio-Economic Indexes for Areas: Index of Advantage-Disadvantage; GUHS, Growing Up Healthy Study; Gen2, Raine Study Generation 2.
*Drunk six drinks or more at once/until vomited.
Bold indicates statistically significant values.

increased. This means that differences seen in the full cohort are potentially not explained by the underlying subfertility. This is supported by the comparison between Gen2 participants born to subfertile versus fertile parents (age 14 years: n = 82 versus n = 806, age 17 years: n = 73 versus n = 693), which did not demonstrate differences for CBCL and YSR outcomes between those born to subfertile and to fertile parents.

The higher rate of depression in the GUHS at age 14 years, which is seen in the full cohort analysis, was not present when comparing GUHS to Gen2 participants born to subfertile parents. This means that this factor (depression) is potentially associated with the underlying subfertility, rather than with the ART treatment itself. This is further underlined by the fact that Gen2 participants born to subfertile parents had a higher prevalence of clinical depression at age 14 than those born to fertile parents (13.7% versus 6.9%, $P = 0.035$), although this difference lost significance after adjustment ($P = 0.098$). At age 17 years there also was a higher rate of clinical depression in those born to subfertile versus those born to fertile parents, although not reaching significance, likely to the result of low power (23.7% versus 15.6%, $P = 0.112$).

Discussion

In this unique study of 14- and 17-year-old adolescents, we demonstrated that adolescents conceived following ART exhibit a pattern of less externalizing behaviours, but showed a greater propensity for internalizing behaviours. We found less adolescent-reported externalizing behaviour and more parent-reported internalizing behaviour at both ages for adolescents conceived using ART than for those conceived without ART. At age 14 years, adolescents conceived after ART had higher mean scores for depression, and more adolescents qualified as having clinically mild and moderate/severe depression. This difference was not detected at age 17 years. We also report a higher prevalence of parent-reported doctor-diagnosed anxiety in the ART cohort at both ages, with no parent-

Table II YSR continuous and binary (cut-off ≥60) T-scores at ages 14 and 17 years, compared between adolescents conceived with ART (GUHS) and without ART (Gen2).

YSR	ART N = 151	Non-ART N = 1557	Univariate *P*-value	ART N = 132	Non-ART N = 1446	Adjusted *P*-value*
Age 14 years						
Internalizing	47 (41–55) [26–83]	47 (41–54) [26–88]	0.778	49 (47–51)	47 (46–49)	0.218
Externalizing	46 (40–53) [25–72]	50 (44–56) [25–86]	**<0.001**	49 (48–52)	51 (50–53)	**0.045**
Total	48 (42–55) [28–80]	50 (44–57) [23–87]	0.073	54 (52–56)	53 (52–55)	0.595
YSR ≥ 60	**N = 151**	**N = 1557**		**N = 132**	**N = 1446**	
Internalizing	18 (11.9%)	157 (10.1%)	0.477	1.52 (0.74–3.12)		0.255
Externalizing	**11 (7.3%)**	**254 (16.3%)**	**0.003**	0.64 (0.32–1.26)		0.195
Total	27 (17.9%)	280 (18.0%)	0.975	0.72 (0.42–1.26)		0.252
Age 17 years						
YSR	**N = 161**	**N = 1232**		**N = 150**	**N = 1188**	
Internalizing	48 (41–57) [26–83]	48 (41–56) [26–92]	0.942	54 (50–58)	53 (49–57)	0.717
Externalizing	46 (38–54) [25–74]	51 (45–58) [25–85]	**<0.001**	49 (45–52)	52 (49–56)	**<0.001**
Total	48 (40–55) [28–74]	51 (45–58) [23–88]	**<0.001**	53 (49–57)	54 (50–58)	0.164
YSR ≥ 60	**N = 161**	**N = 1232**		**N = 150**	**N = 1188**	
Internalizing	25 (15.5%)	198 (16.1%)	0.860	0.89 (0.47–1.68)		0.722
Externalizing	**13 (8.1%)**	**243 (19.7%)**	**<0.001**	0.44 (0.23–0.82)		**0.010**
Total	**23 (14.3%)**	**270 (21.9%)**	**0.002**	0.64 (0.35–1.16)		0.141

Continuous outcomes are presented as medians (25th–75th percentile) (Range) for univariate analyses, and as mean (95% CI) after covariate adjustment. Categorical outcomes are presented as N (%) for univariate analyses, and as odds ratio (95% CI) after covariate adjustment. Univariate comparisons between the cohorts were evaluated using Mann–Whitney *U*-tests for continuous outcomes, and Chi^2 tests for categorical outcomes, adjusted comparisons were performed using generalized estimating equations, with individual families modelled as random effects.

*Adjusted for: non-singleton, primiparity, primary caregiver smoking, family financial problems in preceding year, SES and maternal and paternal age at conception. Bold indicates statistically significant values.

YSR, youth self-report; CBCL, Child behaviour checklist; SES, socio-economic status; GUHS, Growing Up Healthy Study; Gen2, Raine Study Generation 2.

reported differences for doctor-diagnosed attention and behavioural problems, or depression.

These findings are in line with some previous studies, mostly conducted in earlier childhood and adolescence (mean ages 8–13.6 years). These studies have also reported less externalizing problems and more internalizing problems in ART-conceived compared to non-ART conceived children (Colpin and Soenen, 2002; Golombok *et al.*, 2002; Wagenaar *et al.*, 2009a). However, two older studies reported no differences in CBCL assessed behaviour between ICSI, IVF and non-ART conceived children (Cederblad *et al.*, 1996; Barnes *et al.*, 2004). Interestingly, one study reports no differences in behaviour between ICSI and non-ART conceived children, however, when comparing ICSI with IVF, particularly the ICSI-conceived girls reported higher problem scores on all scales, as well as more scores in the (borderline) clinical range than the IVF-conceived offspring (Knoester *et al.*, 2007).

As the report of less externalizing problems in ART-conceived offspring is a consistent finding across multiple studies, and the present study also confirms this even after adjustment for covariates, it is possible that this finding relates to differences in parenting styles or higher SES in families who conceived through ART. In this study, the families who conceived through ART had significantly lower levels of financial hardship in the year preceding the assessments, which was included as a covariate in the analyses. It is, however, possible that other social confounders could not be accounted for, and families with an ART-conceived child may have better resources to support their children, which in turn can affect their behaviour. It is well established that

Table III **CBCL continuous and binary (cut-off ≥60) T-scores at ages 14 and 17 years, compared between adolescents conceived with ART (GUHS) and without ART (Gen2).**

Age 14 years						
CBCL	**ART N = 150**	**Non-ART N = 1781**	**Univariate *P*-value**	**ART N = 133**	**Non-ART N = 1685**	**Adjusted *P*-value***
Internalizing	45 (39–54) [31–79]	47 (39–53) [31–85]	0.740	51 (49–52)	48 (46–50)	**0.027**
Externalizing	43 (36–50) [32–76]	47 (39–56) [32–86]	**<0.001**	48 (46–50)	49 (47–51)	0.293
Total	45 (37–52) [23–84]	46 (39–55) [23–90]	0.098	52 (50–54)	51 (48–53)	0.284
CBCL ≥ 60	**N = 150**	**N = 1781**		**N = 133**	**N = 1685**	
Internalizing	19 (12.7%)	225 (12.6%)	0.991	OR 1.23 (0.61–2.47)		0.570
Externalizing	17 (11.3%)	280 (15.7%)	0.153	OR 0.93 (0.48–1.82)		0.832
Total	19 (12.7%)	251 (14.1%)	0.629	OR 0.84 (0.41–1.73)		0.639
Age 17 years						
CBCL	**N = 160**	**N = 1351**		**N = 149**	**N = 1320**	
Internalizing	47 (41–52) [31–74]	44 (37–50) [31–80]	**<0.001**	50 (45–55)	46 (41–51)	**<0.001**
Externalizing	40 (33–49) [32–80]	43 (37–52) [32–88]	**0.018**	48 (45–52)	47 (43–51)	0.148
Total	42 (37–50) [23–82]	42 (34–51) [23–85]	0.394	48 (43–53)	45 (39–50)	**0.002**
CBCL ≥ 60	**N = 160**	**N = 1351**		**N = 149**	**N = 1320**	
Internalizing	16 (10.0%)	126 (9.3%)	0.782	1.23 (0.59–2.57)		0.581
Externalizing	**8 (5.0%)**	**135 (10.0%)**	**0.041**	0.96 (0.42–2.18)		0.916
Total	10 (6.3%)	115 (8.5%)	0.326	0.78 (0.34–1.77)		0.546

Continuous outcomes are presented as medians (25th–75th percentile) (Range) for univariate analyses, and as mean (95% CI) after covariate adjustment. Categorical outcomes are presented as N (%) for univariate analyses, and as odds ratio (95% CI) after covariate adjustment. Univariate comparisons between the cohorts were evaluated using Mann–Whitney *U*-tests for continuous outcomes, and Chi^2 tests for categorical outcomes, adjusted comparisons were performed using generalized estimating equations, with individual families modelled as random effects.

*Adjusted for: non-singleton, primiparity, primary caregiver smoking, family financial problems in preceding year, SES and maternal and paternal age at conception. Bold indicates statistically significant values.

YSR, Youth Self-Report; CBCL, Child Behaviour Checklist; SES, socio-economic status; GUHS, Growing Up Healthy Study; Gen2, Raine Study Generation 2; OR, odds ratio.

women with a higher SES are more likely to utilize ART (Lehti *et al.*, 2013). Furthermore, as suggested by Wagenaar *et al.*, the lower reported prevalence of externalizing behaviour, could be because IVF parents, after the hardship of trying to conceive, minimize problems and criticize their children less and, therefore, may report less behavioural problems (Golombok *et al.*, 2002; Wagenaar *et al.*, 2009a). A strength of the current study, compared to others in the field, is that we also included reports by adolescents themselves in addition to parent reports. It is encouraging that in our study the participants themselves report lower externalizing problem scores, with fewer participants scoring in the clinically relevant range for externalizing problems.

Contrary to these findings, one study reported more aggression, as well as greater depression and anxiety reported by the 9- to 10-year-old children themselves, in ART versus non-ART offspring (Levy-Shiff *et al.*, 1998). Internalizing problems are harder to detect for parents/teachers, as they are less visible than externalizing behaviour. In our study, adolescents do not report differences in internalizing problems, which is reassuring. Parents of offspring conceived through ART, however, did report higher internalizing problem scores, compared to those parents who conceived without ART in our study, indicative of poorer behaviour. However, no difference in percentage of offspring with internalizing problems above the clinical cut-off was detected in parent reports, which again is reassuring. Other studies have also reported a higher prevalence of anxiety and depression in offspring conceived after ART (Levy-Shiff *et al.*, 1998; Wagenaar *et al.*, 2009a; Beydoun *et al.*, 2010). One of the potential explanations could

Table IV BDI-Y outcomes at ages 14 and 17 years, compared between adolescents conceived with ART (GUHS) and without ART (Gen2).

BDI-Y 14 years	ART N = 151	Non-ART N = 1563	*P*-value univariate	ART N = 131	Non-ART N = 1448	*P*-value adjusted*
Continuous	6 (3–10) [0–39]	4 (1–9) [0–50]	**0.011**	9 (7–10)	6 (5–8)	**0.005**
Binary ≥ 17	19 (12.6%)	133 (8.5%)	0.093	2.37 (1.18–4.77)		**0.016**
BDI-Y categories**						
≤ 16 (normal)	132 (87.4%)	1430 (91.5%)				
17–20 (mild)	5 (3.3%)	71 (4.5%)	**0.009**			
≥ 21 (moderate/severe)	**14 (9.3%)**	**62 (4.0%)**				
BDI-Y 17 years	**N = 161**	**N = 1219**		**N = 132**	**N = 1031**	
Continuous	5 (1–12.5) [0–44]	6 (2–13) [0–57]	0.282	14 (11–18)	14 (11–17)	0.950
Binary ≥ 17	25 (15.5%)	203 (16.7%)	0.718	0.88 (0.49–1.59)		0.671
BDI-Y categories**						
≤16 (normal)	136 (84.5%)	1016 (83.3%)				
17–20 (mild)	13 (8.1%)	97 (8.0%)	0.869			
≥21 (moderate/severe)	12 (7.5%)	106 (8.7%)				

Continuous outcomes are presented as medians (25th–75th percentile) (Range) for univariate analyses, and as mean (95% CI) after covariate adjustment. Univariate comparisons between the cohorts were evaluated using Mann–Whitney *U*-tests for continuous outcomes, and Chi^2 tests for categorical outcomes, adjusted comparisons were performed using generalized estimating equations, with individual families modelled as random effects.
*Adjusted for: non-singleton, primiparity, primary caregiver smoking, family financial problems in preceding year, SES and maternal and paternal age at conception. Categorical outcomes are presented as N (%) for univariate analyses, and as odds ratio (95% CI) after covariate adjustment.
**No adjusted analysis performed owing to sample size limitations. Bold indicates statistically significant values.
BDI-Y, Beck Depression Inventory for Youth; SES, socio-economic status; GUHS, Growing Up Healthy Study; Gen2: Raine Study Generation 2.

Table V Other parent-reported diagnosed mental health conditions at ages 14 and 17 years, compared between adolescents conceived with ART (GUHS) and without ART (Gen2).

Age 14 years	ART N = 150	Non-ART N = 1781	*P*-value
Anxiety	**13 (8.6%)**	**63 (3.5%)**	**0.002**
Attention problem	15 (9.9%)	139 (7.8%)	0.349
Behavioural problem	8 (5.3%)	108 (6.1%)	0.706
Depression	2 (1.3%)	35 (2.0%)	0.590
Age 17 years	**N = 160**	**N = 1351**	
Anxiety	**19 (12.0%)**	**63 (4.5%)**	**<0.001**
Attention problem	7 (4.3%)	98 (7.0%)	0.198
Behavioural problem	5 (3.1%)	58 (4.2%)	0.527
Depression	4 (2.5%)	55 (3.9%)	0.359

Data presented as N (%). Comparisons between the cohorts were evaluated using the Chi^2 test for categorical outcomes. Bold indicates statistically significant values.
GUHS, Growing Up Healthy Study; Gen2, Raine Study Generation 2.

again be a difference in parenting style between parents who conceived their child with and without ART. Parents who conceived through ART may be more overprotective, seeing their children as precious, and also have higher expectations from them, which in turn can affect the child's behaviour (Wagenaar *et al.*, 2009a). Parents who have experienced infertility have also been reported to have lower self-esteem and experience more anxiety, and therefore rely on their network more and present to a health professional more frequently, potentially explaining the higher prevalence of doctor-diagnosed anxiety in the ART cohort in the present study (Segev and van den Akker, 2006). As noted previously, one of the difficulties of studies of this nature is that differences in parenting styles and expectations from children are incredibly difficult to capture, quantify and control for (Hart and Norman, 2013). It could be useful if future studies would apply a siblingship design, including families with ART and non-ART conceived siblings, to control for parent(ing) factors.

The higher prevalence of depression (and moderate/severe depression) in the 14-year-old ART-conceived adolescents in this study is concerning. Although this finding was not supported by parent report of a doctor-diagnosed depression, it is reassuring that this difference was not apparent at age 17 years; however, this was mostly a different sample of ART offspring and therefore reassessment of the 14-yearolds at a later stage, as well as assessment of independent cohorts, is of importance. This need is even more underlined by the fact that other studies have also reported an increase in depression among ART-conceived offspring (Levy-Shiff *et al.*, 1998; Wagenaar *et al.*, 2009a; Beydoun *et al.*, 2010). It is, however, also possible that the higher prevalence of depression at age 17 years is not present anymore, because adolescents perhaps become less sensitive to family and parenting influences over time. The knowledge that ART-conceived offspring may potentially be at an increased risk of developing internalizing behaviours and depression is of importance to (future) parents, offspring and healthcare providers. With this knowledge, healthcare providers can be more alert or perhaps more routinely screen for symptoms of these behaviours. With better knowledge about long-term health outcomes after ART, couples considering ART can be aided in their decision-making by being better informed about the health of their potential future children. For ART-conceived offspring, knowledge about their health outcomes is of importance, both for reassurance and to be aware of potential risks and act accordingly, for example, by recognizing symptoms and seeking help and treatment early.

Apart from differences in parenting styles, another plausible underlying mechanism for differences in health outcomes are epigenetic alterations occurring around the susceptible window of conception during the ART process. Epigenetic changes altering foetal programming of endocrine and metabolic processes can lead to altered activity of the hypothalamic–pituitary–adrenal (HPA) axis, which is involved in regulation of stress and arousal and thereby in regulating emotion and behaviour (Wagenaar *et al.*, 2009a). As stated by Wagenaar *et al.* higher HPA axis activity is associated with behavioural inhibition and thereby more internalizing behaviour, such as depression and anxiety, whereas lower HPA axis activity is associated with more externalizing behaviour (Tyrka *et al.*, 2006; Wagenaar *et al.*, 2009a). This is in line with the findings in the current study.

Another possible explanation is the higher prevalence of obstetric complications, such as PTB and LBW, in ART pregnancies. Prematurity is a known risk factor for behavioural problems (Bhutta *et al.*, 2002). LBW, being small for gestational age and PTB have all been associated with neurodevelopmental disorders (Largo *et al.*, 1989; Hutton *et al.*, 1997; Arcangeli *et al.*, 2012); this may, therefore, explain a higher prevalence of such disorders in ART-conceived offspring. Because adjusting for perinatal factors has been suggested to potentially introduce bias, we have not adjusted our analyses for these factors, but instead ran a separate analysis excluding offspring born preterm and with LBW. These analyses did not alter the results, meaning that the effect seen in this study may not be explained by differences in perinatal factors.

Lastly, it is possible that couples requiring ART differ from fertile couples, and the underlying subfertility or suboptimal health of parents leads to an increase in certain disorders in the offspring (Luke, 2017). Some previous studies report an increased risk of mental health disorders in offspring of subfertile women, or women who conceived after ovulation induction, which was not further exaggerated in those conceived after ART (Bay *et al.*, 2013). In the current study, comparing the ART cohort to a sub-cohort of Gen2 who were born to subfertile parents shows that the CBCL/YSR outcomes were mostly the same, or effects were stronger, indicating that the outcomes of ART offspring do not greatly differ from those of offspring born to subfertile parents without ART. Interestingly, the higher prevalence of depression seen in ART offspring at age 14 years was not present when comparing them to the subset born to subfertile parents, meaning that this increase is potentially explained by the underlying subfertility. This was confirmed by the finding that those Gen2 offspring born to subfertile parents reported a higher prevalence of clinical depression compared to those born to fertile parents. To definitively address the role of subfertility, again the siblingship design could be of use to keep parental factors stable and take the underlying subfertility into account. It would be of interest for future studies to further investigate the mental health of ICSI offspring, as these offspring in particular have a higher risk of congenital malformations, potentially linked to the invasiveness of the procedure and subsequent epigenetic alterations (Davies *et al.*, 2012; Berntsen *et al.*, 2019).

One of the major strengths of this study is the comparison of two representative cohorts from the same geographic region. Selection bias is reduced by inviting all adolescents born after ART over a particular period, instead of recruiting adolescents who are presenting with problems. A further strength is the use of standardized and validated questionnaires, which also capture less severe cases. This contrasts with previous studies, such as Svahn *et al.* (2015), which have used hospital discharge diagnoses, only capturing severe cases. Another strength is that this study assesses adolescents, whereas many studies to date were conducted in earlier life. The main limitation of the study is the time difference between the GUHS and Gen2 assessments. Even though we have adjusted for covariates, other external factors affecting behaviour and mental well-being could have changed, and residual confounding cannot be entirely ruled out. However, when comparing the results from the GUHS to those of a Western Australian youth mental health survey conducted around the same time as the GUHS, the prevalence of depression, anxiety and behavioural problems appears to be higher in the GUHS compared to the ART population survey (Goodsell *et al.*, 2017). In addition, data from the Australian Institute of Health and Welfare state that the prevalence of depression among youth increased between 1998 and 2013 from 2.9% to 5.0%, while the prevalence of any mental disorder, ADHD and conduct disorders remained stable (AIHW, 2021). Therefore, the higher prevalence of depression in the ART cohort could be explained by the increased prevalence over time. The prospective nature of the GUHS only allowed for collection of data at time of assessment, which limited the ability to adjust for certain covariates. For example, no information on maternal pregnancy BMI was available, whereas this has an association with child ADHD. We were unable to differentiate between different types of ART (e.g. IVF versus ICSI), owing to the low number of ICSI cycles at the time of study. Although selection bias is reduced by attempting to recruit all ART births over a 10-year period in the state of Western Australia, it may still be substantial as we do not hold information on total number of patients treated in the fertility clinics during the study time. Numbers of invited participants are lower than the expected ART rate over the study period because of difficulty in contacting families, changes in addresses, declining involvement and non-disclosure to the child regarding their ART status; owing to the study design, it is impossible to compare their characteristics to those of the participants, and therefore participation bias may be substantial. Furthermore, at the 14-year and 17-year Raine Study follow-ups, the cohort family characteristics of participants were compared with Year 2006 Western Australian Population Census data of families living in Western Australia with 15- to 17-year-old children, and were found to be representative. Cohort retention was 65.2% and 60.0% at 14 and 17 years, respectively (Straker *et al.*, 2017), which is comparable to other birth cohorts. A 2017 paper investigated selection bias in the Raine Study cohort, and concluded that despite growing non-participation over time, there was no evidence of significant selection bias based on perinatal characteristics (White *et al.*, 2017). However, other characteristics may differ between participants and those who were lost to follow-up, which could potentially introduce bias. In addition, even though Australian data may be generalizable to other Western countries, the high percentage of offspring of Caucasian descent in both cohorts (GUHS: 96.0% and Raine Study Gen2: 88.9%) reduces generalizability to other ethnicities. Furthermore, as expected with single-country studies, generalizability of the findings to other counties is reduced because of differences between countries in access to ART, the costs involved, and protocols used at the time of study. Generalizability may be increased by using internationally recognized and validated measurement tools, which take age and sex into account. The CBCL parent and YSR participant answers commonly have a weak correlation, which was also true for both cohorts in the present study, and the truth is likely to be lying somewhere in the middle (Robinson *et al.*, 2019). Lastly, the interpretation of results also needs to consider that multiple testing in this study could have led to some chance findings.

Conclusion

In this study of 14- and 17-year-old adolescents conceived with and without ART, at both ages we report fewer externalizing problems and more parent-reported internalizing problems; at age 14 years we observed more (clinical) depression in the ART cohort. Reassuringly, this higher prevalence of depression was not detected at 17 years of age. These findings require replication in independent cohorts and at a later stage in life. If internalizing problems, as well as depression, are more prevalent in ART-conceived offspring, this is of importance for parents, offspring, and healthcare providers alike.

Data availability

The data underlying this article cannot be shared publicly for ethical reasons and privacy protection of the individuals that participated in the study. The data will be shared upon reasonable request to the corresponding author.

Acknowledgements

The authors would like to acknowledge the GUHS participants, families and staff for their cooperation in the study and the NHMRC for their grant support for the GUHS (Hart *et al.*, ID 1042269). PIVET Medical Centre and Concept Fertility Centre for providing IVF clinic data. The Department of Health Western Australia, Midwives Notification System for providing data regarding pregnancy. They would like to acknowledge the Raine Study participants and families for their ongoing participation in the study, and the Raine Study team for study co-ordination and data collection. The NHMRC for their long-term contribution to funding the study over the last 30 years. The core management of the Raine Study is funded by the University of Western Australia, Curtin University, Telethon Kids Institute, Women and Infants Research Foundation, Edith Cowan University, Murdoch University, the University of Notre Dame Australia and the Raine Medical Research Foundation. The 14-year follow-up from the Raine Study was funded by NHMRC Grant: Sly *et al.*, ID 211912, NHMRC Program Grant: Stanley *et al.*, ID 003209 and the Raine Medical Research Foundation. The 17-year follow-up from the Raine Study was funded by NHMRC Program Grant: Stanley *et al.*, ID 353514.

Authors' roles

L.A.W. performed data clean up, data analyses and interpretation of results and wrote the draft manuscript. D.A.D. designed and assisted with statistical analyses. J.A.K., P.B. and J.L.Y. provided interpretation of results and expert opinion. M.R. assisted in data clean up, data scoring, interpretation of results and expert opinion. R.J.H. sought funding for and designed the study, as well as assisted with the interpretation of results and critical discussion. All authors reviewed and agreed to the final version of the manuscript.

Funding

This project was funded by a National Health and Medical Research Council (NHMRC) Grant (Hart *et al.*, ID 1042269). Data analysis was carried out as part of LAW's PhD, conducted at the University of Western Australia, which was supported by an Australian Government Research Training Program Scholarship.

Conflict of interest

R.J.H. is the Medical Director of Fertility Specialists of Western Australia and a shareholder in Western IVF. He has received educational sponsorship from MSD, Merck-Serono and Ferring Pharmaceuticals. P.B. is the Scientific Director of Concept Fertility Centre, Subiaco, Western Australia. J.L.Y. is the Medical Director of PIVET Medical Centre, Perth, Western Australia.

References

Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont, Department of Psychiatry, 1991.

AIHW. *Australia's Youth: Mental Illness*. Australian Institute of Health and Welfare, 2021.

Arcangeli T, Thilaganathan B, Hooper R, Khan KS, Bhide A. Neurodevelopmental delay in small babies at term: a systematic review. *Ultrasound Obstet Gynecol* 2012;**40**:267–275.

Australian Bureau of Statistics. Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA) 2006, 2011, 2016.

Barbuscia A, Myrskyla M, Goisis A. The psychosocial health of children born after medically assisted reproduction: evidence from the UK Millennium Cohort Study. *SSM Popul Health* 2019;**7**:100355.

Barnes J, Sutcliffe AG, Kristoffersen I, Loft A, Wennerholm U, Tarlatzis BC, Kantaris X, Nekkebroeck J, Hagberg BS, Madsen SV *et al.*; European Study. The influence of assisted reproduction on family functioning and children's socio-emotional development: results from a European study. *Hum Reprod* 2004;**19**:1480–1487.

Bay B, Mortensen EL, Hvidtjorn D, Kesmodel US. Fertility treatment and risk of childhood and adolescent mental disorders: register based cohort study. *BMJ* 2013;**347**:f3978.

Beck JS, Beck AT, Jolly JB. *Beck Youth Inventories of Emotional and Social Impairment Manual*. San Antonio: Psychological Corporation, 2001.

Berntsen S, Soderstrom-Anttila V, Wennerholm UB, Laivuori H, Loft A, Oldereid NB, Romundstad LB, Bergh C, Pinborg A. The health of children conceived by ART: 'the chicken or the egg?''. *Hum Reprod Update* 2019;**25**:137–158.

Beydoun HA, Sicignano N, Beydoun MA, Matson DO, Bocca S, Stadtmauer L, Oehninger S. A cross-sectional evaluation of the first cohort of young adults conceived by in vitro fertilization in the United States. *Fertil Steril* 2010;**94**:2043–2049.

Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJ. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA* 2002;**288**:728–737.

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

Colpin H, Soenen S. Parenting and psychosocial development of IVF children: a follow-up study. *Hum Reprod* 2002;**17**:1116–1123.

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

Declercq E, Luke B, Belanoff C, Cabral H, Diop H, Gopal D, Hoang L, Kotelchuck M, Stern JE, Hornstein MD. Perinatal outcomes associated with assisted reproductive technology: the Massachusetts Outcomes Study of Assisted Reproductive Technologies (MOSART). *Fertil Steril* 2015;**103**:888–895.

Farquhar C, Marjoribanks J. Assisted reproductive technology: an overview of Cochrane reviews. *Cochrane Database Syst Rev* 2018; **8**:CD010537.

Fauser BC. Towards the global coverage of a unified registry of IVF outcomes. *Reprod Biomed Online* 2019;**38**:133–137.

Fleming TP, Watkins AJ, Velazquez MA, Mathers JC, Prentice AM, Stephenson J, Barker M, Saffery R, Yajnik CS, Eckert JJ *et al.* Origins of lifetime health around the time of conception: causes and consequences. *Lancet* 2018;**391**:1842–1852.

Golombok S, Brewaeys A, Giavazzi MT, Guerra D, MacCallum F, Rust J. The European study of assisted reproduction families: the transition to adolescence. *Hum Reprod* 2002;**17**:830–840.

Goodsell B, Lawrence D, Ainley J, Sawyer M, Zubrick SR, Maratos J. *Child and Adolescent Mental health and educational outcomes. An analysis of educational outcomes from Young Minds Matter: the second Australian Child and Adolescent Survey of Mental Health and Wellbeing*. Perth, WA: Graduate School of Education, The University of Western Australia, 2017.

Hansen M, Kurinczuk JJ, Milne E, de Klerk N, Bower C. Assisted reproductive technology and birth defects: a systematic review and meta-analysis. *Hum Reprod Update* 2013;**19**:330–353.

Hart R, Norman RJ. The longer-term health outcomes for children born as a result of IVF treatment. Part II—mental health and development outcomes. *Hum Reprod Update* 2013;**19**:244–250.

Heineman KR, Kuiper DB, Bastide-van Gemert S, Heineman MJ, Hadders-Algra M. Cognitive and behavioural outcome of children born after IVF at age 9 years. *Hum Reprod* 2019;**34**:2193–2200.

Hutton JL, Pharoah POD, Cooke RWI, Stevenson RC. Differential effects of preterm birth and small gestational age on cognitive and motor development. *Arch Dis Childhood* 1997;**76**:75–81.

Hvidtjorn D, Schieve L, Schendel D, Jacobsson B, Svaerke C, Thorsen P. Cerebral palsy, autism spectrum disorders, and developmental delay in children born after assisted conception. *Arch Pediatr Adolesc Med* 2009;**163**:72–83.

Jiang Z, Wang Y, Lin J, Xu J, Ding G, Huang H. Genetic and epigenetic risks of assisted reproduction. *Best Pract Res Clin Obstet Gynaecol* 2017;**44**:90–104.

Kallen AJ, Finnstrom OO, Lindam AP, Nilsson EM, Nygren KG, Otterblad Olausson PM. Is there an increased risk for drug treated attention deficit/hyperactivity disorder in children born after in vitro fertilization? *Eur J Paediatr Neurol* 2011;**15**:247–253.

Kallen B, Finnstrom O, Nygren KG, Olausson PO. In vitro fertilization in Sweden: child morbidity including cancer risk. *Fertil Steril* 2005;**84**:605–610.

Knoester M, Helmerhorst FM, van der Westerlaken LA, Walther FJ, Veen S; on behalf of the Leiden Artificial Reproductive Techniques Follow-up Project (L-art-FUP). Matched follow-up study of 5 8-year-old ICSI singletons: child behaviour, parenting stress and child (health-related) quality of life. *Hum Reprod* 2007;**22**:3098–3107.

Largo RH, Pfister D, Molinari L, Kundu S, Lipp A, Duc G. Significance of prenatal, perinatal and postnatal factors in the development of AGA preterm infants at five to seven years. *Dev Med Child Neurol* 1989;**31**:440–456.

Lehti V, Brown AS, Gissler M, Rihko M, Suominen A, Sourander A. Autism spectrum disorders in IVF children: a national case-control study in Finland. *Hum Reprod* 2013;**28**:812–818.

Levy-Shiff R, Vakil E, Dimitrovsky L, Abramovitz M, Shahar N, Har-Even D, Gross S, Lerman M, Levy I, Sirota L *et al*. Medical, cognitive, emotional, and behavioral outcomes in school-age children conceived by in-vitro fertilization. *J Clin Child Psychol* 1998;**27**: 320–329.

Lew M. Three inferential questions, two types of P-value. Supplemental material to The ASA statement on p-values: context, process, and purpose. *Am Stat* 2016;**70**:129–133.

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.

Luke B. Pregnancy and birth outcomes in couples with infertility with and without assisted reproductive technology: with an emphasis on US population-based studies. *Am J Obstet Gynecol* 2017;**217**: 270–281.

Newman JE, Paul RC, Chambers GM. *Assisted Reproductive Technology in Australia and New Zealand 2018*. Sydney, New South

Wales, Australia: National Perinatal Epidemiology and Statistics Unit, the University of New South Wales Sydney, 2020.

Newnham JP, O'Dea MRA, Reid KP, Diepeveen DA. Doppler flow velocity waveform analysis in high risk pregnancies: a randomized controlled trial. *Br J Obstet Gynaecol* 1991;**98**:956–963.

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

Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? *Nat Rev Neurosci* 2008;**9**:947–957.

Pinborg A, Wennerholm UB, Romundstad LB, Loft A, Aittomaki K, Soderstrom-Anttila V, Nygren KG, Hazekamp J, Bergh C. Why do singletons conceived after assisted reproduction technology have adverse perinatal outcome? Systematic review and meta-analysis. *Hum Reprod Update* 2013;**19**:87–104.

Qin J, Sheng X, Wang H, Liang D, Tan H, Xia J. Assisted reproductive technology and risk of congenital malformations: a meta-analysis based on cohort studies. *Arch Gynecol Obstet* 2015;**292**:777–798.

Remes H, Palma Carvajal M, Peltonen R, Martikainen P, Goisis A. The well-being of adolescents conceived through medically assisted reproduction: a population-level and within-family analysis. *Eur J Popul* 2022;https://doi.org/10.1007/s10680-022-09623-6.

Robinson M, Doherty DA, Cannon J, Hickey M, Rosenthal SL, Marino JL, Skinner SR. Comparing adolescent and parent reports of externalizing problems: a longitudinal population-based study. *Br J Dev Psychol* 2019;**37**:247–268.

Robinson M, Oddy WH, McLean NJ, Jacoby P, Pennell CE, de Klerk NH, Zubrick SR, Stanley FJ, Newnham JP. Low-moderate prenatal alcohol exposure and risk to child behavioural development: a prospective cohort study. *BJOG* 2010;**117**:1139–1150.

Roseboom TJ. Developmental plasticity and its relevance to assisted human reproduction. *Hum Reprod* 2018;**33**:546–552.

Schieve LA, Meikle SF, Ferre C, Peterson HB, Jeng G, Wilcox LS. Low and very low birth weight in infants conceived with use of assisted reproductive technology. *N Engl J Med* 2002;**346**:731–737.

Segev J, van den Akker O. A review of psychosocial and family functioning following assisted reproductive treatment. *Clin Effect Nurs* 2006;**9**:e162–e170.

Straker L, Mountain J, Jacques A, White S, Smith A, Landau L, Stanley F, Newnham J, Pennell C, Eastwood P. Cohort profile: the Western Australian Pregnancy Cohort (Raine) Study-Generation 2. *Int J Epidemiol* 2017;**46**:1384–1385j.

Svahn MF, Hargreave M, Nielsen TS, Plessen KJ, Jensen SM, Kjaer SK, Jensen A. Mental disorders in childhood and young adulthood among children born to women with fertility problems. *Hum Reprod* 2015;**30**:2129–2137.

Tyrka AR, Mello AF, Mello MF, Gagne GG, Grover KE, Anderson GM, Price LH, Carpenter LL. Temperament and hypothalamic-pituitary-adrenal axis function in healthy adults. *Psychoneuroendocrinology* 2006;**31**:1036–1045.

Vermeiden JP, Bernardus RE. Are imprinting disorders more prevalent after human in vitro fertilization or intracytoplasmic sperm injection? *Fertil Steril* 2013;**99**:642–651.

Wagenaar K, van Weissenbruch MM, Knol DL, Cohen-Kettenis PT, Delemarre-van de Waal HA, Huisman J. Behavior and socioemotional functioning in 9-18-year-old children born after in vitro fertilization. *Fertil Steril* 2009a;**92**:1907–1914.

Wagenaar K, van Weissenbruch MM, Knol DL, Cohen-Kettenis PT, Delemarre-van de Waal HA, Huisman J. Information processing, attention and visual-motor function of adolescents born after in vitro fertilization compared with spontaneous conception. *Hum Reprod* 2009b;**24**:913–921.

Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat* 2016;**70**:129–133.

White SW, Eastwood PR, Straker LM, Adams LA, Newnham JP, Lye SJ, Pennell CE. The Raine study had no evidence of significant perinatal selection bias after two decades of follow up: a longitudinal pregnancy cohort study. *BMC Pregnancy Childbirth* 2017;**17**:207.

Wilson CL, Fisher JR, Hammarberg K, Amor DJ, Halliday JL. Looking downstream: a review of the literature on physical and psychosocial health outcomes in adolescents and young adults who were conceived by ART. *Hum Reprod* 2011;**26**:1209–1219.

Zubrick SR, Sillburn SR, Gurrin L, Teoh H, Shepherd C, Carlton J, Lawrence D. *Western Australian Child Health Survey; Education, Health and Competence*. Perth, Western Australia, Australia: Australian Bureau of Statistics and the TVW Telethon Institute for Child Health Research, 1997.