PLOS ONE

RESEARCH ARTICLE

Use and validity of child neurodevelopment outcome measures in studies on prenatal exposure to psychotropic and analgesic medications – A systematic review

Sarah Hjorth[1]*, Rebecca Bromley[2,3], Eivind Ystrom[1,4,5], Angela Lupattelli[1], Olav Spigset[6,7], Hedvig Nordeng[1,8]

1 PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway, **2** Division of Evolution and Genomic Science, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, England, **3** Royal Manchester Children's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester, England, **4** Department of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway, **5** PROMENTA Research Center, Department of Psychology, University of Oslo, Oslo, Norway, **6** Department of Clinical Pharmacology, St. Olav's University Hospital, Trondheim, Norway, **7** Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway, **8** Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway

* s.h.andersen@farmasi.uio.no

Citation: Hjorth S, Bromley R, Ystrom E, Lupattelli A, Spigset O, Nordeng H (2019) Use and validity of child neurodevelopment outcome measures in studies on prenatal exposure to psychotropic and analgesic medications – A systematic review. PLoS ONE 14(7): e0219778. https://doi.org/10.1371/journal.pone.0219778

Editor: Andrea Martinuzzi, IRCCS E. Medea, ITALY

Received: November 27, 2018

Accepted: July 1, 2019

Published: July 11, 2019

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by the European Research Council Starting Grant "DrugsInPregnancy" [grant number 639377 to SH] and The PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway. The funders had no role in study design, data collection and

Abstract

In recent years there has been increased attention to child neurodevelopment in studies on medication safety in pregnancy. Neurodevelopment is a multifactorial outcome that can be assessed by various assessors, using different measures. This has given rise to a debate on the validity of various measures of neurodevelopment. The aim of this review was two-fold. Firstly we aimed to give an overview of studies on child neurodevelopment after prenatal exposure to central nervous system acting medications using psychotropics and analgesics as examples, giving special focus on the use and validity of outcome measures. Secondly, we aimed to give guidance on how to conduct and interpret medication safety studies with neurodevelopment outcomes. We conducted a systematic review in the MEDLINE, Embase, PsycINFO, Web of Science, Scopus, and Cochrane databases from inception to April 2019, including controlled studies on prenatal exposure to psychotropics or analgesics and child neurodevelopment, measured with standardised psychometric instruments or by diagnosis of neurodevelopmental disorder. The review management tool Covidence was used for data-extraction. Outcomes were grouped as motor skills, cognition, behaviour, emotionality, or "other". We identified 110 eligible papers (psychotropics, 82 papers, analgesics, 29 papers). A variety of neurodevelopmental outcome measures were used, including 27 different psychometric instruments administered by health care professionals, 15 different instruments completed by parents, and 13 different diagnostic categories. In 23 papers, no comments were made on the validity of the outcome measure. In conclusion, establishing neurodevelopmental safety includes assessing a wide variety of outcomes important for the child's daily functioning including motor skills, cognition,

analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

behaviour, and emotionality, with valid and reliable measures from infancy through to adolescence. Consensus is needed in the scientific community on how neurodevelopment should be assessed in medication safety in pregnancy studies.

Review registration number: CRD42018086101 in the PROSPERO database.

Introduction

Traditionally, studies on medication safety in pregnancy have mainly focused on immediate pregnancy outcomes such as the risk of malformations, low birth weight, and prematurity. In recent years, there has been an increased call for studies on the longer-term safety of prenatal exposure to medications, including studies on child neurodevelopment [1,2].

Psychotropic and analgesic medications share the biological plausibility to affect child neurodevelopment if used prenatally, as they can pass the placenta and bind to targets in the developing brain [3,4]. For these relatively common medications, even small increases in risk of neurodevelopmental delays in offspring could have public health impact [5]. Analgesics have a high frequency of use, with mild analgesics as the most common over the counter medication group in pregnancy with frequencies from 50% to over 70% [6,7]. There has been a marked increase in the use of selective serotonin reuptake inhibitor (SSRI) antidepressants in pregnancy over the last ten years, with the most recent prevalence data ranging from 2–4% in Europe [8,9] to 8% in North America [10]. Other psychotropics are less frequently used in pregnancy, but are important to study given the potential negative impact of a suboptimally treated maternal disorder on the health of mother and child [11,12].

Neurodevelopment comprises a wide range of traits and includes intelligence, language and motor skills, and attentional and executive functioning [13], which all are important for everyday life. When measuring neurodevelopment, some studies use the presence or absence of medical diagnoses, while others use psychometric instruments (questionnaires or tests) completed by parents, teachers, or health care professionals. Recent initiatives have suggested that it is important to consider a spectrum of neurodevelopment, not just diagnostic categories [14]. The question of how to measure neurodevelopmental outcomes in **medication safety studies** was recently raised in an editorial by our research group [2]. There we further call for consensus on how to conduct neurodevelopmental safety studies as part of a future pharmacovigilance framework [2].

Attempts have been made to summarise the literature on specific medication classes as the antidepressants [15], and paracetamol in relation to attention deficit hyperactivity disorder (ADHD, corresponding to the ICD-10 diagnosis hyperkinetic disorder) and autism spectrum disorder (ASD) [16]. A review on antidepressants chose not to present pooled effect estimates given the heterogeneity of the outcome measures [15], whereas a review on paracetamol presented pooled effect estimates for risk of ASD and ADHD, despite clear heterogeneity of outcome measures [16]. In summaries of the evidence, there has been little discussion about the validity and reliability of outcome measures.

Independent of which outcome measures are used, these should be reliable and valid in order for research data to be of value. Evaluations of reliability and validity are important for both medical diagnoses [17] and psychometric instruments [18,19], as both methods to some extent rely on subjective assessments [20]. Reliability is the degree to which the assessment is free from measurement error [18] (see also S1 File for definitions of different types of validity and reliability). Low reliability in an outcome measure mainly introduces random error and can therefore affect the precision of the results. This is particularly problematic in studies with

smaller sample sizes. Validity is defined as the extent to which the outcome measure truly measures the construct (e.g. shyness) it is intended to measure [19]. An outcome measure that is not valid in the context where it is used, will give systematically erroneous results, which is problematic regardless of sample size.

To our knowledge, the use and validity of neurodevelopmental outcome measures in studies on medication safety in pregnancy have not yet been systematically evaluated. Hence the primary aim of this review was to provide an overview of the use and validity of child neurodevelopment outcome measures employed in completed medication safety studies on prenatal exposure to psychotropic or analgesic medications. The secondary aim was aid researchers and clinicians in conducting and interpreting studies on maternal prenatal use of psychotropics and analgesics, and child neurodevelopment.

Methods

A systematic review was conducted in the MEDLINE, Embase, PsycInfo, Scopus, Cochrane, and Web of Science databases from inception to January 17th 2018. The search was updated on April 30th 2019. Reference lists of relevant reviews and included studies were screened to ensure complete coverage of the published literature. The search strategies were developed by the first author and research librarians, with inputs from all authors. Search terms and an example of the search strategy for the MEDLINE database can be found in S2 File. This systematic review was registered in the PROSPERO database (registration number CRD42018086101) and reported according to Preferred Reporting Items for Systematic Reviews (PRISMA).

Studies were considered eligible for inclusion if they fulfilled the criteria for participants, exposures, comparators, outcomes, and study design as described below. Participants were children born to mothers who used psychotropic or analgesic medication in pregnancy. For this review, children were defined as individuals under the age of 18 years. Assessments of neurodevelopment before the child was one month old were not considered, as we wished to exclude transient effects of prenatal exposure to medication. Exposures were maternal antenatal use of analgesic medication (ATC-codes N02 and M01A), antipsychotic medication (ATC-code N05A), anxiolytic medication (ATC-code N05B), hypnotic and sedative medication (ATC-code N05C), or antidepressants (ATC-code N06A) [21]. Antiepileptic medication (ATC-code N03) was not included in the review, as it has been thoroughly investigated in recent reviews [22,23]. Comparators were children born to mothers who did not use the specified medications in pregnancy. Outcomes were all neurodevelopment outcomes that had been assessed either by psychiatric diagnoses, or by standardised psychometric instruments filled in by parents, teachers, or health care professionals. In order to provide an overview, we divided the neurodevelopment outcomes in the following domains:

- Motor skills: Including ICD-10 code F82, specific developmental disorder of motor skills [24]
- Cognition: Including ICD-10 codes F70-79, mental retardation, F80, specific developmental disorder of speech and language, F81, specific developmental disorder of scholastic skills, and F84, pervasive developmental disorders (ASD)
- Behaviour: Including ICD-10 codes F90, hyperkinetic disorders, and F91, conduct disorders
- Emotionality: Including ICD-10 codes F30-39, mood disorders, F40-49, neurotic, stress-related and somatoform disorders (including anxiety), and F93, emotional disorders with onset specific to childhood

- Other: Including sleep disorders (ICD-10 code F51), and tic disorders (ICD-10 code F95)

For all ICD-10 codes, corresponding DSM-5 codes were likewise eligible. Autosomal genetic disorders (Rett syndrome, ICD-10 code F84.2) and unspecific disorders (Other childhood disintegrative disorder, ICD-10 code F84.3) were excluded. As the population for this review was children, we also excluded mental disorders that have their onset in adolescence [25]: bipolar disorder (ICD-10 code F31), schizophrenia (ICD-10 codes F20-29), and substance use disorder (ICD-10 codes F10-19). Randomised controlled trials (RCTs), cohort studies, register-based studies, and case-control studies were considered eligible for inclusion, whereas non-original studies (eg. reviews and editorials), original studies without a comparator group, cross-sectional studies, ecological studies, and animal studies were excluded. No date restrictions were applied, but for resource reasons, the search was limited to peer-reviewed publications in English, French, Italian, Spanish, or one of the Scandinavian languages.

References were imported to the systematic review data management platform Covidence [26]. Title and abstract screening, full text screening, and data extraction were all performed independently by two reviewers (SH and AL). Disagreement was solved by involving a third (HN) and, in case of doubt, a fourth reviewer (RB). When necessary, the authors of the original studies were contacted to provide additional information. Data items extracted from included studies were study design, inclusion and exclusion criteria, study population, duration of follow-up, definition of exposure in pregnancy (including whether gestational age at birth was known, or duration of gestation was estimated), timing, duration and dose of medication, outcome measure used, validity and reliability of outcome measure, covariates and how these were handled, study power, statistical analysis, and effect size. The data extraction forms were developed a priori with inputs from all authors.

Risk of bias in individual studies was assessed using Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidelines [27]. In addition to the items specified in the GRADE guidelines (failure to develop appropriate eligibility criteria, flawed measurement of both exposure and outcome, failure to adequately control confounding, and incomplete follow-up) [27], we assessed how missing data was handled in the studies. The risk of bias assessment was used in determining whether there were other explanations for heterogeneity between study findings than the psychometric properties of the chosen outcome measure. The psychometric properties of the outcome measures in the included studies were assessed on the domains internal consistency, inter-rater and test-retest reliability, construct, content and criterion validity (umbrella term for concurrent and predictive validity) as defined by the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) group and following their recommendations [28]. Risk of publication bias for each medication group was assessed according to GRADE guidelines [29]. Due to the expected heterogeneity between the studies in terms of age of the child at measurement and method of assessment, no meta-analysis was planned. However, to ease comparisons of results from different papers, effect sizes (Cohen's d) [30] were calculated using the metaeff package [31] for Stata [32]. Traditionally, a Cohen's d with an absolute value of 0.2 is considered a small effect, 0.5 a medium effect and 0.8 or above a large effect [33]. The decision to calculate Cohen's d was made post hoc as we became aware of how many different effect measures were used in the different studies.

Data was grouped by type of assessment (diagnoses or psychometric instruments), assessor for psychometric instruments (health care professional, parents, or teachers), and by age of the child.

Results

The literature search yielded 7,527 studies. After removal of duplicate records, 4,331 studies were left for title and abstract screening. Of these, 206 were relevant for full text assessment, and 101 were eligible for inclusion. See Fig 1 for PRISMA flowchart. A further 9 studies [34–42] were identified from reference lists of included studies and relevant reviews. Of the eligible papers, 82 focused on psychotropics (S1, S2, S5 and S6 Tables) and 29 on analgesics (S3 and S7 Tables). Some papers studied more than one medication group. Across all 110 papers, 26 papers used information from databases or national health registries. Neurodevelopment was assessed using 27 different psychometric instruments completed by health care professionals, 15 different psychometric instruments completed by parents, and five different psychometric instruments completed by teachers, not counting different versions of the same instrument (S4 Table). In addition, 13 different diagnostic categories were used. The most commonly used psychometric instrument completed by health care professionals was the Bayley Scales of Infant Development, used in 17 papers. The most common psychometric instrument completed by parents was the Childhood Behaviour Checklist, used in 22 papers. The most common diagnostic category was ASD (ICD-10 code F84) [24], used in 18 papers. However, the outcome measures differed for psychotropics and analgesics. For psychotropics, the most common outcome measure was diagnosis of ASD, used in 16 papers, whereas the most common outcome measures for analgesics were the Ages and Stages Questionnaire and the Child Behaviour Checklist, both used in six papers.

In the following, the use of outcome measures will be described by medication group. Only results for the oldest age band are presented here (and in S1–S3 Tables), if a paper had assessed children at multiple time points using the same outcome measure.

Antidepressants

Antidepressants was the most studied medication group with 66 papers [34, 36–38, 41,43–103]. Children had been assessed from the age of one month to 19 years. Diagnostic codes were used in 23 papers, while psychometric instruments were used in the remaining 43 papers (S1 Table). Most work has been done in the cognitive domain, where focus has been divided between intelligence (IQ), language and risk of ASD diagnosis, and most assessments have been done by health care professionals using psychometric instruments (Fig 2). Contrary to other medications, for antidepressants an additional domain of neurodevelopment other than motor skills, cognition, behaviour, and emotionality was assessed, as two papers reported parent assessed sleep problems.

Anxiolytics

Twenty papers studied neurodevelopment after prenatal exposure to anxiolytics [35,38,40,56,69,79–81,87,92,94,104–112], with ten of the papers published ten or more years ago. Follow-up was available from two months to 15 years (S2 Table). Most papers had used assessment from a health care professional, and only one paper had investigated risk of psychiatric diagnosis [56]. Most work has been done on the cognitive domain; the least has been done on emotionality (Fig 2).

Antipsychotics

Antipsychotics were investigated in seven papers [69,72,73,81,94,113,114]. Children had been followed from two months to 18 years. All papers had used psychometric instruments completed by health care professionals, except one that used ASD diagnosis [71] (Fig 2). Most papers, five of seven, studied motor skills. Only one paper had investigated a behavioural outcome [113] and none had investigated emotionality.

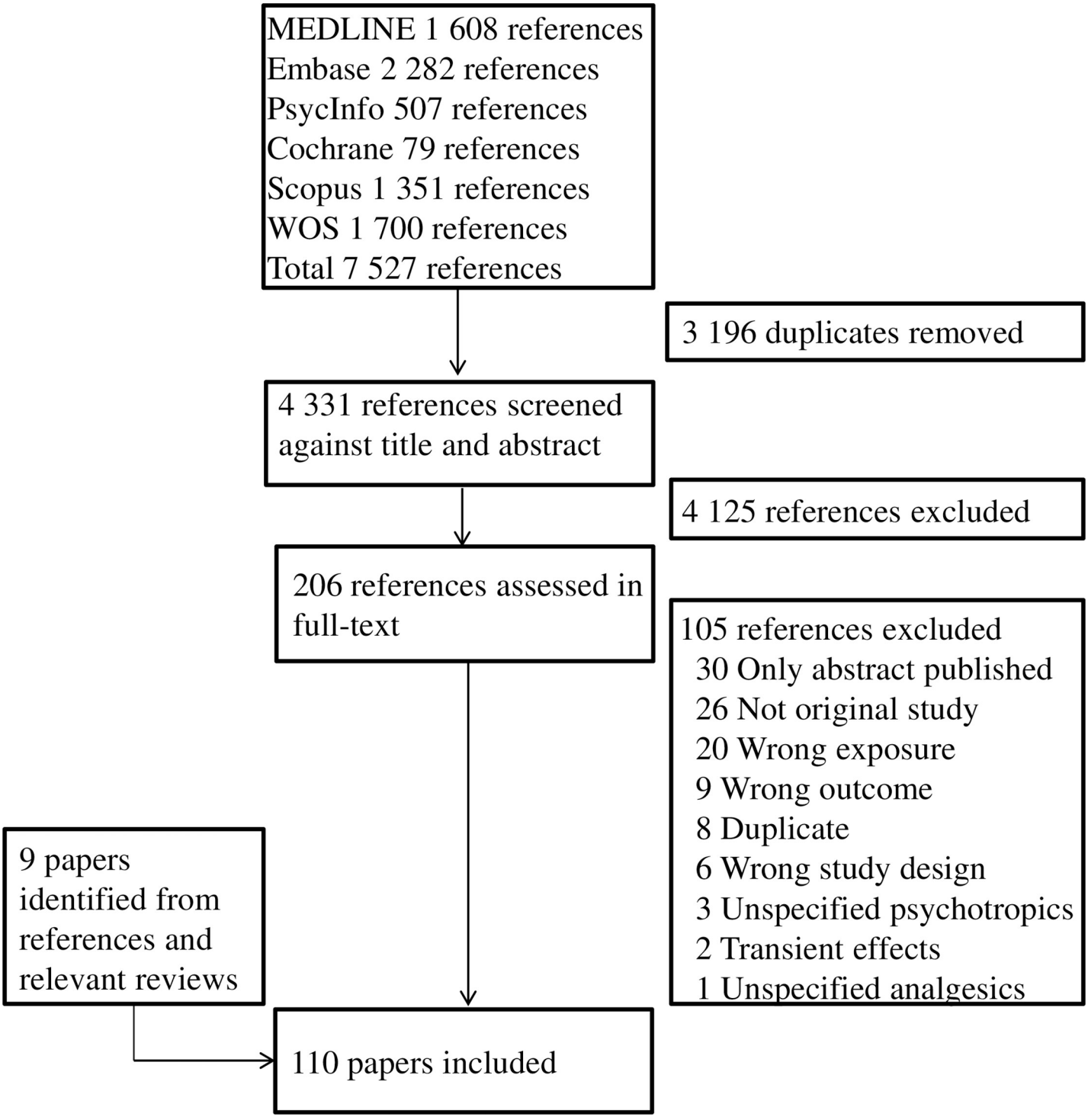

Fig 1. PRISMA flowchart. WOS: Web of Science.

https://doi.org/10.1371/journal.pone.0219778.g001

Hypnotics and sedatives

Six papers studied hypnotics and sedatives [94,104,109,115–117], and three of these papers looked at overdoses taken for suicide attempts. Follow-up was available from 1 year to 5 years.

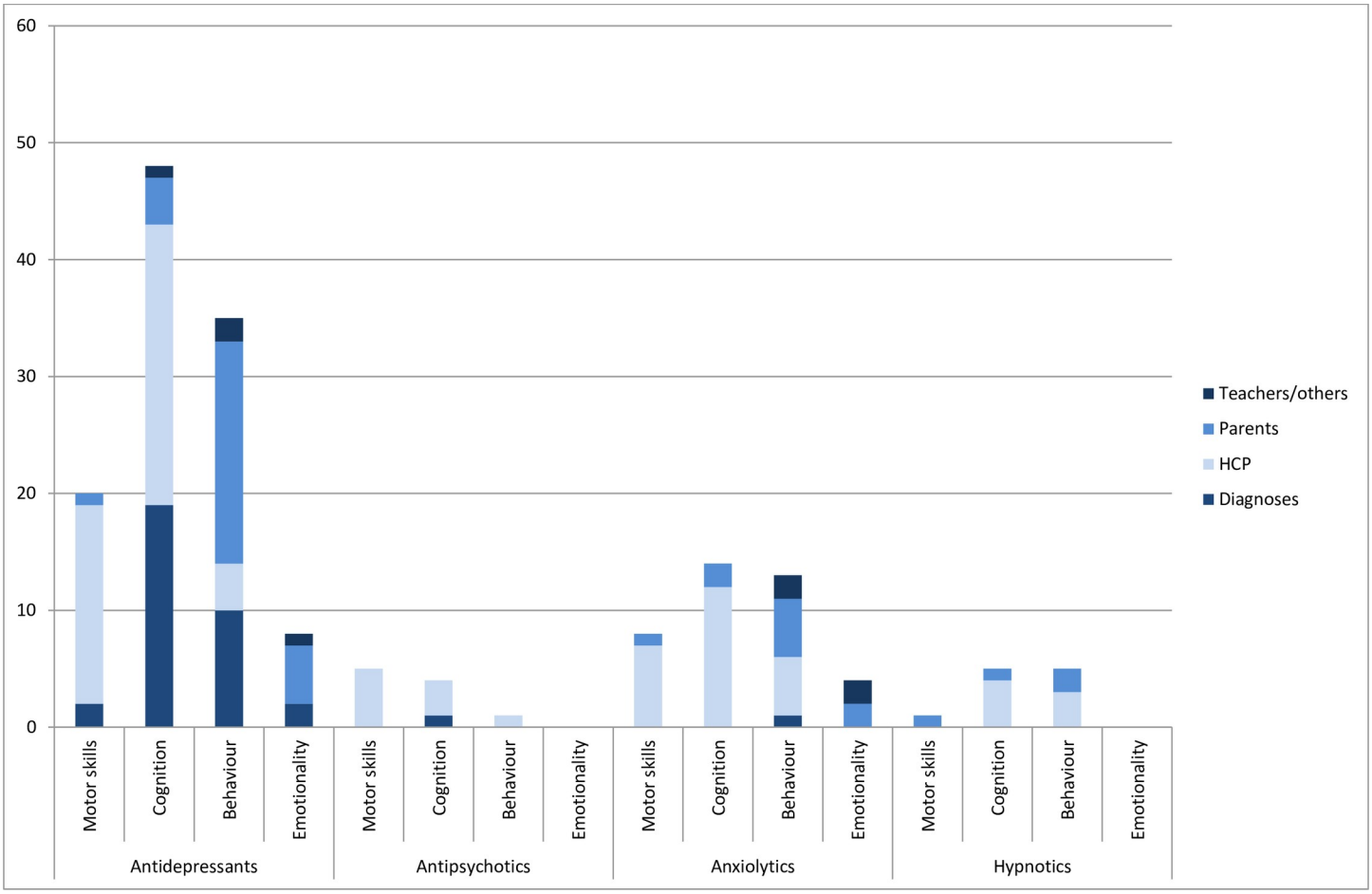

Fig 2. Domains of neurodevelopment evaluated and data sources used in medication safety papers on psychotropics. Some papers had outcomes from more than one domain and assessments from more than one type of assessor. HCP: Health care professionals.

https://doi.org/10.1371/journal.pone.0219778.g002

Three papers did not specify age at follow-up, but other papers from the same study have assessed toddlers. Most papers, four of six, had used psychometric instruments completed by health care professionals (Fig 2). Cognition was studied in four, and behaviour in five papers. One paper investigated motor skills [108], none studied emotionality.

Paracetamol

Due to a surge in papers since 2010, paracetamol was the most investigated analgesic with 18 papers [39,42,72,118–132]. Studies had follow-up between 18 months and 18 years (S3 Table). Most papers studied cognition and behaviour, with motor skills and emotionality investigated by two papers each (Fig 3). The most common data source was parent reporting.

NSAIDs

Five papers investigated NSAIDs [120,133–136]. Three of the papers were more than ten years old. Children were followed from 6 months to 6 years. Three papers studied exposure to indomethacin for treatment of preterm contractions, so the populations consisted of a higher proportion of children born prematurely than the general population. Four papers reported

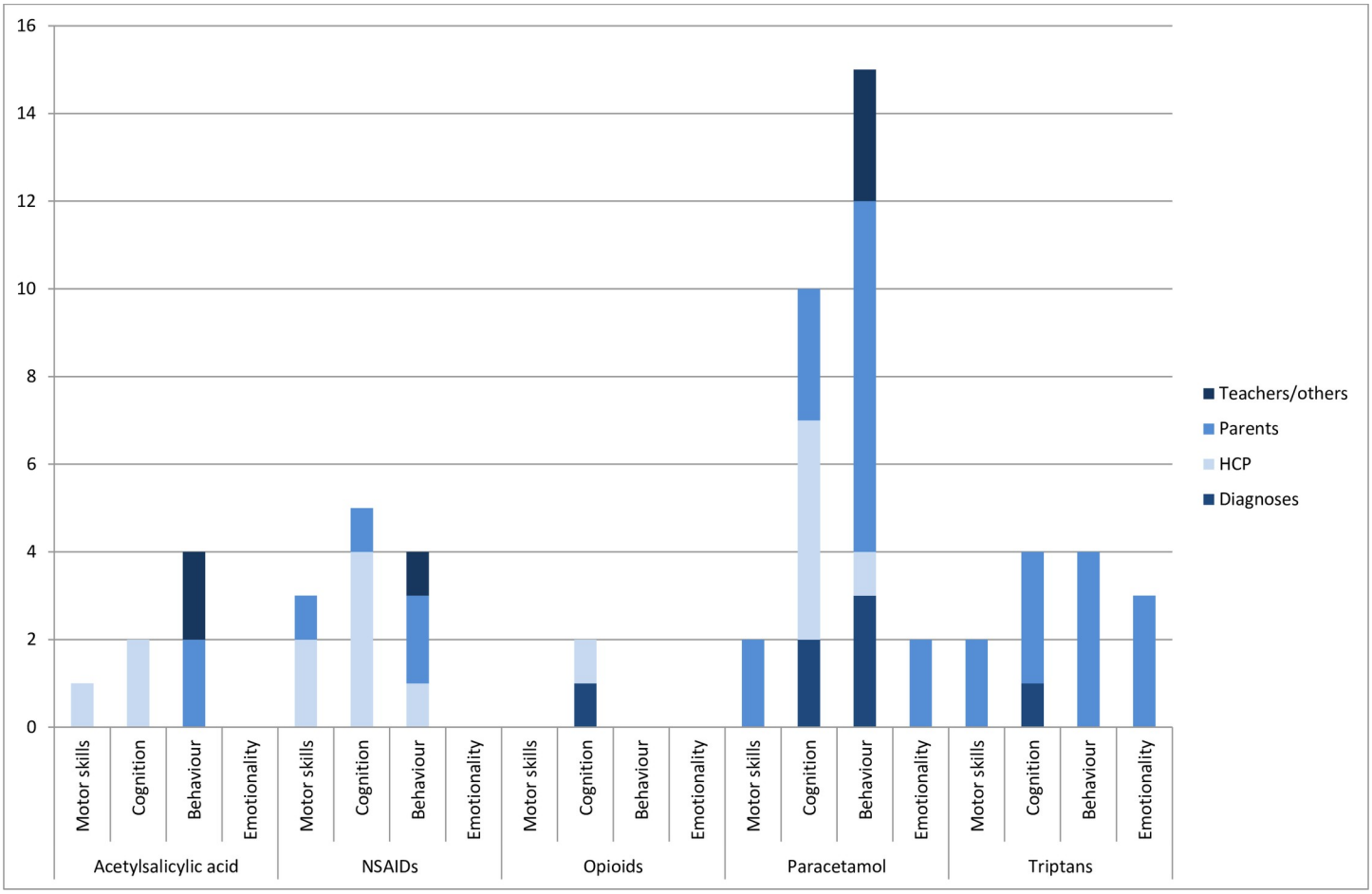

Fig 3. Domains of neurodevelopment evaluated and data sources used in medication safety papers on analgesics. Some papers had outcomes from more than one domain and assessments from more than one type of assessor. HCP: Health care professionals.

https://doi.org/10.1371/journal.pone.0219778.g003

assessments by health care professionals using psychometric instruments, two papers used parent reporting, and one teacher reporting [135]. The papers covered motor skills, cognition and behaviour domains of neurodevelopment, but not emotionality (Fig 3).

Acetylsalicylic acid

Of the five papers on acetylsalicylic acid [39,127,130,137,138], three were more than 20 years old. Children were assessed from 4 to 11 years of age. Three papers used assessments by health care professionals using psychometric instruments. The newest paper used parent and teacher reporting [130]. One paper investigated the motor -, two the cognition -, and two the behaviour domain of neurodevelopment (Fig 3).

Triptans

In four papers from the same research group [139–142], the same cohort of children was followed from 18 months to 5 years of age. All domains of neurodevelopment were assessed and assessors were the children's parents (Fig 3). A fifth paper from a different context had follow-up until age 18 and assessed risk of ASD diagnosis [72].

Analgesic opioids

Many papers have been written on use of illicit opioids, but for this review only analgesic opioids were included, yielding two papers [128,143] (Fig 3). In one paper, children's language development was assessed at 3 years of age by the parents, and in the other, the risk of diagnoses of ASD and developmental delay were assessed in pre-schoolers.

Validity and reliability of neurodevelopmental outcome measures

In 23 of the 110 eligible papers, no comment was made on either reliability or validity of any of the chosen outcome measures (Tables 1–3). The majority, 60 papers, commented qualitatively on reliability and/or validity of at least one of their chosen outcome measures, for instance by writing that the outcome measure was validated in their country. The remaining 27 papers also provided at least one quantitative measure of reliability and/or validity, such as Cronbach's α for internal consistency. Of the papers that commented on specific types of validity and/or reliability, most, 28 papers, commented on concurrent validitywhile content validity was only mentioned in one paper (Fig 4, see Tables 1–3 for more detail). In 37 papers, it was not mentioned what type of validity the authors commented on, but it was rather stated, for example, that the outcome measure was well-validated. Reliability was mentioned in 36 papers, and of these 17 did not specify what type of reliability that was referred to. We have provided an overview of the reliability and validity of the used outcome measures based on information from the psychometric literature (See S4 Table). In the following, the reporting of validity and reliability of outcome measures will be described by medication group.

Antidepressants

In 66 papers on antidepressants, 53 had comments on validity and/or reliability of at least one of their outcome measures. It was most common to report on concurrent validity, done in 22 papers, or not to specify the type of validity that was commented on, which was the case for another 22 papers. One paper mentioned content validity, and none of the papers mentioned structural validity.

Anxiolytics

Ten of the 20 papers on anxiolytics had comments on validity and/or reliability of at least one of their outcome measures. The type of validity that was mentioned most often was construct validity in the form of correlation to the full scale of the instrument. None of the papers mentioned content validity.

Antipsychotics

Of seven papers on antipsychotics, five had comments on validity and/or reliability of at least one of their outcome measures. However, four of these did not mention what type of validity their comment regarded. None of the papers mentioned construct validity, content validity or internal consistency.

Hypnotics and sedatives

Half of the six papers on hypnotics and sedative commented on validity and/or reliability of at least one of their outcome measures. No type of validity or reliability was mentioned in more than one paper. None of the papers mentioned content validity, inter-rater–or test-retest reliability.

Table 1. Validity and reliability of outcome measures as reported by the authors of the study, papers on antidepressants.

		Psychometric properties of outcome measure	
Reference	**Outcome**	**Reliability**	**Validity**
Assessment using psychometric instruments			
i. Assessment by health care professionals			
Infant (<2 years)			
Suri 2011 [99]	BNBAS	Assessors were trained and certified to 0.90 reliability	Sensitive to medication effects. Limited normative base. Results that can be influenced by numerous subtle factors
Mortensen 2003 [81]	Boel test	Not mentioned	Test was performed at children's home, not under standard settings in which the test was developed. Surroundings could be associated with both exposure and outcome
Batton 2013 [44]	Bayley Infant Neurodevelopment Screener	Not mentioned	Validated using clinical and normative standardisation samples
Weikum 2013a [101]	BSID, unspecified edition	Not mentioned	Not mentioned
Gustafsson 2018 [61]	BSID-II	Administrators had completed several practice administrations with repeated feedback prior to engaging in study administration of the BSID	Not mentioned
Oberlander 2004 [87]	BSID-II	Not mentioned	Not mentioned
Reebye 2002 [92]	BSID-II	Not mentioned	Not mentioned
Reebye 2012 [38]	BSID-II	Not mentioned	Not mentioned
Santucci 2014 [93]	BSID-II	Good reliability for infants from 1 to 42 months	Good concurrent validity for infants from 1 to 42 months
Austin 2013 [43]	BSID-III	BSID is considered gold standard. Good reliability and internal consistency	Good validity. Used US norms. BSID is more predictive, the older the child
Hanley 2013 [64]	BSID-III	Not mentioned	Not mentioned
Heikkinen 2002 [67]	Gesell Development scales	Psychometric properties not mentioned (Comment from review authors: Not mentioned which psychometric instrument is used, we have taken this information from the 2003 paper)	
Heikkinen 2003 [36]	Gesell Development scales	Not mentioned	Not mentioned
Johnson 2012 [73]	Infant Neurological International Battery	Interrater reliability 0.97, test-retest 0.95	Sensitivity 90%, specificity 83%, PPV 79%, NPV 93%
de Vries 2013 [53]	Psychomotor assessment according to Prechtl	Interrater reliability of general movement assessment is high (89% to 93%)	Abnormal general movement quality is highly predictive of later neurological impairment, but might not be sensitive enough to detect minor neurological sequelae. The Motor Optimality Score and the concurrent repertoire might be more revealing. Monotonous movement is related to minor neurological dysfunction in preterm infants
Preschool (2–5 years)			
Nulman 2002 [84]	BSID-II, Reynell developmental language scale, MSCA	Not mentioned	Not mentioned
Nulman 1997 [83]	BSID-II	Not mentioned	Not mentioned
Casper 2003 [49]	BSID-II	The two psychologist assessors were reliability certified on the BSID-II annually	Not mentioned
Batton 2013 [44]	BSID-III	Not mentioned	Not mentioned
Galbally 2011 [57]	BSID-III	Test-retest reliability for total score is 0.9 at 12 months	Regarded as the reference standard in the assessment of infant and toddler development. Only moderate predictive validity

(*Continued*)

Table 1. (Continued)

		Psychometric properties of outcome measure	
Reference	**Outcome**	**Reliability**	**Validity**
Hurault-Delarue 2016 [69]	Compulsory medical exam	Not mentioned	Not mentioned
Schechter 2017 [94]	DAS	Fidelity checks every 6-months by a licensed clinical psychologist	Standardised, age-normed, well-validated measure of cognitive ability
Johnson 2016 [74]	DAS-II, Test of early language development, 3rd edition (TELD-3)	Not mentioned	DAS-II: Measure is normalised. TELD-3: Not mentioned
Galbally 2015 [58]	Movement ABC, WPPSI-III	WPPSI, see validity. Motor ABC, reliability not mentioned	WPPSI is regarded as the gold standard for assessing cognitive ability. Motor ABC, validity not mentioned
Mattson 2002 [79]	WPPSI-R	Not mentioned	Not mentioned
School child (6–12 years)			
Nulman 1997 [83]	Reynell developmental language scale, MSCA	Not mentioned	Not mentioned
El Marroun 2017 [55]	SON-R (shortened), NEPSY-II	SON-R: Reliability was 0.73 for Mosaics and 0.71 for Categories. NEPSY-II: Not mentioned	SON-R: Validated in Dutch. Correlation to full scale, r = 0.86. NEPSY-II: Not mentioned
Hermansen 2016 [68]	NEPSY-II (shortened), WPPSI-R	Not mentioned	Not mentioned
Nulman 2012 [85]	WPPSI-III	Not mentioned	Child IQ levels were considerably higher than the normative mean of the general population. High IQs may reflect the Flynn effect. Mothers who contacted Motherisk may have had higher IQs than those who did not. Canadian norm for full-scale IQ is five points higher than the usual norm
Nulman 2015 [86]	WPPSI-III	Not mentioned	Not mentioned
Adolescent (13–18 years)			
Mattson 2002 [79]	WISC-III	Not mentioned	Not mentioned
ii. Assessment by parents			
Infant (<2 years)			
Brandlistuen 2015 [46]	CBCL	Reliability was adequate, Cronbach's α 0.62	The subset of items used in the MoBa study was found to be representative, with a correlation of 0.92 with the full scale
Reebye 2002 [92]	Early infant temperament questionnaire	Not mentioned	Not mentioned
Netsi 2015 [82]	Infant Characteristic Questionnaire, Brief Infant Sleep Questionnaire	Measures are reliable	Questionnaires relied heavily on maternal report and are subject to maternal bias
Nulman 1997 [83]	Toddler temperament scale	Not mentioned	Not mentioned
Nulman 2002 [84]	Toddler temperament scale	Not mentioned	Not mentioned
Preschool (2–5 years)			
Handal 2016a [62]	ASQ	Not mentioned	Validation study was done. For gross motor the estimated Spearman correlation was 0.25 with Mullen scores. For fine motor the correlation was 0.40
El Marroun 2017 [55]	BRIEF	Good test-retest reliability	Good content validity
Galbally 2015 [58]	CBCL, BRIEF	CBCL, see validity. BRIEF, reliability not mentioned	CBCL is reliable, valid and widely used. BRIEF, validity not mentioned

(*Continued*)

Table 1. (Continued)

		Psychometric properties of outcome measure	
Reference	**Outcome**	**Reliability**	**Validity**
Brandlistuen 2015 [46]	CBCL	Reliability was adequate, Cronbach's α 0.74	The subset of items used in the MoBa study was found to be representative, with a correlation of 0.92 with the full scale
Hanley 2015 [65]	CBCL	Not mentioned	Maternal report poses some risk of differential outcome reporting
Johnson 2016 [74]	CBCL	Not mentioned	CBCL is not a diagnostic tool, but children scoring high on the PDD subscale are more likely to evidence behaviours commonly associated with ASD
Misri 2006 [80]	CBCL	Test-retest reliability r = 0.85, interrater reliability r = 0.61	Not mentioned
Nulman 2002 [84]	CBCL	Not mentioned	Not mentioned
Oberlander 2007 [88]	CBCL	Not mentioned	Widely used and well validated. Close relationship between maternal depression and maternal report of child behaviour. Increased level of aggression could represent lower maternal thresholds for tolerating preschool behaviour
Oberlander 2010 [89]	CBCL	Not mentioned	Widely used and well validated
Lupattelli 2018 [76]	CBCL, EAS	CBCL: Not mentioned in the article, but in the supplement internal consistency is reported quantitatively for each subscale. EAS: Moderate internal consistency of Norwegian version. In the supplement internal consistency is reported quantitatively for each subscale	Due to parent reporting misclassification cannot be ruled out, but was probably non-differential. CBCL: Widely used and validated. Predictive validity for adolescent psychiatric disorders showed a sensitivity of 0.71 and a specificity of 0.92. EAS: The short version used was highly correlated to full scale
Handal 2016b [63]	Intelligibility/Complexity of 3-year-old Children's Utterances	Not mentioned	Validation study showed that children categorised by their mothers as having no language delay achieved a higher score on the communication domain of the Vineland Adaptive Behavior Scale than the children categorised as having severe language delay by mothers (Comment from review authors: Vinelands is a structured interview of parents by the HCP)
Skurveit 2014 [96]	Intelligibility/Complexity of 3-year-old Children's Utterances	Not mentioned	Parental self-report is generally a good measure of early expressive vocabulary, especially for severe language delay. The validity of the language grammar rating scale has been described earlier
Pedersen 2013 [90]	SDQ	Not mentioned	Validated in Danish. Not intended to predict underlying psychiatric disorder. Relatively low sensitivity based on single informants
School child (6–12 years)			
Hutchison 2019 [70]	BRIEF	Not mentioned	Maternal perception of her child may have been negatively affected by depression, yet study using performance based measures of executive function found similar results
Hermansen 2016 [68]	CBCL	See validity	Widely used and standardised. Excellent reliability and validity
Nulman 1997 [83]	CBCL	Not mentioned	Not mentioned
Nulman 2012 [85]	CBCL, CPRS	Not mentioned	Maternal perception of her child may have been negatively affected by depression and its associated anxiety and stress

(*Continued*)

Table 1. (Continued)

		Psychometric properties of outcome measure	
Reference	**Outcome**	**Reliability**	**Validity**
Nulman 2015 [86]	CBCL, CPRS-R	Not mentioned	Severity of maternal depression may influence responses when filling out questionnaires. Bias minimised as mothers evaluate both her exposed and unexposed children simultaneously (Comment from review authors: the mother still knows that one child was exposed and the other was not)
El Marroun 2014 [54]	CBCL and Social responsiveness scale (SRS)	Correlations between the CBCL measurements at different ages fell in the expected range, based on a mean correlation (r = 0.60)Crohnbach's α indicated high inter-item reliability for the SRS (a = 0.79)	CBCL: Validated in Dutch. Good predictive validity to identify preschoolers at risk of autism spectrum disorder. Scales could not be normalised, so scores were dichotomised. The SRS correlated well with the pervasive developmental problems scale of the CBCL, r = 0.59
Hanley 2015 [65]	HBQ-P	See validity	Maternal report poses some risk of differential outcome reporting. The HBQ-P has strong psychometric properties
Weikum 2013b [102]	HBQ-P	See validity	The HBQ-P has strong psychometric properties. Mental health scales discriminate groups of children with and without signs of early psychopathology
Grzeskowiak 2016 [60]	SDQ	Not mentioned	Validated in Danish. Excellent discrimination for the identification of emotional (AUC 0.80) and behavioural (AUC 0.89) disorders
iii. Assessment by teachers/others			
Preschool (2–5 years)			
Johnson 2016 [74]	CBCL (other caregiver)	Not mentioned	CBCL is not a diagnostic tool, but children scoring high on the PDD subscale are more likely to evidence behaviours commonly associated with ASDs
Misri 2006 [80]	CBCL (teacher)	Test-retest reliability r = 0.85, interrater reliability r = 0.61	Not mentioned
Oberlander 2007 [88]	CBCL (teacher)	Not mentioned	Widely used and well validated. Close relationship between maternal depression and maternal report of child behaviour. Increased level of aggression could represent lower maternal thresholds for tolerating preschool behaviour
Assessment using medical diagnosis			
Boukhris 2017 [34]	ADHD	Not mentioned	Diagnosis defined from hospital diagnosis or redemption of a prescription for ADHD medication. Diagnoses of ADHD in the cohort were not validated. Sensitivity analysis on children with a diagnosis confirmed by neurologists and psychiatrists was consistent with those of the main analyses
Figueroa 2010 [56]	ADHD	Not mentioned	ADHD identified from diagnoses or treatment from practice, not from any formalised test or direct observation. This could lead to false-positive and false-negative errors. The young age when ADHD was diagnosed represents a limitation. Possible that only the most severe cases of ADHD were identified or that other behavioural problems that resemble ADHD are included
Laugesen 2013 [37]	ADHD	Not mentioned	Detection bias could have led to an overestimation of the association. Children with ADHD were identified based on hospital diagnoses and drug prescriptions. Patients with ADHD diagnosed by private psychiatrists or general practitioners and not prescribed drug treatment would be misclassified as not having ADHD
Man 2017 [78]	ADHD	Not mentioned	The registry contains information from publicly funded healthcare medical records. Does not include data from private medical practitioners or hospitals

(*Continued*)

Table 1. (Continued)

		Psychometric properties of outcome measure	
Reference	**Outcome**	**Reliability**	**Validity**
Castro 2016 [50]	ASD and ADHD	Not mentioned	ICD-9 codes have a high sensitivity and specificity for ASD and ADHD versus generally healthy control. Case definition previously validated and included scores from autism diagnostic observation scale
Clements 2015 [51]	ASD and ADHD	Not mentioned	For ASD diagnosis in registry, sensitivity was 1.00, specificity 0.91. For ADHD, sensitivity was 0.84, specificity 0.90
Sujan 2017 [98]	ASD and ADHD	Not mentioned	Previous research has validated the diagnoses in used registries. Associations were estimated excluding offspring with diagnoses before age 2 years to address concerns about validity of early neurodevelopmental diagnoses. This did not markedly alter results
Wibroe 2017 [103]	ASD and ADHD	Not mentioned	Not mentioned
Malm 2016 [77]	ASD, depression, anxiety, ADHD	Not mentioned	Quality of the registry has been validated and is good for psychiatric diagnoses
Liu 2017 [75]	ASD, F30-39, F40-49, F70-79, F90-99	Not mentioned	Cannot rule out detection bias, but similar associations were observed for all disorders, irrespective of age at onset. Cases were redefined as at least two hospital contacts for psychiatric disorders, but similar results were obtained
Boukhris 2016 [45]	ASD	Not mentioned	Diagnoses of ASD in the cohort were not validated. Sensitivity analysis on children with a diagnosis of ASD confirmed by neurologists and psychiatrists was consistent with those of the main analyses
Brown 2017 [48]	ASD	Not mentioned	ASD was defined as 2 or more outpatient diagnoses by either a paediatrician or psychiatrist, 1 or more diagnoses in hospital databases, after the age of 2 years. A similar definition using US insurance data had a positive predictive value of 87.4%
Croen 2011 [52]	ASD	Not mentioned	Validation study of diagnoses in the registry against the Autism Diagnostic Interview–Revised and the Autism Diagnostic Observation Schedule–Generic; 94% of cases met criteria for ASD on both instruments, and 100% on at least one. A full review of diagnostic information recorded in cohort medical records demonstrated that at least 90% of ASD cases in the cohort meet DSM-IV criteria
Gidaya 2014 [59]	ASD	Not mentioned	Register diagnosis of childhood autism was confirmed 94% of the time with an additional 3% classified with another ASD in validation study
Hviid 2013 [71]	ASD	Not mentioned	A previous study showed that 94% of children registered with autism spectrum disorder diagnoses met diagnostic criteria in chart review. Not all children in the study have been followed throughout childhood. Some may receive a diagnosis of ASD at older ages. Detection bias cannot be ruled out
Janecka 2018 [72]	ASD	Not mentioned	Not mentioned
Rai 2017 [91]	ASD	Not mentioned	Previous validation studies found high validity of the diagnoses recorded in the registers
Sorensen 2013 [97]	ASD, infantile autism	Not mentioned	The quality of the infantile autism diagnosis in the registry has been validated. 94% met the criteria for correct diagnosis
Viktorin 2017b [41]	ASD	Not mentioned	The ASD diagnoses in the register have previously been validated

(*Continued*)

Table 1. (Continued)

		Psychometric properties of outcome measure	
Reference	**Outcome**	**Reliability**	**Validity**
Harrington 2014 [66]	ASD, developmental delay (DD)	Not mentioned	Diagnoses of ASD and DD were confirmed by trained clinicians using validated standardised instruments. Controls were screened and reclassified if appropriate
Brown 2016 [47]	Disorders of speech/language, motor skills, and scholastic skills	Not mentioned	Diagnostic data were based on specialised health services rather than primary care. Some proportion of children with mild dysfunction may have been missed
Simon 2002 [95]	Developmental delay of motor skills, developmental delay of speech	Not mentioned	Diagnosis required both a physician diagnosis and confirmation by a formal developmental evaluation. Examinations at outpatient paediatric visits are relatively crude screening measures. Use of these data may reduce bias in ascertainment of developmental delay, but it sacrifices sensitivity by limiting analyses to abnormalities detected during routine medical care
Viktorin 2017a [100]	Intellectual disability	Not mentioned	Children with ID without clinical care are not captured, so the prevalence estimate of ID in the study may be an underestimate of the true prevalence in the population. Detection bias cannot be ruled out

ADHD: Attention Deficit Hyperactivity Disorder, ASD: Autism Spectrum Disorder, ASQ: Ages and Stages Questionnaire, BNBAS: Brazelton Neonatal Behavoiral Assessment Scale, BRIEF: Behaviour Rating Inventory of Executive Function, BSID: Bayley Scales of Infant Development, CBCL: Child Behaviour Checklist, CPRS: Conners Parent Rating Scale, DAS: Differential ability scales, EAS: Emotionality, Activity, Sociability Temperament Survey, HBQ-P: MacArthur Health and Behaviour Questionnaire, MSCA: McCarthy's scales of children's abilities, SDQ: Strengths and Difficulties Questionnaire, SON-R: Snijders–Oomen Niet-verbale intelligentie Test–Revisie, WISC: Wechsler Intelligence Scale for Children, WPPSI: Wechsler Preschool and Primary Scale of Intelligence.

https://doi.org/10.1371/journal.pone.0219778.t001

Paracetamol

All but one of the 18 papers on paracetamol had comments on validity and/or reliability of at least one of their outcome measures. Eight papers did not specify what type of validity the comment regarded. None of the papers mentioned content validity.

NSAIDs

Three of five papers on NSAIDs had comments on validity and/or reliability of at least one of their outcome measures. All three mentioned cross-cultural validity (a subtype of construct validity), none mentioned criterion validity, content validity, inter-rater–or test-retest reliability.

Acetylsalicylic acid

All five papers commented on validity and/or reliability of at least one of their outcome measures. In three papers, the type of validity was not specified, yet four of the papers commented on a specified type of reliability. None of the papers mentioned construct validity or content validity.

Triptans

Of five papers on triptans, four had comments on validity and/or reliability of at least one of their outcome measures. Most frequently mentioned was predictive validity and cross-cultural validity. None of the papers mentioned structural validity or content validity.

Table 2. Validity and reliability of outcome measures as reported by the authors of the study, papers on psychotropics except antidepressants.

			Psychometric properties of outcome measure	
Reference	**Exposure**	**Outcome measure**	**Reliability**	**Validity**
Assessment using psychometric instruments				
i. Assessment by health care professionals				
Infant (<2 years)				
Platt 1989 [114]	Antipsychotics	BSID	Not mentioned	Measures obtained at different ages may differ in sensitivity. The categorical measure reported was not used in original examination, but derived by combining relevant motor items which showed some variance in study population, and dichotomising resulting scores to maximise drug effects
Peng 2013 [113]	Antipsychotics	BSID-III	Not mentioned	The scale is widely used and has potential to provide clinically relevant information on early neurodevelopment
Johnson 2012 [73]	Antipsychotics	Infant Neurological International Battery	Interrater reliability 0.97, test-retest 0.95	Sensitivity 90%, specificity 83%, PPV 79%, NPV 93%
Mortensen 2003 [81]	Antipsychotics, anxiolytics	Boel test	Not mentioned	Test was performed at children's home, not under standard settings in which the test was developed. Surroundings could be associated with both exposure and outcome
Oberlander 2004 [87]	Anxiolytics (clonazepam combined with SSRI)	BSID-II	Not mentioned	Not mentioned
Reebye 2002 [92]	Anxiolytics (clonazepam combined with SSRI)	BSID-II	Not mentioned	Not mentioned
Reebye 2012 [38]	Anxiolytics (clonazepam combined with SSRI)	BSID-II	Not mentioned	Not mentioned
Viggedal 1993 [40]	Anxiolytics	Griffiths' mental development scale I, Neuropsychological assessment	Griffiths': Not mentioned. Neuropsychological assessment: Deviations from normal activity and attention was diagnosed when present at two independent observations	Griffiths': The high IQ in the reference group is considered normal as the average IQ nowadays is about 110. Neuropsychological assessment: Not mentioned
Gidai 2008a [105]	Anxiolytics (alprazolam)	Hungarian development test, Behavioural style questionnaire	Not mentioned	Not mentioned
Gidai 2008b [106]	Anxiolytics (medazepam)	Hungarian development test, Behavioural style questionnaire	Not mentioned	Not mentioned
Gidai 2008c [107]	Anxiolytics (chlordiazepoxide)	Hungarian development test, Behavioural style questionnaire	Not mentioned	Not mentioned
Timmermann 2008b [112]	Anxiolytics (meprobamate)	Hungarian development test, Behavioural style questionnaire	Not mentioned	Not mentioned
Laegreid 1992 [108]	Anxiolytics	Touwen Neurologic Assessment, Clinical neurologic assessment	Not mentioned	The test showed a fair differentiation and an evident developmental sequence
Petik 2008a [115]	Hypnotics (glutethimide)	Hungarian development test, Behavioural style questionnaire	Not mentioned	Not mentioned
Petik 2008b [116]	Hypnotics (Amobarbital)	Hungarian development test, Behavioural style questionnaire	Not mentioned	Not mentioned

(*Continued*)

Table 2. (Continued)

			Psychometric properties of outcome measure	
Reference	**Exposure**	**Outcome measure**	**Reliability**	**Validity**
Timmermann 2008a [117]	Hypnotics (barbital, hexobarbital, butobarbital)	Hungarian development test, Behavioural style questionnaire	Not mentioned	Not mentioned
Preschool (2–5 years)				
Hurault-Delarue 2016 [69]	Antipsychotics, anxiolytics and hypnotics	Compulsory medical exam	Not mentioned	Not mentioned
Schechter 2017 [94]	Antipsychotics, anxiolytics, hypnotics	DAS	Fidelity checks were done approximately every 6-months by a licensed clinical psychologist	The DAS is a standardised, age-normed, well-validated measure of cognitive ability
Hartz 1975 [35]	Anxiolytics (meprobamate, chlordiazepoxide)	Stanford Binet Intelligence Scale	Not mentioned	Not mentioned
Mattson 2002 [79]	Anxiolytics	WPPSI-R	Not mentioned	Not mentioned
School child (6–12 years)				
Platt 1989 [114]	Antipsychotics	Paediatric neurologic assessment	Not mentioned	Not mentioned
Adolescent (13–18 years)				
Mattson 2002 [79]	Anxiolytics	WISC-III	Not mentioned	Not mentioned
ii. Assessment by parents				
Infant (<2 years)				
Reebye 2002 [92]	Anxiolytics (clonazepam combined with SSRI)	Early Infancy Temperament Questionnaire	Not mentioned	Not mentioned
Preschool (2–5 years)				
Lupattelli 2019 [109]	Anxiolytics, hypnotics	ASQ, CPRS-R	ASQ: Internal consistency was 0.6 to 0.7. CPRS-R: Internal consistency 0.9	Widely used and validated. In the supplement, the authors present associations between ASQ and diagnosis of motor delay (Beta 1.96 to 4.48) or language impairment (Beta 2.06 to 2.45), and between CPRS-R and parental ADHD symptoms (Beta 0.12 to 0.30 for paternal symptoms and 0.41 to 0.76 for maternal symptoms)
Brandlistuen 2017 [104]	Anxiolytics, hypnotics	CBCL (shortened)	Not mentioned	Validated, representative of full scale, the domains predict later psychopathology. High factor loadings
Misri 2006 [80]	Anxiolytics (clonazepam combined with SSRI)	CBCL	Test-retest reliability r = 0.85, interrater reliability r = 0.61	Not mentioned
Odsbu 2015 [110]	Anxiolytics	Intelligibility/Complexity of 3-year-old Children's Utterances	Not mentioned	Parental self-report is a good measure of early expressive vocabulary, especially for severe language delay. Validity of the language grammar rating scale in the cohort has been described previously
School child (6–12 years)				
Radojčić 2017 [111]	Anxiolytics	CBCL	Cronbach's alphas for all scales were the same in 6 year-old children and in older children, indicating that problems were reliably measured in children older than 6 years of age	Validated in the Netherlands
iii. Assessment by teachers/others				
Preschool (2–5 years)				

(*Continued*)

Table 2. (Continued)

			Psychometric properties of outcome measure	
Reference	**Exposure**	**Outcome measure**	**Reliability**	**Validity**
Misri 2006 [80]	Anxiolytics (clonazepam combined with SSRI)	CBCL	Test-retest reliability r = 0.85, interrater reliability r = 0.61	Not mentioned
School child (6–12 years)				
Radojčić 2017 [111]	Anxiolytics	CBCL	Cronbach's alphas for all scales were the same in 6 year-old children and in older children, indicating that problems were reliably measured in children older than 6 years of age	Validated in the Netherlands
Assessment using medical diagnosis				
Figueroa 2010 [56]	Anxiolytics	ADHD	Not mentioned	ADHD identified from diagnoses or treatment from practice, not from any formalised test or direct observation. This could lead to false-positive and false-negative errors. The young age when ADHD was diagnosed represents a limitation. It is possible that only the most severe cases of ADHD were identified or that other behavioural problems that resemble ADHD are included
Janecka 2018 [72]	Lithium	ASD	Not mentioned	Not mentioned

ADHD: Attention Deficit Hyperactivity Disorder, ASD: Autism Spectrum Disorder, ASQ: Ages and Stages Questionnaire, BSID: Bayley Scales of Infant Development, CBCL: Child Behaviour Checklist, CPRS:R: Conners Parent Rating Scale, revised, DAS: Differential ability scales, NPV: Negative predictive value, PPV: Positive predictive value, WISC: Wechsler Intelligence Scale for Children, WPPSI: Wechsler Preschool and Primary Scale of Intelligence.

https://doi.org/10.1371/journal.pone.0219778.t002

Opioids

Both papers on opioids commented on the validity of at least one of their outcome measures, one on concurrent validity, the other did not specify type of validity. None of the papers commented on reliability, construct validity or content validity.

Discussion

In the 110 papers on neurodevelopment after prenatal exposure to psychotropics (66 papers on antidepressants, 27 papers on other psychotropics) or analgesics (29 papers) identified from a systematic review of the literature, 47 different psychometric instruments and 13 different diagnostic categories were used to measure neurodevelopment. Twenty-three papers did not mention the reliability or validity of any of the neurodevelopment outcome measures. Among the papers that did mention psychometric properties, 37 papers did not specify on what type of validity they were commenting.

Strengths and limitations

Strengths of this review include a comprehensive search in six different databases, an interdisciplinary research team, compliance with PRISMA guidelines, and assessment of study eligibility and quality, as well as data extraction, done in double.

Limitations include that only published papers in predefined languages were included in the review, though no papers in other languages that fulfilled the remaining eligibility criteria turned up in our search. Further, search strategy could have been optimised by inclusion of

Table 3. Validity and reliability of outcome measures as reported by the authors of the study, papers on analgesics.

			Psychometric properties of outcome measure	
Reference	**Exposure**	**Outcome measure**	**Reliability**	**Validity**
Assessment using psychometric instruments				
i. Assessment by health care professionals				
Infants (<2 years)				
Salokorpi 1996 [136]	Indomethacin	Autti-Rämö neurodevelopmental test battery	Not mentioned	Standardised for Finnish children
Amin 2008 [134]	Indomethacin	BSID-II, mental development index	Not mentioned	Not mentioned
Al-Alaiyan 1996 [133]	Indomethacin	Gesell development scales, revised	Not mentioned	Not mentioned
Avella-Garcia 2016 [118]	Paracetamol	BSID, unspecified ed.	Not mentioned in the paper, but in supplement. Cronbach's α 0.70 (good to moderate)	Not mentioned
Preschool (2–5 years)				
Barr 1990 [137]	ASA	Items from Gross Motor Scale (University of Oregon Medical School), Gesell–and Bayley Scales, Wisconsin Fine Motor Steadiness Battery, and Halstead Reitan Neuropsychological Battery	Monthly reliability checks revealed good interrater reliability	For three variables, the best predictors were exam conditions. None of these measures are normally used with children this young
Klebanoff 1988 [138]	ASA	Stanford Binet Intelligence Scale	3 months test-retest, and in addition inter-rater, reliability was 0.83	Test focuses on verbal abilities. Highly correlated to school performance
Streissguth 1987 [39]	ASA	WPPSI	Not mentioned	IQ scores by age 4 years have a good predictive validity for later intellectual function
Avella-Garcia 2016 [118]	Paracetamol	McCarthy Scales of Children's Abilities, CAST*	McCarthy Scale of Children's Abilities: Cronbach's α 0.90. CAST: Cronbach's α 0.64 (good to moderate)	Only mentioned for CAST: Sensitivity 100%, specificity 97% for ASD
Bornehag 2018 [119]	Paracetamol	Swedish language development scale†	Not mentioned	Validated in Sweden
Liew 2016b [124]	Paracetamol	TEACh-5	The authors do not specify the reliability or validity but refers readers interested in psychometric properties to two papers on the psychometrics of the instrument	
Liew 2016c [125]	Paracetamol	WPPSI-R, shortened	Not mentioned	Full-scale IQ was derived from the selected items. Danish WPPSI-R norms were not available, so Swedish norms were used. Therefore the distribution of IQ scores in the sample does not have a mean of 100 and SD of 15
Child (6–12 years)				
Markovic 2019 [135]	NSAIDs	SON-R	Reliability of the tests used in the study was 0.73 and 0.71	Correlation to full scale was r = 0.86. Validated in Dutch
Laue 2019 [121]	Paracetamol	WISC-IV	Not mentioned	WISC is objective and less biased than parental report
ii. Assessment by parents				
Infants (<2 years)				
Vlenterie 2016 [42]	Paracetamol	Motor milestone questionnaire, ASQ, CBCL, EAS	Motor milestone is believed to be objective and therefore a reliable maternal report on motor development. ASQ: Not mentioned. CBCL: Not mentioned. EAS: The short form was as reliable and precise as the full scale	Motor milestones: Not mentioned. ASQ: Validated in Norway. CBCL and EAS: Parent-reported behaviour outcomes can suffer from differential misclassification

(*Continued*)

Table 3. (Continued)

			Psychometric properties of outcome measure	
Reference	**Exposure**	**Outcome measure**	**Reliability**	**Validity**
Wood 2016a [140]	Triptans	ASQ, CBCL, EAS	ASQ: The questions had excellent test–retest reliability and agreement between parents and professional examiners. CBCL: Not mentioned. EAS: In a Norwegian sample, internal consistency (α) within each scale ranged from 0.48 to 0.79	The ASQ is predictive of school performance. The short version has been validated in Norway and in young children. CBCL: The shortened CBCL has been validated in Norway and in young children. Parents are better reporters of externalising symptoms, whereas children are better reporters of internalising symptoms. EAS has been validated in children as young as those studied
Preschool (2–5 years)				
Markovic 2019 [135]	NSAIDs	CBCL	Internal consistency (α) 0.68	Good validity, validated in Dutch. The subscales had good fit in international studies in diverse societies
Brandlistuen 2013 [120]	NSAIDs and paracetamol	Motor milestone questionnaire, ASQ, CBCL, EAS	Maternal reports of gross motor milestone attainment have been reported to be highly reliable. ASQ: Not mentioned. CBCL: Not mentioned. EAS: The short form was as reliable and precise as the full scale	Motor milestones: Not mentioned. ASQ: Validated in Norway. Average factor loading 0.61 for fine motor and 0.75 for gross motor items, adequate reliability. Average factor loading 0.82 for communication, good reliability. (Comment from review authors: According to COSMIN principles, factor loadings are considered part of construct validity). CBCL: Correlation to CBCL full scale was 0.92. Average factor loading 0.58 for externalising and 0.52 for internalising behaviour scales, adequate reliability. EAS: Factor loading for emotionality 0.71, activity 0.68, for sociability 0.58, for shyness 0.69
Liew 2014 [122]	Paracetamol	SDQ	Reliable screening tool	With the cut-off used, the scale has high specificity for ADHD-like behaviours and 17% of children with problems on the SDQ have received a diagnosis of HKD
Liew 2016b [124]	Paracetamol	BRIEF	The authors do not specify the reliability or validity but refers readers interested in psychometric properties to two papers on the psychometrics of the instrument	
Skovlund 2017 [128]	Paracetamol and opioids	Intelligibility/Complexity of 3-year-old Children's Utterances & ASQ	Not mentioned	Parental report is considered a valid measure and validation against clinical assessment has been described for the instrument in the cohort
Wood 2016b [141]	Triptans	CBCL	Not mentioned	The shortened CBCL has been validated in Norway. The domains used, have been shown to predict later psychopathology in children and adolescents
Wood 2016c [142]	Triptans	ASQ, EAS	Not mentioned	The ASQ is predictive of school performance. This short version has been validated in Norway. EAS: Not mentioned
Harris 2018 [139]	Triptans	ASQ, CBCL, EAS	Not mentioned in the article, but in the supplement internal consistency is reported quantitatively for each subscale	Not mentioned
Child (6–12 years)				
Stergiakouli 2016 [129]	Paracetamol	SDQ	SDQ is a validated and reliable screening instrument	
Tovo-Rodrigues 2018 [131]	Paracetamol	SDQ	Not mentioned	Validated for a Brazilian population and for the studied age group

(*Continued*)

Table 3. (Continued)

			Psychometric properties of outcome measure	
Reference	**Exposure**	**Outcome measure**	**Reliability**	**Validity**
Thompson 2014 [130]	Paracetamol and ASA	SDQ, CPRS:R-L (only for paracetamol)	SDQ has a test-retest reliability of 0.62 after 4 to 6 months. Internal consistencies of the subscales range from 0.62 to 0.75	Self-reported problem behaviour has been shown to be a more valid indicator of mental and physical health than parent-reported problems
Ruisch 2018 [127]	Paracetamol and ASA	Development and Well-Being Assessment	Inter-rater differences between maternal and teacher assessments could reflect different behaviours in different settings or bias in the assessment	Good validity
iii. Assessment by teachers/others				
Preschool (2–5 years)				
Avella-Garcia 2016 [118]	Paracetamol	California Preschool Social Competence Scale, ADHD, DSM-IV form list	California Preschool Social Competence Scale: Cronbach's α 0.89. ADHD, DSM-IV form list: Cronbach's α 0.90	Not mentioned
Liew 2016b [124]	Paracetamol	BRIEF	The authors do not specify the reliability or validity but refers readers interested in psychometric properties to two papers on the psychometrics of the instrument	
Child (6–12 years)				
Markovic 2019 [135]	NSAIDs	CBCL	Not mentioned	Not mentioned
Thompson 2014 [130]	Paracetamol and ASA	SDQ (children)	SDQ has a test-retest reliability of 0.62 after 4 to 6 months. Internal consistencies of the subscales range from 0.62 to 0.75	Self-reported problem behaviour has been shown to be a more valid indicator of mental and physical health than parent-reported problems
Ruisch 2018 [127]	Paracetamol and ASA	Development and Well-Being Assessment	Inter-rater differences between maternal and teacher assessments could reflect different behaviours in different settings or bias in the assessment	Good validity
Assessment by medical diagnosis				
Rubenstein 2019 [143]	Opioids	Developmental delay and ASD	Not mentioned	All children in the study were screened for ASD and had general developmental evaluations by a clinician. Children with positive screenings had comprehensive ASD evaluations
Janecka 2018 [72]	Triptans and paracetamol	ASD	Not mentioned	Not mentioned
Liew 2016a [123]	Paracetamol	Infantile autism and ASD	Not mentioned	Diagnoses of ASD were ascertained from the general and psychiatric hospital registries in Denmark using standardised diagnostic criteria. Diagnoses of infantile autism in the psychiatric registry have previously been shown to have high validity
Liew 2014 [122]	Paracetamol	HKD	Not mentioned	Children who received diagnoses solely prior to 5 years of age were not considered as having HKD due to higher diagnostic uncertainty at younger ages. 79% of all children diagnosed with HKD had redeemed medications at least twice
Liew 2019 [126]	Paracetamol	ADHD	Diagnoses of ADHD were ascertained through maternal report. This method is reliable	In a validation study, all girls with maternal report of ADHD scored above 90% on ADHD Rating Scale-IV, as did 64% of boys. ADHD prevalence in the study was comparable to estimates by Centers for Disease Control and Prevention

(*Continued*)

Table 3. (Continued)

			Psychometric properties of outcome measure	
Reference	**Exposure**	**Outcome measure**	**Reliability**	**Validity**
Ystrom 2017 [132]	Paracetamol	ADHD	Not mentioned	Children who received diagnosis before 3 years of age were excluded. Cases were children with one or more diagnoses from specialist health care

*Structured interview of parents by the health care professional

† Mixture of parental questionnaires and nurse observation

ADHD: Attention Deficit Hyperactivity Disorder, ASA: Acetylsalicylic acid, ASD: Autism Spectrum Disorder, ASQ: Ages and Stages Questionnaire, BRIEF: Behaviour Rating Inventory of Executive Function, BSID: Bayley Scales of Infant Development, CAST: Childhood Autism Spectrum Test, CBCL: Child Behaviour Checklist, CPRS: R-L: Conners' Parent Rating Scale, revised, long format, EAS: Emotionality, Activity, Sociability Temperament Survey, HKD: Hyperkinetic disorder, SDQ: Strengths and Difficulties Questionnaire, SON-R: Snijders–Oomen Niet-verbale intelligentie Test–Revisie, TEACh-5: Test of everyday attention, 5 years, WISC: Wechsler Intelligence Scale for Children, WPPSI: Wechsler Preschool and Primary Scale of Intelligence.

https://doi.org/10.1371/journal.pone.0219778.t003

the names of maternal illnesses that are indications for use of the studied medications. Publication bias could not be ruled out for most of the medication groups. Although this may affect the interpretation of effect sizes and risks of using the medications in pregnancy, we consider it unlikely to affect how authors report the psychometric properties of the instruments they use. Finally, the reviewers were not blinded to study authors when assessing study eligibility and quality. This could be a limitation as many of the included studies were done in our research group. However, the use of predefined criteria for eligibility and quality assessment should decrease the risk of bias.

Points for consideration

There are many factors influencing study design in this area, however a lack of current consensus leads to incompatible data across studies which undoubtedly prolongs the period of time it takes to confirm safety or risks to the foetus. Based on the systematic literature review and our experiences as researchers and clinicians, we provide five points that should be considered for the conduct and interpretation of studies on maternal prenatal use of medications, and child neurodevelopment.

A wide variety of outcomes must be assessed to establish neurodevelopmental safety. The human brain is complex and its functions are diverse. Whilst there is comorbidity between neurodevelopmental disorders [24,144], all domains of neurodevelopment (e.g. IQ, language efficiency, attention etc.) are important for children's daily living, and it should be considered that they may be differentially impacted upon by teratogen exposure. Therefore a call for complete consensus on how to measure neurodevelopment, for instance by selecting one domain of neurodevelopment as the priority in medication safety research, is not reasonable. A complete consensus on choice of outcome measure may also be difficult to obtain, as a single outcome measure does not exist which can reliably measure all diverse aspect of neurodevelopment, and different populations may require different measures due to variables such as age and geographical region. All outcome measures have strengths and weaknesses that we will elaborate on below, and validation of psychometric instruments and diagnoses is done using other psychometric instruments or diagnostic categories as references, making the discussion of validity relativistic. Despite these challenges, to be able to build on each other's research and pool results in meta-analyses, it is necessary that some agreement should be reached regarding a core outcome set for teratology studies investigating neurodevelopment,

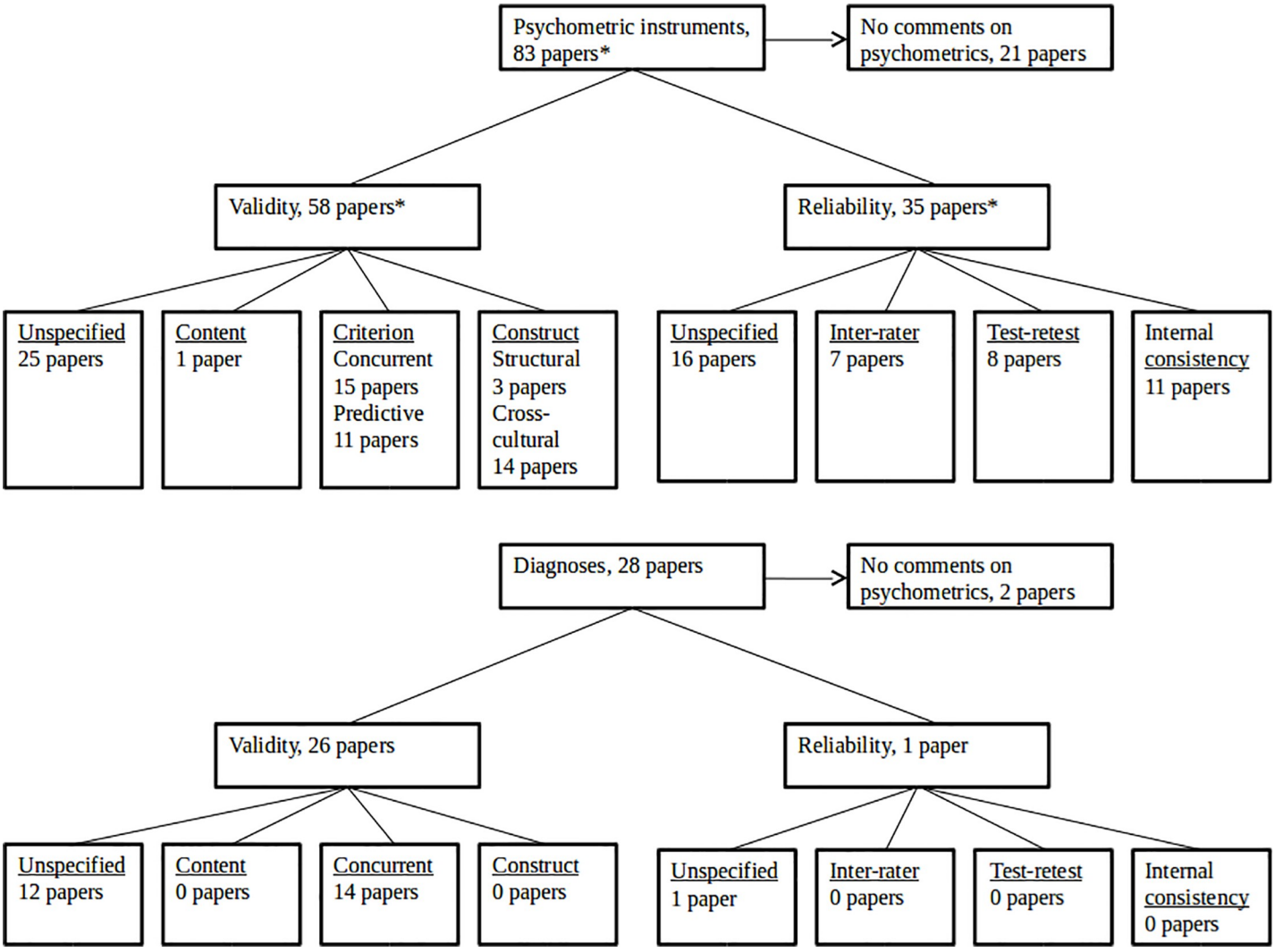

Fig 4. Psychometric properties of the neurodevelopment outcome measure mentioned in medication safety papers on psychotropics and analgesics. *Some papers commented on both validity and reliability, and those papers that commented on specific types of validity or reliability could comment on more than on type. One study used both diagnoses and psychometric instruments. Therefore the numbers add up to more than the total number of papers.

https://doi.org/10.1371/journal.pone.0219778.g004

so studies assessing the same domain of neurodevelopment and using the same data source also would use compatible outcome measures and report results in a uniform manner.

Previous literature, both animal and human, should inform choice of outcomes to be measured. In order to bring about a more uniform approach to the study of central nervous system acting medications in pregnancy and the potential impact on the developing brain, new research should select primary outcomes guided by previous literature from both animal and human studies. Reviews of the pre-clinical research, or knowledge of the literature on in-vitro or animal studies, are necessary along with those of the already available human literature. Prenatal exposure to the antiepileptic medication valproate demonstrates this point. Despite early case reports noting impaired human neurodevelopment alongside major congenital malformations, Pregnancy Registers around the world were established to assess major congenital malformation risk but not neurodevelopment [145]. Further, an early review of the pre-clinical research data would have added further weight for the requirement to study

neurodevelopment following valproate exposure with the same gravitas as major congenital malformations. As early as 1996 an association between valproate and ASD like behaviours was noted [146], whilst ASD in exposed children was not the focus of a prospective investigation until much later [22].

Data sources have different strengths and weaknesses and should complement each other. Standardised psychometric instruments completed blinded to exposure status by health care professionals are considered the gold standard to assess certain areas of neurodevelopment, e.g. Bayley Scales of Infant Development to assess early development [57], or Wechsler Preschool and Primary Scale of Intelligence to assess child IQ [58]. Such clinical assessments are often detailed, providing comprehensive information on a number of neurodevelopmental outcomes. A strength of assessment by health care professionals for research purposes is that blinding of the assessors can reduce unconscious bias. It is therefore important that blinding takes place when possible and that authors state whether the assessors were blinded when reporting a study. In this review, blinding status was reported in half of the 52 papers that used assessments by health care professionals (S5–S7 Tables). A limitation to the use of licenced psychometric instruments is costliness, training of assessors, and the amount of time spent on each assessment. In addition, families will be required to give up time for the assessment. Therefore we see in this review that studies using psychometric instruments completed by health care professionals often are small in size and in some cases only powered to detect the largest of group differences. In small studies, where random error may impact results, it is important to report on the reliability of the outcome measure used. For all psychometric instruments, the concurrent and predictive validity should be considered, as well as content validity to enable clinicians and researchers to determine whether the symptoms or traits evaluated by the scales are in fact clinically relevant to the specific population being investigated.

There are limited opportunities for health care professionals to measure child behaviour and emotionality in addition to cognitive development, as a clinic will rarely provide natural settings to observe these domains. In addition, emotionality is dependent on both situation and relation to the assessor. Therefore these domains will often rely on parent, teacher, or child reporting. In settings where few children are looked after in day care centres, parents will often be the only ones who see preschool children on a sufficiently regular basis to provide assessments of behaviour or emotionality.

Often less burdensome for families, and, in some cases, available to research groups that do not have access to licenced psychologists, psychometric instruments completed by parents or teachers can be used in large samples. There are two main weaknesses in parent reporting. One is the lack of blinding to exposure status, which is particularly important for medications that have received media attention as having unfavourable effects on the developing brain. The other is specific to psychotropics, namely distortion bias, the influence of maternal mood on the assessment. Whether maternal mood will affect reporting is at present disputed [64]. Some studies indicate that mothers with no emotional disorders will underrate child problems, while mothers with emotional disorders will overrate [147,148], whereas other studies do not find clinically significant effects of maternal emotional disorder [149,150].

Teacher reporting may be blinded to exposure status. However, the expectations of children in a classroom setting may be different from what is expected from children elsewhere. One review concluded that teacher reporting results in a higher prevalence of ADHD than if the disease is classified according to diagnostic criteria [151]. In addition, not all domains of neurodevelopment may be assessed with equal ease in a classroom setting. Hence in a meta-analysis of 119 studies, parent reporting of emotional problems correlated better than teacher reporting with children's own assessments [152]. When parent and teacher reporting only

show moderate correlations, it is important to consider that they represent assessment of the child in very different settings. Some children have problems at school that are not present in the home-environment. Hence one assessment is not necessarily more correct than the other. In both parent and teacher reporting, the concurrent and predictive validity should be considered, as well as content validity to enable clinicians and researchers to determine whether the symptoms or traits evaluated by the scales are in fact clinically relevant problems.

When using the presence or absence of a diagnosis (i.e. ASD or ADHD) to assess neurodevelopment, we clearly only examine the most affected individuals. In countries or regions with health registries, this data source is comparatively cheap and fast. However, detection bias is always possible, and blinding of assessors rating the outcome is not possible. For instance, women exposed to a suspected teratogen or women with a history of mental illness may be more likely to get their children examined for mental illnesses. Further, the presence or absence of diagnoses can be a somewhat crude measure and may differ by region, country, or version of the diagnostic manual in their criteria. Often registries contain data from public secondary services, wherefore children managed in primary care or in private hospitals may be misclassified. In addition, not all children with clinically relevant problems will fulfil the diagnostic criteria for a certain disorder. As an example, the known developmental neurobehavioural teratogen valproate increases the prevalence of ASD in children from 1.8% in general population controls to 8–15% in prenatally exposed children [22]. However, the clinical picture in these cases of ASD is atypical [153]. It is possible that the medications we investigate will increase the risk of a syndrome that is not caught by common diagnostic criteria. Lessons learned from congenital malformations show that minor deviations should not be overlooked as they might be part of specific diagnosis [154]. When using the presence or absence of diagnoses as a neurodevelopmental outcome, the authors should address both the validity of the diagnostic criteria, and the validity of the recording of diagnoses in the registries the data stem from. The specificity of diagnoses from registries can be increased by requiring that a diagnosis should be present at least twice in a child's medical records before the child is considered as having that diagnosis [155]. This will exclude the instances where a child is evaluated to rule out a diagnosis.

As the different data sources have different strengths and weaknesses, they can be used to complement each other. So far, a minority of studies have used a mixture of data sources, and only one used assessment by both diagnoses and psychometric instruments [122]. Future studies should to a greater extent use more than one data source to measure neurodevelopment if the expertise within the research group allows. For an example of how this could be done, see the paper by Liew and colleagues [122].

In meta-analyses, the use of different data sources can be challenging. Currently, review authors differ on whether to combine outcome measures from different data sources in meta-analyses [16,156–158], or summarise the evidence qualitatively [15,159]. Until a consensus is reached on which outcome measures to use and which to combine, we would like to caution over the combining of data from different research methodologies. Further, given the diversity of neurodevelopmental outcomes and their measurement there is a requirement of in-depth knowledge of the various outcome measures, as they are often based on different constructs reflecting different neurodevelopmental domains at different ages. Finally, standard approaches to meta-analysis of data including publication bias and heterogeneity in outcome measures between studies should be taken into account per outcome using standardised methods such as funnel plots.

The outcome measure should be age appropriate. Standardised tests and questionnaires are often validated for a certain age group. If the outcome measure is to be used in a different age group, it should first be validated for use in that age group [19]. Researchers using

diagnoses as outcomes should be aware that certain diagnoses are not valid below a certain age. For example, the American Academy of Pediatrics recommends that children should be 4 years old before DSM-IV diagnostic criteria of ADHD can be used [144]. Children should not be considered as having a diagnosis, if they only have a diagnosis recorded at an age where it is considered implausible that a correct diagnosis can be made. Length of follow-up should be guided by the average age at diagnosis for the specific disorder in the country where the research is carried out.

When planning new research on medication safety for neurodevelopment, it should also be considered that brain development continues into early adulthood [25], and that some difficulties in certain cognitive domains will not be detectable until the teenage years, when more complex cognitive processing is required [13]. For example, very different levels of inhibition or reasoning ability are expected from a 3-year-old and a 13-year-old. Longer follow-up would also allow investigation into mental disorders that may have developmental origins, but that have their onset in adolescence, such as schizophrenia. Another way to take into account the continuing development of the child brain is to investigate trajectories of development. This method is common in psychology [160], and could be employed in medication safety studies as well, if children are assessed at several points in time.

Reliability and validity of the outcome measure should be reported. The use of several different outcome measures across studies makes it difficult for readers to be familiar with all the different measures. This increases the responsibility of authors to provide information on validity and reliability. Many of the studies in this review were large, including several hundred exposed pregnancies, thus limiting the risk of random error. In these studies, the most important psychometric property to report is validity, as invalid measures may introduce bias.

In psychology, construct validity is often considered the most important form of validity, as there are no objective criteria or "gold standards" to compare to [161]. Quite surprisingly, only three papers mentioned structural validity, one of which provided quantitative measures, and no papers explicitly mentioned hypothesis testing. Construct validity can be evaluated using statistical methods, and therefore numbers ought to be reported.

Criterion validity can also be evaluated using statistical methods. About a third of the papers provided a comment on criterion validity for at least one of their outcome measures, however only 15 reported a quantitative measure of criterion validity. Concurrent validity, the performance of the psychometric instrument or diagnosis against a gold standard, was reported for 28 papers. Specifically for diagnoses in registry-based studies, concurrent validity can both refer to the validity of the diagnostic criteria, and to the validity of the registration of the investigated diagnoses in the particular registry the data stems from. However, only the latter was mentioned in the papers identified in this review. Predictive validity is mainly relevant to psychometric instruments, and was only mentioned by 11 papers. Predictive validity is important for measures used in children, as children may grow into or out of difficulties [22].

Only one paper mentioned content validity [55], the degree to which the questions or tasks that make up a psychometric instrument, or the criteria that make up a diagnosis, are relevant and comprehensive measures of the domain of neurodevelopment that the outcome measure is used to investigate [19]. Content validity cannot be evaluated with the use of statistical methods. As such it is more subjective than the other forms of validity, which may make authors hesitant to comment on it. However, expert group evaluation of content validity has been done for diagnostic criteria and for some psychometric instruments [162,163], and could be reported. Another option is to make the questions or tasks of an outcome measure available to readers in a supplement (if copyright allows), so readers can assess content validity.

If the content or criterion validity is low or unknown, this should be reflected in the language used in the paper. For example, if a study uses a psychometric instrument to assess the

presence of ADHD, and the instrument has not shown acceptable validity when tested against diagnostic interviews for ADHD, authors should write that they have assessed "symptoms of ADHD" and not "ADHD" as such.

In studies where a psychometric instrument is used in a different language or for another population than that in which it was developed, cross-cultural validity will tell us the extent to which the instrument measures the same as the original instrument. Without validation it cannot be assumed that the outcome measure is valid in a different population from the one for which it was developed [19]. Yet, cross-cultural validity alone will not be a sufficient measure of validity, as it does not provide any information on whether the original psychometric instrument measures what is intended.

We recognise that many journals have word limits for articles, making it difficult to include detailed information on reliability and validity. However, given the importance of this information it is suggested that this is prioritised. Today many journals allow online supplements, where psychometric properties can be described. For an example of how this could be done, see the online supplement to the paper by Avella-Garcia and colleagues [118]. Many papers identified in this review only referred to the manual of the outcome measure they use, which is often not accessible to the readers.

Finally, other methodological issues than use and validity of outcome measure can introduce between-study heterogeneity and should be considered. Some of these issues are the study limitations in observational studies according to GRADE [27], as assessed in S5–S7 Tables. Other examples include choice of appropriate comparator group to handle confounding by indication of medication use [156], analyses of direct and indirect medication effects by taking into account postnatal factors [164], and analysis of medication use by timing, dose and/or duration [165]. Interested readers are referred to a recent review on these methodological issues in medication safety studies with central nervous system outcomes [166].

Conclusions

Studies have used several outcome measures including diagnoses and psychometric instruments completed by health care professionals, parents, or teachers to assess child neurodevelopment, yet few studies reported adequately on the reliability and validity of their outcomes. In order to establish neurodevelopmental safety of prenatal exposure to a medication, it is necessary to assess several domains of neurodevelopment until adolescence using age appropriate outcome measures. For medications where an animal model exists, this should inform which outcomes are assessed first. Authors should use reliable and valid outcome measures to assess neurodevelopment. We encourage reporting on the validity and reliability of the outcome measures used. In addition, results should be interpreted in light of the reliability and validity of the outcome measure that is used. Consensus is required on which outcome measure to use for each age group and data source in each domain of neurodevelopment. Until such consensus is in place, researchers should to a larger extent combine different data sources in one study, and authors of meta-analyses should be aware that in-depth knowledge of the various outcome measures is necessary when deciding which outcomes can be combined.

Supporting information

S1 File. Definitions of different types of reliability and validity.
(PDF)

S2 File. Search terms and search strategy, example for the MEDLINE database.
(PDF)

S3 File. PRISMA checklist.
(PDF)

S1 Table. Study characteristics, papers on antidepressants.
(PDF)

S2 Table. Study characteristics, papers on psychotropics except for antidepressants.
(PDF)

S3 Table. Study characteristics, papers on analgesics.
(PDF)

S4 Table. Reliability and validity of the outcome measures used in studies on neurodevelopmental safety of prenatal exposure to psychotropics and analgesics.
(PDF)

S5 Table. Risk of bias assessments, papers on antidepressants.
(PDF)

S6 Table. Risk of bias assessments, papers on psychotropics except for antidepressants.
(PDF)

S7 Table. Risk of bias assessments, papers on analgesics.
(PDF)

Author Contributions

Conceptualization: Sarah Hjorth, Rebecca Bromley, Eivind Ystrom, Angela Lupattelli, Olav Spigset, Hedvig Nordeng.

Data curation: Sarah Hjorth.

Formal analysis: Sarah Hjorth.

Funding acquisition: Hedvig Nordeng.

Investigation: Sarah Hjorth.

Supervision: Hedvig Nordeng.

Validation: Sarah Hjorth, Rebecca Bromley, Angela Lupattelli, Hedvig Nordeng.

Visualization: Sarah Hjorth.

Writing – original draft: Sarah Hjorth, Rebecca Bromley.

Writing – review & editing: Rebecca Bromley, Eivind Ystrom, Angela Lupattelli, Olav Spigset, Hedvig Nordeng.

References

1. EUROmediCAT Steering Group. EUROmediCAT Recommendations for European Pharmacovigilance concerning safety of medication use in pregnancy: ABSTRACT. Pharmacoepidemiol Drug Saf. 2015 Oct; 24:3–7. https://doi.org/10.1002/pds.3866 PMID: 26395593
2. Nordeng H, Lupattelli A, Wood M. Prenatal exposure to antidepressants and increased risk of psychiatric disorders. BMJ. 2017 Sep 6;j3950. https://doi.org/10.1136/bmj.j3950 PMID: 28877910
3. Oberlander TF, Gingrich JA, Ansorge MS. Sustained neurobehavioral effects of exposure to SSRI antidepressants during development: molecular to clinical evidence. Clin Pharmacol Ther. 2009; 86 (6):672–7. https://doi.org/10.1038/clpt.2009.201 PMID: 19890255

4. de Fays L, Van Malderen K, De Smet K, Sawchik J, Verlinden V, Hamdani J, et al. Use of paracetamol during pregnancy and child neurological development. Dev Med Child Neurol. 2015; 57(8):718–24. https://doi.org/10.1111/dmcn.12745 PMID: 25851072
5. Broussard CS, Frey MT, Hernandez-Diaz S, Greene MF, Chambers CD, Sahin L, et al. Developing a systematic approach to safer medication use during pregnancy: summary of a Centers for Disease Control and Prevention–convened meeting. Am J Obstet Gynecol. 2014 Sep; 211(3):208–214.e1. https://doi.org/10.1016/j.ajog.2014.05.040 PMID: 24881821
6. Lupattelli A, Spigset O, Twigg MJ, Zagorodnikova K, Mårdby AC, Moretti ME, et al. Medication use in pregnancy: a cross-sectional, multinational web-based study. BMJ Open. 2014 Feb; 4(2):e004365. https://doi.org/10.1136/bmjopen-2013-004365 PMID: 24534260
7. Werler MM, Mitchell AA, Hernandez-Diaz S, Honein MA, the National Birth Defects Prevention Study. Use of over-the-counter medications during pregnancy. Am J Obstet Gynecol. 2005 Sep; 193(3):771–7.
8. Bakker MK, Kölling P, van den Berg PB, de Walle HEK, de Jong van den Berg LTW. Increase in use of selective serotonin reuptake inhibitors in pregnancy during the last decade, a population-based cohort study from the Netherlands. Br J Clin Pharmacol. 2008 Apr; 65(4):600–6. https://doi.org/10.1111/j.1365-2125.2007.03048.x PMID: 17953715
9. Jimenez-Solem E, Andersen JT, Petersen M, Broedbaek K, Andersen NL, Torp-Pedersen C, et al. Prevalence of antidepressant use during pregnancy in Denmark, a nation-wide cohort study. PloS One. 2013; 8(4):e63034. https://doi.org/10.1371/journal.pone.0063034 PMID: 23638179
10. Mitchell AA, Gilboa SM, Werler MM, Kelley KE, Louik C, Hernández-Díaz S. Medication use during pregnancy, with particular focus on prescription drugs: 1976–2008. Am J Obstet Gynecol. 2011 Jul; 205(1):51.e1-51.e8.
11. Jarde A, Morais M, Kingston D, Giallo R, MacQueen GM, Giglia L, et al. Neonatal Outcomes in Women With Untreated Antenatal Depression Compared With Women Without Depression: A Systematic Review and Meta-analysis. JAMA Psychiatry. 2016 Aug 1; 73(8):826. https://doi.org/10.1001/jamapsychiatry.2016.0934 PMID: 27276520
12. Tosato S, Albert U, Tomassi S, Iasevoli F, Carmassi C, Ferrari S, et al. A Systematized Review of Atypical Antipsychotics in Pregnant Women: Balancing Between Risks of Untreated Illness and Risks of Drug-Related Adverse Effects. J Clin Psychiatry. 2017 May; 78(5):e477–89. https://doi.org/10.4088/JCP.15r10483 PMID: 28297592
13. Bromley RL, Baker GA. Fetal antiepileptic drug exposure and cognitive outcomes. Seizure. 2017 Jan; 44:225–31. https://doi.org/10.1016/j.seizure.2016.10.006 PMID: 27784632
14. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 2013 May 14; 11:126. https://doi.org/10.1186/1741-7015-11-126 PMID: 23672542
15. El Marroun H, White T, Verhulst FC, Tiemeier H. Maternal use of antidepressant or anxiolytic medication during pregnancy and childhood neurodevelopmental outcomes: a systematic review. Eur Child Adolesc Psychiatry. 2014; 23(10):973–92. https://doi.org/10.1007/s00787-014-0558-3 PMID: 24863148
16. Masarwa R, Levine H, Gorelik E, Reif S, Perlman A, Matok I. Prenatal Exposure to Acetaminophen and Risk for Attention Deficit Hyperactivity Disorder and Autistic Spectrum Disorder: A Systematic Review, Meta-Analysis, and Meta-Regression Analysis of Cohort Studies. Am J Epidemiol. 2018 Aug; 187(8):1817–27. https://doi.org/10.1093/aje/kwy086 PMID: 29688261
17. Kraemer HC, Kupfer DJ, Clarke DE, Narrow WE, Regier DA. DSM-5: How Reliable Is Reliable Enough? Am J Psychiatry. 2012 Jan; 169(1):13–5. https://doi.org/10.1176/appi.ajp.2011.11010050 PMID: 22223009
18. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Reliability. In: Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press; 2011. p. 96–149.
19. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Validity. In: Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press; 2011. p. 150–201.
20. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Concepts, theories and models, and types of measurements. In: Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press; 2011. p. 7–29.
21. WHO. WHOCC—ATC/DDD Index [Internet]. WHO Collaborating Centre for Drug Statistics Methodology. 2017 [cited 2018 Jul 6]. Available from: https://www.whocc.no/atc_ddd_index/
22. Bromley R, Weston J, Adab N, Greenhalgh J, Sanniti A, McKay AJ, et al. Treatment for epilepsy in pregnancy: neurodevelopmental outcomes in the child. In: The Cochrane Collaboration, editor. Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd; 2014.

23. Veroniki AA, Rios P, Cogo E, Straus SE, Finkelstein Y, Kealey R, et al. Comparative safety of antiepileptic drugs for neurological development in children exposed during pregnancy and breast feeding: a systematic review and network meta-analysis. BMJ Open. 2017 Jul; 7(7):e017248. https://doi.org/10.1136/bmjopen-2017-017248 PMID: 28729328

24. WHO. ICD-10 Version:2016 [Internet]. International Statistical Classification of Diseases and Related Health Problems 10th Revision. 2016 [cited 2017 Sep 19]. Available from: http://apps.who.int/classifications/icd10/browse/2016/en

25. Andersen SL. Trajectories of brain development: point of vulnerability or window of opportunity? Neurosci Biobehav Rev. 2003 Jan; 27(1–2):3–18. PMID: 12732219

26. Covidence. Covidence—Accelerate your systematic review [Internet]. 2017 [cited 2017 Oct 11]. Available from: https://www.covidence.org/

27. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol. 2011 Apr; 64(4):407–15. https://doi.org/10.1016/j.jclinepi.2010.07.017 PMID: 21247734

28. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2010 May; 19(4):539–49.

29. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. J Clin Epidemiol. 2011 Dec; 64(12):1277–82. https://doi.org/10.1016/j.jclinepi.2011.01.011 PMID: 21802904

30. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J: L. Erlbaum Associates; 1988. 567 p.

31. Kontopantelis E, Reeves D. METAEFF: Stata module to perform effect sizes calculations for meta-analyses [Internet]. 2011 [cited 2018 Jul 6]. Available from: http://EconPapers.repec.org/RePEc:boc:bocode:s457072

32. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LP; 2017.

33. Fritz CO, Morris PE, Richler JJ. Effect size estimates: Current use, calculations, and interpretation. J Exp Psychol Gen. 2012; 141(1):2–18. https://doi.org/10.1037/a0024338 PMID: 21823805

34. Boukhris T, Sheehy O, Bérard A. Antidepressant Use in Pregnancy and the Risk of Attention Deficit with or without Hyperactivity Disorder in Children. Paediatr Perinat Epidemiol. 2017 Jul; 31(4):363–73. https://doi.org/10.1111/ppe.12378 PMID: 28640459

35. Hartz SC, Heinonen OP, Shapiro S, Siskind V, Slone D. Antenatal Exposure to Meprobamate and Chlordiazepoxide in Relation to Malformations, Mental Development, and Childhood Mortality. N Engl J Med. 1975 Apr 3; 292(14):726–8. https://doi.org/10.1056/NEJM197504032921405 PMID: 1113782

36. Heikkinen T, Ekblad U, Palo P, Laine K. Pharmacokinetics of fluoxetine and norfluoxetine in pregnancy and lactation. Clin Pharmacol Ther. 2003 Apr; 73(4):330–7. https://doi.org/10.1016/s0009-9236(02)17634-x PMID: 12709723

37. Laugesen K, Olsen MS, Telén Andersen AB, Frøslev T, Sørensen HT. In utero exposure to antidepressant drugs and risk of attention deficit hyperactivity disorder: a nationwide Danish cohort study. BMJ Open. 2013 Sep 20; 3(9):e003507. https://doi.org/10.1136/bmjopen-2013-003507 PMID: 24056487

38. Reebye PN, Ng TWC, Misri S, Stikarovska I. Affect Expression and Self-Regulation Capacities of Infants Exposed in utero to Psychotropics. Front Psychiatry. 2012; 3.

39. Streissguth AP, Treder RP, Barr HM, Shepard TH, Bleyer WA, Sampson PD, et al. Aspirin and acetaminophen use by pregnant women and subsequent child IQ and attention decrements. Teratology. 1987 Apr; 35(2):211–9. https://doi.org/10.1002/tera.1420350207 PMID: 3603404

40. Viggedal G, Hagberg BS, Laegreid L, Aronsson M. Mental development in late infancy after prenatal exposure to benzodiazepines—a prospective study. J Child Psychol Psychiatry. 1993 Mar; 34(3):295–305. PMID: 8463369

41. Viktorin A, Uher R, Reichenberg A, Levine SZ, Sandin S. Autism risk following antidepressant medication during pregnancy. Psychol Med. 2017 Dec; 47(16):2787–96. https://doi.org/10.1017/S0033291717001301 PMID: 28528584

42. Vlenterie R, Wood ME, Brandlistuen RE, Roeleveld N, van Gelder MMHJ, Nordeng H. Neurodevelopmental problems at 18 months among children exposed to paracetamol in utero: a propensity score matched cohort study. Int J Epidemiol. 2016 Aug 31;dyw192.

43. Austin M-P, Karatas JC, Mishra P, Christl B, Kennedy D, Oei J. Infant neurodevelopment following in utero exposure to antidepressant medication. Acta Paediatr Oslo Nor 1992. 2013; 102(11):1054–9.

44. Batton B, Batton E, Weigler K, Aylward G, Batton D. In utero antidepressant exposure and neurodevelopment in preterm infants. Am J Perinatol. 2013; 30(4):297–301. https://doi.org/10.1055/s-0032-1324697 PMID: 22893558

45. Boukhris T, Sheehy O, Mottron L, Berard A. Antidepressant Use During Pregnancy and the Risk of Autism Spectrum Disorder in Children. JAMA Pediatr. 2016; 170(2):117–24. https://doi.org/10.1001/jamapediatrics.2015.3356 PMID: 26660917

46. Brandlistuen RE, Ystrom E, Eberhard-Gran M, Nulman I, Koren G, Nordeng H. Behavioural effects of fetal antidepressant exposure in a Norwegian cohort of discordant siblings. Int J Epidemiol. 2015 Aug; 44(4):1397–407. https://doi.org/10.1093/ije/dyv030 PMID: 25873178

47. Brown AS, Gyllenberg D, Malm H, McKeague IW, Hinkka-Yli-Salomaki S, Artama M, et al. Association of Selective Serotonin Reuptake Inhibitor Exposure During Pregnancy With Speech, Scholastic, and Motor Disorders in Offspring. JAMA Psychiatry. 2016; 73(11):1163–70. https://doi.org/10.1001/jamapsychiatry.2016.2594 PMID: 27732704

48. Brown HK, Ray JG, Wilton AS, Lunsky Y, Gomes T, Vigod SN. Association Between Serotonergic Antidepressant Use During Pregnancy and Autism Spectrum Disorder in Children. JAMA. 2017; 317 (15):1544–52. https://doi.org/10.1001/jama.2017.3415 PMID: 28418480

49. Casper RC, Fleisher BE, Lee-Ancajas JC, Gilles A, Gaylor E, DeBattista A, et al. Follow-up of children of depressed mothers exposed or not exposed to antidepressant drugs during pregnancy. J Pediatr. 2003; 142(4):402–8. https://doi.org/10.1067/mpd.2003.139 PMID: 12712058

50. Castro VM, Kong SW, Clements CC, Brady R, Kaimal AJ, Doyle AE, et al. Absence of evidence for increase in risk for autism or attention-deficit hyperactivity disorder following antidepressant exposure during pregnancy: a replication study. Transl Psychiatry. 2016; 6(101562664):e708.

51. Clements CC, Castro VM, Blumenthal SR, Rosenfield HR, Murphy SN, Fava M, et al. Prenatal antidepressant exposure is associated with risk for attention-deficit hyperactivity disorder but not autism spectrum disorder in a large health system. Mol Psychiatry. 2015; 20(6):727–34. https://doi.org/10.1038/mp.2014.90 PMID: 25155880

52. Croen LA, Grether JK, Yoshida CK, Odouli R, Hendrick V. Antidepressant use during pregnancy and childhood autism spectrum disorders. Arch Gen Psychiatry. 2011; 68(11):1104–12. https://doi.org/10.1001/archgenpsychiatry.2011.73 PMID: 21727247

53. de Vries NKS, van der Veere CN, Reijneveld SA, Bos AF. Early Neurological Outcome of Young Infants Exposed to Selective Serotonin Reuptake Inhibitors during Pregnancy: Results from the Observational SMOK Study. Scott JG, editor. PLoS ONE. 2013 May 28; 8(5):e64654.

54. El Marroun H, White TJH, van der Knaap NJF, Homberg JR, Fernandez G, Schoemaker NK, et al. Prenatal exposure to selective serotonin reuptake inhibitors and social responsiveness symptoms of autism: population-based study of young children. Br J Psychiatry J Ment Sci. 2014; 205(2):95–102.

55. El Marroun H, White TJ, Fernandez G, Jaddoe VW, Verhulst FC, Stricker BH, et al. Prenatal exposure to selective serotonin reuptake inhibitors and non-verbal cognitive functioning in childhood. J Psychopharmacol Oxf Engl. 2017; 31(3):346–55.

56. Figueroa R. Use of antidepressants during pregnancy and risk of attention-deficit/hyperactivity disorder in the offspring. J Dev Behav Pediatr JDBP. 2010; 31(8):641–8. https://doi.org/10.1097/DBP.0b013e3181e5ac93 PMID: 20613624

57. Galbally M, Lewis AJ, Buist A. Developmental outcomes of children exposed to antidepressants in pregnancy. Aust N Z J Psychiatry. 2011; 45(5):393–9. https://doi.org/10.3109/00048674.2010.549995 PMID: 21314237

58. Galbally M, Lewis AJ, Buist A. Child developmental outcomes in preschool children following antidepressant exposure in pregnancy. Aust N Z J Psychiatry. 2015; 49(7):642–50. https://doi.org/10.1177/0004867415569800 PMID: 25698806

59. Gidaya NB, Lee BK, Burstyn I, Yudell M, Mortensen EL, Newschaffer CJ. In utero exposure to selective serotonin reuptake inhibitors and risk for autism spectrum disorder. J Autism Dev Disord. 2014; 44 (10):2558–67. https://doi.org/10.1007/s10803-014-2128-4 PMID: 24803368

60. Grzeskowiak LE, Morrison JL, Henriksen TB, Bech BH, Obel C, Olsen J, et al. Prenatal antidepressant exposure and child behavioural outcomes at 7 years of age: a study within the Danish National Birth Cohort. BJOG Int J Obstet Gynaecol. 2016; 123(12):1919–28.

61. Gustafsson HC, Goodman SH, Feng T, Choi J, Lee S, Newport DJ, et al. Major depressive disorder during pregnancy: Psychiatric medications have minimal effects on the fetus and infant yet development is compromised. Dev Psychopathol. 2018 Aug; 30(3):773–85. https://doi.org/10.1017/S0954579418000639 PMID: 30068426

62. Handal M, Skurtveit S, Furu K, Hernandez-Diaz S, Skovlund E, Nystad W, et al. Motor development in children prenatally exposed to selective serotonin reuptake inhibitors: a large population-based pregnancy cohort study. BJOG Int J Obstet Gynaecol. 2016; 123(12):1908–17.

63. Handal M, Skurtveit S, Roth C, Hernandez-Diaz S, Selmer R. Prenatal Exposure to Folic Acid and Antidepressants and Language Development: A Population-Based Cohort Study. J Clin Psychopharmacol. 2016; 36(4):333–9. https://doi.org/10.1097/JCP.0000000000000519 PMID: 27285658

64. Hanley GE, Brain U, Oberlander TF. Infant developmental outcomes following prenatal exposure to antidepressants, and maternal depressed mood and positive affect. Early Hum Dev. 2013; 89(8):519–24. https://doi.org/10.1016/j.earlhumdev.2012.12.012 PMID: 23384962

65. Hanley GE, Brain U, Oberlander TF. Prenatal exposure to serotonin reuptake inhibitor antidepressants and childhood behavior. Pediatr Res. 2015; 78(2):174–80. https://doi.org/10.1038/pr.2015.77 PMID: 25897539

66. Harrington RA, Lee L-C, Crum RM, Zimmerman AW, Hertz-Picciotto I. Prenatal SSRI use and offspring with autism spectrum disorder or developmental delay. Pediatrics. 2014; 133(5):e1241–8. https://doi.org/10.1542/peds.2013-3406 PMID: 24733881

67. Heikkinen T, Ekblad U, Kero P, Ekblad S, Laine K. Citalopram in pregnancy and lactation. Clin Pharmacol Ther. 2002; 72(2):184–91. https://doi.org/10.1067/mcp.2002.126181 PMID: 12189365

68. Hermansen TK, Roysamb E, Augusti E-M, Melinder A. Behavior and inhibitory control in children with prenatal exposure to antidepressants and medically untreated depression. Psychopharmacology (Berl). 2016; 233(8):1523–35.

69. Hurault-Delarue C, Damase-Michel C, Finotto L, Guitard C, Vayssiere C, Montastruc J-L, et al. Psychomotor developmental effects of prenatal exposure to psychotropic drugs: a study in EFEMERIS database. Fundam Clin Pharmacol. 2016; 30(5):476–82. https://doi.org/10.1111/fcp.12209 PMID: 27324073

70. Hutchison SM, Masse LC, Glier MB, Brain U, Devlin AM, Oberlander TF. Impact of Prenatal Selective Serotonin Reuptake Inhibitor Antidepressant Exposure and Maternal Mood on Physical Activity, Dietary Intake, and Markers of Adiposity at Age 6 Years. J Dev Behav Pediatr JDBP. 2019 Mar;

71. Hviid A, Melbye M, Pasternak B. Use of selective serotonin reuptake inhibitors during pregnancy and risk of autism. N Engl J Med. 2013; 369(25):2406–15. https://doi.org/10.1056/NEJMoa1301449 PMID: 24350950

72. Janecka M, Kodesh A, Levine SZ, Lusskin S, Viktorin A, Rahman R, et al. Association of Autism Spectrum Disorder With Prenatal Exposure to Medication Affecting Neurotransmitter Systems. Jama Psychiatry. 2018 Dec; 75(12):1217–24. https://doi.org/10.1001/jamapsychiatry.2018.2728 PMID: 30383108

73. Johnson KC, LaPrairie JL, Brennan PA, Stowe ZN, Newport DJ. Prenatal antipsychotic exposure and neuromotor performance during infancy. Arch Gen Psychiatry. 2012; 69(8):787–94. https://doi.org/10.1001/archgenpsychiatry.2012.160 PMID: 22474072

74. Johnson KC, Smith AK, Stowe ZN, Newport DJ, Brennan PA. Preschool outcomes following prenatal serotonin reuptake inhibitor exposure: differences in language and behavior, but not cognitive function. J Clin Psychiatry. 2016; 77(2):e176–82. https://doi.org/10.4088/JCP.14m09348 PMID: 26930533

75. Liu X, Agerbo E, Ingstrup KG, Musliner K, Meltzer-Brody S, Bergink V, et al. Antidepressant use during pregnancy and psychiatric disorders in offspring: Danish nationwide register based cohort study. Bmj-Br Med J. 2017 Sep 6; 358:j3668.

76. Lupattelli A, Wood M, Ystrom E, Skurtveit S, Handal M, Nordeng H. Effect of Time-Dependent Selective Serotonin Reuptake Inhibitor Antidepressants During Pregnancy on Behavioral, Emotional, and Social Development in Preschool-Aged Children. J Am Acad Child Adolesc Psychiatry. 2018 Mar; 57 (3):200–8. https://doi.org/10.1016/j.jaac.2017.12.010 PMID: 29496129

77. Malm H, Brown AS, Gissler M, Gyllenberg D, Hinkka-Yli-Salomaki S, McKeague IW, et al. Gestational Exposure to Selective Serotonin Reuptake Inhibitors and Offspring Psychiatric Disorders: A National Register-Based Study. J Am Acad Child Adolesc Psychiatry. 2016; 55(5):359–66. https://doi.org/10.1016/j.jaac.2016.02.013 PMID: 27126849

78. Man KKC, Chan EW, Ip P, Coghill D, Simonoff E, Chan PKL, et al. Prenatal antidepressant use and risk of attention-deficit/hyperactivity disorder in offspring: population based cohort study. BMJ. 2017; 357(8900488, bmj, 101090866):j2350.

79. Mattson SN, Calarco KE, Chambers CD, Jones KL. Interaction of maternal smoking and other in-pregnancy exposures: analytic considerations. Neurotoxicol Teratol. 2002; 24(3):359–67. PMID: 12009491

80. Misri S, Reebye P, Kendrick K, Carter D, Ryan D, Grunau RE, et al. Internalizing behaviors in 4-year-old children exposed in utero to psychotropic medications. Am J Psychiatry. 2006; 163(6):1026–32. https://doi.org/10.1176/ajp.2006.163.6.1026 PMID: 16741203

81. Mortensen JT, Olsen J, Larsen H, Bendsen J, Obel C, Sorensen HT. Psychomotor development in children exposed in utero to benzodiazepines, antidepressants, neuroleptics, and anti-epileptics. Eur J Epidemiol. 2003; 18(8):769–71. PMID: 12974552

82. Netsi E, Evans J, Wulff K, O'Mahen H, Ramchandani PG. Infant outcomes following treatment of antenatal depression: Findings from a pilot randomized controlled trial. J Affect Disord. 2015 Dec; 188:252–6. https://doi.org/10.1016/j.jad.2015.08.055 PMID: 26372945

83. Nulman I, Rovet J, Stewart DE, Wolpin J, Gardner HA, Theis JGW, et al. Neurodevelopment of Children Exposed in Utero to Antidepressant Drugs. N Engl J Med. 1997 Jan 23; 336(4):258–62. https://doi.org/10.1056/NEJM199701233360404 PMID: 8995088

84. Nulman I, Rovet J, Stewart DE, Wolpin J, Pace-Asciak P, Shuhaiber S, et al. Child development following exposure to tricyclic antidepressants or fluoxetine throughout fetal life: a prospective, controlled study. Am J Psychiatry. 2002; 159(11):1889–95. https://doi.org/10.1176/appi.ajp.159.11.1889 PMID: 12411224

85. Nulman I, Koren G, Rovet J, Barrera M, Pulver A, Streiner D, et al. Neurodevelopment of children following prenatal exposure to venlafaxine, selective serotonin reuptake inhibitors, or untreated maternal depression. Am J Psychiatry. 2012; 169(11):1165–74. https://doi.org/10.1176/appi.ajp.2012.11111721 PMID: 23128923

86. Nulman I, Koren G, Rovet J, Barrera M, Streiner DL, Feldman BM. Neurodevelopment of children prenatally exposed to selective reuptake inhibitor antidepressants: Toronto sibling study. J Clin Psychiatry. 2015; 76(7):e842–7. https://doi.org/10.4088/JCP.14m09240 PMID: 26231010

87. Oberlander TF, Misri S, Fitzgerald CE, Kostaras X, Rurak D, Riggs W. Pharmacologic factors associated with transient neonatal symptoms following prenatal psychotropic medication exposure. J Clin Psychiatry. 2004; 65(2):230–7. PMID: 15003078

88. Oberlander TF, Reebye P, Misri S, Papsdorf M, Kim J, Grunau RE. Externalizing and attentional behaviors in children of depressed mothers treated with a selective serotonin reuptake inhibitor antidepressant during pregnancy. Arch Pediatr Adolesc Med. 2007; 161(1):22–9. https://doi.org/10.1001/archpedi.161.1.22 PMID: 17199063

89. Oberlander TF, Papsdorf M, Brain UM, Misri S, Ross C, Grunau RE. Prenatal effects of selective serotonin reuptake inhibitor antidepressants, serotonin transporter promoter genotype (SLC6A4), and maternal mood on child behavior at 3 years of age. Arch Pediatr Adolesc Med. 2010; 164(5):444–51. https://doi.org/10.1001/archpediatrics.2010.51 PMID: 20439795

90. Pedersen LH, Henriksen TB, Bech BH, Licht RW, Kjaer D, Olsen J. Prenatal antidepressant exposure and behavioral problems in early childhood—a cohort study. Acta Psychiatr Scand. 2013; 127(2):126–35. https://doi.org/10.1111/acps.12032 PMID: 23126521

91. Rai D, Lee BK, Dalman C, Newschaffer C, Lewis G, Magnusson C. Antidepressants during pregnancy and autism in offspring: population based cohort study. BMJ. 2017; 358(8900488, bmj, 101090866): j2811.

92. Reebye PN, Morison SJ, Panikkar H, Misri S, Grunau RE. Affect expression in prenatally psychotropic exposed and nonexposed mother-infant dyads. Infant Ment Health J. 2002 Jul; 23(4):403–16.

93. Santucci AK, Singer LT, Wisniewski SR, Luther JF, Eng HF, Dills JL, et al. Impact of prenatal exposure to serotonin reuptake inhibitors or maternal major depressive disorder on infant developmental outcomes. J Clin Psychiatry. 2014; 75(10):1088–95. https://doi.org/10.4088/JCP.13m08902 PMID: 25373117

94. Schechter JC, Brennan PA, Smith AK, Stowe ZN, Newport DJ, Johnson KC. Maternal Prenatal Psychological Distress and Preschool Cognitive Functioning: the Protective Role of Positive Parental Engagement. J Abnorm Child Psychol. 2017 Feb; 45(2):249–60. https://doi.org/10.1007/s10802-016-0161-9 PMID: 27150387

95. Simon GE, Cunningham ML, Davis RL. Outcomes of prenatal antidepressant exposure. Am J Psychiatry. 2002; 159(12):2055–61. https://doi.org/10.1176/appi.ajp.159.12.2055 PMID: 12450956

96. Skurtveit S, Selmer R, Roth C, Hernandez-Diaz S, Handal M. Prenatal exposure to antidepressants and language competence at age three: results from a large population-based pregnancy cohort in Norway. BJOG Int J Obstet Gynaecol. 2014 Dec; 121(13):1621–31.

97. Sorensen MJ, Gronborg TK, Christensen J, Parner ET, Vestergaard M, Schendel D, et al. Antidepressant exposure in pregnancy and risk of autism spectrum disorders. Clin Epidemiol. 2013; 5 (101531700):449–59.

98. Sujan AC, Rickert ME, Oberg S, Quinn PD, Hernandez-Diaz S, Almqvist C, et al. Associations of Maternal Antidepressant Use During the First Trimester of Pregnancy With Preterm Birth, Small for Gestational Age, Autism Spectrum Disorder, and Attention-Deficit/Hyperactivity Disorder in Offspring. Jama-J Am Med Assoc. 2017 Apr 18; 317(15):1553–62.

99. Suri R, Hellemann G, Stowe ZN, Cohen LS, Aquino A, Altshuler LL. A prospective, naturalistic, blinded study of early neurobehavioral outcomes for infants following prenatal antidepressant exposure. J Clin Psychiatry. 2011; 72(7):1002–7. https://doi.org/10.4088/JCP.10m06135 PMID: 21672498

100. Viktorin A, Uher R, Kolevzon A, Reichenberg A, Levine SZ, Sandin S. Association of Antidepressant Medication Use During Pregnancy With Intellectual Disability in Offspring. JAMA Psychiatry. 2017; 74 (10):1031–8. https://doi.org/10.1001/jamapsychiatry.2017.1727 PMID: 28700807

101. Weikum WM, Mayes LC, Grunau RE, Brain U, Oberlander TF. The impact of prenatal serotonin reuptake inhibitor (SRI) antidepressant exposure and maternal mood on mother-infant interactions at 3 months of age. Infant Behav Dev. 2013; 36(4):485–93. https://doi.org/10.1016/j.infbeh.2013.04.001 PMID: 23728194

102. Weikum WM, Brain U, Chau CMY, Grunau RE, Boyce WT, Diamond A, et al. Prenatal serotonin reuptake inhibitor (SRI) antidepressant exposure and serotonin transporter promoter genotype (SLC6A4) influence executive functions at 6 years of age. Front Cell Neurosci. 2013; 7(101477935):180.

103. Wibroe MA, Mathiasen R, Pagsberg AK, Uldall P. Risk of impaired cognition after prenatal exposure to psychotropic drugs. Acta Psychiatr Scand. 2017; 136(2):177–87. https://doi.org/10.1111/acps.12754 PMID: 28561934

104. Brandlistuen RE, Ystrom E, Hernandez-Diaz S, Skurtveit S, Selmer R, Handal M, et al. Association of prenatal exposure to benzodiazepines and child internalizing problems: A sibling-controlled cohort study. PloS One. 2017; 12(7):e0181042. https://doi.org/10.1371/journal.pone.0181042 PMID: 28746341

105. Gidai J, Acs N, Banhidy F, Czeizel AE. An evaluation of data for 10 children born to mothers who attempted suicide by taking large doses of alprazolam during pregnancy. Toxicol Ind Health. 2008; 24 (1–2):53–60. https://doi.org/10.1177/0748233708089017 PMID: 18818181

106. Gidai J, Acs N, Banhidy F, Czeizel AE. A study of the effects of large doses of medazepam used for self-poisoning in 10 pregnant women on fetal development. Toxicol Ind Health. 2008; 24(1–2):61–8. https://doi.org/10.1177/0748233708089016 PMID: 18818182

107. Gidai J, Ács N, Bánhidy F, Czeizel A. A study of the teratogenic and fetotoxic effects of large doses of chlordiazepoxide used for self-poisoning by 35 pregnant women. Toxicol Ind Health. 2008 Feb; 24(1–2):41–51. https://doi.org/10.1177/0748233708089018 PMID: 18818180

108. Laegreid L, Hagberg G, Lundberg A. Neurodevelopment in late infancy after prenatal exposure to benzodiazepines—a prospective study. Neuropediatrics. 1992; 23(2):60–7. https://doi.org/10.1055/s-2008-1071314 PMID: 1351263

109. Lupattelli A, Chambers CD, Bandoli G, Handal M, Skurtveit S, Nordeng H. Association of Maternal Use of Benzodiazepines and Z-Hypnotics During Pregnancy With Motor and Communication Skills and Attention-Deficit/Hyperactivity Disorder Symptoms in Preschoolers. JAMA Netw Open. 2019 Apr 5; 2(4):e191435. https://doi.org/10.1001/jamanetworkopen.2019.1435 PMID: 30951155

110. Odsbu I, Skurtveit S, Selmer R, Roth C, Hernandez-Diaz S, Handal M. Prenatal exposure to anxiolytics and hypnotics and language competence at 3 years of age. Eur J Clin Pharmacol. 2015; 71 (3):283–91. https://doi.org/10.1007/s00228-014-1797-4 PMID: 25547568

111. Radojcic MR, El Marroun H, Miljkovic B, Stricker BHC, Jaddoe VWV, Verhulst FC, et al. Prenatal exposure to anxiolytic and hypnotic medication in relation to behavioral problems in childhood: A population-based cohort study. Neurotoxicol Teratol. 2017 Jun; 61:58–65. https://doi.org/10.1016/j.ntt.2017.02.005 PMID: 28259732

112. Timmermann G, Acs N, Banhidy F, Czeizel AE. A study of teratogenic and fetotoxic effects of large doses of meprobamate used for a suicide attempt by 42 pregnant women. Toxicol Ind Health. 2008; 24(1–2):97–107. https://doi.org/10.1177/0748233708089015 PMID: 18818186

113. Peng M, Gao K, Ding Y, Ou J, Calabrese JR, Wu R, et al. Effects of prenatal exposure to atypical antipsychotics on postnatal development and growth of infants: a case-controlled, prospective study. Psychopharmacology (Berl). 2013; 228(4):577–84.

114. Platt JE, Friedhoff AJ, Broman SH, Bond R, Laska E, Lin SP. Effects of prenatal neuroleptic drug exposure on motor performance in children. Hum Psychopharmacol Clin Exp. 1989 Sep; 4(3):205–13.

115. Petik D, Ács N, Bánhidy F, Czeizel A. A study of the effects of large doses of glutethimide that were used for self-poisoning during pregnancy on human fetuses. Toxicol Ind Health. 2008 Feb; 24(1–2):69–78. https://doi.org/10.1177/0748233708089014 PMID: 18818183

116. Petik D, Timmermann G, Czeizel AE, Acs N, Banhidy F. A study of the teratogenic and fetotoxic effects of large doses of amobarbital used for a suicide attempt by 14 pregnant women. Toxicol Ind Health. 2008; 24(1–2):79–85. https://doi.org/10.1177/0748233708093314 PMID: 18818184

117. Timmermann G, Czeizel A, Bánhidy F, Ács N. A study of the teratogenic and fetotoxic effects of large doses of barbital, hexobarbital and butobarbital used for suicide attempts by pregnant women. Toxicol Ind Health. 2008 Feb; 24(1–2):109–19. https://doi.org/10.1177/0748233708089004 PMID: 18818187

118. Avella-Garcia CB, Julvez J, Fortuny J, Rebordosa C, Garcia-Esteban R, Galan IR, et al. Acetaminophen use in pregnancy and neurodevelopment: attention function and autism spectrum symptoms. Int J Epidemiol. 2016; 45(6):1987–96. https://doi.org/10.1093/ije/dyw115 PMID: 27353198

119. Bornehag C-G, Reichenberg A, Hallerback MU, Wikstrom S, Koch HM, Jonsson BA, et al. Prenatal exposure to acetaminophen and children's language development at 30 months. Eur Psychiatry. 2018 Jun; 51:98–103. https://doi.org/10.1016/j.eurpsy.2017.10.007 PMID: 29331486

120. Brandlistuen RE, Ystrom E, Nulman I, Koren G, Nordeng H. Prenatal paracetamol exposure and child neurodevelopment: a sibling-controlled cohort study. Int J Epidemiol. 2013 Dec 1; 42(6):1702–13. https://doi.org/10.1093/ije/dyt183 PMID: 24163279

121. Laue HE, Cassoulet R, Abdelouahab N, Serme-Gbedo YK, Desautels A-S, Brennan KJM, et al. Association Between Meconium Acetaminophen and Childhood Neurocognitive Development in GESTE, a Canadian Cohort Study. Toxicol Sci. 2019 Jan; 167(1):138–44. https://doi.org/10.1093/toxsci/kfy222 PMID: 30202886

122. Liew Z, Ritz B, Rebordosa C, Lee P-C, Olsen J. Acetaminophen use during pregnancy, behavioral problems, and hyperkinetic disorders. JAMA Pediatr. 2014; 168(4):313–20. https://doi.org/10.1001/jamapediatrics.2013.4914 PMID: 24566677

123. Liew Z, Ritz B, Virk J, Olsen J. Maternal use of acetaminophen during pregnancy and risk of autism spectrum disorders in childhood: A Danish national birth cohort study: Acetaminophen and Autism Spectrum Disorders. Autism Res. 2016 Sep; 9(9):951–8. https://doi.org/10.1002/aur.1591 PMID: 26688372

124. Liew Z, Bach CC, Asarnow RF, Ritz B, Olsen J. Paracetamol use during pregnancy and attention and executive function in offspring at age 5 years. Int J Epidemiol. 2016; 45(6):2009–17. https://doi.org/10.1093/ije/dyw296 PMID: 28031314

125. Liew Z, Ritz B, Virk J, Arah OA, Olsen J. Prenatal Use of Acetaminophen and Child IQ: A Danish Cohort Study. Epidemiology. 2016 Nov; 27(6):912–8. https://doi.org/10.1097/EDE.0000000000000540 PMID: 27479646

126. Liew Z, Kioumourtzoglou M-A, Roberts AL, O'Reilly EJ, Ascherio A, Weisskopf MG. Use of Negative Control Exposure Analysis to Evaluate Confounding: An Example of Acetaminophen Exposure and Attention-Deficit/Hyperactivity Disorder in Nurses' Health Study II. Am J Epidemiol. 2019 Apr; 188(4):768–75. https://doi.org/10.1093/aje/kwy288 PMID: 30923825

127. Ruisch IH, Buitelaar JK, Glennon JC, Hoekstra PJ, Dietrich A. Pregnancy risk factors in relation to oppositional-defiant and conduct disorder symptoms in the Avon Longitudinal Study of Parents and Children. J Psychiatr Res. 2018 Jun; 101:63–71. https://doi.org/10.1016/j.jpsychires.2018.02.020 PMID: 29550610

128. Skovlund E, Handal M, Selmer R, Brandlistuen RE, Skurtveit S. Language competence and communication skills in 3-year-old children after prenatal exposure to analgesic opioids. Pharmacoepidemiol Drug Saf. 2017; 26(6):625–34. https://doi.org/10.1002/pds.4170 PMID: 28168770

129. Stergiakouli E, Thapar A, Davey Smith G. Association of Acetaminophen Use During Pregnancy With Behavioral Problems in Childhood: Evidence Against Confounding. JAMA Pediatr. 2016; 170(10):964–70. https://doi.org/10.1001/jamapediatrics.2016.1775 PMID: 27533796

130. Thompson JMD, Waldie KE, Wall CR, Murphy R, Mitchell EA, ABC study group. Associations between acetaminophen use during pregnancy and ADHD symptoms measured at ages 7 and 11 years. Robinson E BD Wild C, editor. PloS One. 2014; 9(9):e108210. https://doi.org/10.1371/journal.pone.0108210 PMID: 25251831

131. Tovo-Rodrigues L, Schneider BC, Martins-Silva T, Del-Ponte B, de Mola CL, Schuler-Faccini L, et al. Is intrauterine exposure to acetaminophen associated with emotional and hyperactivity problems during childhood? Findings from the 2004 Pelotas birth cohort. Bmc Psychiatry. 2018 Nov 20; 18:368. https://doi.org/10.1186/s12888-018-1942-1 PMID: 30458756

132. Ystrom E, Gustavson K, Brandlistuen RE, Knudsen GP, Magnus P, Susser E, et al. Prenatal Exposure to Acetaminophen and Risk of ADHD. Pediatrics. 2017 Nov; 140(5):e20163840. https://doi.org/10.1542/peds.2016-3840 PMID: 29084830

133. al-Alaiyan S, Seshia MM, Casiro OG. Neurodevelopmental outcome of infants exposed to indomethacin antenatally. J Perinat Med. 1996; 24(4):405–11. PMID: 8880639

134. Amin SB, Kamaluddeen M, Sangem M. Neurodevelopmental outcome of premature infants after exposure to antenatal indomethacin. Am J Obstet Gynecol. 2008; 199(1):41.e1–8.

135. Markovic M, Swanson SA, Stricker BH, Jaddoe VWV, Verhulst FC, Tiemeier H, et al. Prenatal exposure to non-steroidal anti-inflammatory drugs (NSAIDs) and neurodevelopmental outcomes in children. Pharmacoepidemiol Drug Saf. 2019 Apr; 28(4):452–9. https://doi.org/10.1002/pds.4625 PMID: 30838712

136. Salokorpi T, Eronen M, von Wendt L. Growth and development until 18 months of children exposed to tocolytics indomethacin or nylidrin. Neuropediatrics. 1996; 27(4):174–7. https://doi.org/10.1055/s-2007-973782 PMID: 8892364

137. Barr HM, Streissguth AP, Darby BL, Sampson PD. Prenatal exposure to alcohol, caffeine, tobacco, and aspirin: Effects on fine and gross motor performance in 4-year-old children. Dev Psychol. 1990; 26(3):339–48.

138. Klebanoff MA, Berendes HW. Aspirin exposure during the first 20 weeks of gestation and IQ at four years of age. Teratology. 1988; 37(3):249–55. https://doi.org/10.1002/tera.1420370310 PMID: 3368878

139. Harris G-ME, Wood M, Ystrom E, Nordeng H. Prenatal triptan exposure and neurodevelopmental outcomes in 5-year-old children: Follow-up from the Norwegian Mother and Child Cohort Study. Paediatr Perinat Epidemiol. 2018 May; 32(3):247–55. https://doi.org/10.1111/ppe.12461 PMID: 29569251

140. Wood ME, Frazier JA, Nordeng HME, Lapane KL. Longitudinal changes in neurodevelopmental outcomes between 18 and 36 months in children with prenatal triptan exposure: findings from the Norwegian Mother and Child Cohort Study. BMJ Open. 2016; 6(9):e011971. https://doi.org/10.1136/bmjopen-2016-011971 PMID: 27625061

141. Wood ME, Lapane K, Frazier JA, Ystrom E, Mick EO, Nordeng H. Prenatal Triptan Exposure and Internalising and Externalising Behaviour Problems in 3-Year-Old Children: Results from the Norwegian Mother and Child Cohort Study: Prenatal triptan exposure and externalising symptoms. Paediatr Perinat Epidemiol. 2016 Mar; 30(2):190–200. https://doi.org/10.1111/ppe.12253 PMID: 26525300

142. Wood ME, Frazier JA, Nordeng HME, Lapane KL. Prenatal triptan exposure and parent-reported early childhood neurodevelopmental outcomes: an application of propensity score calibration to adjust for unmeasured confounding by migraine severity. Pharmacoepidemiol Drug Saf. 2016; 25(5):493–502. https://doi.org/10.1002/pds.3902 PMID: 26554750

143. Rubenstein E, Young JC, Croen LA, DiGuiseppi C, Dowling NF, Lee L-C, et al. Brief Report: Maternal Opioid Prescription from Preconception Through Pregnancy and the Odds of Autism Spectrum Disorder and Autism Features in Children. J Autism Dev Disord. 2019 Jan; 49(1):376–82. https://doi.org/10.1007/s10803-018-3721-8 PMID: 30132098

144. Subcommittee on Attention-Deficit/Hyperactivity Disorder, Steering Committee on Quality Improvement and Management. ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. PEDIATRICS. 2011 Nov 1; 128(5):1007–22. https://doi.org/10.1542/peds.2011-2654 PMID: 22003063

145. Tomson T, Battino D, French J, Harden C, Holmes L, Morrow J, et al. Antiepileptic drug exposure and major congenital malformations: The role of pregnancy registries. Epilepsy Behav. 2007 Nov; 11 (3):277–82. https://doi.org/10.1016/j.yebeh.2007.08.015 PMID: 17996635

146. Rodier PM, Ingram JL, Tisdale B, Nelson S, Romano J. Embryological origin for autism: developmental anomalies of the cranial nerve motor nuclei. J Comp Neurol. 1996 Jun 24; 370(2):247–61. https://doi.org/10.1002/(SICI)1096-9861(19960624)370:2<247::AID-CNE8>3.0.CO;2-2 PMID: 8808733

147. Müller JM, Furniss T. Correction of distortions in distressed mothers' ratings of their preschool children's psychopathology. Psychiatry Res. 2013 Nov; 210(1):294–301. https://doi.org/10.1016/j.psychres.2013.03.025 PMID: 23648281

148. Najman JM, Williams GM, Nikles J, Spence S, Bor W, O'Callaghan M, et al. Bias influencing maternal reports of child behaviour and emotional state. Soc Psychiatry Psychiatr Epidemiol. 2001 Apr; 36 (4):186–94. PMID: 11518032

149. van der Toorn SLM, Huizink AC, Utens EMWJ, Verhulst FC, Ormel J, Ferdinand RF. Maternal depressive symptoms, and not anxiety symptoms, are associated with positive mother-child reporting discrepancies of internalizing problems in children: a report on the TRAILS study. Eur Child Adolesc Psychiatry. 2010 Apr; 19(4):379–88. https://doi.org/10.1007/s00787-009-0062-3 PMID: 19823897

150. Sawyer MG, Streiner DL, Baghurst P. The influence of distress on mothers' and fathers' reports of childhood emotional and behavioral problems. J Abnorm Child Psychol. 1998 Dec; 26(6):407–14. PMID: 9915648

151. Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. Int J Epidemiol. 2014 Apr; 43(2):434–42. https://doi.org/10.1093/ije/dyt261 PMID: 24464188

152. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychol Bull. 1987; 101(2):213–32. PMID: 3562706

153. Rasalam A, Hailey H, Williams J, Moore S, Turnpenny P, Lloyd D, et al. Characteristics of fetal anticonvulsant syndrome associated autistic disorder. Dev Med Child Neurol. 2005 Jul 14; 47(8):551–5. PMID: 16108456

154. Pinsky L. Informatic Morphogenetic Variants. In: Issue and reviews in teratology: 3. New York: Plenum Press; 1985. p. 135–70.

155. Bateman BT, Hernandez-Diaz S, Fischer MA, Seely EW, Ecker JL, Franklin JM, et al. Statins and congenital malformations: cohort study. BMJ. 2015 Mar 17; 350(mar17 10):h1035–h1035.

156. Morales DR, Slattery J, Evans S, Kurz X. Antidepressant use during pregnancy and risk of autism spectrum disorder and attention deficit hyperactivity disorder: systematic review of observational studies and methodological considerations. BMC Med. 2018 Dec; 16(1).

157. Mezzacappa A, Lasica P-A, Gianfagna F, Cazas O, Hardy P, Falissard B, et al. Risk for Autism Spectrum Disorders According to Period of Prenatal Antidepressant Exposure: A Systematic Review and Meta-analysis. [Review]. JAMA Pediatr. 2017; 171(6):555–63. https://doi.org/10.1001/jamapediatrics.2017.0124 PMID: 28418571

158. Kobayashi T, Matsuyama T, Takeuchi M, Ito S. Autism spectrum disorder and prenatal exposure to selective serotonin reuptake inhibitors: A systematic review and meta-analysis. Reprod Toxicol Elmsford N. 2016; 65(be4, 8803591):170–8.

159. Hermansen TK, Melinder A. Prenatal SSRI exposure: Effects on later child development. Child Neuropsychol. 2015 Sep 3; 21(5):543–69. https://doi.org/10.1080/09297049.2014.942727 PMID: 25089614

160. Nagin DS, Odgers CL. Group-Based Trajectory Modeling in Clinical Research. Annu Rev Clin Psychol. 2010 Mar; 6(1):109–38.

161. Shadish WR, Cook TD, Campbell DT. Construct validity and external validity. In: Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2001. p. 64–102.

162. Regier DA, Kaelber CT, Roper MT, Rae DS, Sartorius N. The ICD-10 clinical field trial for mental and behavioral disorders: results in Canada and the United States. Am J Psychiatry. 1994 Sep; 151 (9):1340–50. https://doi.org/10.1176/ajp.151.9.1340 PMID: 8067491

163. Brown T, Lalor A. The Movement Assessment Battery for Children—Second Edition (MABC-2): A Review and Critique. Phys Occup Ther Pediatr. 2009 Jan; 29(1):86–103. https://doi.org/10.1080/01942630802574908 PMID: 19197761

164. Ananth CV, VanderWeele TJ. Placental abruption and perinatal mortality with preterm delivery as a mediator: disentangling direct and indirect effects. Am J Epidemiol. 2011 Jul 1; 174(1):99–108. https://doi.org/10.1093/aje/kwr045 PMID: 21430195

165. Grzeskowiak LE, Gilbert AL, Morrison JL. Investigating outcomes associated with medication use during pregnancy: A review of methodological challenges and observational study designs. Reprod Toxicol. 2012 Jun; 33(3):280–9. https://doi.org/10.1016/j.reprotox.2012.01.006 PMID: 22329969

166. Wang Z, Ho PWH, Choy MTH, Wong ICK, Brauer R, Man KKC. Advances in Epidemiological Methods and Utilisation of Large Databases: A Methodological Review of Observational Studies on Central Nervous System Drug Use in Pregnancy and Central Nervous System Outcomes in Children. Drug Saf. 2018 Nov 13.