# **REVIEW**

**Open Access** 

# Intelligence, motoric and psychological outcomes in children from different ART treatments: a systematic review and meta-analysis



Tono Djuwantono<sup>1,2\*</sup>, Jenifer Kiem Aviani<sup>2</sup>, Wiryawan Permadi<sup>1</sup>, Danny Halim<sup>3</sup>, Tri Hanggono Achmad<sup>3,4</sup> and Meita Dhamayanti<sup>5\*</sup>

# Abstract

**Background** Subtle abnormalities in children's intelligence, motor skills, and psychology from various assisted reproductive treatments (ARTs) might be underdiagnosed. Understanding the prognosis of intelligence, motor skills, and psychology in children from ART would provide parents with reasonable expectations and enable them to plan relevant support to achieve the optimum potential in ART children.

**Methods** We searched PubMed, EMBASE, Ovid, Google Scholar, and Scopus databases until April 13, 2021, to identify relevant studies. Thirty-four studies met the inclusion and exclusion criteria. The meta-analysis employed a standardized mean difference model. The outcome of this study is to compare intelligence quotient (IQ), motoric ability, and behavioral problems between all ARTs, in vitro fertilization (IVF), intracytoplasmic sperm injection (ICSI) to naturally conceived (NC) children. Subdomains of intelligence based on the Cattell, Horn, and Carroll Model (CHC Model) of cognitive architecture, including fluid reasoning, short-term and working memory, processing speed, visual-spatial ability, long-term memory retrieval, and crystalized intelligence (knowledge), were evaluated and summarized in details. Motor skill was stratified into two domains: gross motoric and fine motoric. Behavioral problem was categorized as externalizing and internalizing behavior.

**Results** Meta-analysis showed that verbal intelligence score in IVF toddlers is significantly lower than NC toddlers (p = 0.02); conversely, ICSI toddlers scored significantly higher verbal intelligence score compared to NC toddlers (p = 0.005). Toddlers born after ART had significantly lower non-verbal intelligence score (p = 0.047). IVF toddlers scored significantly lower fine motor score (p = 0.01) compared to naturally conceived toddlers. Based on parent's CBCL, NC toddlers had higher total (p = 0.01) and externalizing behavior (p = 0.001) scores compared to ART toddlers. Evaluation of full scale IQ and all domains of intelligence in preschool and primary school children revealed that no significant differences exist between ART and NC children. Based on preschool and primary school parents' CBCL, IVF children had significantly lower externalizing behavior score compared to NC children (p = 0.04). Meta-analyses of studies

\*Correspondence: Tono Djuwantono tono.djuwantono@unpad.ac.id Meita Dhamayanti meita.dhamayanti@unpad.ac.id Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/A.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

on young adolescents revealed that ART young adolescents scored higher academically than their NC counterparts, including on mathematics (p < 0.00001) and reading or language (p < 0.00001).

**Conclusions** Despite differences in certain aspects, this finding suggests that ART is unlikely to cause negative impacts on children's neurodevelopment.

**Keywords** Assisted reproductive treatment, In vitro fertilization, Intracytoplasmic sperm injection, Children neurodevelopment, Intelligence quotient, Motoric skills, Behavioral problems, Toddlers, Preschool and primary school children, Young adolescents

# Introduction

Over the past few decades, assisted reproductive technology (ART) has been integrated into the standard protocols to treat infertility. In 2014, there were 1,929,905 ART cycles from 2,746 centers in 76 countries. From 2010 to 2014, the number of reported non-donor aspirations and frozen embryo transfer cycles increased by 37.3% and 67.5%, respectively. The proportion of fresh non-donor single embryo transfers increased from 30.0% in 2010 to 40.0% in 2014 [1].

Since its inception, numerous ART methods have been developed to address a variety of etiologies. Ovulation induction refers to ovarian follicle stimulation by fertility drugs to reverse anovulation or oligoovulation. Gamete intrafallopian transfer (GIFT) involves removing eggs from a woman's ovaries and placing them in one of the Fallopian tubes along with the man's sperm. It is used when the fertility problem is caused by sperm dysfunction or idiopathic (unknown cause) infertility. In vitro fertilization is a technique that allows male and female gametes (sperm and egg) to fertilize outside of the female body. This technique is indicated mainly for tubal factor infertility or if the previous methods have failed. Intracytoplasmic sperm injection (ICSI) is a solution to acquire pregnancy(-ies) if most sperms are immotile. The technique involves sperm injection directly into the cytoplasm of a mature oocyte, thus bypassing many natural barriers that prevent natural conception. Despite the superiority of this technique, concerns about preventing defective sperm from fertilizing mature oocytes are frequently raised [2].

Despite the wide use of ART, there are still concerns regarding its safety. How various assisted conception techniques to affect children's neurodevelopmental outcomes is still unclear. Increased risks of multiple births, preterm birth, and low birth body weight have been described in ART compared to spontaneous pregnancies [3, 4]. Those risks are also associated with neuromotor development disturbances [5]. Our previous meta-analysis showed that children born after ART attain a higher risk for neurodevelopmental disorders, especially cerebral palsy (risk ratio [RR] 1.82, [1.41, 2.34]; P=0.00001) [6]. However, a question regarding subtle clinical manifestations, i.e., intelligence, motor, and mental developments, remains unanswered and less studied. A limited number of studies with various timing of follow-ups, different ART methods, and methodological shortcomings are the major limitations for neurodevelopmental risk interpretation.

There were inconsistent results regarding the neurophysiological and behavioral outcomes of children born after ART. Many of these studies only focused on mental and psychomotor development in the first 3 years of life. Children at preschool to early adolescent ages, when cognitive demand increases, motoric skills are well developed, while socioemotional and behavioral changes are marked, have been insufficiently studied [7].

This study aimed to conclude studies on neurodevelopmental outcomes (intelligence, motoric, and behavior) in children born after different ART treatments compared to naturally conceived (NC) children at every developmental stage: toddlers (1–3 years), preschool to school age (4–8 years) and young adolescents (8–18 years).

# Methods

# Literature search and identification

This meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) [8] reporting guidelines. Pub-Med, EMBASE, Ovid, Google Scholar, and Scopus databases were used to collect publications up to April 13, 2021. The following search terms were applied: (reproductive techniques OR assisted reproductive OR in vitro fertilization) AND (psychomotor performance OR intelligence test OR intelligence quotient OR child behavior OR behavioral test OR temperament).

# Inclusion and exclusion criteria

Studies were included if they (1) reported singleton-born children; (2) reported neurodevelopmental outcome scores on intelligence, language development, motoric skill, socioemotional, or behavior; (3) reported children born from ART techniques; and (4) reported naturally conceived children as control. Studies were excluded if they (1) did not include original data, such as reviews, systematic reviews, comments, or editorial letters; (2) did not include a control group (e.g., case reports); (3) could not ascertain the use of fertility treatment; (4) was not written in English; (5) reported children aged < 12 months; (6) used unstandardized instruments for assessment; (7) reported children born after donor insemination, oocyte donation, or sperm donation; (8) included children with serious health problems or neurodevelopmental disorders.

# Data collection and analysis

Three authors (TD, JKA, DH) reviewed the title and abstract of every article independently. The full-text article was thoroughly read if the abstract met the inclusion criteria. Screening through the reference lists was performed to identify publications that were previously unidentified but relevant to this study. The following information was retrieved: author, country, publication year, number of participants, method of conception, domain, and methods of neurodevelopmental assessment. Newcastle–Ottawa Scale (NOS) was applied to assess the risk of bias in the studies [9].

# Data synthesis

A rigorous review was done by stratifying the result based on age groups, as these groups represent different developmental milestones. In this review, children were grouped into toddler (1–3 years), preschool and primary school age (4–11 years), and young adolescent (12–18 years).

At the age of 1- to 3-year-old, toddlers are advancing their sensorimotor to preoperational intelligence, where they are progressing from learning objects and environment by touch to the development of language and communication. Gross motor skill quickly develops when the transition from crawling to walking and standing occurs. Fine motor skills in this age group are limited to refinements in reaching, grasping, and manipulating small objects. During this period, children are also learning to socialize mainly through playing activity, where they learn cooperation, empathy, and develop friendships with others [10].

Primary school is the first stage of basic education. It bridges early childhood education to formal school education. The programs are typically designed to provide students with fundamental skills in literacy (reading and writing) and mathematics, and to establish a solid foundation for learning. According to ISCED classification, primary education typically starts between the ages of 5to 8-year-old (1st to 3rd grade). However, in many countries, primary school starts from 4- to 12-year-old (1st to 6th grade). Gross motor is already well developed, and complex fine motoric tasks such as writing and typing can already be performed [11].

At young adolescent ages, children are usually already attending secondary school. Secondary education prepares students for tertiary or higher education and/or provides skills relevant to employment. In this stage, the competencies achieved in primary school are developed in more detail [12]. School grades can be used as a measurement tool of academic intelligence.

By referring to the previously mentioned developmental milestones at different stages of life, three domains of development were assessed: intelligence, motor development, and behavior (social skills). In addition to full scale IQ, two domains of intelligence were assessed: verbal and non-verbal intelligence (Performance IQ). When possible, the subdomains of intelligence based on the Cattell, Horn, and Carroll Model (CHC Model) of Cognitive Architecture, including quantitative intelligence, fluid reasoning, short-term memory and processing speed, visual-spatial ability, long-term memory retrieval, and executive function were evaluated and summarized in details. Verbal intelligence is the ability to understand and reason using concepts framed in words. Verbal IQ is related to crystalized or comprehension knowledgeability in the CHC model. Fluid intelligence is the ability to solve novel reasoning problems and is correlated with essential skills, such as comprehension, problem-solving, and learning. Short-term memory is the capacity for holding a small amount of information in an active, readily available state for a short interval. Processing speed is the ability to perform simple repetitive cognitive tasks quickly and fluently. Visuospatial intelligence is the ability to perceive, analyze, and understand visual information. Long-term memory retrieval is a process of accessing stored memory gained from the learning process [12].

Motoric skills were analyzed in 2 domains: gross motoric and fine motoric. Gross motor (physical) skill is the ability to move the whole body, which involves core stabilizing muscles to perform everyday functions, such as standing, walking, dressing, etc. Fine motor skill is the ability to move minor muscles such as the wrist, hand, fingers, feet, and toes to perform small movements such as picking up objects, gripping, tool manipulation, etc. [13].

Behavioral problems were categorized as externalizing and internalizing behavior. The externalizing spectrum incorporates a variety of disinhibited or externally-focused behavioral symptoms, including aggression, conduct problems, delinquent behavior, oppositionality, hyperactivity, and attention problems. In contrast, the internalizing spectrum includes a variety of over-inhibited or internally-focused symptoms, including anxiety, fear, sadness/depression, social withdrawal, and somatic complaints [14].

Another evaluated aspect was executive function. Executive function is defined as a set of cognitive processes that is necessary for selecting and successfully monitoring behaviors that facilitate the attainment of chosen goals. There are three basic executive function components: inhibition, working memory, and cognitive flexibility. Inhibition is the self-control of attention, behavior, thoughts, and/or emotions to override a strong internal predisposition or external lure and do what is more important. The second aspect is working memory. Working memory is related to the act of holding information (perceptual input) in mind and manipulating or connecting it to bring conceptual knowledge. Working memory is also related to selective, focused attention as the brain will focus on the information held in the mind, turning out irrelevant thoughts. Cognitive flexibility is the third element of executive function. One aspect of cognitive flexibility is being able to change perspective spatially or interpersonally, which is related to inhibition or previous perspective. Higherorder executive functions require the simultaneous use of multiple basic executive functions, including planning and fluid intelligence (e.g., reasoning and problemsolving) [15].

# Statistical analysis

Random effect standardized mean difference (SMD) with a 95% confidence interval was used in the meta-analysis for continuous data. This type of data analysis was used to summarize studies that reported the same outcomes measured in a variety of psychometric scales. Nonetheless, we were aware that this method might be unable to identify real scale differences. RevMan version 5.3 software (Cochrane Collaboration) was used for these purposes. The inconsistency index ( $I^2$ ) test, which ranges from 0 to 100%, was performed to evaluate heterogeneity across studies. *P* value < 0.05 or values above 50% indicate a significant heterogeneity. The risk of bias was evaluated by the Cochrane Risk of Bias Assessment tool (Cochrane Collaboration).

# Results

The literature searches identified 2503 studies, with the addition of 32 studies identified through reference screening (Fig. 1). Following a review of 96 full-text articles, 57 were excluded for failing to meet the inclusion criteria. Five studies were excluded because they focused on infants under 1 year of age (1 study) and reported duplication of cohort and data in four other studies. Only 34 studies [16–49] were ultimately included in the meta-analyses. The quality of the included studies that were assessed by the Newcastle–Ottawa Scale is shown in Supplemental Table S1 for cohort studies reporting intelligence outcomes, Supplemental Table S2 for case–control studies reporting intelligence outcomes, Supplemental Table S3 for cohort studies reporting motoric outcomes, Supplemental Table S4 for case–control reporting motoric outcome, Supplemental Table S5 for cohort studies reporting motoric outcome, Supplemental Table S4 for case–control reporting motoric outcome, Supplemental Table S5 for cohort studies reporting behavioral outcomes.

# **Characteristic of participants**

Table 1 shows pooled analysis of the background characteristics of the children. There were no differences in children's gender proportion, mother education level, and family socioeconomic background in all age groups born from all types of ART conceptions compared to naturally conceived control.

# Toddler (1- to 3-year-old)

# Intelligence outcome

Four studies used Bayley's Mental Development Index to measure cognitive development in the toddler age group [16, 19, 20, 26]. There were no significant differences in the mental development of assisted reproductive technology (ART)-born compared to naturally conceived (NC) toddlers (p = 0.16). There was no evidence of publication bias (p-Egger=0.506), and the data exhibited good homogeneity ( $I^2$ =0%, p=0.94) (Fig. 2A). Supplemental Table S6 summarizes the statistics for the meta-analysis.

The McArthur Bates Language Inventory [17], British Naming Ability [16], Receptive Expressive Emergent Language-II (REEL-2) [19], McArthur Communicative Developmental Inventories (N-CDI) [21], Brunet-Lezine language sub-scores [22, 25], and Griffith hearing and speech sub-scores [23, 24] were used to measure language development or verbal intelligence. There were no significant differences in language development between ART-born and NC toddlers (p=0.76). Although there was significant heterogeneity ( $I^2 = 71\%$ , p = 0.0003), the pooled analysis did not indicate publication bias (p-Egger=0.118) (Fig. 2B). Since the method of conception might affect heterogeneity, separate subgroup analyses were performed. Good homogeneities were identified in the analyses on IVF vs NC and ICSI vs NC (p > 0.05); high heterogeneity was only detected in the analysis on ART vs NC group which included studies that did not specify the mode of conception ( $I^2 = 92\%$ , p = 0.00005). The language development score of toddlers born after IVF was significantly lower than NC toddlers (p = 0.02); meanwhile, ICSI toddlers' score was significantly higher compared to NC toddlers (p = 0.005).



Fig. 1 Flow diagram of included study in meta-analysis

Non-verbal intelligence was reported in 3 studies that used Bayley-III cognitive [15] and Griffith performance sub-scores [23, 24]. Pooled analyses showed that non-verbal intelligence in ART toddlers is significantly lower compared to the NC toddlers (p=0.047) (Fig. 2C). Good homogeneity ( $I^2$ =10%, p=0.34) and lack of publication bias (p-Egger=0.703) were both displayed in these studies.

# Motoric outcome

Bayley-II Psychomotoric Development Index (PDI) [16, 19, 20], Bayley-III motor composite score [15], Brunet-Lezine posture and coordination [22, 25], and Griffith locomotor and eye-hand coordination [22, 24] were utilized to assess the total motor skill outcome. Pooled analysis showed no significant difference in total motor score between tod-dlers born via ART and naturally conceived toddlers (p=0.27) (Fig. 3A). There were no evidence of data heterogeneity ( $I^2 = 6\%$ , p = 0.38) and publication

bias (p-Egger = 0.575). Similarly, subgroup analyses also revealed the insignificant differences of total motor score in toddlers born from ART, IVF, or ICSI compared to NC toddlers, with good homogeneity and no publication bias (p > 0.05).

The gross motor score was obtained from the Griffith locomotor [22, 25] and the Brunet-Lezine posture subtests [23, 24], in both pooled analysis (p=0.79) and subgroup analyses based on the method of conception (IVF, p=0.93; ICSI, p=0.83) (Fig. 3B). Significant heterogeneity between studies was identified ( $I^2=61\%$ , p=0.03), especially in the ICSI subgroup ( $I^2=76\%$ , p=0.006), suggesting that factors other than conception mode might also influenced how children developed their motor skills.

The fine motor score in ART and NC toddlers was similar (p=0.055) based on Brunet-Lezine's coordination [, ] and Griffith's ey-hand coordination [, ] assessments. The analyses showed low heterogeneity ( $I^2=35\%$ , p=0.17)

Characteristics	Age group	Reporting	No. of c	hildren			Ratio, <i>p</i> value	Heterogeneity (l <sup>2</sup> )
		studies	IVF	F ICSI All A		Control		(I <sup>2</sup> )
Male (%)	Toddler	16, 21, 25		225/497		303/621	0.97 [0.88, 1.07], 0.54	0%, 0.78
		17, 20			129/278	1032/2174	0.93 [0.75, 1.15], 0.52	0%, 0.72
		Summary			381/831	1335/2795	0.96 [0.88, 1.05]. 0.41	0%, 0.94
	Preschool	31, 34, 35, 42		122/250		129/250	0.89 [0.63, 1.27], 0.53	0%, 0.91
		34,35	81/164			94/170	0.79 [0.51, 1.21], 0.28	0%, 1.00
		27, 28, 30, 32, 33			156/326	4080/7983	0.89 [0.69, 1.15], 0.38	0%, 0.84
		Summary			359/740	4303/8403	0.87 [0.73, 1.05], 0.15	0%, 0.99
	Young adoles- cent	44–46, 48, 49			6410/12,495	772,649/1,509,646	1.00 [0.98, 1.02], 0.90	0%, 0.78
Mother's higher education (uni-	Toddler	16, 21, 25		248/515		296/679	1.04 [0.92, 1.17], 0.58	0%, 0.82
versity or above)	ersity or above)				74/154	3868/10,661	1.11 [0.48, 2.55], 0.81	93%, 0.0002
		Summary			322/669	4164/11,340	1.07 [0.82, 1.41], 0.60	83%, < 0.00001
	Preschool	31, 34–37, 42		422/779		409/763	0.93 [0.75, 1.16], 0.52	78%, 0.0001
		34, 35	65/164			73/170	0.92 [0.66, 1.29], 0.64	41%, 0.19
		28, 30, 32, 33			139/255	2917/6406	0.93 [0.80, 1.08], 0.33	32%, 0.22
		Summary			626/1198	3399/7339	0.94 [0.83, 1.07], 0.37	66%, 0.00005
	Young adoles- cent	45, 46, 48, 49			2907/11,293	359,623/1,505,567	1.00 [0.81, 1.23], 0.99	87%, < 0.00001
Family financial co	ondition							
Low	Toddler	16, 25		76/480		149/626	0.83 [0.32, 2.13], 0.70	91%, 0.001
		17, 18			53/377	2414/12,662	0.54 [0.13, 2.26], 0.40	90%, 0.001
		Summary			129/857	2563/13,288	0.72 [0.41, 1.25], 0.24	85%, 0.0002
	Preschool	34, 35, 38		17/418		15/429	1.17 [0.60, 2.28], 0.64	0%, 0.99
		34, 35	20/173			14/170	1.40 [0.73, 2.69]	0%, 0.99
		Summary			37/591	29/599	1.29 [0.81, 2.05], 0.29	0%, 1.00
	Young adoles- cent	NR	NR	NR	NR	NR	NR	NR

# Table 1 Pooled analysis of characteristics of the children from included studies

Characteristics	Age group	Reporting	No. of c	hildren			Ratio, <i>p</i> value	Heterogeneity
		studies	IVF	ICSI	All ART	Control		( <i>I</i> <sup>2</sup> )
Middle	Toddler	16, 17, 25			344/758	947/2714	1.09 [0.90, 1.31], 0.39	67%, 0.05
	Preschool	34, 35, 38		151/427		166/429	0.97 [0.72, 1.29], 0.89	45%, 0.16
		34, 35	44/164			45/170	1.02 [0.49, 2.12], 0.97	75%, 0.04
		Summary			195/591	211/599	0.99 [0.76, 1.29], 0.93	50%, 0.09
	Young adoles- cent	NR	NR	NR	NR	NR	NR	NR
High	Toddler	16, 17, 25			274/758	1167/2714	1.10 [0.78, 1.54], 0.60	83%, 0.003
	Preschool	34, 35, 38		270/418		276/429	0.95 [0.78, 1.17], 0.64	80%, 0.007
		34, 35	126/173			139/170	0.90 [0.75, 1.08] 0.25	57%, 0.13
		Summary			396/591	415/599	0.93 [0.84, 1.04], 0.22	63%, 0.03
	Young adoles- cent	NR	NR	NR	NR	NR	NR	NR

# Table 1 (continued)

NR Not reported

and no publication bias (p-Egger = 0.322). The subgroup analyses revealed that toddlers born after IVF had a noticeably lower fine motor score (p = 0.01) than NC toddlers. No significant disparity was noticed in ICSI toddlers compared to NC toddlers (p = 0.28) (Fig. 3C).

# Behavior and social outcomes

According to three studies, NC mothers reported behavioral issues more frequently than ART mothers, as assessed using Achenbach's Child Behavioral Checklist (CBCL) [20, 21, 26]. Compared to ART children, NC children showed higher total (p=0.01) and externalizing behavior scores (p=0.001) (Fig. 4A, C). No significant difference was noted in internalizing behavior score between the two groups (p=0.09) (Fig. 4B). The data showed good homogeneities ( $I^2$ =0%, p>0.05) and no publication biases.

There was no statistically significant difference observed in the social skills of ART and NC toddlers as assessed using Griffith's social [23, 24], Brunet-sociability Lezine's [21, 25], and Vineland Adaptive Behavior socialization [16, 19] (p=0.22) (Fig. 4D). Likewise, there were no significant differences observed in the subgroup analyses between the IVF (p=0.57) and ICSI (p=0.28) toddlers compared to NC toddlers. No heterogeneity (I<sup>2</sup> = 0%, p >0.05), and publication bias (p-Egger>0.05) were found in the analyses.

# Preschool and primary school ages (4- to 11-year-old) Intelligence outcome

Weschler Preschool and Primary School Intelligence-Revised version (WPPSI-R) [22, 27, 31, 33, 39, 40, 42], Weschler Abbreviated Scale of Intelligence (WASI) [30], Weschler Intelligence Scale for Children (WISC) [36, 37], Kauffman Assessment Battery for Children (K-ABC) [38, 41], and Revised Amsterdam Child Intelligence Test (RAKIT) [35] were used to measure intelligence. There was no significant difference in the overall full-scale IQ of ART schoolers compared to NC schoolers (p=0.31). There was significant heterogeneity observed among the studies ( $l^2 = 50\%$ , p = 0.01), but no evidence of publication bias was detected (p-Egger = 0.438). ICSI subgroup analysis also demonstrated significant heterogeneity ( $I^2 = 59\%$ , p = 0.01). Across the three subgroups, the results consistently indicated that there was no significant differencebetween ART and NC schoolers.

The verbal intelligence quotient was calculated from WPPSI-R [22, 27, 31, 33, 39, 40, 42], WASI [32], and WISC [36, 37] verbal IQ, K-ABC Knowledge subtest [41], RAKIT verbal meaning, learning names, and idea production subtests [35], British Ability Scale (BAS) vocabulary subtest [28], and Ages and Stages Questionnaire communication subtest [30] scores. In the subgroup and overall analyses, no significant differences were found (Fig. 5B). Nonetheless, both the overall ( $l^2$ =78%,

# A. Toddler Intelligence: Mental Development Index

	ART NC						\$	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Joengbloed-Pereboom 2011	98	13.1	66	100.8	15	100	30.7%	-0.20 [-0.51, 0.12]		
Barnes 2004	110.9	18.9	51	113.4	16.8	51	19.8%	-0.14 [-0.53, 0.25]		
Gibson 1998	102.4	8.3	65	103.2	7	62	24.6%	-0.10 [-0.45, 0.24]		$\textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \end{array} \end{array} \\ \hline \begin{tabular}{c} \textcircled{\begin{tabular}{c}} \end{array} \end{array} \\ \hline \begin{tabular}{c} \textcircled{\begin{tabular}{c}} \end{array} \end{array} \\ \hline \begin{tabular}{c} \hline \end{array} \end{array} \\ \begin{tabular}{c} \hline \begin{tabular}{c} \hline \end{array} \end{array} \\ \begin{tabular}{c} \hline \begin{tabular}{c} \hline \end{array} \end{array} \\ \begin{tabular}{c} \hline \begin{tabular}$
Agarwal 2005	92.2	19	41	92.9	13.5	147	24.9%	-0.05 [-0.39, 0.30]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \bullet$
Total (95% CI)			223			360	100.0%	-0.12 [-0.30, 0.05]	•	
Heterogeneity: Tau <sup>2</sup> = 0.00; Cl	hi² = 0.4	1, df =	3 (P =	0.94); l <sup>2</sup>	* = 0%			ł	1 05 0 05	
Test for overall effect: Z = 1.41	(P = 0.	16)							Higher in NC Higher in ART	1

# B. Toddler Intelligence: Language Development (Verbal Intelligence) Score

	-	ART	-	-	NC		·	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
1.4.1 ART vs NC										
Balayla 2017	53.9	23.6	175	55.6	24.4	1345	13.8%	-0.07 [-0.23, 0.09]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet ? ? \bullet \bullet$
Carson 2009	81.67	14.8557	99	76	14.8557	10574	12.7%	0.38 [0.18, 0.58]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet ? ? ? \bullet \bullet$
Subtotal (95% CI)			274			11919	26.6%	0.15 [-0.29, 0.59]		
Heterogeneity: Tau <sup>2</sup> =	0.09; Chi	<sup>2</sup> = 12.24,	df = 1 (	P = 0.00	05); l <sup>2</sup> = 9	2%				
Test for overall effect:	Z = 0.67	(P = 0.50)								
1.4.2 IVF vs NC										
Gibson 1998	91.9	7.6	65	95	8.8	62	8.7%	-0.38 [-0.73, -0.02]		
Sutcliffe 1995	106.18	21.24	68	112.33	13.87	83	9.3%	-0.35 [-0.67, -0.03]		
Place 2003	101	16.81	32	100.1	17.82	40	6.4%	0.05 [-0.41, 0.52]		
Subtotal (95% CI)			165			185	24.4%	-0.27 [-0.50, -0.03]		
Heterogeneity: Tau <sup>2</sup> =	0.01; Chi	$r^2 = 2.40, c$	lf = 2 (F	p = 0.30);	$l^2 = 17\%$					
lest for overall effect:	Z = 2.25	(P = 0.02)								
1.4.3 ICSI vs NC										
Nekkebroeck 2008	53 11	20.24	35	56 42	20.12	53	7 1%	-0 11 [-0 54 0 31]		
Place 2003	99.11	12 68	46	100.42	17.82	40	7.1%	-0.06 [-0.48, 0.37]		
Sutcliffe 2003	106 47	14.4	208	105 15	14 16	221	13.0%	0.09 [-0.10, 0.28]		
Vo 2021	103.2	10.6	421	101 1	9.8	421	14.4%	0 21 [0 07 0 34]	_ <b>_</b>	
Sutcliffe 2003	108.9	17.8	56	104.8	17.5	39	7.4%	0.23 [-0.18, 0.64]		
Subtotal (95% CI)			766			774	49.0%	0.14 [0.04, 0.24]	•	
Heterogeneity: Tau <sup>2</sup> =	0.00: Chi	<sup>2</sup> = 3.51. c	f = 4 (F)	e = 0.48);	$ ^2 = 0\%$					
Test for overall effect:	Z = 2.79	(P = 0.005)	5)	,,						
Total (95% CI)			1205			12878	100.0%	0.02 [-0.13, 0.17]	<b>•</b>	
Heterogeneity: Tau <sup>2</sup> =	0.04; Chi	<sup>2</sup> = 30.65,	df = 9 (	P = 0.00	03); l <sup>2</sup> = 7	1%				
Test for overall effect:	Z = 0.31	(P = 0.76)							Higher in NC Higher in ART	1
Test for subgroup diffe	erences: (	Chi <sup>2</sup> = 10.1	2, df =	2 (P = 0.	006), l <sup>2</sup> =	80.2%				

### C. Toddler Intelligence: Non-Verbal Intelligence Score



Risk of bias legend

(A) Representativeness of the exposed cohort (B) Selection of the non-exposed cohort

(B) Selection of the non-expose

(C) Ascertainment of exposure

(D) Demonstration that outcome of interest was not present at start of study
 (E) Comparability of cohort on the basis of perinatal outcomes

(F) Comparability of cohort on the basis of permata outcol (F) Comparability of cohort on the basis of other factors

(G) Assessment of intelligence outcome

(H) Assessment of motoric outcome

(I) Assessment of social skill

(I) Assessment of social skill

(J) Assessment of behavioral outcome

(K) Period of follow-up

(L) Adequacy of follow-up of cohort

Fig. 2 Intelligence outcome of ART-born compared to naturally conceived (NC) toddlers as assessed with A Mental Development Index, B Language Development (Verbal Intelligence) Score, and C Non-verbal Intelligence Score

p = 0.0001) and subgroups analyses ( $I^2 = 48-85\%$ , p < 0.05) demonstrated significant heterogeneity. There were no evidence of publication biases in all groups (p-Egger >0.05).

Quantitative intelligence was extracted from WPPSI, WISC, and WASI arithmetic subtests [31, 36, 37, 39, 40]. There was no significant difference in arithmetic subtest

# A. Toddler Motoric: Total Motor Score

		ART			NC		5	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
2.1.1 ART vs NC										
Joengbloed-Pereboom 2011	93	15.3	56	95.63	15.2305	86	5.5%	-0.17 [-0.51, 0.17]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet ? \bullet \bullet \bullet$
Balayla 2017	100.8	9.8	175	101.8	12.2	1345	22.4%	-0.08 [-0.24, 0.07]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet ? ? \bullet \bullet$
Subtotal (95% CI)			231			1431	28.0%	-0.10 [-0.24, 0.04]		
Heterogeneity: Tau <sup>2</sup> = 0.00; Ch	ni² = 0.21	, df = 1 (F	P = 0.64	l); l <sup>2</sup> = 0	%					
Test for overall effect: Z = 1.36	(P = 0.1	7)								
2.1.2 IVF vs NC										
Sutcliffe 1995	102.7	16.3	68	105.2	13.4	83	6.1%	-0.17 [-0.49, 0.15]		
Place 2003	105.9	15.1	32	107.8	14.4	40	3.0%	-0.13 [-0.59, 0.34]		
Gibson 1998	90.4	14.4	65	89.5	15.5	62	5.2%	0.06 [-0.29, 0.41]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \bullet \bullet \bullet$
Subtotal (95% CI)			165			185	14.3%	-0.08 [-0.29, 0.13]		
Heterogeneity: Tau <sup>2</sup> = 0.00; Ch	ni² = 0.95	i, df = 2 (F	P = 0.62	2); I <sup>2</sup> = 0	%					
Test for overall effect: Z = 0.71	(P = 0.4	8)								
2.1.3 ICSI vs NC										
Place 2003	102.9	14.1	46	107.8	14.4	40	3.5%	-0.34 [-0.77, 0.09]		
Agarwal 2005	94.8	12.9998	41	96.5	12.9998	147	5.3%	-0.13 [-0.48, 0.22]		•••••
Sutcliffe 2003	96	15.4	208	97.9	14.3	221	16.2%	-0.13 [-0.32, 0.06]		
Vo 2021	103.5	9.8	421	102.6	9.3	421	28.9%	0.09 [-0.04, 0.23]	+	•••••••••• <u>?</u> ••
Sutcliffe 2003	103.8	15.9	56	101.4	13.9	39	3.8%	0.16 [-0.25, 0.57]		
Subtotal (95% CI)			772			868	57.7%	-0.04 [-0.19, 0.12]	-	
Heterogeneity: Tau <sup>2</sup> = 0.01; Ch	hi² = 7.16	6, df = 4 (F	P = 0.13	3); l <sup>2</sup> = 4	4%					
Test for overall effect: Z = 0.49	(P = 0.6	52)								
T-1-1 (05% OI)			4400			0404	400.00/	0.0510.40.0.041		
Total (95% CI)			1168			2484	100.0%	-0.05 [-0.13, 0.04]	🖣 .	
Heterogeneity: Tau <sup>2</sup> = 0.00; Ch	$n^2 = 9.62$	2, df = 9 (F	P = 0.38	$(3); I^2 = 6$	%				-1 -0.5 0 0.5	1
Test for overall effect: Z = 1.10	(P = 0.2	27)							Higher in NC Higher in ART	
Test for subgroup differences:	$Chi^2 = 0$	.31, df = 2	(P = 0	.85), l² =	= 0%					

# B. Toddler Motoric: Gross Motor Score

	ART				NC			Std. Mean Difference	Std. Mean Difference	e Risk of Bias		
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% Cl	I IV, Random, 95% CI	ABCDEFGHIJKL		
2.3.1 IVF vs NC												
Place 2003	111	19.46	32	111.5	16.87	40	10.4%	-0.03 [-0.49, 0.44]		$\bullet \bullet \circ \circ \circ \circ \circ \circ \circ \circ \circ$		
Sutcliffe 1995	104.54	17.95	68	104.1	11.73	83	16.0%	0.03 [-0.29, 0.35]		$\bigcirc \bigcirc $		
Subtotal (95% CI)			100			123	26.3%	0.01 [-0.25, 0.28]				
Heterogeneity: Tau <sup>2</sup> =	0.00; Chi <sup>s</sup>	<sup>2</sup> = 0.04	, df = 1	(P = 0.3)	84); I <sup>2</sup> =	0%						
Test for overall effect:	Z = 0.08 (	(P = 0.9	3)									
2.3.2 ICSI vs NC												
Place 2003	106.5	17.79	46	111.5	16.87	40	11.6%	-0.29 [-0.71, 0.14]				
Sutcliffe 2003	93.1	16.05	201	95.39	14.34	221	23.3%	-0.15 [-0.34, 0.04]	_ <b>-</b> +			
Vo 2021	106.6	9.9	421	104.7	9.6	421	26.6%	0.19 [0.06, 0.33]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \circ \circ \circ \circ \circ \circ \circ$		
Sutcliffe 2003	101.5	15.9	56	96.4	15.4	39	12.1%	0.32 [-0.09, 0.73]	+	— •••••••••••		
Subtotal (95% CI)			724			721	73.7%	0.03 [-0.23, 0.28]	-			
Heterogeneity: Tau <sup>2</sup> =	0.05; Chi <sup>;</sup>	<sup>2</sup> = 12.4	8, df =	3 (P = 0	.006); F	² = 76%						
Test for overall effect:	Z = 0.21 (	(P = 0.8	3)									
Total (95% CI)			824			844	100.0%	0.02 [-0.16, 0.21]	•			
Heterogeneity: Tau <sup>2</sup> =	0.03; Chi <sup>i</sup>	<sup>2</sup> = 12.7	0. df =	5 (P = 0	.03); l <sup>2</sup>	= 61%			1 1 1 1	<u>j</u>		
Test for overall effect:	Z = 0.26 (	P = 0.7	9)		,.				-1 -0.5 0 0.5			
Test for subgroup diffe	rences: C	chi² = 0.	01, df =	= 1 (P =	0.93), l <sup>i</sup>	² = 0%			Figher in NC Higher In /			

### C. Toddler Motoric: Fine Motor Score

		ART			NC			Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% C	IV, Random, 95% Cl	ABCDEFGHIJKL
2.2.1 IVF vs NC										
Sutcliffe 1995	100.85	14.53	68	106.2	15.11	83	13.6%	-0.36 [-0.68, -0.04]		$\bullet \bullet $
Place 2003	100.9	10.81	32	104.1	11.84	40	7.5%	-0.28 [-0.74, 0.19]		$\bullet \bullet ? \bullet \bullet$
Subtotal (95% CI)			100			123	21.1%	-0.33 [-0.60, -0.07]		
Heterogeneity: Tau <sup>2</sup> =	0.00; Chi	<sup>2</sup> = 0.08	, df = 1	(P = 0.7	8); I <sup>2</sup> = (	0%				
Test for overall effect:	Z = 2.45	(P = 0.0	1)							
2.2.2 ICSI vs NC										
Place 2003	99.2	10.29	46	104.1	11.84	40	8.7%	-0.44 [-0.87, -0.01]		$\bullet \bullet \circ \circ \circ \circ \circ \circ \circ \circ$
Sutcliffe 2003	98.92	14.73	208	100.41	14.21	221	26.2%	-0.10 [-0.29, 0.09]		
Sutcliffe 2003	106.1	16	56	106.3	12.3	39	9.4%	-0.01 [-0.42, 0.40]		
Vo 2021	100.4	9.7	421	100.4	8.9	421	34.6%	0.00 [-0.14, 0.14]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet$
Subtotal (95% CI)			731			721	78.9%	-0.07 [-0.21, 0.06]	-	
Heterogeneity: Tau <sup>2</sup> =	0.00; Chi	<sup>2</sup> = 4.01	, df = 3	(P = 0.2	6); l² = 2	25%				
Test for overall effect:	Z = 1.08	(P = 0.2	8)							
Total (95% CI)			831			844	100.0%	-0.14 [-0.27, 0.00]	•	
Heterogeneity: Tau <sup>2</sup> =	0.01; Chi	<sup>2</sup> = 7.68	, df = 5	(P = 0.1	7); l² = 3	35%				1
Test for overall effect:	Z = 1.92	(P = 0.0	5)						Higher in NC Higher in ART	1
Test for subgroup diffe	rences: C	Chi² = 2.	88, df =	: 1 (P = 0	0.09), I²	= 65.39	%			
Risk of bias legend										
(A) Representativenes	s of the e	xposed	cohort							
(B) Selection of the no	n-expose	d cohor	t							
(C) Ascertainment of e	xposure									
(D) Demonstration that	outcome	e of inter	rest was	s not pre	sent at	start of	study			
(E) Comparability of co	hort on th	ne basis	of peri	natal out	comes					
(F) Comparability of co	hort on th	ne basis	of othe	er factors						
(G) Assessment of inte	lligence o	outcome	9							
(H) Assessment of mol	oric outc	ome								
<ol> <li>Assessment of social</li> </ol>	al skill									
(J) Assessment of beh	avioral o	utcome								
(K) Period of follow-up										
(L) Adequacy of follow	<ul> <li>up of col</li> </ul>	hort								

Fig. 3 Motoric outcome of ART-born compared to naturally conceived (NC) toddlers as assessed with A Total Motor Score, B Gross Motor Score, and C Fine Motor Score

# A. Toddler: Parent's CBCL Total Behavior Problems Score

	ART			NC		:	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Nekkebroeck 2008	46.46 8.9	34	50	9.02	53	5.5%	-0.39 [-0.83, 0.04]		• • • • • • • • • • ? ? • • •
Joengbloed-Pereboom 2011	46.1 8.1	66	49.11	11.3864	86	9.9%	-0.30 [-0.62, 0.03]		
Barnes 2004	48.9 9	301	49.9	9.5	310	40.9%	-0.11 [-0.27, 0.05]		
Barnes 2004	49.1 9.1	345	49.9	9.5	310	43.7%	-0.09 [-0.24, 0.07]		$\textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \end{array} \end{array} \\ \hline \begin{tabular}{c} \textcircled{\begin{tabular}{c}} \end{array} \end{array} \\ \hline \begin{tabular}{c} \textcircled{\begin{tabular}{c}} \end{array} \end{array} \\ \hline \begin{tabular}{c} \hline \hline \end{array} \end{array} \\ \hline \begin{tabular}{c} \hline \end{array} \\ \hline \begin{tabular}{c} \hline \begin{tabular}{c} \hline \end{array} \end{array} \\ \hline \begin{tabular}{c} \hline \begin{tabular}{c} \hline \end{array} \end{array} \\ \hline \begin{tabular}{c} \hline tabul$
Total (95% CI)		746			759	100.0%	-0.13 [-0.23, -0.03]	•	
Heterogeneity: $Tau^2 = 0.00$ ; Cl Test for overall effect: $Z = 2.56$	hi² = 2.80, df 6 (P = 0.01)	= 3 (P =	= 0.42);	l² = 0%			ł	-1 -0.5 0 0.5 Higher in NC Higher in ART	

# B. Toddler: Parent's CBCL Internalizing Behavior Score

	AR	г		NC		:	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean S	D Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Nekkebroeck 2008	45.21 9.	3 33	48.43	8.97	53	7.4%	-0.35 [-0.79, 0.09]		
Joengbloed-Pereboom 2011	43.4	9 66	46.57	10.4025	86	12.8%	-0.32 [-0.64, 0.00]		
Barnes 2004	47.4 1	345	48	10	310	40.7%	-0.06 [-0.21, 0.09]		
Barnes 2004	47.6 9.	1 301	48	10	310	39.0%	-0.04 [-0.20, 0.12]		• • • • • • ? ? ? • • •
Total (95% CI)		745			759	100.0%	-0.11 [-0.23, 0.02]	•	
Heterogeneity: Tau <sup>2</sup> = 0.00; Cl Test for overall effect: Z = 1.72	hi² = 3.83, c ? (P = 0.09)	f = 3 (P :	= 0.28);	l² = 22%			-	1 -0.5 0 0.5 Higher in NC Higher in ART	1

# C. Toddler: Parent's CBCL Externalizing Behavior Score

		ART			NC			Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Nekkebroeck 2008	48.06	8.98	33	51.74	8.98	53	5.4%	-0.41 [-0.85, 0.03]		
Joengbloed-Pereboom 2011	49.7	7.6	66	52.35	12.4667	86	10.0%	-0.25 [-0.57, 0.07]		
Barnes 2004	49	8.3	301	50.4	8.7	310	40.9%	-0.16 [-0.32, -0.01]		• • • • • • ? ? ? • • •
Barnes 2004	49.3	8.6	345	50.4	8.7	310	43.8%	-0.13 [-0.28, 0.03]		$\textcircled{\bullet} \textcircled{\bullet} \textcircled{\bullet} \textcircled{\bullet} \textcircled{\bullet} \textcircled{\bullet} \textcircled{\bullet} \textcircled{\bullet} $
Total (95% CI)			745			759	100.0%	-0.17 [-0.27, -0.07]	•	
Heterogeneity: Tau <sup>2</sup> = 0.00; Ch	ni² = 1.6	4, df =	3 (P =	0.65); l <sup>2</sup>	= 0%			F 1		
Test for overall effect: Z = 3.27	(P = 0.	001)						-1	Higher in NC Higher in ART	

# **D. Toddler: Social Score**

	ART				NC		5	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
3.1.1 IVF vs NC										
Gibson 1998	100.8	3	65	101.4	3.1	62	6.6%	-0.20 [-0.54, 0.15]		$\textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \end{array} \end{array} \\ \hline \begin{tabular}{c} \textcircled{\begin{tabular}{c}} \end{array} \end{array} \\ \hline \begin{tabular}{c} \textcircled{\begin{tabular}{c}} \end{array} \end{array} \\ \hline \begin{tabular}{c} \hline \end{array} \end{array} \\ \hline \begin{tabular}{c} \end{array} \\ \hline \begin{tabular}{c} \end{array} \end{array} \\ \hline \begin{tabular}{c} \hline \end{array} \end{array} \\ \hline \begin{tabular}{c} \end{array} \\ \begin{tabular}{c} \end{array} \\ \ \begin{tabular}{c} \end{array} \\ \ \begin{tabular}{c} \end{array} \\ \ \begin{tabular}{c} \end{array} \\ \begin{tabular}{c} \end{array} \begin{tabular}{c} \end{array} \ \begin{tabular}{c} \end{array} \begin{tabular}{c} \end{array} \\ \begin{tabular}{c} \end{array} \ \begin{tabular}{c} \end{array} \ \begin{table} \begin{tabular}{c}$
Sutcliffe 1995	108.78	17.91	68	109.18	12.12	83	7.8%	-0.03 [-0.35, 0.29]		$\bullet \bullet \circ \circ \bullet \bullet$
Place 2003	98.6	9.98	32	97.5	10.3	40	3.7%	0.11 [-0.36, 0.57]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \circ \circ \circ \circ \circ \circ \circ$
Subtotal (95% CI)			165			185	18.0%	-0.06 [-0.27, 0.15]	-	
Heterogeneity: Tau <sup>2</sup> =	0.00; Chi	<sup>2</sup> = 1.12	, df = 2	(P = 0.5)	7);   <sup>2</sup> = (	0%				
Test for overall effect:	Z = 0.57	(P = 0.5	7)							
3.1.2 ICSI vs NC										
Sutcliffe 2003	102.5	14.1	56	103.8	14	39	4.8%	-0.09 [-0.50, 0.32]		
Place 2003	96.7	12.95	46	97.5	10.3	40	4.4%	-0.07 [-0.49, 0.36]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \bullet \circ \circ \circ \circ \circ \circ$
Sutcliffe 2003	95.14	15.41	208	96.06	15.29	221	22.3%	-0.06 [-0.25, 0.13]		
Vo 2021	102.8	11.7	421	103.4	9.4	421	43.8%	-0.06 [-0.19, 0.08]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \circ \circ \circ \circ \circ \circ \circ$
Agarwal 2005	98.6	12.2	41	98.5	10	147	6.7%	0.01 [-0.34, 0.36]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \bullet \bullet$
Subtotal (95% CI)			772			868	82.0%	-0.05 [-0.15, 0.04]	•	
Heterogeneity: Tau <sup>2</sup> =	0.00; Chi	$^{2} = 0.17$	, df = 4	(P = 1.0)	0); l <sup>2</sup> = (	0%				
Test for overall effect	Z = 1.08	(P = 0.2	8)							
Total (95% CI)			937			1053	100.0%	-0.06 [-0.15, 0.03]	•	
Heterogeneity: Tau <sup>2</sup> =	0.00; Chi	<sup>2</sup> = 1.29	, df = 7	(P = 0.9)	9); l² = (	0%			-1 -0.5 0 0.5	1
Test for overall effects	Z = 1.22	(P = 0.2	2)						Higher in NC Higher in ART	
Test for subgroup diff	erences: C	$Chi^2 = 0.$	00, df =	1 (P = 0	.96), l <sup>2</sup>	= 0%			right in the right in the	
Risk of bias legend										
(A) Representativene	ss of the e	exposed	cohort							

(B) Selection of the non-exposed cohort

(C) Ascertainment of exposure

(D) Demonstration that outcome of interest was not present at start of study

(E) Comparability of cohort on the basis of perinatal outcomes

(F) Comparability of cohort on the basis of other factors (G) Assessment of intelligence outcome

(H) Assessment of motoric outcome

(I) Assessment of social skill

(J) Assessment of behavioral outcome (K) Period of follow-up

(L) Adequacy of follow-up of cohort

Fig. 4 Behavior and social outcomes of ART-born toddlers compared to naturally conceived (NC) toddlers as assessed using A Total Behavior Problems Score, **B** Internalizing Behavior Score, **C** Externalizing Behavior Score, and **D** Social Score

# A. Preschool and Primary School: Full Scale Intelligence Quotient

/		ART	••••		NC		.gones	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% C	I IV, Random, 95% CI	ABCDEFGHIJKL
4.1.1 ART vs NC										
Bay 2014	102.8	21.3207	71	105.6	12.9	1577	7.8%	-0.21 [-0.45, 0.03]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet ? ? \bullet \bullet \bullet$
Heineman 2019	116.29	15.8297	57	115.1	15.8297	66	4.9%	0.07 [-0.28, 0.43]		
Heijligers 2018	109.55	14.18	51	106.06	9.94	35	3.6%	0.27 [-0.16, 0.71]		•••••
Subtotal (95% CI)			179			1678	16.3%	0.00 [-0.28, 0.29]	$\bullet$	
Heterogeneity: Tau <sup>2</sup> = 0.03; C	chi <sup>2</sup> = 4.3	5, df = 2 (F	P = 0.11	); I <sup>2</sup> = 54	%					
Test for overall effect: Z = 0.0	12 (P = 0.9)	98)								
4.1.2 IVF vs NC										
Place 2003	101.9	11.98	17	108.4	11.95	15	1.6%	-0.53 [-1.24, 0.18]	• • • • • • • • • • • • • • • • • • •	
Knoester 2008	107	16.0123	82	110	16.0123	85	5.9%	-0.19 [-0.49, 0.12]		
Ponjaert-Kristoffersen 2005	108.36	13.37	399	109.25	13.33	454	11.6%	-0.07 [-0.20, 0.07]		
Subtotal (95% CI)			498			554	19.1%	-0.10 [-0.22, 0.02]	-	
Heterogeneity: Tau <sup>2</sup> = 0.00; C	$chi^2 = 1.96$	5, df = 2 (F	<sup>o</sup> = 0.38	(); $I^2 = 0\%$	6					
Test for overall effect: Z = 1.6	60 (P = 0.7)	11)								
4.1.3 ICSI vs NC										
Place 2003	99.7	12	17	108.4	11.95	15	1.6%	-0 71 [-1 43 0 01]	← − − − − − − − − − − − − − − − − − − −	
Knoester 2008	103	16.0123	86	110	16.0123	85	5.9%	-0.44 [-0.74, -0.13]		
Winter 2014	115.55	14.41	49	118.85	12.7	48	4.1%	-0.24 [-0.64, 0.16]		
Faramarzi 2016	105.9	14	28	107.2	12.9	32	2.8%	-0.10 [-0.60, 0.41]		
Leunens 2008	107.9	12.6	109	108.9	12.8	90	6.5%	-0.08 [-0.36, 0.20]		
Ponjaert-Kristoffersen 2005	109.11	13.89	483	109.25	13.33	454	11.9%	-0.01 [-0.14, 0.12]	-	• • • • ? • • • ? ? • •
Ludwig 2009	103.07	10.7	276	102.67	10.73	273	10.3%	0.04 [-0.13, 0.20]		
Ponjaert-Kristoffersen 2004	108.6	15.2	239	108	14.2	265	10.0%	0.04 [-0.13, 0.22]		
Sanchez Albisua 2011	104.93	12.3	35	103.8	11	37	3.3%	0.10 [-0.37, 0.56]		
Leunens 2006	110.6	15	151	106	13.75	153	8.1%	0.32 [0.09, 0.55]		
Subtotal (95% CI)			1473			1452	64.6%	-0.04 [-0.17, 0.09]	•	
Heterogeneity: Tau <sup>2</sup> = 0.02; 0	chi <sup>2</sup> = 21.7	70, df = 9	(P = 0.0)	10); l <sup>2</sup> =	59%					
Test for overall effect: Z = 0.5	4 (P = 0.5)	59)								
Total (95% CI)			2150			3684	100.0%	-0.05[-0.14.0.05]	•	
Hotorogonoity: $Tau^2 = 0.02$ : (	bi2 - 20 (	02 df - 15	(P = 0	01): 12 -	50%	0004	100.070	0.00 [-0.14, 0.00]		4
Test for overall effect: $7 = 1.02$ , C	1 (P = 0)	2, ui = 10 21)	(r = 0	01), 1	50 %				-1 -0.5 0 0.5	1
	0112 - 0.	70 -16 - 0	(D - 0	70) 12 -	00/				Higher in NC Higher in ART	

Test for subgroup differences:  $Chi^2 = 0.72$ , df = 2 (P = 0.70),  $l^2 = 0\%$ 

# B. Preschool and Primary School: Verbal Intelligence Score

		ART			NC		3	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
4.2.1 ART vs NC										
Bay 2014	102	21.3732	71	104.8	10.8	1577	7.1%	-0.24 [-0.48, -0.01]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ$
Fallah 2013	49.9	6.02	61	50.08	6.8	61	5.6%	-0.03 [-0.38, 0.33]		••••••••••
Heineman 2019	116.92	17.2687	57	116.7	17.2687	66	5.6%	0.01 [-0.34, 0.37]		
Carson 2011	112.5	13.891	96	110.5	13.891	6244	7.5%	0.14 [-0.06, 0.35]	<u>+</u>	
Heijligers 2018	109.16	13.75	51	105.57	9.91	35	4.8%	0.29 [-0.14, 0.72]		
Subtotal (95% CI)			336			7983	30.6%	0.01 [-0.17, 0.20]	-	
Heterogeneity: Tau <sup>2</sup> = 0.02; C	$chi^2 = 7.72$	2, df = 4 (F	P = 0.10	); I <sup>2</sup> = 48	%					
Test for overall effect: Z = 0.1	5 (P = 0.8	38)								
4.2.2 IVF vs NC										
Knoester 2008	110.67	10.3829	82	116.22	10.3829	85	6.2%	-0.53 [-0.84, -0.22]		• • • • • • • • ? ? ? • •
Place 2003	104.2	14.98	17	109.9	12.69	15	2.7%	-0.40 [-1.10, 0.30]	<	
Ponjaert-Kristoffersen 2005	108.46	12.74	416	109.44	13.42	469	8.2%	-0.07 [-0.21, 0.06]		
Subtotal (95% CI)			515			569	17.1%	-0.30 [-0.66, 0.07]		
Heterogeneity: Tau <sup>2</sup> = 0.07; C	chi <sup>2</sup> = 7.6	1, df = 2 (F	9 = 0.02	?); l <sup>2</sup> = 74	%					
Test for overall effect: Z = 1.6	61 (P = 0.1	11)								
4.2.3 ICSI vs NC										
Knoester 2008	106.22	10.3829	83	116.22	10.3829	85	6.1%	-0.96 [-1.28, -0.64]	←	
Place 2003	102.3	11.97	15	109.9	12.69	15	2.6%	-0.60 [-1.33, 0.13]	<	
Winter 2014	113.35	12.95	49	115.63	10.96	48	5.2%	-0.19 [-0.59, 0.21]		
Ponjaert-Kristoffersen 2005	108.75	13.77	483	109.44	13.42	469	8.2%	-0.05 [-0.18, 0.08]		• • • • ? • • • • • ? • •
Leunens 2008	109.3	12.4	109	109.4	14	90	6.5%	-0.01 [-0.29, 0.27]		
Faramarzi 2016	109.6	13.4	28	109.7	14.5	32	4.1%	-0.01 [-0.51, 0.50]		
Sanchez Albisua 2011	101.56	14.2	35	100.73	12.09	37	4.5%	0.06 [-0.40, 0.52]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \circ \bullet \circ \bullet \bullet \bullet \bullet \bullet \bullet \bullet$
Ponjaert-Kristoffersen 2004	109	13.7	293	107	15	265	7.9%	0.14 [-0.03, 0.31]		
Leunens 2006	113.3	13.7	151	108.2	12.3	153	7.2%	0.39 [0.16, 0.62]		
Subtotal (95% CI)			1246			1194	52.2%	-0.10 [-0.34, 0.14]		
Heterogeneity: Tau <sup>2</sup> = 0.10; C	$chi^2 = 52.6$	63, df = 8 (	P < 0.0	00001); l <sup>2</sup>	= 85%					
Test for overall effect: Z = 0.8	4 (P = 0.4	40)								
Total (95% CI)			2097			9746	100.0%	-0.09 [-0.23, 0.05]	•	
Heterogeneity: Tau <sup>2</sup> = 0.06; C	chi² = 72.4	43, df = 16	(P < 0	.00001);	l² = 78%					
Test for overall effect: Z = 1.2	6 (P = 0.2	21)	101 10	,					-1 -0.5 0 0.5 Higher in NC Higher in APT	1
Test for subgroup differences	: Chi <sup>2</sup> = 2	.36, df = 2	(P = 0.1)	.31), l <sup>2</sup> =	15.3%				Higher III NO Higher III ART	

# C. Preschool and Primary School: Quantitative Intelligence/ Arithmetics Score

								-		
		ART			NC		:	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Ponjaert-Kristoffersen 2005	10.24	2.47	483	10.54	2.54	469	29.4%	-0.12 [-0.25, 0.01]		• • • • ? • • • ? ? • •
Ponjaert-Kristoffersen 2005	10.62	2.6	416	10.54	2.54	469	28.0%	0.03 [-0.10, 0.16]		
Leunens 2008	9.7	2.8	151	9.6	2.9	153	12.0%	0.03 [-0.19, 0.26]	_ <b>_</b> _	
Ponjaert-Kristoffersen 2004	11.02	2.79	293	10.75	2.85	265	19.8%	0.10 [-0.07, 0.26]	+	
Leunens 2006	10	2.4	109	9.7	2.4	90	8.1%	0.12 [-0.15, 0.40]		
Faramarzi 2016	8.2	2.5	28	7.8	2.6	32	2.6%	0.15 [-0.35, 0.66]		•••••
Total (95% CI)			1480			1478	100.0%	0.01 [-0.07, 0.09]	•	
Heterogeneity: Tau <sup>2</sup> = 0.00; C	chi <sup>2</sup> = 6.0	)9, df =	= 5 (P =	0.30);	<sup>2</sup> = 18	%				
Test for overall effect: Z = 0.2	5 (P = 0	.80)							Higher in NC Higher in ART	1

Fig. 5 Intelligence outcome of ART-born compared to naturally conceived (NC) pre and primary schoolers as measured with A Full Scale IQ, B Verbal IQ, C Quantitative Intelligence/Arithmetics, D Performance IQ, E Fluid Intelligence, F Short-term Memory and Processing Speed, and G Visual-spatial Intelligence, H Long-term Memory Retrieval/Learning Ability I Executive Function

# D. Preschool and Primary School: Non-Verbal Intelligence Score

		ART			NC		:	Std. Mean Difference		Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% Cl	:	IV, Random, 95% CI	ABCDEFGHIJKL
4.3.1 ART vs NC											
Bay 2014	102	25.5648	71	105.1	16.2	1577	7.1%	-0.19 [-0.42, 0.05]			• • • • • • • • ? ? • • •
Carson 2011	81.9	14.3867	96	83.1	14.3867	6224	8.0%	-0.08 [-0.28, 0.12]		-+	$\bullet \bullet \bullet \bullet \bullet \bullet \bullet ? ? ? \bullet \bullet$
Heineman 2019	112.19	16.6479	57	110.8	16.6479	66	4.8%	0.08 [-0.27, 0.44]		_ <del></del>	$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet ? \bullet \bullet \bullet \bullet$
Heijligers 2018 Subtotal (95% CI)	106.8	13.13	51 275	104.37	9.16	35 7902	3.7% 23.7%	0.21 [-0.23, 0.64]		•	
Heterogeneity: $Tau^2 = 0.00^{\circ}$	Chi <sup>2</sup> = 3 20	) df = 3 (F	P = 0.36	b) $l^2 = 69$	6					- -	
Test for overall effect: $Z = 0.8$	35 (P = 0.3	39)	0100	,,,, , , , , , , , , , , , , , , , , , ,	•						
4.3.2 IVF vs NC											
Place 2003	99.2	13.09	17	105	12.68	15	1.8%	-0.44 [-1.14, 0.27]	+		
Knoester 2008	109.56	10.39	82	112.67	10.39	85	5.7%	-0.30 [-0.60, 0.01]			• • • • • • • • ? ? ? • •
Ponjaert-Kristoffersen 2005	106.06	14.32	415	106.41	13.05	469	9.8%	-0.03 [-0.16, 0.11]		-	• • • • ? • • • ? ? • •
Subtotal (95% CI)			514			569	17.2%	-0.15 [-0.38, 0.08]			
Heterogeneity: Tau <sup>2</sup> = 0.02; 0	Chi² = 3.6	1, df = 2 (F	P = 0.16	5); I² = 45	%						
Test for overall effect: Z = 1.2	28 (P = 0.2	20)									
4.3.3 ICSI ve NC											
4.5.5 1051 VS 140	104.67	10.20	00	110.07	10.20	05	E E0/	0 77 [ 4 09 0 45]			
Rhoester 2008	07.2	10.39	03	112.07	10.39	00 15	0.0%	-0.77 [-1.08, -0.45]	-		
Face 2003	97.5	10.3	28	102.0	12.00	35	3.0%	-0.26[-0.76_0.24]			
Leunens 2008	104.4	13.3	100	102.3	12.6	90	6.2%	-0.20 [-0.70, 0.24]			
Poniaert-Kristoffersen 2004	104.4	16	293	106.2	14.6	265	8.9%	-0.03[-0.19]0.14]			
Winter 2014	114.34	14.13	49	114 34	14.13	49	4.2%	0.00[-0.40, 0.40]			
Ludwig 2009	103.08	10.77	276	102.67	10.73	273	8.9%	0.04 [-0.13, 0.21]			
Ponjaert-Kristoffersen 2005	106.96	14.17	483	106.41	13.05	469	9.9%	0.04 [-0.09, 0.17]		- <b>-</b> -	
Sanchez Albisua 2011	106.62	11.4	35	105.25	10.46	37	3.4%	0.12 [-0.34, 0.59]		<del></del>	
Leunens 2006	107.9	16.3	151	103.8	15.2	153	7.4%	0.26 [0.03, 0.49]		— <b>-</b>	
Subtotal (95% CI)			1522			1471	59.1%	-0.08 [-0.24, 0.08]			
Heterogeneity: Tau <sup>2</sup> = 0.04; 0	Chi² = 32.8	37, df = 9 (	P = 0.0	001); l² =	= 73%						
Test for overall effect: Z = 0.9	98 (P = 0.3	33)									
Total (95% CI)			2311			9942	100.0%	-0.07 [-0.18, 0.03]		. •	
Heterogeneity: Tau <sup>2</sup> = 0.02; 0	Chi <sup>2</sup> = 40.6	65, df = 16	(P = 0	.0006); I <sup>2</sup>	= 61%				-1	-0.5 0 0.5	
Test for overall effect: Z = 1.4	I4 (P = 0.1	15)							- F	ligher in NC Higher in ART	
Test for subgroup differences	: Chi <sup>2</sup> = 0	.44, df = 2	(P = 0.	.80), l <sup>2</sup> =	0%						

# E. Preschool and Primary School: Fluid Intelligence Score

		ART			NC		;	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% Cl	ABCDEFGHIJKL
4.8.1 IVF vs NC										
Ponjaert-Kristoffersen 2005	10.62	2.65	416	11.1	2.66	469	16.1%	-0.18 [-0.31, -0.05]		
Knoester 2008	16.8	6.1613	82	16.3	4.9595	85	9.4%	0.09 [-0.21, 0.39]		• • • • • • • • ? ? ? • •
Subtotal (95% CI)			498			554	25.5%	-0.08 [-0.34, 0.17]	-	
Heterogeneity: Tau <sup>2</sup> = 0.02; C	hi² = 2.5	55, df = 1	(P = 0.	11); l <sup>2</sup> =	= 61%					
Test for overall effect: Z = 0.6	3 (P = 0	.53)								
4.8.2 ICSI vs NC										
Leunens 2008	9.85	2.7	109	10.45	3.7	90	10.2%	-0.19 [-0.47, 0.09]		
Knoester 2008	15.4	4.9595	83	16.3	4.9595	85	9.4%	-0.18 [-0.48, 0.12]		• • • • • • • • ? ? ? • •
Ponjaert-Kristoffersen 2004	11.12	2.48	300	11.22	2.51	266	14.7%	-0.04 [-0.21, 0.13]		
Winter 2014	12.44	2.65	49	12.51	2.555	48	6.8%	-0.03 [-0.42, 0.37]		
Ponjaert-Kristoffersen 2005	11.09	2.57	485	11.1	2.66	469	16.3%	-0.00 [-0.13, 0.12]	+	• • • • ? • • • ? ? • •
Faramarzi 2016	10.9	2.6	28	10.4	1.9	32	4.8%	0.22 [-0.29, 0.73]		•••••
Leunens 2006	9.7	3.3	151	8.4	3	153	12.1%	0.41 [0.18, 0.64]		
Subtotal (95% CI)			1205			1143	74.5%	0.02 [-0.13, 0.17]	+	
Heterogeneity: Tau <sup>2</sup> = 0.02; C	hi² = 16	.47, df =	6 (P = 0	0.01); l <sup>2</sup>	= 64%					
Test for overall effect: Z = 0.2	8 (P = 0	.78)								
Total (95% CI)			1703			1697	100.0%	-0.00 [-0.13, 0.12]	<b>•</b>	
Heterogeneity: Tau <sup>2</sup> = 0.02; C	hi² = 23	.33, df =	8 (P = 1	0.003);	$ ^2 = 66\%$			H	1 05 0 05	
Test for overall effect: Z = 0.0	7 (P = 0	.94)							Higher in NC Higher in ART	
Tast for subgroup differences	Chi2 -	0.46 df -	- 1 (D -	0.50)	12 - 0.0%				ingite interestingiter invite	

Test for subgroup differences:  $Chi^2 = 0.94$ ) Test for subgroup differences:  $Chi^2 = 0.46$ , df = 1 (P = 0.50),  $I^2 = 0\%$ 

# F. Preschool and Primary School: Short-Term Memory and Processing Speed Score

		ART			NC			Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% C	IV, Random, 95% CI	ABCDEFGHIJKL
Winter 2014	10.86	2.73	49	11.38	2.42	48	7.4%	-0.20 [-0.60, 0.20]		
Heijligers 2018	103.22	13.11	51	105.69	13.91	35	6.4%	-0.18 [-0.61, 0.25]		••••
Leunens 2008	11.1	2.75	109	11.55	3.05	90	15.1%	-0.16 [-0.43, 0.12]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ ? ? \bullet \bullet$
Ludwig 2009	101.9	11.53	276	101.71	11.72	273	42.2%	0.02 [-0.15, 0.18]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet ? ? \bullet \bullet$
Leunens 2006	11.06667	3.033333	151	10.8	3.033333	153	23.4%	0.09 [-0.14, 0.31]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ ? ? \bullet \bullet$
Sanchez Albisua 2011	104.94	12.76	35	103.81	11.04	37	5.5%	0.09 [-0.37, 0.56]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet ? \bullet \bullet \bullet \bullet$
Total (95% CI) Heterogeneity: Tau <sup>2</sup> = 0. Test for overall effect: Z	00; Chi² = 3 = 0.31 (P =	.51, df = 5 ( 0.76)	671 P = 0.6	2); l² = 0	%	636	100.0%	-0.02 [-0.13, 0.09]	-1 -0.5 0 0.5 Higher in NC Higher in ART	-1 1

Fig. 5 continued

### G. Preschool and Primary School: Visual-Spatial Score

		ART			NC			Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
4.5.1 ART vs NC										
Knoester 2008	16.25	1.5616	82	17.2	5.1706	85	5.2%	-0.25 [-0.55, 0.06]		• • • • • • • • ? ? ? • •
Carson 2011	86.1	19.8442	96	90	19.8442	6244	9.8%	-0.20 [-0.40, 0.01]		
Ponjaert-Kristoffersen 2005	10.62667	3.2	416	10.74	3.083333	469	16.0%	-0.04 [-0.17, 0.10]	-	• • • • ? • • • ? ? • •
Subtotal (95% CI)			594			6798	31.0%	-0.12 [-0.25, 0.01]	•	
Heterogeneity: Tau <sup>2</sup> = 0.00; C	chi <sup>2</sup> = 2.66, d	f = 2 (P = 0)	).26); l <sup>2</sup>	= 25%						
Test for overall effect: Z = 1.8	0 (P = 0.07)									
4.5.2 ICSI vs NC										
Winter 2014	11.82	2.95	49	13.15	3.06	48	3.2%	-0.44 [-0.84, -0.04]		
Knoester 2008	15.85	5.1706	86	17.2	5.1706	85	5.3%	-0.26 [-0.56, 0.04]		• • • • • • • • ? ? ? • •
Leunens 2008	10.8	2.833333	109	11.16667	4.466667	90	6.0%	-0.10 [-0.38, 0.18]		
Ponjaert-Kristoffersen 2004	10.73	3.316667	293	10.94	2.983333	265	12.5%	-0.07 [-0.23, 0.10]		
Faramarzi 2016	10.7	4.35	28	10.95	3.65	32	2.1%	-0.06 [-0.57, 0.45]		
Ludwig 2009	104.57	11.03	276	104.59	10.91	273	12.4%	-0.00 [-0.17, 0.17]		
Ponjaert-Kristoffersen 2005	10.85	3.23	483	10.74	3.083333	469	16.6%	0.03 [-0.09, 0.16]	-	$\bullet \bullet \bullet \bullet ? \bullet \bullet \bullet ? ? \bullet \bullet$
Sanchez Albisua 2011	108.3	10.04	35	106.86	9.87	37	2.5%	0.14 [-0.32, 0.61]		
Leunens 2006	11.63333	3.166667	151	11.1	3.033333	153	8.4%	0.17 [-0.05, 0.40]		
Subtotal (95% CI)			1510			1452	69.0%	-0.03 [-0.12, 0.06]	•	
Heterogeneity: Tau <sup>2</sup> = 0.01; C	$hi^2 = 11.24$ ,	df = 8 (P =	0.19);	<sup>2</sup> = 29%						
Test for overall effect: Z = 0.6	3 (P = 0.53)									
Total (95% CI)			2104			8250	100.0%	-0.06 [-0.13, 0.02]	•	
Heterogeneity: Tau <sup>2</sup> = 0.00; C	hi <sup>2</sup> = 15.66.	df = 11 (P	= 0.15)	$l^2 = 30\%$					1. J. 1. J.	4
Test for overall effect: Z = 1.4	9 (P = 0.14)		,						-1 -0.5 0 0.5 1	
Test for subgroup differences	: Chi <sup>2</sup> = 1.18	, df = 1 (P =	= 0.28).	l <sup>2</sup> = 15.1%					Higher In NC Higher In ART	

### H. Preschool and Primary School: Long-term Memory Retrieval/ Learning Ability Score



## I. Preschool and Primary School: Executive Function Score



(J) Assessment of behavioral outcome

(K) Period of follow-up (L) Adequacy of follow-up of cohort

Fig. 5 continued

score between ART and NC schoolers (p = 0.80) (Fig. 5C). The data exhibited good homogeneity ( $l^2 = 18\%$ , p = 0.30) and no evidence of publication bias (p-Egger = 0.338).

Non-verbal intelligence score was obtained from WPPSI-R [22, 27, 31, 33, 39, 40, 42], WASI [32], WISC [36, 37] Performance IQ subtest, K-ABC total score excluding knowledge score [38, 41], RAKIT recognizes figure exclusion, discs, and hidden figures [35], and British Ability Scale (BAS) non-verbal ability [28] scores. According to the total pooled analysis (p=0.15) and subgroup analyses (p=0.20-0.39) (Fig. 5D), ART schoolers had comparable non-verbal score to NC schoolers. Significant heterogeneities were noticed in the pooled  $(I^2 = 61\%)$ , p = 0.0006) and ICSI subgroup (I<sup>2</sup>=73%, p < 0.0001) analyses. There were no indications of publication biases in all groups (p-Egger > 0.05) (Supplemental Table S6).

Fluid intelligence score was derived from the picture concepts, picture completion, and matrix reasoning subtests of the WPPSI, WISC, and WASI [31, 36, 37, 39, 40, 42], the K-ABC planning subtest [38], and the RAKIT recognize figure exclusion subtest [35] scores. The categorization of CHC intelligence model was based on Keith et al. (2006) [50] for Weschler, Gallagher et al. (2011) [51] for K-ABC, and Jan te Nijenhuis et al. (2004) [52] for RAKIT subtests. As seen in Figure 5E, there were no differences of the fluid intelligence score between IVF (p=0.53) and ICSI (p=0.78) schoolers compared to NC schoolers. The data exhibited moderate heterogeneity ( $I^2 = 61-64\%$ ); however, no publication biases were observed (p-Egger > 0.05).

Short-term memory and processing speed scores were obtained from the WPPSI, WISC, and WASI picture memory, sequencing, and digit span, coding, and substitution subtests [36, 37], K-ABC sequential processing [38, 41], and automated working memory assessment (AWMA) [33] scores. ART schoolers had comparable short-term memory and processing speed scores with NC schoolers (p=0.76) (Fig. 5F). The data showed homogeneity ( $l^2$ =0%, p=0.62) and indicated no publication bias (p-Egger=0.554).

Visual-spatial intelligence score was determined from the WPPSI, WISC, WASI block design, geometric design, and maze subtests [31, 36, 37, 39, 40, 42], the K-ABC simultaneous processing [38, 41], the RAKIT disks and hidden figures [35], and the British Ability Scale II (BAS-II) spatial ability [28] subtests. There were no discernible differences of visual-spatial intelligence score between ART and NC schoolers, as indicated by total (p=0.14) and subgroups analyses (p ART = 0.07; p ICSI = 0.53) (Fig. 5G). The data exhibited homogeneity and indicated no publication bias.

Long-term memory retrieval/ learning ability score was obtained from WPPSI, WISC, WASI animal pegs, and zoo location [31, 39, 40], K-ABC learning ability [38], and NEPSY domain memory and learning [32] subtests. ART schoolers exhibited equal learning ability to NC schoolers (p=0.53) (Fig. 5H). The data were homogenous ( $I^2$ =0%, p=0.54), and indicated no publication bias (p-Egger=0.443).

Executive function score was obtained from the ASQ problem-solving [30], the Behavior Rating Inventory of Executive Function (BRIEF) general executive composite [27, 33], and A Developmental NEuroPSYchological Assessment (NEPSY) domain attention and executive function [32] scores. There was no discernible difference in the executive function score between ART and NC schoolers (p=0.37) (Fig. 5I). Significant heterogeneity was noted ( $I^2$ =74%, p=0.010), but there was no evidence of publication bias (p-Egger=0.533).

# Motoric outcome

Total motor score was assessed using the Kauffman ABC Motoric Scale [36, 37, 42], Peabody Development Motor Scale [39], McCarthy Scales of Children's Ability (MSCA) motor scale index [40], Zimmer/Volkamer Motor Test MOT 4–6 [38], and the ASQ fine and gross motor score [30]. ART schoolers had comparable total motor score with NC schoolers (p=0.50), although high heterogeneity was identified ( $I^2$ =75% (p=0.0002) (Fig. 6A), with no evidence of publication bias (p-Egger=0.399).

Only five studies reported the gross and fine motor sub-scores. The gross motor score was meta-analyzed from Kauffman ABC Motoric Scale ball and balance scores [36, 37, 42], the Peabody Development Motor Scale gross motor quotient [39], and the ASQ gross motor score [30]. The fine motor score was meta-analyzed from K-ABC motoric scale manual score [36, 37, 42], Peabody Development Motor Scale fine motor quotient [39], and ASQ fine motor score [30]. There were no differences in the gross and fine motor scores between ART-born and NC schoolers (p=0.72 and 0.25, respectively). Although there were significant heterogeneities ( $I^2$ =82–84%), there were no evidence of publication biases detected (p-Egger > 0.05).

# Behavior and social outcome

In five studies, preschool and primary schoolers' mothers reported internalizing, externalizing, and total behavioral problems by completing Achenbach's Child's Behavior Checklist [22, 29, 32, 33, 39]. Externalizing behavior was also reported in one study using the German behavioral questionnaire for preschoolers, Verhaltensbeurteilungsbogen für Vorschulkinder (VBV), aggressive/ oppositional, hyperactivity, and attention subtests [41]. Pooled analysis indicated that NC schoolers exhibited higher total behavior problems score behavioral issues  $[(p=0.02), I^2=50\% (p=0.05)]$  (Fig. 7A). Internalizing behavior score was not significantly different between the two groups  $[(p=0.06), I^2=0\%, (p=0.44)]$  (Fig. 7B). However, externalizing behavior score was significantly higher in NC schoolers than ART schoolers [(p=0.001, $I^2 = 0\% (p = 0.59)$ ] (Fig. 7C).

In three studies, teachers also reported the behavioral problems using the Teacher Report Form (TRF) [29, 32, 33]. Total behavior (p = 0.64), internalizing behavior (p = 0.61), and externalizing behavior (p = 0.20) were not differ between NC and ART schoolers (Fig. 7D-F). There were moderate data heterogeneities ( $I^2$  = 0–57%) and no evidence of publication bias (p-Egger > 0.05).

Three studies reported social skills based on the ASQ personal-social [30], NEPSY social cognition domain

# A. Preschool and Primary School: Total Motor Score

		ART			NC		:	Std. Mean Difference		Std. Mean Diffe	rence	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	8	IV, Random, 9	5% CI	ABCDEFGHIJKL
Fallah 2013	42.35	10.2	61	46.18	8.4	61	8.4%	-0.41 [-0.77, -0.05]		<u> </u>		
Ponjaert-Kristoffersen 2004	94.7	15.6	293	99.5	14.85	265	14.6%	-0.31 [-0.48, -0.15]				
Ponjaert-Kristoffersen 2005	52.33	8.88	409	53.41	9.11	452	15.8%	-0.12 [-0.25, 0.01]				
Leunens 2008	5.84	4.2	109	6.1	4.3	90	10.7%	-0.06 [-0.34, 0.22]				
Ponjaert-Kristoffersen 2005	52.84	9.77	479	53.41	9.11	452	16.0%	-0.06 [-0.19, 0.07]				
Ludwig 2009	107.12	13.31	276	106.15	14.03	273	14.6%	0.07 [-0.10, 0.24]				
Leunens 2006	1.43	1.89	151	1.24	1.76	153	12.5%	0.10 [-0.12, 0.33]			-	
Winter 2014	9.85	5.21	49	7.03	3.94	48	7.3%	0.60 [0.20, 1.01]			<b>···</b> →	$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \bullet \bullet$
Total (95% CI)			1827			1794	100.0%	-0.05 [-0.19, 0.09]		•		
Heterogeneity: $Tau^2 = 0.03$ ; C Test for overall effect: $Z = 0.6$	hi <sup>2</sup> = 27.7 7 (P = 0.5	(0, df =	7 (P = 0	0.0002);	l² = 75%	Ď			⊢ -1	-0.5 0	0.5 1	
	. (	- /								Higher in NC High	er in AR I	

### B. Preschool and Primary School: Gross Motor Score

		ART			NC		5	Std. Mean Difference	Std. Mean Difference Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% C	I IV, Random, 95% CI A B C D E F G H I J K L
Leunens 2008	1.29	1.4	109	1.76	1.7	90	20.3%	-0.30 [-0.58, -0.02]	
Ponjaert-Kristoffersen 2004	96.1	14.3	293	98.6	14.2	265	23.3%	-0.18 [-0.34, -0.01]	
Fallah 2013	52.2	7.7	61	52.05	7.5	61	18.1%	0.02 [-0.34, 0.37]	
Leunens 2006	0.19	0.58	151	0.13	0.36	153	21.9%	0.12 [-0.10, 0.35]	┿╍╌╴╴
Winter 2014	2.87	2.88	49	1.09	1.63	48	16.4%	0.75 [0.34, 1.17]	
Total (95% CI)			663			617	100.0%	0.05 [-0.23, 0.33]	
Heterogeneity: Tau <sup>2</sup> = 0.08; C	hi² = 22	.25, df	= 4 (P	= 0.000	2); l <sup>2</sup> =	= 82%			
Test for overall effect: Z = 0.3	6 (P = 0	.72)							-1 -0.5 0 0.5 I Higher in NC. Higher in ART
	- (	/							Higher in NC Higher in ART

### C. Preschool and Primary School: Fine Motor Score



(K) Period of follow-up

(L) Adequacy of follow-up of cohort

(L) Adequacy of follow-up of conort

Fig. 6 Motoric outcome ART-born compared to naturally conceived (NC) pre and primary schoolers as assessed using A Total Motor Score, B Gross Motor Score, and C Fine Motor Score

[32], and VBV social skill subtest [41]. The differences of social scores between ART schoolers and NC schoolers are insignificant [(p=0.08),  $I^2=0\%$  (p=0.61)] (Fig. 7D), with no evidence of publication bias (p-Egger=0.611).

# Young adolescent (12–18 years)

# Intelligence outcome

Intelligence in the young adolescent age group was measured from school subject's test scores [43–46]. ART students scored significantly higher than NC students in reading or language (only from native language score) (p=0.00001), although significant heterogeneity was acknowledged ( $l^2 = 94\%$ , p = 0.00001) (Fig. 8A). Similarly, meta-analysis also revealed that ART students scored significantly higher in mathematics (p = 0.00001), although significant heterogeneity was also identified ( $l^2 = 90\%$  (p = 0.0001) (Fig. 8B). Publication bias was detected in the analysis on mathematics score (p-Egger = 0.025), but not in the analysis on language score (p-Egger = 0.104).

# **Behavioral outcome**

The Achenbach Children Behavior Checklist completed by parents and the Achenbach Youth Self-Report were used to measure behavioral outcomes in

# A. Preschool and Primary School: Parents' CBCL Total Behavior Problems Score

	iiiiai	ART		1 41 6	lon-ART		. i otai	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% C	I IV, Random, 95% CI	ABCDEFGHIJKL
6.1.1 ART vs NC										
Heineman 2019	45.9	10.81	46	49.1	9.54	66	4.8%	-0.32 [-0.69, 0.06]		
Heijligers 2018 Subtotal (95% CI)	45.51	9.61	51 97	44.55	8.73	33 99	3.6% 8.5%	0.10 [-0.34, 0.54] -0.12 [-0.53, 0.29]		
Heterogeneity: Tau <sup>2</sup> = 0.04; C	hi² = 2.00	D, df = 1 (	P = 0.1	6); I <sup>2</sup> = 5	50%					
Test for overall effect: Z = 0.5	8 (P = 0.5	56)								
6.1.2 IVE ve NC										
	40.0		07	04.0	44.0	00	0.5%	0.001.0.75.0.001		
Colpin 2002	18.3	11.1	37	21.2	14.6	23	2.5%	-0.23 [-0.75, 0.29]		
Barnes 2004 Subtotal (95% CI)	48.9	9	301	49.9	9.5	310	27.6%	-0.11 [-0.27, 0.05]	-	
Hotorogonoity: $Tou^2 = 0.00$ ; C	$hi^2 = 0.10$	0 df = 1 /	D = 0.6	7). 12 - (	00/	555	50.170	-0.12 [-0.27, 0.00]	-	
Test for overall effect: $Z = 1.5$	2 (P = 0.1)	9, UI – I ( 13)	P - 0.0	(7), I <sup>_</sup> = (	176					
	2 (1 - 0.	13)								
6.1.3 ICSI vs NC										
Ponjaert-Kristoffersen 2004	18.2	13	293	20	13.4	265	25.1%	-0.14 [-0.30, 0.03]		
Barnes 2004	49.1	9.1	345	49.9	9.5	310	29.5%	-0.09 [-0.24, 0.07]		
Knoester 2007	22	18.9293	78	20	18.9293	75	6.9%	0.11 [-0.21, 0.42]		$\bullet \bullet \bullet \bullet \bullet \bullet ? ? ? \bullet \bullet \bullet$
Subtotal (95% CI)			716			650	61.5%	-0.09 [-0.19, 0.02]	◆	
Heterogeneity: Tau <sup>2</sup> = 0.00; C	hi² = 1.7	5, df = 2 (	P = 0.4	2); I <sup>2</sup> = (	)%					
Test for overall effect: Z = 1.5	7 (P = 0.1	12)								
Total (95% CI)			1151			1082	100.0%	-0.10 [-0.18, -0.02]	•	
Heterogeneity: Tau <sup>2</sup> = 0.00; C	chi² = 4.12	2, df = 6 (	P = 0.6	6); I <sup>2</sup> = (	)%				-1 -0.5 0 0.5	
Test for overall effect: Z = 2.3	4 (P = 0.0	02)							Higher in Non-ART Higher in ART	
Test for subgroup differences	: Chi <sup>2</sup> = 0	.14, df =	2 (P =	0.93), I²	= 0%					

# B. Preschool and Primary School: Parents' CBCL Internalizing Behavior Score

		ART			NC		:	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
6.2.1 ART vs NC										
Heineman 2019	46.5	9.39	46	50.7	10.56	66	4.5%	-0.41 [-0.79, -0.03]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \bullet \bullet \bullet \bullet \bullet$
Heijligers 2018	48	10.78	51	46.39	9.32	33	3.4%	0.16 [-0.28, 0.59]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet ? ? \bullet \bullet \bullet$
Subtotal (95% CI)			97			99	7.9%	-0.14 [-0.70, 0.42]		
Heterogeneity: Tau <sup>2</sup> = 0.12; 0	Chi² = 3.6	9, df = 1	(P = 0.	.05); l <sup>2</sup> =	- 73%					
Test for overall effect: Z = 0.4	9 (P = 0.	62)								
6.2.2 IVF vs NC										
Knoester 2007	4	4.6051	70	5.5	4.6051	75	6.1%	-0.32 [-0.65, 0.00]		
Barnes 2004	47.4	10	301	48	10	310	26.0%	-0.06 [-0.22, 0.10]		
Colpin 2002	5.2	3.5	27	5.1	4.7	23	2.1%	0.02 [-0.53, 0.58]		
Subtotal (95% CI)			398			408	34.2%	-0.11 [-0.27, 0.05]	-	
Heterogeneity: Tau <sup>2</sup> = 0.00; 0	Chi² = 2.2	3, df = 2	(P = 0	.33); I² =	= 10%					
Test for overall effect: Z = 1.3	5 (P = 0.	18)								
6.2.3 ICSI vs NC										
Ponjaert-Kristoffersen 2004	4	4	293	4.3	4.1	265	23.7%	-0.07 [-0.24, 0.09]		
Barnes 2004	47.6	9.1	345	48	10	310	27.8%	-0.04 [-0.20, 0.11]		
Knoester 2007	5.5	6.3098	78	5.5	6.3098	75	6.5%	0.00 [-0.32, 0.32]		• • • • • • ? ? ? • • •
Subtotal (95% Cl)			716			650	57.9%	-0.05 [-0.16, 0.06]	•	
Heterogeneity: Tau <sup>2</sup> = 0.00; 0	Chi <sup>2</sup> = 0.1	9, df = 2	(P = 0	.91); l² =	• 0%					
Test for overall effect: Z = 0.9	93 (P = 0.	35)								
									•	
Total (95% Cl)			1211			1157	100.0%	-0.08 [-0.16, 0.00]	•	
Heterogeneity: Tau <sup>2</sup> = 0.00; 0	Chi² = 6.8	6, df = 7	(P = 0	.44); l² =	= 0%				-1 -0.5 0 0.5	1
Test for overall effect: Z = 1.8	87 (P = 0.	.06)							Higher in NC Higher in ART	•
Test for subgroup differences	:: Chi² = (	0.43, df =	= 2 (P =	0.80), I	<sup>2</sup> = 0%				0	

### C. Preschool and Primary School: Parents' CBCL Externalizing Behavior Score

Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI IV, Random, 95% CI A B C D E F G H I J K L	
	_
6.3.1 ART vs NC	
Heineman 2019 44.7 10.76 46 47.6 9.62 66 5.0% -0.29 [-0.66, 0.09] 🛛 😽 🕑 🕑 🕑 🐨 🐨 🖓 🐨 🕒 🐨 🖓	
Heijligers 2018 44.75 9.72 51 43.97 8.66 33 3.8% 0.08 [-0.36, 0.52]	
Subtotal (95% Cl) 97 99 8.8% -0.12 [-0.48, 0.24]	
Heterogeneity: Tau <sup>2</sup> = 0.02; Chi <sup>2</sup> = 1.55, df = 1 (P = 0.21); l <sup>2</sup> = 36%	
Test for overall effect: Z = 0.65 (P = 0.52)	
6.3.2 IVF vs NC	
Colpin 2002 5.9 5.3 27 7.9 6.1 23 2.3% -0.35 [-0.91, 0.21]	
Knoester 2007 6 6.1401 70 7 6.1401 75 6.8% -0.16 [-0.49, 0.16]	
Barnes 2004 49.3 8.6 301 50.4 8.7 310 28.6% -0.13 [-0.29, 0.03] 📑 😯 😌 😌 😌 😌 😌 😌 😌	
Subtotal (95% Cl) 398 408 37.6% -0.15 [-0.29, -0.01]	
Heterogeneity: Tau <sup>2</sup> = 0.00; Chi <sup>2</sup> = 0.56, df = 2 (P = 0.76); l <sup>2</sup> = 0%	
Test for overall effect: Z = 2.08 (P = 0.04)	
6.3.3 [CS] ve NC	
Darliest Club 40.5 / .0 193 40.0 0.7 / 75 1.7 - 0.25 [0.40, -0.05] -	
Salicitez Aluista 2011 5.25 1.07 55 5.1 1.7 57 5.47% 0.06 [0.39, 0.04]	
Nilosteri 2007 6 9.4047 76 7 9.4047 79 7.276 0.11 [0.21, 0.42]	
$\frac{1}{2} \frac{1}{2} \frac{1}$	
$\frac{1}{100} \frac{1}{1000} \frac{1}{1000} \frac{1}{1000} \frac{1}{10000} \frac{1}{10000000000000000000000000000000000$	
rest tot overall effect. Z = 1.57 (r = 0.12)	
Total (95% Cl) 1094 1059 100.0% -0.14 [-0.23, -0.06]	
Heterogeneity: Tau <sup>2</sup> = 0.00; Chi <sup>2</sup> = 6.51, df = 8 (P = 0.59); l <sup>2</sup> = 0%	
Test for overall effect: Z = 3.27 (P = 0.001)	
Test for subgroup differences: Chi <sup>2</sup> = 0.07, df = 2 (P = 0.97), l <sup>2</sup> = 0%	

Fig. 7 Behavior and social outcomes of ART-born compared to naturally conceived (NC) pre and primary schoolers assessed using **A** Parents' CBCL Total Behavior Problems Score, **B** Parents' CBCL Internalizing Behavior Score, **C** Parents' CBCL Externalizing Behavior Score, **D** Teachers' TRF Total Behavior Problems Score, **E** Teachers' TRF Internalizing Behavior Score, **F** Teachers' TRF Externalizing Behavior Score, and **G** Social Score

### D. Preschool and Primary School: Teachers' TRF Total Behavior Problems Score

	А	RT		NC				Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Heijligers 2018	43.93	7.2	51	46.63	7.33	33	37.4%	-0.37 [-0.81, 0.07]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \circ \circ \bullet \bullet \bullet$
Heineman 2019	48.6	8.22	46	50.4	7.79	66	41.6%	-0.22 [-0.60, 0.15]		
Colpin 2002	29.9	17.4	17	19.4	16.2	11	20.9%	0.60 [-0.18, 1.38]		$\mathbf{b} \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \circ \circ \bullet \bullet$
Total (95% CI)			114			110	100.0%	-0.11 [-0.55, 0.34]		
Heterogeneity: Tau <sup>2</sup> = Test for overall effect:	0.08; Chi Z = 0.47 (	<sup>2</sup> = 4.0 (P = 0	64, df = ).64)	= 2 (P =	0.10);	l² = 57	%		-1 -0.5 0 0.5 Higher in NC Higher in ART	1

### E. Preschool and Primary School: Teachers' TRF Internalizing Behavior Score

		ART NC						Std. Mean Difference	Std. Mean Difference Risk of Bias		
	Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% C	IV, Random, 95% CI A B C D E F G H I J K L	
	Heijligers 2018	43.4	8.15	51	46.09	7.66	33	37.5%	-0.33 [-0.78, 0.11]		
	Heineman 2019	49.6	8.48	46	49.4	9.01	66	48.3%	0.02 [-0.35, 0.40]		
	Colpin 2002	9	8.7	17	7	4.9	11	14.2%	0.26 [-0.50, 1.02]		
	Total (95% CI)			114			110	100.0%	-0.08 [-0.37, 0.22]	-	
Heterogeneity: Tau <sup>2</sup> = 0.01; Chi <sup>2</sup> = 2.33, df = 2 (P = 0.31); I <sup>2</sup> = 14%							l² = 14	%			
	Test for overall effect:	Z = 0.51	(P = 0	0.61)						-1 -0.5 0 0.5 I Higher in NC Higher in ART	

### F. Preschool and Primary School: Teachers' TRF Externalizing Behavior Score

	ART NC					:	Std. Mean Difference	Std. Mean Difference	Risk of Bias	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Heijligers 2018	45.68	6.4	51	47.94	6.98	33	37.0%	-0.34 [-0.78, 0.10]		
Heineman 2019	48.9	7.07	46	50.1	7.41	66	50.6%	-0.16 [-0.54, 0.21]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \bullet \bullet \bullet \bullet$
Colpin 2002	6.9	7.5	17	4.9	6.3	11	12.4%	0.27 [-0.49, 1.04]		$\rightarrow$
Total (95% CI)			114			110	100.0%	-0.17 [-0.44, 0.09]	-	
Heterogeneity: $Tau^2 = 0.00$ ; $Chi^2 = 1.86$ , $df = 2$ (P = 0.39); $I^2 = 0\%$									-1 -0.5 0 0.5	
rest for overall effect:	: Z = 1.27 (P = 0.20)								Higher in NC Higher in ART	

### G. Preschool and Primary School: Social Score

		ART			NC		:	Std. Mean Difference	Std. Mean Difference	Risk of Bias	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL	
Heineman 2019	2	7.5376	57	1.5	6.1018	66	38.9%	0.07 [-0.28, 0.43]		$\textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} \textcircled{\begin{tabular}{c}} & & & & \\ \hline \end{tabular} \end{array} \end{array} \\ \hline \end{tabular} \begin{array}{c} & & & \\ \hline \end{tabular} \end{array} \end{array} \\ \hline \end{tabular} \begin{array}{c} & & & \\ \hline \end{tabular} \end{array} \end{array} \\ \hline \end{tabular} \begin{array}{c} & & & \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \begin{array}{c} & & & \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \begin{array}{c} & & \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \begin{array}{c} & & \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \begin{array}{c} & & \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \end{array} \\ \hline \end{tabular} \begin{array}{c} & & \\ \hline \end{tabular} \end{array} \\ \end{tabular} \\ \end{tabular} \end{array} \\ \end{tabular} \end{array} \\ \end{tabular} \end{array} \\ \end{tabular} \end{tabular} \end{array} \\ \end{tabular} \end{array} \\ \end{tabular} \end{array} \\ \end{tabular} \end{tabular} \\ \end{tabular} \\ \end{tabular} \end{array} \\ \end{tabular} \end{tabular} \\ \end{tabular} t$	
Fallah 2013	51.5	7.4	61	49.7	8.4	61	38.6%	0.23 [-0.13, 0.58]			
Sanchez Albisua 2011	5.9	1.9	35	5.2	1.9	37	22.5%	0.36 [-0.10, 0.83]		$\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet ? \bullet \bullet \bullet \bullet$	
Total (95% CI)			153			164	100.0%	0.20 [-0.02, 0.42]	•		
Heterogeneity: Tau <sup>2</sup> = 0.	00; Chi <sup>2</sup>	= 0.99, c	df = 2 (F	P = 0.61	); I <sup>2</sup> = 0%					1	
Test for overall effect: Z	Test for overall effect: Z = 1.75 (P = 0.08) -1 -0.5 0 0.5 1 Higher in NC Higher in ART										
Risk of bias legend											
(A) Representativeness	of the ex	posed co	ohort								
(B) Selection of the non-	exposed	l cohort									
(C) Ascertainment of exp	osure										
(D) Demonstration that of	utcome	of interes	st was r	not pres	ent at sta	art of st	udy				
(E) Comparability of coh	ort on th	e basis o	of perina	atal outo	omes						
(F) Comparability of cohe	ort on the	e basis o	fother	factors							
(G) Assessment of intelli	gence o	utcome									
(H) Assessment of moto	ric outco	me									
(I) Assessment of social	skill										

(J) Assessment of behavioral outcome

(K) Period of follow-up(L) Adequacy of follow-up of cohort

Fig. 7 continued

the young adolescent group [47–49]. No significant differences between ART and NC young adolescents were identified on total behavioral problems [(p=0.20),  $I^2=0\%$  (p=0.58)] (Fig. 9A) and (p=0.59),  $I^2=0\%$  (p=0.33) (Fig. 9D)], internalizing behavior [(p=0.42),  $I^2=55\%$  (p=0.14)] (Fig. 9B) and (p=0.84),  $I^2=28\%$  (p=0.24)] (Fig. 9E), and externalizing behavior [(p=0.11),  $I^2=0\%$  (p=0.80) (Fig. 9C) and (p=0.81),  $I^2=0\%$  (p=0.41) (Fig. 9F)], as reported by parents and the young adolescents themselves respectively.

# Obstetrics and neonatal characteristics

Table 2. shows obstetrics and neonatal characteristics in ART and NC groups. According to the data from all age

groups, babies born after ART typically have lower gestational ages. They also had a 1.58 to 2.34 times higher risk of preterm birth (gestational age < 37 weeks) and 2.44 to 4.48 times higher risk of low birth weight (birth weight < 2500 g).

# Discussion

This meta-analysis acknowledged that verbal IQ is significantly lower in IVF toddlers, but higher in ICSI toddlers, compared to NC toddlers. Furthermore, non-verbal intelligence is significantly lower in ART compared to NC toddlers. There are no discernible differences in all areas of intelligence between ART and NC preschool and primary schoolers. Interestingly,

# A. Young adolescent: Reading or Language Scores

		ART			NC		5	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean SD Total Mean SD				SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Luke 2020	1,566	115.07	3307	1,550	109.48	11985	25.5%	0.14 [0.11, 0.18]		
Spangmose 2017	7.11	2.44	2544	6.72	2.47	4985	24.8%	0.16 [0.11, 0.21]		
Luke 2021	1,701.2	122.88	1165	1,681	82	3979	23.1%	0.22 [0.15, 0.28]	-	
Norrman 2018	14.2	4.4	8232	12.8	4.9	1499667	26.6%	0.29 [0.26, 0.31]		•••••
Total (95% CI)			15248			1520616	100.0%	0.20 [0.12, 0.28]	•	
Heterogeneity: Tau <sup>2</sup> = Test for overall effect:	0.01; Chi <sup>2</sup> Z = 4.82 (	= 52.35, P < 0.000	df = 3 ( 001)	P < 0.00	0001); I²	= 94%			-1 -0.5 0 0.5 Higher in NC Higher in ART	-  1

### **B. Young Adolescent: Mathematics Score**



(B) Selection of the non-exposed cohort

(C) Ascertainment of exposure

(D) Demonstration that outcome of interest was not present at start of study

(E) Comparability of cohort on the basis of perinatal outcomes

(F) Comparability of cohort on the basis of other factors (G) Assessment of intelligence outcome

(H) Assessment of motoric outcome

(I) Assessment of social skill

(J) Assessment of behavioral outcome

(K) Period of follow-up

(L) Adequacy of follow-up of cohort

Fig. 8 Intelligence outcome of ART-born compared to naturally conceived (NC) young adolescents as assessed using A Reading/Language Score and **B** Mathematics Score

meta-analyses showed that ART young adolescents had higher intelligence scores compared to NC young adolescents. Fine motor score in IVF toddlers is significantly lower; nonetheless, there were no differences in the ICSI group or total group analysis compared to NC toddlers. In preschool and primary school groups, no differences were found in total motor, gross motor, and fine motor scores between ART and NC children.

We hypothesize that there are several factors that might affect these outcomes. First, in the toddler group, IVF conception was only reported in 3 studies [19, 22, 23], and 2 of them [19, 23] were reported in 1995 and 1998, respectively. We speculate that changes in protocols in IVF might play roles in determining the children's development. For example, before 2001, there was no preimplantation genetic screening. Improvements in IVF, freeze-thawing, and oocyte retrieval methods have resulted in higher pregnancy and assured higher quality of implanted embryos [53].

While non-verbal intelligence involves parietal lobes and is linked to white matter microstructure, verbal intelligence is related to cortical structure and thickness of the temporal lobes and temporal pole lateral areas. Lower white matter tract integrity has a significant negative impact on general intelligence [54]. The lateral rostral medulla region of the brain stem controls fine motor function [55]. Recent research discovered that single nucleotide polymorphisms have functional effects on neurogenesis, neuronal differentiation, or the structure or activity of synapses [54]. To avoid any genetic defects, the quality of the transferred embryo is crucial in the ART procedure.

However, a study by Zhang et al. [56] revealed that singleton children born following a poor-quality embryo transfer had comparable full-scale, verbal, and performance intelligence as measured with the Weschler Preschool and Primary Scale of Intelligence in comparison to children born following a good-quality transfer. Thus, other factors might have a more significant role in intelligence and motoric ability development.

Second, as shown in Table 2., prematurity and low birth weight were inexplicably more common in ART children. According to a study by Nagy et al. [57], children who were born preterm and those who were underweight at birth performed worse on tests of intelligence and executive function than children who were born full-term, although their results were still within the normal range on average [57]. According to Casey et al., low cortical volume and surface area are related to low birth weight [58]. Advanced imaging techniques revealed that

# A. Young adolescent: Parents' CBCL Total Behavior Problems Score

	ART NC				Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean SE	Total Mea	n SD Tota	I Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Colpin 2008	14.71 12.7	24 18.7	5 13.73 2	13.7%	-0.30 [-0.90, 0.30]		
Wagenaar 2009	48.92 9.36	130 50.0	9 10.33 14	86.3%	-0.12 [-0.36, 0.12]		$\bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \circ \bullet \bullet \bullet \bullet$
Total (95% CI)	154 162				-0.14 [-0.36, 0.08]	•	
Heterogeneity: Tau <sup>2</sup> = Test for overall effect:	0.00; Chi² = 0 Z = 1.27 (P =	0.31, df = 1 (P 0.20)	= 0.58); l <sup>2</sup> = 0%			-1 -0.5 0 0.5 Higher in NC Higher in ART	1 1

### B. Young Adolescent: Parents' CBCL Internalizing Behavior Score

-	AF	RT NC				-	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Colpin 2008	7.25 7	.33 24	11.26	7.78	19	33.4%	-0.52 [-1.14, 0.09]		<b>•</b> • • • • • ? ? ? • • •
Wagenaar 2009	51.55 9	.99 130	51.77	10	142	66.6%	-0.02 [-0.26, 0.22]		$\bullet \bullet \bullet \bullet \bullet \bullet ? ? ? \bullet \bullet \bullet$
Total (95% CI)		154			161	100.0%	-0.19 [-0.65, 0.27]		
Heterogeneity: Tau <sup>2</sup> = 0.07; Chi <sup>2</sup> = 2.23, df = 1 (P = 0.14); I <sup>2</sup> = 55%							-		1
Test for overall effect: Z = 0.80 (P = 0.42)							-1	Higher in NC Higher in ART	1

# C. Young Adolescent: Parents' CBCL Externalizing Behavior Score

	ART	RT NC			:	Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean SD	Total M	lean SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Colpin 2008	5.29 5.77	24 6	6.75 5.62	20	13.8%	-0.25 [-0.85, 0.34]		
Wagenaar 2009	46.92 8.77	130 48	8.52 9.95	142	86.2%	-0.17 [-0.41, 0.07]		$\bullet \bullet \bullet \bullet \bullet \bullet \circ ? ? ? \bullet \bullet \bullet$
Total (95% CI)	154 162				100.0%	-0.18 [-0.40, 0.04]	•	
Heterogeneity: Tau <sup>2</sup> =	0.00; Chi <sup>2</sup> = 0	.06, df = 1	(P = 0.80);	$ ^2 = 0\%$	5			4
Test for overall effect: Z = 1.60 (P = 0.11)							-1 -0.5 0 0.5 1 Higher in NC Higher in ART	

### D. Young adolescent: Youth Self-Report Total Behavior Problems Score

-	ART NC						Std. Mean Difference	Std. Mean Difference	Risk of Bias	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Colpin 2008	29.96	20.91	24	37	18.77	19	18.6%	-0.35 [-0.95, 0.26]		<b>•</b> • • • • • ? ? ? • • •
Wagenaar 2011	48.33	8.55	86	48.41	9.53	97	81.4%	-0.01 [-0.30, 0.28]		$\bullet \bullet \bullet \bullet \bullet \bullet \circ \circ \circ \circ \bullet \bullet \bullet \bullet$
Total (95% CI)	110 116						100.0%	-0.07 [-0.33, 0.19]	-	
Heterogeneity: Tau <sup>2</sup> = 0.00; Chi <sup>2</sup> = 0.96, df = 1 (P = 0.33); I <sup>2</sup> = 0% Test for overall effect: Z = 0.54 (P = 0.59)									-1 -0.5 0 0.5 Higher in NC Higher in ART	H 1

# E. Young Adolescent: Youth Self-Report Internalizing Behavior Score

0								<u>v</u>		
		ART	RT NC					Std. Mean Difference	Std. Mean Difference	Risk of Bias
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	IV, Random, 95% CI	ABCDEFGHIJKL
Colpin 2008	4.13	3.99	24	5.85	6.4	20	27.7%	-0.32 [-0.92, 0.27]		
Wagenaar 2011	48.47	9.78	86	47.73	9.69	97	72.3%	0.08 [-0.21, 0.37]		$\bullet \bullet \bullet \bullet \bullet \bullet \circ \circ$
Total (95% CI)			110			117	100.0%	-0.04 [-0.39, 0.32]		
Heterogeneity: Tau <sup>2</sup> =	= 0.02; Chi <sup>2</sup> = 1.39, df = 1 (P = 0.24); l <sup>2</sup> = 28 <sup>6</sup>						%			1
Test for overall effect:	verall effect: $Z = 0.20$ (P = 0.84)					-1 -0.5 0 0.5				
					Higher in NC Higher in ART					

### F. Young Adolescent: Youth Self-Report Externalizing Behavior Score

-	ART NC				Std. Mean Difference			Std. Mean Difference	Risk of Bias		
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% C	1	IV, Random, 95% CI	ABCDEFGHIJKL
Colpin 2002	9.75	7.4	24	11.79	8.03	19	18.7%	-0.26 [-0.87, 0.34]	_		<b>•</b> • • • • • ? ? ? • • • •
Wagenaar 2011	48.8	8.14	86	48.62	9.53	97	81.3%	0.02 [-0.27, 0.31]			$\bullet \bullet \bullet \bullet \bullet \bullet ??? \bullet \bullet \bullet$
Total (95% CI)			110			116	100.0%	-0.03 [-0.29, 0.23]			
Heterogeneity: Tau <sup>2</sup> =	0.00; Ch	$ni^2 = 0.$	67, df =	= 1 (P =	0.41);	$ ^2 = 0\%$	D		<u> </u>		—!
Test for overall effect:	7 = 0.24	(P = 0)	0.81)						-1	-0.5 0 0.5	1
		(								Higher in NC Higher in AR	
Risk of bias legend											

- (A) Representativeness of the exposed cohort
- (B) Selection of the non-exposed cohort
- (C) Ascertainment of exposure
- (D) Demonstration that outcome of interest was not present at start of study
- (E) Comparability of cohort on the basis of perinatal outcomes
- (F) Comparability of cohort on the basis of other factors
- (G) Assessment of intelligence outcome
- (H) Assessment of motoric outcome
- (I) Assessment of social skill
- (J) Assessment of behavioral outcome
- (K) Period of follow-up
- (L) Adequacy of follow-up of cohort

Fig. 9 Behavior outcome of ART-born compared to naturally conceived (NC) young adolescents as assessed using **A** Parents' CBCL Total Behavior Problems Score, **B** Parents' CBCL Internalizing Behavior Score, **C** Parents' CBCL externalizing Behavior Score, **D** Youth Self-report Total Behavior Problems Score, **E** Youth Self-report Internalizing Behavior Score, and **F** Youth Self-report Externalizing Behavior Score

Characteristics	Age Group	Reporting studies	No. of Children (cases / total)			Effect Size, <i>p</i> - value	Heterogeneity (I <sup>2</sup> ), <i>p</i> -value
			ICSI	All ART	Control		
Gestational Age [mean $\pm$ st. dev	Toddler	14, 15, 19, 21		38.33 ± 2.19 (422)	38.98 ± 1.88 (2,371)	-0.57 [-0.81, -0.34], < <b>0.0001</b>	34%, 0.21
(no. of children)]	Pre-school	35 – 39, 42	39.56 ± 1.47 (971)		39.61 <u>+</u> 1.55 (915)	-0.04 [-0.19, 0.10], 0.55	0%, 0.58
		27, 32, 33		39.79 <u>+</u> 1.32	40.19 <u>+</u> 1.39	-0.40 [-0.64, -0.15], <b>0.002</b>	0%, 0.74
		Summary		39.62 <u>+</u> 1.43 (1,139)	39.75 <u>+</u> 1.51 (2,539)	-0.13 [-0.25, -0.01] <b>, 0.04</b>	22%, 0.25
	Young Adoles- cent	48, 49		38.93 <u>+</u> 2.48 (225)	39.60 <u>+</u> 1.80 (240)	-0.67 [-1.07, -0.28], <b>0.00009</b>	0%, 0.85
Preterm birth (Gestational age	Toddler	14, 23	10/76		13/200	2.11 [0.95, 4.66], <b>0.04</b>	0%, 0.94
< 37 weeks)		16, 18		26/165	695/10,661	2.23 [1.02, 4.87], <b>0.04</b>	57%, 0.13
		Summary		36/341	708/ 10,861	2.34 [1.65, 3.33], <b>&lt;0.00001</b>	0%, 0.46
	Pre-school	31,32, 34-37, 39, 42		62/852	38/825	1.58 [1.07, 2.32], <b>0.02</b>	57%, 0.02
	Young Adoles- cent	44 – 46, 48, 49		1,236/ 12,484	81,341/ 1,507,453	1.90 [1.80, 2.01], <b>&lt;0.00001</b>	92%, <0.00001
Birthweight [mean $\pm$ st. dev (no. of children)]	Toddler	14, 19	3,073.65 <u>+</u> 608.15 (76)		3,129.3 <u>+</u> 554.40 (200)	-55.65 [-220.91, 109.60], 0.51	0%, 0.96
		15, 18, 21		3,304.89 ± 655.84 (412)	3,455.56 ± 601.44 (2,258)	-150.48 [-275.36, -25.60], <b>0.02</b>	43%, 0.17
		Summary		3,255.96 <u>+</u> 640. 54 (488)	3,359.01 ± 557.51 (2,458)	-103.00 [-167.48, -38.53], <b>0.002</b>	0%, 0.42
	Pre-school	31, 35 – 39, 42	2,791.99 <u>+</u> 558.17 (999)		3,414.624 <u>+</u> 529.12 (947)	-621.97 [-672.43, -571.51], <b>&lt;0.00001</b>	100%, <0.00001
		27, 32, 33, 35		3,358.63 <u>+</u> 595.67 (251)	3,572.94 <u>+</u> 531.94 (1,763)	-213.61 [-303.09, -124.14] <b>&lt;0.00001</b>	0%, 0.93
		Summary		2,911.08 ± 558.76 (1,167)	3,447.96 <u>+</u> 525.37 (2,625)	-538.10 [-583.05, -493.14], <b>&lt;0.00001</b>	100%, <0.00001
	Young Adoles- cent	48, 49		3,254.18 <u>+</u> 633.29 (225)	3,413.05 <u>+</u> 476.93 (240)	-158.89 [-261.37, -56.41], <b>0.002</b>	0%, 0.45
Low Birthweight (Birthweight <u>&lt;</u>	Toddler	14, 16	20/140		572/ 10,721	2.44 [1.57, 3.79], <b>&lt;0.0001</b>	0%, 0.56
2,500g)	Pre-school	30, 32, 34, 35		29/275	7/297	4.48 [1.99, 10.09], <b>0.0003</b>	0%, 0.90
	Young Adoles- cent	45, 46, 48, 49		956/ 11,349	31,653 / 1,502,580	3.40 [3.18, 3.64], < <b>0.00001</b>	95%, <0.00001

# Table 2 Pooled analysis of obstetric and neonatal characteristics

the sensory-motor pathway matured more quickly in preterm infants; however, areas of injury and disturbed development are also visible in their parietal white matter. The corpus callosum left inferior longitudinal fasciculus, and left dorsal visual stream mature more slowly in preterm infants. However, if the infant is healthy, these areas will eventually develop more quickly [59]. Lastly, external factors might also contribute, especially to children's intelligence. According to the findings, the development of ART children at later stages of life is arguably superior to that of NC children. Since this study only included singletons, ART children were probably the first to be born and may have had fewer siblings. Additionally, the likelihood of their parents cohabiting, remaining married, being employed, and having higher socioeconomic, occupational, and educational levels [43–46] helped to improve early cognitive stimulation, which impacts academic performance.

According to parental reports, NC children in the toddler, preschool, and primary school age groups had more behavioral issues.

In contrast, according to their teachers, there were no discernible differences. There were no differences between the young adolescent group's self-reports and those of their parents. As it solely depends on parents' perceptions regarding the question related to their children's behavior, this self-reporting questionnaire method may introduce potential methodological bias.

Lower birth weight, which is more common in ART children, had a significant impact on limbic network connectivity, which is in charge of emotion regulation and internally generated thoughts [60]. However, since all of the children in these studies had scores within normal ranges, we surmise that their limbic development was normal based on the most recent results. The influence of parenting factors on a child's externalizing and internalizing behavior may be more significant. Parenting stress impact externalizing behavior, whereas parenting negative engagement impact internalizing behavior [61]. Compared to naturally fertile mothers, ART mothers express more warmth and positive feelings toward their children and greater parental competence [26, 62, 63]. These results may indicate a tendency to report socially acceptable responses, given that those behavior problems were assessed using a self-reported questionnaire [64].

# Limitations

The evidence is arguably weak because the current study is a systematic review based on a limited number of studies. A type II statistical error or false negative may result from a small sample size. This occurs when the null hypothesis—which claims no differences between the two groups being compared—is incorrect but still accepted [65]. The second drawback stems from the fact that the analyzed studies used a variety of instruments with various scales, resulting in the evaluation of distinct areas of motoric and intellectual development. This restriction may have introduced bias due to heterogeneity. Subtest categorization and standardized mean differences based on tested theory can overcome this drawback. Third, the widely used method for evaluating children's behavioral issues is based on self-reports, which may have information bias. Fourth, the included studies did not mention any additional pediatric medical conditions that might impact the results of their neurodevelopmental studies. For instance, none of the studies mentioned bronchopulmonary dysplasia, a condition frequently associated with brain abnormalities in very preterm infants [66].

# Conclusion

This meta-analysis identified differences on certain aspects of intelligence between ART and NC children. The non-verbal intelligence score of ART toddlers was significantly lower than that of NC toddlers; however, preschool and primary school ART children showed comparable results in all areas of intelligence compared to their NC counterparts. Interestingly, ART young adolescents scored significantly higher academic scores than NC young adolescents. ART toddlers had significantly lower fine motor skills. Parents of naturally born toddlers and school-age children reported more overall behavioral problems. However, behavior scores of young adolescents from both groups were comparable. These results may be influenced by both internal and external variables, including the year of ART procedures, prevalence of prematurity and low birth weight, family socioeconomic background, and parenting style.

# Abbreviations

ART	Assisted reproductive treatment
ASQ	Ages and Stages Questionnaire
BAS	British Ability Scale
BRIEF	Behavior Rating Inventory of Executive Function
CBCL	Child's Behaviour Check List
CHC	Cattell, Horn, and Carroll
ICSI	Intracytoplasmic sperm injection
IQ	Intelligence quotient
IVF	In vitro fertilization
K-ABC	Kauffman's Assessment Battery for Children
MDI	Mental Development Index
MSCA	McCarthy Scales of Childrens Ability
NC	Naturally conceived
NEPSY	A Developmental NEuroPSYchological Assessment
PDI	Psychomotor Development Index
PIQ	Performance intelligence quotient
RAKIT	Revised Amsterdam Child Intelligence Test
REEL	Receptive Expressive Emergent Language
SMD	Standardized mean difference
VBV	Verhaltensbeurteilungsbogen für Vorschulkinder
VIQ	Verbal intelligence quotient
WASI	Weschler Abbreviated Scale of Intelligence
WISC	Weschler Intelligence Scale for Children
WPPSI	Weschler Preschool and Primary Scale of Intelligence

# **Supplementary Information**

The online version contains supplementary material available at https://doi. org/10.1186/s11689-023-09490-0.

Additional file 1: Supplement Table 1. Newcastle-Ottawa Scale for Cohort Studies: Intelligence, School Performance, Language Development. Supplement Table 2. Newcastle-Ottawa Scale for Cohort Studies: Motoric Development. Supplement Table 3. Newcastle-Ottawa Scale of Cohort Studies: Behavioral and Social Development. Supplement Table 4. Newcastle Ottawa Scale for the Case-Control Studies. Supplement Table 5. Characteristic of the Included Studies. Supplement Table 6. Summary of Meta-analysis.

# Acknowledgements

Not applicable.

# Authors' contributions

TD, WP, MD, THA, and DH conceptualized and designed the study. TD, JKA, and DH screened the literature, assessed study eligibility and quality, and analyzed the data. JKA and DH did the statistical analysis. TD, MD, THA, and DH directed the discussion. TD, JKA, WP, THA, MD, and DH reviewed and revised the manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

### Funding

This research did not receive any grants from funding agencies in the public, commercial, or not-for-profit sectors.

# Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

# Declarations

**Ethics approval and consent to participate** Not applicable.

## **Consent for publication**

Net available

Not applicable.

### **Competing interests**

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Obstetrics and Gynecology, Faculty of Medicine, Universitas Padjadjaran/Dr. Hasan Sadikin General Hospital, Bandung, West Java, Indonesia. <sup>2</sup>Bandung Fertility Center, Limijati Mother and Child Hospital, Bandung, West Java, Indonesia. <sup>3</sup>Research Center for Medical Genetics, Faculty of Medicine, Universitas Padjadjaran, Bandung, West Java, Indonesia. <sup>4</sup>Department of Basic Medical Science, Faculty of Medicine, Universitas Padjadjaran, Bandung, West Java, Indonesia. <sup>5</sup>Department of Child Health, Faculty of Medicine, Universitas Padjadjaran/Dr. Hasan Sadikin General Hospital, West Java, Bandung, Indonesia.

# Received: 31 May 2022 Accepted: 27 June 2023 Published online: 22 August 2023

### References

- Chambers GM, Dyer S, Zegers-Hochschild F, de Mouzon J, Ishihara O, Banker M, et al. International committee for monitoring assisted reproductive technologies world report: assisted reproductive technology, 2014. Hum Rep. 2011;36(11):2921–34. https://doi.org/10.1093/humrep/ deab198.
- Palermo GD, O'Neill CL, Chow S, Cheung S, Parrella A, Pereira N, Rosenwaks Z. Intracytoplasmic sperm injection: state of the art in humans. Reproduction. 2017;154(6):F93–110. https://doi.org/10.1530/REP-17-0374.

- Olivennes F, Schenider Z, Remy V, Blanchet V, Kerbrat V, Fanchin R, et al. Perinatal outcome and follow-up of 82 children aged 1–9 conceived from cryopreserved embryos. Hum Rep. 1996;11(7):1565–8.
- Helmerhorst FM, Perquin DAM, Donker D, Keirse MJNC. Perinatal outcome of singletons and twins after assisted conception: a systematic review of controlled studies. BMJ. 2004;328:261.
- Ludwig AK, Sutcliffe AG, Diedrich K, Ludwig M. Postneonatal 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.
- Marlow N, Rose AS, Rands CE, Draper ES. Neuropsychological and educational problems at school age associated with neonatal encephalopathy. Arch Dis Child Fetal Neonatal Ed. 2005;90(5):F380–7.
- Bay B, Mortensen EL, Kesmodel US. Assisted reproduction and child neurodevelopmental outcomes: a systematic review. Fertil Steril. 2013;100(3):844–53. https://doi.org/10.1016/j.fertnstert.2013.05.034.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.
- Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-analyses. 2014. http://www.ohri.ca/programs/ clinical\_epidemiology/oxford.asp.
- Colson ER, Dworkin PH. Toddler development. Pediatr Rev. 1997;18(8):255–9.
- United Nations Educational, Scientific, and Cultural Organization. International standard classification of education. 2011. Available at: https://web. archive.org/web/20130124032233/http://www.uis.unesco.org/Educa tion/Documents/isced-2011-en.pdf. Accessed 30 Sept 2021.
- Ichise R. An analysis of the CHC model for comparing cognitive architectures. Proc Comp Sci. 2016;88:239–44. https://doi.org/10.1016/j.procs. 2016.07.431.
- Levac DE, Huber ME, Sternad D. Learning and transfer of complex motor skills in virtual reality: a perspective review. J Neuroeng Rehabil. 2019;16(1):121. https://doi.org/10.1186/s12984-019-0587-8.
- Willner CJ, Gatzke-Kopp LM, Bray BC. The dynamics of internalizing and externalizing comorbidity across the early school years. Dev Psychopathol. 2016;28(4 Pt1):1033–52.
- Diamond A. Executive functions. Ann Rev Psychol. 2013;64:135–68. https://doi.org/10.1146/annurev-psych-113011-143750.
- Agarwal P, Loh SK, Lim SB, et al. Two-year neurodevelopmental outcome in children conceived by intracytoplasmic sperm injection: prospective cohort study. BJOG. 2005;112(10):1376–83. https://doi.org/10.1111/j. 1471-0528.2005.00663.x.
- Balayla J, Sheehy O, Fraser WD, et al. Neurodevelopmental outcomes after assisted reproductive technologies. Obstet Gynecol. 2017;129(2):265–72. https://doi.org/10.1097/AOG.00000000001837.
- Carson C, Kurinczuk JJ, Sacker A, et al. Cognitive development following ART: effect of choice of comparison group, confounding and mediating factors. Hum Reprod. 2010;25(1):244–52. https://doi.org/10.1093/ humrep/dep344.
- Gibson FL, Ungerer JA, Leslie GI, Saunders DM, Tennant CC. Development, behaviour and temperament: a prospective study of infants conceived through in-vitro fertilization. Hum Reprod. 1998;13(6):1727–32. https:// doi.org/10.1093/humrep/13.6.1727.
- Jongbloed-Pereboom M, Middelburg KJ, Heineman MJ, Bos AF, Hadders-Algra M. The Groningen assisted reproductive technologies cohort study: developmental status and behavior at 2 years. Fertil Steril. 2011;95(7):2283–9. https://doi.org/10.1016/j.fertnstert.2011.03.034.
- Nekkebroeck J, Bonduelle M, Desmyttere S, Van den Broeck W, Ponjaert-Kristoffersen I. Socio-emotional and language development of 2-year-old children born after PGD/PGS, and parental well-being. Hum Reprod. 2008;23(8):1849–57. https://doi.org/10.1093/humrep/den179.
- Place I, Englert Y. A prospective longitudinal study of the physical, psychomotor, and intellectual development of singleton children up to 5 years who were conceived by intracytoplasmic sperm injection compared with children conceived spontaneously and by in vitro fertilization. Fertil Steril. 2003;80(6):1388–97. https://doi.org/10.1016/j.fertn stert.2003.06.004.

- Sutcliffe AG, D'Souza SW, Cadman J, Richards B, McKinlay IA, Lieberman B. Outcome in children from cryopreserved embryos. Arch Dis Child. 1995;72(4):290–3. https://doi.org/10.1136/adc.72.4.290.
- Sutcliffe AG, Saunders K, McLachlan R, et al. A retrospective case-control study of developmental and other outcomes in a cohort of Australian children conceived by intracytoplasmic sperm injection compared with a similar group in the United Kingdom. Fertil Steril. 2003;79(3):512–6. https://doi.org/10.1016/s0015-0282(02)04701-5.
- Vo MT, Le TMC, Le TQ, Do DV, Ngo MX. Comparison of psychomotor development among children conceived through icsi in-vitro-fertilisation and naturally at 5 through 30 months of age, Vietnam. Eur J Obstet Gynecol Reprod Biol. 2021;258:157–61. https://doi.org/10.1016/j.ejogrb.2020.12.027.
- Barnes J, Sutcliffe AG, Kristoffersen I, et al. The influence of assisted reproduction on family functioning and children's socio-emotional development: results from a European study. Hum Reprod. 2004;19(6):1480–7. https://doi.org/10.1093/humrep/deh239.
- Bay B, Mortensen EL, Kesmodel US. Fertility treatment and child intelligence, attention, and executive functions in 5-year-old singletons: a cohort study. BJOG. 2014;121(13):1642–51. https://doi.org/10.1111/1471-0528.12907.
- Carson C, Kelly Y, Kurinczuk JJ, Sacker A, Redshaw M, Quigley MA. Effect of pregnancy planning and fertility treatment on cognitive outcomes in children at ages 3 and 5: longitudinal cohort study. BMJ. 2011;343:d4473. https://doi.org/10.1136/bmj.d4473.
- Colpin H, Soenen S. Parenting and psychosocial development of IVF children: a follow-up study. Hum Reprod. 2002;17(4):1116–23. https://doi. org/10.1093/humrep/17.4.1116.
- Fallah R, Akhavan Karbasi S, Galalian MT, Dehghani-Firouzabadi R. Comparison of developmental status of 5-year-old singleton children born through assisted and natural conceptions. Iran J Reprod Med. 2013;11(5):365–70.
- Faramarzi M, Golsorkhtabaramiri M, Esmaeilzadeh S, Ghofrani F, Sorkhi H. Are children born through Intra-Cytoplasmic Sperm Injection (ICSI) have a lower intelligence quotient? Mid East Fert Soc J. 2015;21:16–21. https:// doi.org/10.1016/j.mefs.2015.06.001.
- 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 [published correction appears in Hum Reprod. 2020 January 1;35(1):256]. Hum Reprod. 2019;34(11):2193–200. https://doi.org/10.1093/ humrep/dez202.
- Heijligers M, Verheijden LMM, Jonkman LM, et al. The cognitive and socio-emotional development of 5-year-old children born after PGD. Hum Reprod. 2018;33(11):2150–7. https://doi.org/10.1093/humrep/ dev302.
- Knoester M, Helmerhorst FM, van der Westerlaken LA, Walther FJ, Veen S, 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(12):3098–107. https://doi.org/10.1093/humrep/dem261.
- Knoester M, Helmerhorst FM, Vandenbroucke JP, et al. Cognitive development of singletons born after intracytoplasmic sperm injection compared with in vitro fertilization and natural conception. Fertil Steril. 2008;90(2):289–96. https://doi.org/10.1016/j.fertnstert.2007.06.090.
- Leunens L, Celestin-Westreich S, Bonduelle M, Liebaers I, Ponjaert-Kristoffersen I. Cognitive and motor development of 8-year-old children born after ICSI compared to spontaneously conceived children. Hum Reprod. 2006;21(11):2922–9. https://doi.org/10.1093/humrep/del266.
- Leunens L, Celestin-Westreich S, Bonduelle M, Liebaers I, Ponjaert-Kristoffersen I. Follow-up of cognitive and motor development of 10-yearold singleton children born after ICSI compared with spontaneously conceived children. Hum Reprod. 2008;23(1):105–11. https://doi.org/10. 1093/humrep/dem257.
- Ludwig A, Katalinic A, Thyen U, Sutcliffe AG, Diedrich K, Ludwig M. Neuromotor development and mental health at 5.5 years of age of singletons born at term after intracytoplasmatic sperm injection ICSI: results of a prospective controlled single-blinded study in Germany. Fertil Steril. 2009;91(1):125–32. https://doi.org/10.1016/j.fertnstert.2007.11.030.
- Ponjaert-Kristoffersen I, Tjus T, Nekkebroeck J, et al. Psychological followup study of 5-year-old ICSI children. Hum Reprod. 2004;19(12):2791–7. https://doi.org/10.1093/humrep/deh511.

- Ponjaert-Kristoffersen I, Bonduelle M, Barnes J, et al. International collaborative study of intracytoplasmic sperm injection-conceived, in vitro fertilization-conceived, and naturally conceived 5-year-old child outcomes: cognitive and motor assessments. Pediatrics. 2005;115(3):e283–9. https://doi.org/10.1542/peds.2004-1445.
- Sanchez-Albisua I, Lidzba K, Borell-Kost S, Mau-Holzmann UA, Licht P, Krägeloh-Mann I. Medical, psychological and intellectual development of 5-year-old children born after intracytoplasmic sperm injection. Neuropediatrics. 2011;42(3):104–9. https://doi.org/10.1055/s-0031-1280796.
- Winter C, Van Acker F, Bonduelle M, Desmyttere S, De Schrijver F, Nekkebroeck J. Cognitive and psychomotor development of 5- to 6-year-old singletons born after PGD: a prospective case-controlled matched study. Hum Reprod. 2014;29(9):1968–77. https://doi.org/10.1093/humrep/ deu165.
- Luke B, Brown MB, Ethen MK, et al. Third grade academic achievement among children conceived with the use of in vitro fertilization: a population-based study in Texas. Fertil Steril. 2020;113(6):1242-1250.e4. https://doi.org/10.1016/j.fertnstert.2020.01.015.
- Luke B, Brown MB, Ethen MK, et al. Sixth grade academic achievement among children conceived with IVF: a population-based study in Texas, USA. J Assist Reprod Genet. 2021;38(6):1481–92. https://doi.org/10.1007/ s10815-021-02170-9.
- Norrman E, Petzold M, Bergh C, Wennerholm UB. School performance in singletons born after assisted reproductive technology. Hum Reprod. 2018;33(10):1948–59. https://doi.org/10.1093/humrep/dey273.
- Spangmose AL, Malchau SS, Schmidt L, et al. Academic performance in adolescents born after ART-a nationwide registry-based cohort study. Hum Reprod. 2017;32(2):447–56. https://doi.org/10.1093/humrep/ dew334.
- Colpin H, Bossaert G. Adolescents conceived by IVF: parenting and psychosocial adjustment. Hum Reprod. 2008;23(12):2724–30. https://doi.org/ 10.1093/humrep/den297.
- 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. 2009;92(6):1907–14. https://doi.org/10.1016/j.fertnstert.2008.09. 026.
- Wagenaar K, van Weissenbruch MM, van Leeuwen FE, et al. Self-reported behavioral and socioemotional functioning of 11- to 18-year-old adolescents conceived by in vitro fertilization. Fertil Steril. 2011;95(2):611–6. https://doi.org/10.1016/j.fertnstert.2010.04.076.
- Keith T, Fine J, Taub G, Reynolds M, Kranzler J. Higher order, multi sample, confirmatory factor analysis of the Weschler Intelligence Scale for Children-Fourth Edition: what does it measure? School Psychol Rev. 2006;35(1):108–27.
- Gallagher SL, Sullivan AL. Chapter 30: Kaufman assessment battery for children. In: Davis A, editor. Handbook of pediactric neuropsychology. 2nd ed. New York: Springer; 2011. p. 343–52.
- te Nijenhuis J, Tolboom E, Resing W, Bleichrodt N. Does cultural background influence the intellectual performance of children from immigrant groups? Eur J Psychol Assess. 2004;20(1):10–26. https://doi.org/10. 1027/1015-5759.20.1.10.
- Wang J, Sauer MV. In vitro fertilization (IVF): a review of 3 decades of clinical innovation and technological advancement. Ther Clin Risk Manag. 2006;2(4):355–64. https://doi.org/10.2147/tcrm.2006/2.4.355.
- Goriounova NA, Mansvelder HD. Genes, cells and brain areas of intelligence. Front Hum Neurosci. 2019;13:44. https://doi.org/10.3389/fnhum. 2019.00044.
- Ruder L, Schina R, Knoida H, Valencia-Garcia S, Pivetta C, Arber S. A functional map for diverse forelimb actions with brain stem circuitry. Nature. 2021;590:445–50. https://doi.org/10.1038/s41586-020-03080-z.
- 56. Zhang CX, Xue JL, Zhao W, et al. Embryo morphologic quality in relation to the metabolic and cognitive development of singletons conceived by in vitro fertilization and intracytoplasmic sperm injection: a matched cohort study. Am J Obstet Gynecol. 2022;227(3):479.e1-479.e23. https:// doi.org/10.1016/j.ajog.2022.05.019.
- Nagy A, Kalmár M, Beke AM, Gráf R, Horváth E. Intelligence and executive function of school-age preterm children in function of birth weight and perinatal complication. Appl Neuropsychol Child. 2022;11(3):400–11. https://doi.org/10.1080/21622965.2020.1866571.

- Casey KF, Levesque ML, Szyf M, Ismaylova E, Verner M, Suderman M, et al. Birth weight discordance, DNA methylation, and cortical morphology of adolescent monozygotic twins. Hum Brain Mapp. 2017;38(4):2037–50. https://doi.org/10.1002/hbm.23503.
- Ream MA, Lehwald L. Neurologic consequences of preterm birth. Curr Neurol Neurosci Rep. 2018;18:48. https://doi.org/10.1007/ s11910-018-0862-2.
- Hayward DA, Pomares F, Casey KF, Ismaylova E, Levesque M, Greenlaw K, Vitaro F, et al. Birth weight is associated with adolescent brain development: a multimodal imaging study in monozygotic twins. Hum Brain Mapp. 2020;41(18):5228–39. https://doi.org/10.1002/hbm. 25188.
- Ding N, Gao H, Jiang J, Zhai M, Shao H, Fu L, et al. The characteristics and factors of the internalizing and externalizing behaviors of children at high risk for autism spectrum disorder. BMC Psychiatry. 2021;21:523. https:// doi.org/10.1186/s12888-021-03479-6.
- 62. Golombok S, Cook R, Bish A, Murray C. Families created by the new reproductive technologies: quality of parenting and social and emotional development of the children. Child Dev. 1995;66(2):285–98. https://doi.org/10.1111/j.1467-8624.1995.tb00871.x.
- van Balen F. Child-rearing following in vitro fertilization. J Child Psychol Psychiatry. 1996;37(6):687–93. https://doi.org/10.1111/j.1469-7610.1996. tb01460.x.
- Colpin H, Demyttenaere K, Vandemeulebroecke L. New reproductive technology and the family: the parent-child relationship following in vitro fertilization. J Child Psychol Psychiatry. 1995;36(8):1429–41. https://doi. org/10.1111/j.1469-7610.1995.tb01673.x.
- Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochem Med (Zagreb). 2021;31(1):010502. https://doi.org/10.11613/BM.2021.010502.
- Twilhaar ES, Wade RM, de Kieviet JF, van Goudoever JB, van Elburg RM, Oosterlaan J. Cognitive outcomes of children born extremely or very preterm since the 1990s and associated risk factors: a meta-analysis and meta-regression. JAMA Pediatr. 2018;172(4):361–7. https://doi.org/10. 1001/jamapediatrics.2017.5323.

# **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

### At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

