Department of Health and Human Services
Public Health Service
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Surveillance and Epidemiology (OSE)
Office of Pharmacovigilance and Epidemiology (OPE)

Epidemiology: Review of Published Studies

Date:	July 15, 2022
Reviewers:	Danielle Abraham, PhD, MPH Division of Epidemiology I (DEPI-I) Office of Surveillance and Epidemiology (OSE) Andrew D. Mosholder, MD, MPH DEPI-I, OSE
Team Leader:	Kira Leishear White, PhD, MS DEPI-I, OSE
Acting Director:	CAPT Sukhminder K. Sandhu, PhD, MPH, MS DEPI-I, OSE
Subject:	Functional Neurobehavioral Outcomes and Urogenital Outcomes Associated with Prenatal Acetaminophen Exposure
Drug Name:	Acetaminophen
Applicant/sponsor:	Multiple
OSE RCM #:	2016-1871
SS ID #:	1001355

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Acetaminophen (APAP), an analgesic and antipyretic, is available in both prescription and over-the-counter formulations. The Division of Epidemiology I (DEPI-I) conducted several literature reviews between 2014 and 2019 on the association between prenatal APAP use and adverse neurobehavioral and urogenital outcomes. A recent literature review published in 2021 by Bauer et al., which received media attention, also examined these associations. Due to frequent APAP use in pregnancy, even weak associations between APAP use and these outcomes, if causal, may have substantial public health impacts. The purpose of this review is for DEPI-I to review recently published articles, including those in the Bauer et al. (2021) review, on the association between prenatal APAP exposure and both neurobehavioral and urogenital outcomes.

DEPI-I reviewed a total of 24 studies that were not included in prior DEPI-I literature reviews. Twenty-two studies examined the association between prenatal APAP use and functional neurobehavioral outcomes. Two studies examined associations with urogenital outcomes, specifically anogenital distance and pubertal development.

Studies reviewed, including meta-analyses, suggest a consistent association between prenatal APAP exposure, particularly high levels of or long durations of exposure, and attention deficit hyperactivity disorder. However, associations based on trimester-specific use were not consistent. Associations were seen for other neurobehavioral outcomes; however, findings were mixed and inconsistent across reviewed studies. The most methodologically rigorous study found a weak association, at 11 years of age, between any prenatal APAP exposure and behavioral problems reported by parents and children. When examining the relationship by duration of exposure (assessed by trimesters or weeks), the only other significant association in both reporting sources was for >10 cumulative weeks of exposure, and the association was also weak. (b) (5)

(b) (5)

(b) (5)

. In the meantime, it may be prudent, as a precautionary measure, for FDA to (b) (5) FDA's prior drug safety communication from January 9, 2015, on APAP in pregnancy focused on the treatment of pain. The communication did not discuss amount of use. Untreated fevers during pregnancy are associated with poor pregnancy outcomes (b) (5)

(b) (5) (b) (5)

1 INTRODUCTION

Based on a growing number of research studies, a recently published literature review in *Nature Review Endocrinology* suggests that there are reproductive and neurodevelopmental safety concerns related to prenatal acetaminophen, or paracetamol, (APAP) use (1).

Between 2014 and 2016, DEPI-I conducted several reviews examining neurobehavioral outcomes associated with prenatal APAP exposure:

- Taylor LG, Wang C. Review of study of acetaminophen use in pregnancy and risks of ADHD in offspring. May 15, 2014. Silver Spring (MD), U.S. Food and Drug Administration. DARRTS Reference ID: 3507534.

- Mosholder AD, Taylor LG, Pinheiro SP. Acetaminophen use in pregnancy and ADHD in offspring. March 18, 2015. Silver Spring (MD), U.S. Food and Drug Administration. DARRTS Reference ID: 3718011.

- Mosholder AD, Taylor LG, Wang C. Neurodevelopmental outcomes following prenatal acetaminophen exposure. October 14, 2016. Silver Spring (MD), U.S. Food and Drug Administration. DARRTS Reference ID: 3999031.

In the prior reviews, seven out of eight studies found an association between prenatal APAP exposure and ADHD, ADHD behaviors, or other adverse neurobehavioral outcomes. Associations, although consistent, were weak to modest in magnitude with some evidence of a dose-response relationship. When exposure timing was examined, associations were stronger when there was APAP exposure during multiple trimesters. However, studies reviewed may be limited broadly by residual confounding. Studies also typically did not have adequate information on dosage, number of pills, and duration of use to appropriately examine dose-response relationships. Studies also suggested that APAP may prevent the negative effects of fever.

In 2019, DEPI-I conducted a review examining urogenital outcomes associated with prenatal APAP exposure:

- Mosholder AD, Leishear K, Sandhu SK. Urogenital outcomes with in utero acetaminophen exposure. January 7, 2019. Silver Spring (MD), U.S. Food and Drug Administration. DARRTS Reference ID: 4372262.

In the prior review, DEPI-I reviewed eight studies of the association between prenatal APAP exposure and urogenital outcomes. Prenatal APAP was associated with cryptorchidism (three studies), reduced penile width (one study), and mean anogenital distance (AGD) (one study); however, three studies found no associations with urogenital outcomes. One study found that

APAP attenuated the impact of fever on hypospadias. Studies may have been limited by residual confounding and insufficient capture of APAP exposure during important developmental windows.

Based on the media attention received by the Bauer et al. (2021) review, OSE/OPE deemed it prudent for DEPI-I to update their prior literature reviews on the association between prenatal APAP exposure and both neurobehavioral and urogenital outcomes.

1.1 Background

APAP is an analgesic and antipyretic available alone or in combination products in both over-the-counter and prescription formulations.[1] Werler et al. (2005) found that at least 65.5% of women in the U.S. report prenatal APAP use (2). In Bandoli et al. (2020), 62% of women in the U.S. or Canada reported APAP use during pregnancy with only 3% of use stemming from prescription formulations (3). The prevalence of APAP use during pregnancy differs by geographical region; however, APAP is typically the most commonly used over-the-counter analgesic/antipyretic during pregnancy (4). APAP use during pregnancy is more common in the first and second trimesters, typically used short-term, and most often used for headaches (3). Individuals with longer durations of use differ from those with short-term use (3). Women who used APAP for 45 days or more were more likely to be obese, suffering from anxiety or depression, and users of tobacco or antidepressants during pregnancy, compared to those with shorter durations of APAP use (3). Individuals with 45 days or more of APAP use were also more likely to use APAP for sleep as well as pain or injury indications (3).

Nonsteroidal anti-inflammatory drugs (NSAIDs) and opioids are associated with an increased risk of birth defects, compared to APAP (5). These other analgesics are also associated with adverse pregnancy and birth outcomes (6, 7). UpToDate lists APAP as the preferred analgesic and antipyretic in pregnant women (8). However, APAP crosses the placental barrier (9). Mechanistically, APAP may also impact neurodevelopment and have endocrine disrupting effects (4). Given APAP's frequent use in pregnancy, even weak associations with adverse outcomes, if causal, may have substantial public health implications.

1.2 Regulatory History

On January 9, 2015, the U.S. Food and Drug Administration (FDA) released a drug safety communication regarding the use of pain medicines, including APAP, during pregnancy.[1] The FDA found that the published studies, at the time, had limitations and discordant results and, thus, made the decision not to modify recommendations about pain medication use during pregnancy. The communication encouraged women to consult with their health care professionals to discuss the risks and benefits of pain medicine use and advised health care

[1] Drugs@FDA: FDA-Approved Drugs. Silver Spring (MD), U.S. Food and Drug Administration. Accessed January 13, 2022 at: https://www.accessdata fda.gov/scripts/cder/daf/index.cfm.

professionals to follow drug label recommendations for prescribing pain medicines to pregnant patients. The communication made no recommendations about dosage or amount of use.

2 REVIEW METHODS AND MATERIALS

The review followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (10). The review sought to identify newer studies of prenatal APAP exposure and functional neurobehavioral outcomes or urogenital outcomes not previously reviewed. We conducted a PubMed literature search restricting to original, epidemiologic studies published from January 1, 2016, through December 1, 2021. Articles were limited to human species studies and published in English. The literature search was performed using the following search terms:

- (Acetaminophen OR Paracetamol) AND pregnancy
- (Acetaminophen OR Paracetamol) AND birth outcomes
- (Acetaminophen OR Paracetamol) AND neurodevelopment
- (Acetaminophen OR Paracetamol) AND urogenital
- (Acetaminophen OR Paracetamol) AND prenatal

Additionally, a reverse, reference search was conducted examining individual studies from the reviews and meta-analyses from Bauer et al. (2021) and relevant reviews and meta-analyses identified in the PubMed search. Abstracts from articles identified through all methods (published between January 1, 2016, and December 1, 2021) were manually reviewed by DEPI-I for inclusion.

After review of the eligible articles, a *post hoc* list was generated of higher quality study criteria (see Appendix A). One reviewer applied the criteria to each article reviewed.

3 REVIEW RESULTS

Of the 44 epidemiologic articles (not including meta-analyses and review articles) in the Bauer et al. (2021) review, 29 had not been previously assessed by DEPI-I. Nine of these articles were excluded: one focused on adult APAP use and fertility, one studied genetic risk factors for APAP use, one examined cerebral palsy as an outcome, and six were published prior to 2016. In total, 20 articles from Bauer et al. (2021), not previously reviewed by DEPI-I, were flagged for review.

The PubMed search identified 436 articles. One reviewer screened the abstracts of all 436 articles for review inclusion. Several articles examined the relationship between APAP and other infant and childhood outcomes, including fetal toxicity, birth outcomes (e.g., preterm birth, low birthweight, small-for-gestational age), neurologic congenital defects, gastroschisis, heart defects, cerebral palsy, asthma/wheezing, atopic dermatitis, celiac disease, immune function, and obesity. The final review excluded these articles as they were outside the scope of the review. Screening also led to the exclusion of descriptive studies of APAP use in pregnancy, studies focused on APAP use around labor and delivery, studies of APAP use in infancy/childhood

alone, and studies not focused on use in pregnancy. The PubMed search yielded four additional studies and the reverse search yielded two additional studies.

After reading the full manuscript texts, one article from Bauer et al. (2021) was an epigenetic study and one article from the PubMed search was a methods paper; these two studies were further excluded. Overall, DEPI-I reviewed 24 studies identified by the three selection strategies.

Five relevant meta-analyses were identified in the Bauer et al. (2021) article and the PubMed search.

3.1 Common Pregnancy Outcomes Cohorts

Most studies reviewed were cohort studies and extracted study samples from, or were nested within, large, geographically defined, cohorts (11-31). Table 1 displays a list of the cohorts. Three studies did not conduct population-based cohort studies. Chen et al. (2019) conducted a case-control study using the Taiwan Longitudinal Health Insurance Database (32). Parker et al. (2020) used controls from a North American case-control study of craniofacial malformations (33). Lastly, there was one case-control study of approximately 200 Canadian children (34). Appendix B provides additional details of all studies reviewed. The data collection schedules are highly heterogeneous across cohorts.

Table 1. List of Cohorts Used by Reviewed Studies

Cohort Name	Geography	Reviewed Studies
Gestation and the Environment Cohort	Sherbrooke Québec, Canada	(11, 31)
Project Viva Cohort	Eastern Massachusetts, USA	(12, 13)
Pelotas Birth Cohort	Brazil	(12, 14, 15)
Swedish Environmental Longitudinal, Mother and Child, Asthma and Allergy (SELMA) Study	Värmland, Sweden	(16)
Norwegian Mother and Child Cohort (MoBa)	Norway	(17-21)
Avon Longitudinal Study of Parents and Children (ALSPAC)	Avon, United Kingdom	(22, 23)
Danish National Birth Cohort (DNBC)	Denmark	(24-26)
Boston Birth Cohort (BBC)	Boston Medical Center	(27, 28)
Nurses' Health Study*	United States	(29)
Nutrition in Early Life and Asthma (NELA) Study	Murica, Spain	(30)

*Not a pregnancy or birth cohort study

3.2 Covariates

Individual studies each adjusted for a multitude of covariates; however, the covariates selected varied by study. Covariates used for adjustment or stratification in each study are detailed in both Appendix B and Appendix C. Appendix B provides a summary for each study, including a list of covariates used in adjustment. Appendix C lists all covariates considered across all studies with a listing of each study that used a given variable in adjustment or stratification. The most common covariates included child sex and maternal age, education, body mass index (BMI), parity, alcohol use, and smoking status (Appendix C). Unless otherwise noted, results presented will be from the fully adjusted models.

3.3 Functional Neurobehavioral Outcomes

In total, 22 studies examined the association between prenatal APAP use and functional neurobehavioral outcomes, including attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), other developmental disorders, behavioral problems, temperament, attention, cognition/executive function/neurodevelopment, and language and communication.

Recently published meta-analyses find some significant associations between prenatal APAP use and various outcomes including autism spectrum disorder (ASD) (35-37), attention deficit hyperactivity disorder (ADHD) (35, 36, 38), conduct disorder (36), and hyperactivity symptoms (36). In Masarwa et al. (2018), the study found a significant association between APAP and ADHD with a pooled relative risk (RR)=1.34 (95% confidence interval [CI]: 1.21, 1.47; I^2=72%, n=6 studies). There were also significant associations between APAP and ASD (RR=1.19; 95% CI: 1.14, 1.25; I^2=14%, n=5 studies), hyperactivity (RR=1.24; 95% CI: 1.04, 1.43; I^2=93%, n=5 studies), and conduct disorder (RR=1.23; 95% CI: 1.04, 1.42; I^2=93%; n=4 studies). Significant heterogeneity was present for all outcomes, except ASD. The meta-analysis examined several covariates as modifiers through meta-regression including child's age at follow-up; maternal fever, age, smoking, socio-economic status; country latitude; study methodological quality; and duration of APAP exposure. Significant modifiers included child's age (stronger association with greater age), duration of exposure (stronger association with longer duration), and maternal age (decreased association with younger age) (36). Masarwa et al. (2020) examined the pooled association between APAP and ADHD, pooling data from seven studies (38). The meta-analysis found a significant association pooling data from all seven studies (RR=1.31; 95% CI: 1.23, 1.39, I^2=48%) and when restricting to four studies used in a bias analysis (RR=1.34; 95% CI: 1.09, 1.59; I^2=63%) (38). The bias analysis assessed the impact of unmeasured confounding, selection bias, and exposure misclassification on the association between APAP and ADHD (38). Alemany et al. (2021) pooled individual-level data from 73,881 mother-child pairs captured in European birth cohorts (35). The meta-analysis examined prenatal and postnatal APAP and considered the following covariates: maternal age at delivery, maternal education/socio-occupational status, maternal BMI, alcohol, smoking, mental health problems during pregnancy, fever during pregnancy, infections during pregnancy, child sex, child age at outcome assessment, child cold within first two years, and child respiratory infections within first two years. The study found a significant association between prenatal APAP and ADHD or borderline/clinical ADHD symptoms (adjusted odds ratio [aOR]=1.21; 95% CI: 1.07, 1.36; I^2=0.0%, n=6 studies).

When stratified by child sex, the association was similar in boys (aOR=1.23; 95% CI: 1.05, 1.44) and girls (aOR=1.18; 95% CI: 0.97, 1.44). For ASD or borderline/clinical ASD symptoms, there was a significant association with prenatal APAP (aOR=1.19; 95% CI: 1.07, 1.33; I^2=2.4%, n=6 studies). When stratified by child sex, the association was slightly stronger in boys (aOR=1.28; 95% CI: 1.12, 1.46), compared to girls (aOR=1.06; 95% CI; 0.82, 1.36). There were no associations between postnatal APAP and ADHD outcomes or ASD outcomes (35). The study found no effect modification by age at child assessment and found similar associations when restricting to those with diagnoses in a hospital register (35). Gou et al. (2019) pooled adjusted associations from eight studies and found a significant association between prenatal APAP and ADHD (adjusted RR [aRR]=1.25; 95% CI; 1.17, 1.34; I^2=26%) (37). The study also considered several stratified analyses[2]. When stratified by duration of use, Gou et al. (2019) found a significant association for those who used APAP for 28 days or more during pregnancy (RR=1.63, 95% CI: 1.23, 2.16; I^2=85%, n=3 studies), but not for those who used APAP for less than 28 days (RR=1.11 (95% CI: 0.96, 1.28; I^2=69%, n=3 studies) (37). The meta-analysis also found that associations were strongest for APAP exposure in the third trimester (RR=1.26; 95% confidence interval: 1.08, 1.47; I^2=0, n=3 studies) and found an association with first trimester exposure (RR=1.21; 95% CI: 1.01, 1.45, I^2=37%, n=2 studies); there was no significant association between second trimester use and ADHD (RR=1.06; 95% CI: 0.95, 1.17; I^2=0%, n=3 studies) (37). The study found no difference by whether outcomes were obtained through parental report or standard assessments (37).

3.3.1 Attention Deficit Hyperactivity Disorder (ADHD)

Six studies assessed the relationship between maternal APAP use and ADHD. In a seventh study, the main association of interest was prenatal fever and ADHD, with APAP use studied as an effect measure modifier. In general, there is an association between APAP use and ADHD diagnoses with some evidence of a dose- or duration-response relationship. Results suggesting a trimester-specific risk window are inconsistent.

Chen et al. (2019) used a national registry to examine the association[3] between prescription, prenatal APAP use and ADHD diagnosis, based on International Classification of Diseases (ICD) codes from psychiatry visits (32). In this case-control study, the authors included Taiwanese births between 1998 and 2008 with ADHD diagnoses assessed through 2013. Overall, there was a significant association between any prenatal APAP use and ADHD (aOR=1.20; 95% CI: 1.01, 1.42). There was no association between first trimester (aOR=1.09; 95% CI: 0.92, 1.28) or third trimester (aOR=0.97; 95% CI: 0.83, 1.13) APAP use and ADHD; however, there was a significant association with second trimester use (aOR=1.19; 95% CI: 1.00, 1.40). Cumulative, log-transformed APAP use was not associated with ADHD diagnosis. As a negative control exposure (NCE), APAP use three months prior to pregnancy had no association with ADHD diagnosis (aOR=1.06; 95% CI: 0.90, 1.25) (32).

[2] It was unclear if the pooled, stratified RR were adjusted.

[3] Study adjusted for mother/child age, sex, income, gestational infections, maternal mental health disorders, comorbid perinatal conditions, level of urbanization

Liew et al. (2019) used data from the Nurses' Health Study to assess the relationship[4] between self-reported regular APAP use (≥ twice per week or ≥ once per week) and self-reported child ADHD diagnosis (29). As part of the Nurses' Health Study, women completed biannual questionnaires, irrespective of pregnancy. Questionnaire dates were cross-walked to child dates of birth and served as a proxy for prenatal use. NCE periods included questionnaires completed four years before and four years after the pregnancy flagged questionnaire. Regular APAP use at the time of pregnancy was associated with ADHD (aOR=1.35; 95% CI: 1.07, 1.71); however, there was no significant association in the before (aOR=1.12; 95% CI: 0.91, 1.38) or after (aOR=1.05; 95% CI: 0.88, 1.26) NCE periods. Restricting to the sub-sample of individuals who reported completing the questionnaire while pregnant, a similar pattern was seen; however, the time of pregnancy aOR was no longer significant (aOR=1.39; 95% CI: 0.99, 1.95) (29).

Ystrom et al. (2017) assessed the relationship[5] between APAP use and ADHD diagnosis in the Norwegian Mother and Child Cohort (MoBa) study (21). ICD codes for hyperkinetic disorder in the Norwegian Patient Registry were used to capture outcomes. NCEs included both maternal and paternal pre-pregnancy APAP use. There was no significant association between maternal pre-pregnancy use and ADHD; however, there was an association with paternal use (adjusted hazard ratio [aHR]=1.27; 95% CI: 1.08, 1.49). When examining paternal days of APAP use, a dose-response was seen. Prenatal use was associated with a significant increase in hazard of ADHD diagnosis (aHR=1.12; 95% CI: 1.02, 1.24). No significant associations were noted for APAP use during any one trimester or all three trimesters. A significant association was noted for any two trimesters; when stratified, the association was only significant for those with APAP use in both the first and second trimester. Analysis by days of use[6] found a dose-response pattern in the magnitude of associations; however, the only statistically significant association was for ≥29 days or more of APAP use (aOR=2.20; 95% CI: 1.50, 3.24). Stratified by indication, significant associations were noted for those with fever and infections with 22 to 28 days of use (aHR=6.15; 95% CI: 1.71, 22.05), as well as pain conditions with ≥29 days of use (aHR=2.56; 95% CI: 1.54, 4.25) (21).

Baker et al. (2020) examined the association[7] between APAP detected in meconium and parent report or medical record documentation of a physician diagnosis of ADHD (11). ADHD was assessed at 6-7 years of age. Among 345 children, 57.7% had APAP detected in meconium. After inverse probability treatment weighting (IPTW) with propensity scores, there was a significant, adjusted association between APAP detection in meconium and ADHD diagnosis (aOR=2.43; 95% CI: 1.41, 4.21). This association persisted when excluding children of mothers who received APAP during delivery (aOR=2.38, 95% CI: 1.35, 4.21). Significant associations

[4] Study adjusted for maternal age at child's birth, child's birth order, birth year, maternal gestational diabetes, preeclampsia, self-reported regular maternal use of NSAIDs or aspirin during pregnancy

[5] Study adjusted for parental symptoms of ADHD, maternal alcohol use, maternal smoking, anxiety and depression symptoms during pregnancy, maternal education, marital status, BMI, maternal age, parity

[6] Specific analysis adjusted for year of birth, maternal age, parity, comedication within each indication, APAP use (pre-pregnancy or postpartum depending on trimester)

[7] Study adjusted for child sex, familial income, maternal age at delivery, maternal education, maternal pre-pregnancy body mass index, maternal smoking, alcohol use during pregnancy

were noted when APAP levels were analyzed categorically for those with high (aOR=4.10; 95% CI: 2.41, 6.95) but not low (aOR=1.44; 95% CI: 0.79, 2.63) levels of APAP, compared to those with no APAP detected in meconium. When APAP in meconium was treated as a continuous variable, each doubling of APAP was associated with a 10% increase in the odds of ADHD diagnosis (aOR=1.10; 95% CI: 1.02, 1.20). A sensitivity analysis adjusting for maternal ADHD had minimal impact on findings; however, data on maternal ADHD were only available for a subset (n=155) of mothers (11).

Ji et al. (2018) obtained venous blood samples from women enrolled in the Boston Birth Cohort (BBC) one to three days after delivery and assessed for APAP metabolite (unchanged APAP, 3-(*N*-Acetyl-L-cystein-*S*-yl) APAP, and APAP glucuronide) levels (27). Levels of all three metabolites were summed to generate a measure of overall APAP burden. Metabolites were detected in all samples. ADHD was captured by ICD codes in electronic medical record data. Compared to those in the lowest tertile, children of mothers with APAP levels in the second and third tertiles had significantly higher odds[8] of ADHD for all metabolites, except for the second tertile for APAP glucuronide. The magnitude of association was higher for the third tertiles, compared to the second tertiles, for all metabolites. Examining overall APAP burden, the risk of ADHD among children of those with levels above the median (aOR=1.88; 95% CI: 1.18, 3.00) and below the median (aOR=1.58, 95% CI: 1.02, 2.46) were significantly higher than children of mothers with non-detectable APAP (27).

A similar study was conducted in the BBC using cord plasma to assess the same APAP metabolites (28). In this study by Ji et al. (2020), unchanged APAP and APAP burden were classified into tertiles while the other two metabolites were dichotomized by detection (yes/no). Outcomes included the following four, mutually exclusive ICD code-based diagnostic groups: ADHD (n=257), ASD (n=66), ADHD and ASD (n=42), and other developmental disorders (n=304). There were significant associations[9] between all metabolites and ADHD. There was a dose-response relationship noted for APAP burden (second tertile aOR=2.26; 95% CI: 1.40, 3.69; third tertile aOR=2.86; 95% CI: 1.77, 4.67). No significant associations were noted for APAP metabolites and ADHD-ASD co-diagnosis, except for the third tertile of unchanged APAP (aOR=3.38; 95% CI: 1.25, 9.85) (28).

Gustavson et al. (2019) examined the association[10] between fever and ADHD diagnosis (ICD codes for hyperkinetic disorder) (17). Women were asked to report fever episodes, their timing during pregnancy, and whether APAP was used to manage the fever episode. Fever during pregnancy overall and during the first trimester was significantly associated with ADHD diagnosis. Although the study did not examine the direct association between APAP and

[8] Study adjusted for maternal age at delivery, maternal race/ethnicity, maternal education, smoking, alcohol, pre-pregnancy body mass index, parity, child's sex, delivery type, preterm birth, birthweight, maternal fever during pregnancy, intrauterine infection/inflammation during pregnancy, breastfeeding

[9] Study adjusted for maternal age at delivery, maternal race/ethnicity, maternal education level, marital status, stress during pregnancy, smoking before or during pregnancy, alcohol use before or during pregnancy, maternal body mass index, parity, child's sex, delivery type, preterm birth, low birth weight

[10] Study adjusted for maternal body mass index, education level, smoking, history of psychological/psychiatric problems, maternal age, parity, child year of birth, maternal ADHD symptoms

ADHD, the study did stratify the association between fever and ADHD by APAP use. Among those with fever who did not take APAP, the association with ADHD diagnosis was statistically significant (aOR=1.32; 95% CI: 1.01, 1.71); among those with fever who did take APAP, the association had a similar magnitude but was no longer statistically significant (aOR=1.35; 95% CI: 0.96, 1.90) (17).

3.3.2 Autism Spectrum Disorder (ASD)

Three studies examined the association between prenatal APAP use and ASD. In a small case-control study (n=215) of children five months to ten years of age, there was no difference in ASD prevalence by dichotomous APAP use (Chi-square p-value=0.657); no adjusted analysis was performed (34). The two BBC studies of APAP metabolites in maternal venous blood (27) and cord plasma (28) also examined ASD outcomes, identified through ICD codes. There were no significant associations[11] in the maternal venous blood study (27). In the cord plasma study, associations[12] were only noted for the highest tertile for both unchanged APAP (aOR=3.72; 95% CI: 1.70, 8.55) and APAP burden (aOR=3.62; 95% CI: 1.62, 8.60) as well as any detection for APAP glucuronide (aOR=2.29; 95% CI: 1.06, 4.85) (28). Associations with comorbid ADHD and ASD were discussed previously in Section 3.2.1 (28).

A fourth study focused on fever and ASD (18). Women were asked to report past fever and antipyretic use at gestational week 17, gestational week 30, and at six months after delivery. Children were screened and evaluated for ASD at three, five, and seven years. For children not screened, ICD codes were used to determine ASD. Fever was significantly associated[13] with ASD overall and in a dose-dependent manner (18). When timing of fever was assessed, the statistically significant association was restricted to the second trimester (18). Use of APAP for fever episodes attenuated all measures of association. In the second trimester, fever with no APAP use was associated with a significant, increased odds of ASD (aOR=1.44; 95% CI: 1.02, 2.03). For those with APAP use, the association was no longer significant (aOR=1.37; 95% CI: 0.98, 1.90) (18).

3.3.3 Other Developmental Disorders

The relationship between prenatal APAP and other developmental disorders was not commonly assessed in the studies reviewed. In the two BBC studies, there were no significant associations

[11] Study adjusted for maternal age at delivery, maternal race/ethnicity, maternal education, smoking, alcohol, pre-pregnancy body mass index, parity, child's sex, delivery type, preterm birth, birthweight, maternal fever during pregnancy, intrauterine infection/inflammation during pregnancy, breastfeeding

[12] Study adjusted for maternal age at delivery, maternal race/ethnicity, maternal education level, marital status, stress during pregnancy, smoking before or during pregnancy, alcohol use before or during pregnancy, maternal body mass index, parity, child's sex, delivery type, preterm birth, low birth weight

[13] Study adjusted for maternal age, smoking, parity, parental education, birth month, birth year, fever in other trimesters

between APAP metabolites in maternal venous blood[14] (27) or cord plasma[15] (28) and other developmental disorders. In these studies, other developmental disorders included developmental delays; intellectual disabilities; or other mental, behavioral, or neurodevelopmental disorders.

3.3.4 Behavioral Problems

Behavioral problems in children were assessed with several questionnaires including the Development and Well-Being Assessment (DAWBA) (23), the Child Behavior Checklist (CBCL) (15, 20, 33), the Teacher Report Form (TRF) (33), and the Strengths and Difficulties Questionnaire (SDQ) (13, 14, 24). Results were inconsistent across studies with respect to the relationship between APAP and total scores on the questionnaires as well as sub-scale scores.

In the study by Ruisch et al. (2018), women reported previous prenatal APAP use and other pregnancy factors at gestational week 18 (23). The DAWBA was conducted at seven years of age with a parent interview and teacher-completed questionnaire. Based on the DAWBA, children were classified as having symptoms of oppositional-defiant disorder (ODD) or conduct disorder (CD). The study used an α=0.017 to account for multiple hypothesis testing. In adjusted models[16], including adjustment for genetic risk factors, there was no association between prenatal APAP use and parent or teacher reported CD symptoms. For ODD symptoms, there was a significant association for teacher report (adjusted incidence rate ratio [aIRR]=1.22 98.3% CI: 1.00, 1.49; p=0.015) but not maternal report (aIRR=1.00; 98.3% CI: 0.86, 1.15) (23).

Three studies used the CBCL as an outcome measure (15, 20, 33). Tovo-Rodrigues et al. (2020) found no association[17] between prenatal APAP and clinically dichotomized CBCL T-scores, per the CBCL scoring manual, in children at 48 months of age (15). There was no association between APAP and total CBCL scores (aRR=0.99; 95% CI: 0.85, 1.16) or any subscale score (15). There was no effect measure modification by sex (15). Trønnes et al. (2020) examined the

[14] Study adjusted for maternal age at delivery, maternal race/ethnicity, maternal education, smoking, alcohol, pre-pregnancy body mass index, parity, child's sex, delivery type, preterm birth, birthweight, maternal fever during pregnancy, intrauterine infection/inflammation during pregnancy, breastfeeding

[15] Study adjusted for maternal age at delivery, maternal race/ethnicity, maternal education level, marital status, stress during pregnancy, smoking before or during pregnancy, alcohol use before or during pregnancy, maternal body mass index, parity, child's sex, delivery type, preterm birth, low birth weight

[16] Study adjusted offspring sex, socio-economic status, young maternal age, single parent status during pregnancy, offspring ADHD symptom scores, maternal smoking, alcohol use, cannabis/hard drug use, life events stress score, anxiety symptom score, depression symptom score, genetic risk factors, comorbid disruptive behavior symptomatology

[17] Study adjusted for Family Wealth Index Quotient, maternal schooling, age, skin color, marital status, parity, smoking during pregnancy, alcohol consumption during pregnancy, mood symptoms during pregnancy, pre-gestational BMI, prenatal care, infectious diseases during pregnancy, high blood pressure during pregnancy, gestational diabetes/treatment, child sex, other analgesic use

relationship[18] between APAP use by trimester and CBCL internalizing problems and internalizing problems subscale scores (20). Scores were dichotomized based on T-scores (clinically significant T-score ≥63). There were no associations between the number of trimesters of exposure and behavior problems at five years of age, except that those exposed to APAP in three trimesters had 1.36 times the risk (95% CI: 1.02, 1.80) of internalizing problems, than those with no exposure. When grouped by trimester of exposure (none, first trimester, second/third trimester), there were no significant associations with behavior problems (20). CBCL total and sub-scale T-scores were treated as continuous as well as dichotomous (T-scores >60, indicative of clinical and borderline clinical scores) in Parker et al. (2020) (33). No significant associations[19] were noted between overall APAP use and total CBCL scores (continuous or dichotomous) or sub-scale scores (continuous or dichotomous). No associations were found when prenatal APAP use was dichotomized as short-term (<28 uses) or long-term (≥28 uses) (33). Parker et al. (2020) also assessed the same associations with behavior problems reported by teachers with the TRF. Again, no significant associations were found with continuous T-scores or dichotomized scores (33). Emphasizing the impact of confounding by indication, in adjusted models without stabilized inverse probability weighting (SIPW) and without adjustment for indication there was a significant association between APAP and total maternal-reported behavioral problems T-scores (any APAP adjusted mean difference [aMD]=2.2 [95% CI: 0.3, 4.1], short-term APAP aMD=2.0 [95% CI: 0.0, 4.0], long-term APAP aMD=2.8 [95% CI: 0.1, 5.5]). No associations were noted for dichotomous total behavioral problems outcomes or teacher-reported total behavioral outcomes (33).

The SDQ was used to capture behavioral disorders by three of the studies reviewed. In general, there were some significant associations with worse behavior problems among those exposed to APAP; however, the results were dependent on which SDQ score was considered, how the exposure was operationalized, and the reporting source of the SDQ.

Tovo-Rodrigues et al. (2018) assessed prenatal APAP exposure after delivery (14). Psychologists administered the SDQ to parents/caregivers when children were six and eleven years of age. Established, screening cut-offs specific to the Brazilian population were used to dichotomize SDQ total and sub-scale scores. There was evidence of statistical interaction by sex for some outcomes at six years of age, but not eleven at years of age. There were no associations[20] between APAP use and total behavioral difficulties, emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, or prosocial behavior for girls at six years, girls at eleven years, and boys at eleven years of age. There was an increased odds of

[18] Study adjusted for maternal age at delivery, marital status, education level, parity, pre-pregnancy BMI, folic acid supplement, smoking habits, alcohol use, symptoms of anxiety and depression, maternal health conditions during pregnancy (headache/migraine, pain, fever/infections), concomitant medication use, child sex

[19] Study analysis weighted by stabilized inverse probability weights and adjusted for maternal age, race, education, marital status, smoking during early pregnancy, drinking during early pregnancy, headache, fever, pain, and upper respiratory infection without fever

[20] Study adjusted for economic indicator, sex, mother's educational level, mother's age, mother's skin color, mother's parity, smoking during pregnancy, alcohol consumption during pregnancy, mood issues during pregnancy, infections during pregnancy, pre-pregnancy body mass index, other nonsteroidal anti-inflammatory drug use during pregnancy

emotional symptoms (aOR=1.47; 95% CI: 1.07, 2.02) and hyperactivity/inattention (aOR=1.42; 95% CI: 1.06, 1.92) for boys at six years of age only (14). In a sensitivity analysis there was no association between prenatal NSAID use and outcomes (14).

Inoue et al. (2021) examined the relationship[21] between overall and trimester-specific prenatal APAP exposure and SDQ total, composite sub-scale, and individual sub-scale scores at eleven years of age (24). The SDQ was administered to both parents and children when the children were eleven years of age; scores were dichotomized based on recommended cut-offs or the 95% percentile of score distributions, if no recommended cut-off was available. The correlation coefficient for child and parent total scores was 0.58. There was a significant association between prenatal APAP use and worse total behavioral difficulties reported by both parents (aRR=1.14; 95% CI: 1.01, 1.29) and children (aRR=1.40; 95% CI: 1.20, 1.63). For parent-reported outcomes, difficulties were significantly worse among those exposed to APAP for the internalizing composite sub-scale as well as emotional symptoms and hyperactivity sub-scales. For child-reported outcomes, significantly worse difficulties were noted for the internalizing and externalizing composite sub-scale as well as emotional symptoms, conduct problems, and hyperactivity. Trimester-specific results were inconsistent. For parent-reported total SDQ behavioral difficulties, there was a significant association (more difficulties) with two trimesters of APAP exposure (aRR=1.22; 95% CI: 1.03, 1.45). For child-reported difficulties, there was a significant association (more difficulties) with first trimester exposure only (aRR=1.50; 95% CI: 1.19, 1.88), third trimester exposure only (aRR=1.33; 95% CI: 1.03, 1.71), and exposure in all three trimesters (aRR=1.53; 95% CI: 1.19, 1.96). However, there was a dose-response relationship by weeks of exposure for both parent-reported (p for trend=0.03) and child-reported (p for trend <0.01) total behavioral difficulties. The study also examined postnatal exposure in the first 18 months and found no significant associations with parent-reported (aRR=1.18; 95% CI: 0.95, 1.48) or child-reported (aRR=0.97; 95% CI: 0.73, 1.30) SDQ total difficulties (24).

The third study by Rifas-Shiman et al. (2020) also examined the impact[22] of both prenatal and postnatal APAP exposure on mother and teacher-reported SDQ behavioral difficulties (13). Prenatal APAP exposure was only captured in the early and middle stages of pregnancy; postnatal APAP was assessed from birth to one year of age. Based on frequency of APAP use, utilization was classified into four ordinal categories or dichotomous categories. SDQ total scores and prosocial sub-scale scores were calculated; all outcomes remained continuous. Significantly higher, or worse, scores for parent-reported total difficulties were noted for each categorical increase in APAP exposure (adjusted beta-coefficient [aβ]=0.24; 95% CI: 0.02, 0.46) and those with any (versus no) prenatal APAP exposure (aβ=0.75; 95% CI: 0.17, 1.32) but not when use was categorized as ≥10 versus <10 uses. Maternal-reported prosocial scores were significantly better, or higher, among those exposed to prenatal APAP ≥10 times, compared to those exposed <10 times. For teacher-reported outcomes, greater APAP exposure, using any

[21] Study adjusted for mother's age at childbirth, parity, maternal socio-economic status, pre-pregnancy body mass index, birth year, maternal smoking, alcohol use during pregnancy, maternal muscle/joint disease during pregnancy, maternal fever during pregnancy, maternal infection/inflammation during pregnancy, nonsteroidal anti-inflammatory drug use during pregnancy

[22] Study adjusted for maternal age, education, parity, pregnancy smoking status, household income, HOME score, gestational age, child sex, child race/ethnicity, antibiotic use during pregnancy, antidepressant use during pregnancy, mother depressive symptoms mid-pregnancy

categorization, was significantly associated with worse total behavioral difficulties scores. No significant associations were noted for prosocial behavior. Postnatal APAP use was significantly associated only with parent-reported total behavioral difficulties (13). When prenatal and postnatal APAP use were examined together, results suggested an additive relationship (13). The study also examined the impact of prenatal and infant use of ibuprofen and found some significant associations with behavior outcomes (13).

One study by Golding et al. (2020) explored the relationship between prenatal APAP exposure (between 18 to 32 weeks gestation) and 135 different outcomes, including behavioral problems, assessed at different ages and with differing reporting sources (22). The study examined the crude association between APAP use (dichotomized as yes/no) and each of the 135 outcomes, operationalized continuously. Associations with a p-value <0.0001 were further explored in adjusted analysis. Some significant associations were found in the adjusted analysis ($p<0.05$).[23] After adjustment there were significant associations for the seven SDQ outcomes and two DAWBA outcomes. Focusing on total scores, the mean SDQ total behavioral difficulties scores at 42 months (aMD=0.54; 95% CI: 0.26, 0.82) and 47 months (aMD=0.31; 95% CI: 0.08, 0.53) were significantly higher among those with APAP exposure between 18-32 weeks of gestation (22).

3.3.5 Temperament

Trønnes et al. (2020) studied the association between trimester-specific prenatal APAP exposure and several outcomes, including temperament, at five years of age (20). Temperament was assessed with the short version of the Emotionality, Activity, and Shyness Temperament Questionnaire (EAS), specifically emotionality, activity, sociability, and shyness sub-scales (20). The EAS has no recommended cut-off score. Weighted, adjusted models[24] demonstrated no association between prenatal APAP use and emotionality, activity, and sociability T-scores. There was a significant association between prenatal APAP use and shyness, but only when comparing those with two trimesters of exposure, compared to no exposure. Individuals with two trimesters of exposure had significantly lower shyness T-scores (aβ= -0.62; 95% CI: -1.05, -0.19) (20). APAP use prior to pregnancy (NCE) was only significantly associated with activity (lower sub-scale scores in those exposed) (20). Golding et al. (2020) also examined temperament outcomes, operationalized with continuous scales (22).[25] Adaptability scores at six months

---

[23] Associations were adjusted for history of asthma, history of indigestion, history of back pain, history of migraine, pre-pregnancy BMI, in poor health (18-32 weeks), had a cold (18-32 weeks), had flu (18-32 weeks), had other infection (18-32 weeks), had headache (18-32 weeks), healthy diet score, processed diet score, drank no alcohol, domestic chemical score, parity

[24] Adjusted for maternal age at delivery, marital status, education level, parity, pre-pregnancy BMI, folic acid supplement, smoking habits, alcohol use, symptoms of anxiety and depression, maternal health conditions during pregnancy (headache/migraine, pain, fever/infections), concomitant medication use, child sex

[25] Associations were adjusted for history of asthma, history of indigestion, history of back pain, history of migraine, pre-pregnancy BMI, in poor health (18-32 weeks), had a cold (18-32 weeks), had flu (18-32 weeks), had other infection (18-32 weeks), had headache (18-32 weeks), healthy diet score, processed diet score, drank no alcohol, domestic chemical score, parity

(aMD=0.33; 95% CI: 0.05, 0.60) and persistence scores at 24 months (aMD=0.36; 95% CI: 0.12, 0.60) were significantly higher among those with APAP exposure between 18 and 32 weeks of gestation. As discussed in the behavioral disorders section, Golding et al. (2020) used an exploratory approach and screened many continuous outcomes (22).

3.3.6 Attention

Liew et al. (2016) assessed attention in children at five years of age with the Test of Everyday Attention for Children at Five (TEACh-5) (25). An overall attention score was calculated along with a selective attention sub-scale score and a sustained attention sub-scale score. All scores were obtained by re-standardizing (mean=0, standard deviation [SD]=1) the mean of standardized, relevant sub-scores. There were no significant differences in scores by ever (versus never) prenatal use. Use of APAP during the first trimester was associated[26] with significantly lower, or worse, overall attention scores (aMD= -0.34; 95% CI: -0.63, -0.05). There were no other trimester-specific associations. When the TEACh-5 scores were dichotomized (1 SD below the mean), ever use was associated with an increased odds of subnormal overall attention (aOR=1.5; 95% CI: 1.0, 2.5) and selective attention (aOR=1.5; 95% CI: 1.0, 2.4) but not sustained attention (aOR=1.3; 95% CI: 0.8, 2.1). There were no trimester-specific associations for overall attention. A dose-response was noted between the number of weeks of prenatal APAP exposure and subnormal overall attention (p for trend=0.05). There were no trends noted for selective or sustained attention. Results did not change substantially when restricted to those without certain indications (25).

3.3.7 Cognition/Executive Function/Neurodevelopment

Several studies examined the relationship between APAP exposure and various constructs related to cognition (12), executive function (13, 25), and neurodevelopment (15). Similar to other constructs assessed with questionnaire scales, findings were mixed.

The exploratory study by Golding et al. (2020), which included cognitive outcomes, found a significant, adjusted association[27] only for one cognitive outcome— APAP use between 18-32 weeks gestation and freedom from distractibility at eight years (aMD= -0.35; 95% CI: -0.69, -0.00) (22). Further details of this study are presented in the sections on behavioral problems and temperament.

In Bertoldi et al. (2020), cognitive outcomes in children were assessed in a Massachusetts cohort (Project Viva) with both the Peabody Picture Vocabulary Test (PPVT-III) and the Wide Range

[26] Study adjusted for maternal age at child's birth, child's birth order, birth year, maternal gestational diabetes, preeclampsia, self-reported regular maternal use of NSAIDs or aspirin during pregnancy

[27] Associations were adjusted for history of asthma, history of indigestion, history of back pain, history of migraine, pre-pregnancy BMI, in poor health (18-32 weeks), had a cold (18-32 weeks), had flu (18-32 weeks), had other infection (18-32 weeks), had headache (18-32 weeks), healthy diet score, processed diet score, drank no alcohol, domestic chemical score, parity

Achievement of Visual Motor Abilities (WRAVMA)[28] (12). In the same publication, cognitive outcomes in a Brazilian cohort (Pelotas Cohort) were assessed with the INTERGROWTH-21st Neurodevelopment Assessment (INTER-NDA)[29] (12). All PPVT-III, WRAVMA, and INTER-NDA scores were standardized. Outcomes were assessed, on average, at 3.3 years of age in the Massachusetts sample and 2.0 years of age in the Brazilian sample. There were no associations between prenatal APAP use and PPVI-III outcomes and WRAVMA total scores. The only significant association between prenatal APAP use and WRAVMA sub-scale scores was between APAP use in the first and second trimester and WRAVMA drawing scores (aβ= -1.51; 95% CI: -2.92, -0.10). APAP use in the first and second trimesters (INTER-NDA aβ=0.08; 95% CI: 0.01, 0.16) as well as in the first, second, and third trimesters were associated with better total cognitive development (INTER-NDA aβ=0.10; 95% CI: 0.01, 0.18). There were no associations between APAP use in infancy (≤ one year of age) and cognitive outcomes (12).

Two studies examined executive function outcomes assessed with the Behavior Rating Inventory of Executive Function (BRIEF) (13, 25). Rifas-Shiman et al. (2020) assessed executive function in children in mid-childhood (around eight years of age) (13). Prenatal APAP use, regardless of operationalization (categorical, dichotomous), was associated[30] with significantly higher, or worse, T-scores for parent-rated versions of the BRIEF Global Executive Composite, the BRIEF Behavior Regulation Index, and the BRIEF Metacognition Index. For example, those exposed to any prenatal APAP had 1.60 points higher adjusted T-scores (95% CI: 0.47, 2.73) on the parent-rated BRIEF Global Executive Composite, compared to those with no exposure. Early and mid-pregnancy APAP use was associated with significantly worse teacher-rated BRIEF Global Executive Composite and Behavior Regulation Index T-scores, with one exception (any versus never APAP exposure and Behavior Regulation Index). There were no significant associations between early and mid-pregnancy APAP exposure and BRIEF Metacognition Index T-Scores (13). There were significant associations between APAP use in infancy and worse parent-rated BRIEF Composite and Index T-scores, except for categorical APAP use and BRIEF Behavior Regulation Index scores. There were no significant associations between infant APAP use and teacher-rated BRIEF scores (13). Some significant associations were also noted for prenatal and infant ibuprofen exposure (13). There was evidence of an additive effect of APAP and ibuprofen (13). Liew et al. (2016) also assessed executive function with the BRIEF completed by both a parent and teacher; scores were assessed at five years of age and treated as both continuous T-scores (derived from sample-specific distributions) and dichotomous scores (1 SD above the mean) (25). Each week increase in prenatal APAP exposure was associated[31] with an increased

[28] Analysis of prenatal APAP associations in Project Viva adjusted for maternal age, race/ethnicity, education, household income, parity, smoking/alcohol intake during pregnancy, pre-pregnancy body mass index, antibiotic use during pregnancy, depressive symptoms during pregnancy, ibuprofen use during pregnancy

[29] Analysis of prenatal APAP associations in Pelotas adjusted for maternal age, self-reported skin color, education, household income, parity, smoking/alcohol intake during pregnancy, pre-pregnancy body mass index, antibiotic use during pregnancy, depression/anxiety during pregnancy, ibuprofen use during pregnancy

[30] Study adjusted for maternal age, education, parity, pregnancy smoking status, household income, HOME score, gestational age, child sex, child race/ethnicity, antibiotic use during pregnancy, antidepressant use during pregnancy, mother depressive symptoms mid-pregnancy

[31] Study adjusted for maternal age at child's birth, child's birth order, birth year, maternal gestational diabetes, preeclampsia, self-reported regular maternal use of NSAIDs or aspirin during pregnancy

odds of parent-rated metacognition difficulty (aOR=1.07; 95% CI: 1.01, 1.14) and a significant dose-response was noted (p=0.02). There were no other significant associations between overall prenatal APAP use, trimester-specific prenatal APAP use, or cumulative (by week) prenatal APAP use and BRIEF General Executive Composite, BRIEF Regulation Index, or BRIEF Metacognition Index outcomes, except teacher-rated Metacognition Index for 2-5 weeks of use (aOR=2.4; 95% CI: 1.04, 4.4). No significant associations were found when stratifying by indication or sex (25).

One Brazilian study examined neurodevelopmental milestones with Battelle's Developmental Inventory (BDI) assessments at 24 months (15). BDI total and domain scores were dichotomized based on a threshold used in a prior study (10th percentile being indicative of low BDI performance). APAP exposure was assessed at a single perinatal assessment. Prenatal APAP use was not associated[32] with low neurodevelopmental performance in adjusted models overall (aRR=1.00; 95% CI: 0.78, 1.28) or with respect to BDI domains (15). There was no evidence of differences by sex (15).

The study by Laue et al. (2018) sought to determine if prenatal APAP exposure, measured by meconium APAP concentration, is correlated with subtest scores on the Wechsler Intelligence Scale for Children at age 6-8 years (31). There was no clear pattern associating[33] decreased subtest scores with APAP exposure. Limitations included a relatively small sample (n=188 children total, 31 of whom had meconium APAP levels in the higher category) and the fact that full scale intelligence quotient was not analyzed (31).

3.3.8 Language and Communication

A study by Skovlund et al. (2017) focused on the association between prenatal analgesic opioid use and language competence and communication skills at three years of age (19). However, the study also examined the relationship[34] between prenatal APAP use and the aforementioned outcomes. Language competence and communication skills were determined by a language grammar rating scale and six skill items from the Ages and Stages Questionnaire (ASQ), respectively. Skills ability was categorized based on prior studies or validations. Prenatal APAP exposure was reported over three periods: six months pre-pregnancy through gestational weeks 17-18, gestational week 19 through gestational week 29, and gestational week 30 to birth. There was no association between number of periods of prenatal APAP use and lower language competence. Compared to no APAP exposure, there was a significant association between prenatal APAP use and low communication skills for those with two periods (aOR=1.08; 95%

[32] Study adjusted for Family Wealth Index Quotient, maternal schooling, age, skin color, marital status, parity, smoking during pregnancy, alcohol consumption during pregnancy, mood symptoms during pregnancy, pre-gestational BMI, prenatal care, infectious diseases during pregnancy, high blood pressure during pregnancy, gestational diabetes/treatment, child sex, other analgesic use

[33] Study adjusted for maternal age at delivery maternal body mass index, parity, maternal intelligence quotient, sex, gestational age, birth weight, 5-minute Apgar score, age at follow-up, maternal education, family income

[34] APAP analysis adjusted for opioid use, pain, maternal work situation, paternal education, maternal body mass index, parity, maternal smoking, benzodiazepine, selective serotonin reuptake inhibitor use during pregnancy

CI: 1.03, 1.14) and three periods (aOR=1.16; 95% CI: 1.04, 1.28) but not one period (aOR=1.00; 95% CI: 0.96, 1.05) of APAP exposure (19). For opioids, there was no association between the number of periods of prenatal opioid use and lower language competence or communication skills. For example, the association between two or three periods of opioid use and lower language competence was aOR=0.92 (95% CI: 0.62, 1.35) and lower communication skills was aOR=1.13 (95% CI; 0.83, 1.53) (19).

Trønnes et al. (2020) assessed communication problems among over 32,000 children at five years of age with the ASQ (20). Maternal-reported scores were dichotomized based on T-scores (T-score ≥65 suggesting a communication problem, per prior screening validation). There were no significant associations[35] between number of trimesters of APAP use or specific trimesters of use and communication problems. The study used the pre-pregnancy period as a NCE. There was a signfiicant, adjusted association between pre-pregnancy APAP use and communication problems (aRR=1.19; 95% CI: 1.02, 1.38) (20).

In the study by Bornehag et al. (2018), prenatal APAP exposure was determined by self-report of use between conception and gestational weeks eight to thirteen (16). A maternal urinary sample to determine APAP levels was collected between gestational weeks eight to thirteen. Language delay was assessed at 30 months of age. Among those with both APAP measures, there was a correlation between self-reported number of pills and urinary concentration of APAP. Among boys, there were no associations[36] between either self-reported APAP (n=388) or urinary APAP concentration (n=63) and language delay. Among girls, there was a significantly increased odds of language delay among those reporting any APAP use (aOR=4.64; 95% CI: 1.02, 21.05; n=366). Compared to those reporting no pills, there was only a significant association in those reporting the highest number of pills (>6) (aOR=5.92; 95% CI: 1.10, 31.94). In the urinary sample subset (n=48), the association was non-significant for logarithmic APAP use (aOR=1.50; 95% CI: 0.95, 2.34) and only significant for the highest (fourth) quartile of APAP concentration, compared to the lowest (first) quartile (aOR=10.34; 95% CI: 1.37, 77.86) (16).

3.4 Urogenital Outcomes

Only two new studies assessed the relationship between APAP exposure and urogenital outcomes. One study assessed AGD (30) and the other assessed pubertal development (26).

A meta-analysis conducted by Gurney et al. (2016) examined the association between prenatal analgesic use and cryptorchidism (39). The study found a weak association between use of analgesia and cryptorchidism (crude OR=1.11; 95% CI: 1.00, 1.23; I^2=0.0%; n=10 studies). There were no significant associations when stratified by study type (case control [n=5], cohort [n=5]). Of note, the meta-analysis considered all analgesics, not just APAP (39).

---

[35] Study adjusted for maternal age at delivery, marital status, education level, parity, pre-pregnancy body mass index, folic acid supplement, smoking habits, alcohol use, symptoms of anxiety and depression, maternal health conditions during pregnancy (headache/migraine, pain, fever/infections), concomitant medication use, child sex

[36] Study adjusted for maternal weight, education, smoking, and week of enrollment

3.4.1 Anogenital Distance (AGD)

Navarro-Lafuente et al. (2021) found no association[37] between any prenatal APAP use and anus-scrotum AGD Z-scores (aβ= - 0.003; 95% CI: -0.23, 0.24) or anus-penis AGD Z-scores (aβ=0.03; 95% CI: -0.2, 0.26) (30). There were no significant associations between APAP exposure in the first, second, or third trimester and either anus-scrotum or anus-penis AGD (30). There were also no associations when prenatal APAP was operationalized as the number of days of use overall or in the first trimester (30).

3.4.2 Pubertal Development

Ernst et al. (2018) studied[38] individuals enrolled in Danish Puberty Cohort, which is nested within the DNBC (26). Pubertal milestones and stage (Tanner stage) were assessed at 11.5 years of age and re-assessed every six months (26). Among boys, there were no significant differences in age of pubertal milestones (genital Tanner stage, pubic hair Tanner stage, axillary hair, acne, voice break, adult voice, and first ejaculation) among those exposed and not exposed to prenatal APAP. For girls, the only significant associations were between prenatal APAP exposure and axillary hair development and acne development; there were no associations with breast Tanner stage, pubic hair Tanner stage, axillary hair, acne, and menarche. Girls with prenatal APAP exposure had significantly early ages of axillary hair development (aMD= -1.2 months; 95% CI: -2.2, -0.2) and acne development (aMD= -1.7 months; 95% CI: -2.7, -0.6). Among girls, differences in age of pubertal milestones were most pronounced for those with the highest amount of cumulative exposure (>12 weeks) and those exposed in the third trimester; however, these patterns were not universal (26).

4 DISCUSSION

The reviewed articles address some of the study criticisms raised in prior DEPI-I reviews of the relationship between APAP and neurodevelopmental and urogenital outcomes. For example, more studies address possible residual confounding as well as the role of APAP use in early childhood. However, there are several methodological considerations and limitations that persist and make synthesizing the overall evidence challenging.

4.1 Exposure Assessment

[37] Study adjusted for previous pre-term deliveries and gestational age

[38] Study stratified by sex and adjusted for pre-pregnancy body mass index, alcohol units per week in 1st trimester, # of cigarettes/day 1st trimester, time to pregnancy, highest social class of parents, maternal age at menarche, maternal age at delivery, parity, fever during pregnancy, muscle or joint disease during pregnancy, inflammation or infection during pregnancy

Most studies relied on maternal self-report of prenatal APAP use (12-26, 29, 30, 33, 34). All studies relying on maternal self-report, except for Saunders et al. (2019) (34), collected prenatal APAP exposure information prior to outcomes. Saunders et al. (2019) conducted a case-control study and assessed prenatal exposure information through interviews and medical records when children ranged in age from five months to ten years (34). Prospective data collection strategies reduce the risk of information bias and, specific to case-control studies, recall bias. As a result, any misclassification of the exposure is likely to be non-differential and bias results towards the null. The meta-analysis by Masarwa et al. (2020) conducted several bias analyses and examined the impact of exposure misclassification on meta-analysis findings. Meta-analysis corrected estimates suggest that correction for exposure misclassification results in a substantially stronger magnitudes of association (38). For example, in individual studies in the meta-analysis, correction for exposure misclassification, compared to unadjusted estimates, increased the relative risk by between 0.11 and 0.79 (38).

Three studies reviewed only collected self-reported prenatal APAP exposure information after delivery. Two studies collected information at a post-birth perinatal evaluation (14, 15) while one collected this information an average of 12 months after delivery (33). Two studies only captured self-reported prenatal APAP exposure at one time point—18 weeks gestation (23) and 32 weeks gestation (30). Although the former study also captured prior medication use in the medical record at birth (23). The majority of studies reviewed captured self-reported prenatal APAP use at multiple time points. Although it varied by study, interviews or questionnaires were typically conducted in early pregnancy, middle pregnancy, and after delivery with women asked to report APAP use over a designated prior period (12, 13, 16-22, 24-26). Compared to studies that conducted only one post-delivery interview about prenatal APAP use, these data collection schedules reduced the amount of time participants needed to remember prior APAP use. Reducing recall time may limit exposure misclassification, especially with an over-the-counter drug that women may be less likely to remember taking, compared to a prescription drug (40). Additionally, this approach emphasizes weekly- and trimester-specific use, which may be pertinent if studies are interested in plausible vulnerability or critical development periods during pregnancy.

Studies differed in how participants were asked about APAP use. Although some studies may have captured more granular information on prenatal APAP use (e.g., number of days used, dosage, number of tablets used, number of weeks used), for primary analyses, most studies typically dichotomized or categorized prenatal APAP use overall or by trimester. Several studies operationalized APAP exposure multiple ways. For example, Inoue et al. (2021), operationalized APAP exposure by trimester, by cumulative number of trimesters, by number of weeks of exposure (categorical), and by any exposure (24). These more nuanced operationalizations allow for dose-responses, cumulative exposures, and plausible critical development periods to be assessed. However, dosage information was not captured, and specific timing information was often unavailable in these studies.

One study asked women to report, on a bi-annual basis, if they regularly used APAP over the prior two years (29). Questionnaire dates were cross-walked with birth years and APAP use was imputed to the pregnancy. A subset of individuals did complete the questionnaire while pregnant (29). Misclassification is likely in this study, unless pregnancy has no impact on APAP use.

Several studies suggest that misclassification of self-reported prenatal APAP exposure would likely be non-differential and bias results towards the null (14, 25, 29, 30).

Three studies additionally captured postnatal exposure to APAP; exposure was assessed at three months (12), six months (24), twelve months (12, 13), and eighteen months (24) of age. However, APAP exposure through breastmilk was not directly captured (13).

Six studies reviewed did not use maternal self-report to capture prenatal APAP use. Chen et al. (2019) captured prescription claims for prenatal APAP use from medical records (32). The study was able to determine the cumulative dosage by trimester; however, actual adherence and over-the-counter use was not captured (32). Five studies directly measured APAP or metabolite levels to assess prenatal APAP exposure. One study used urine collected once at enrollment (between eight and thirteen weeks of gestation) (16). This study did find a correlation between urinary APAP concentrations and maternal self-report of use. However, all samples had some detectable APAP, possibly reflecting some environmental sources of APAP, such as aniline, and urinary concentrations only reflect very recent use (prior 36-48 hours) (16, 41). Baker et al. (2020) and Laue et al. (2018) assessed APAP levels in meconium (11, 31); however, meconium only captures prenatal use in the last two-thirds of pregnancy (11, 31). In Ji et al. (2018), APAP metabolites were assessed in a maternal, venous blood sample taken one to three days postpartum; however, this biomarker reflects only very recent use (APAP has a half-life of 2-3 hours) and is impacted by metabolism (27, 42). Ji et al. (2020) assessed APAP metabolites in cord plasma, which again only reflects recent use due to APAP's short half-life (28, 43). Because these two studies only capture very recent use, the studies may not be capturing prenatal or peripartum use at all. These two metabolite studies created post hoc metabolite categories and dichotomies based on statistical cut-points related to study sample distribution percentiles (27, 28). APAP assessed in the perinatal period may also be impacted by APAP use during delivery. Only Baker et al. (2020) conducted a sensitivity analysis excluding women with APAP use during delivery (11). Overall, these directly measured APAP levels, although free from bias because they are objective measures (e.g., not self-report), they may not be a valid reflection of total prenatal APAP use during pregnancies.

4.2 Outcome Assessment

For the AGD study, assessments were conducted soon after birth (30). For puberty milestones, outcomes were assessed from eleven through eighteen years (26). It is possible that there were some children with puberty onset prior to eleven years in the milestones study; these individuals would have their outcomes of interest missed due to left truncation (26). For functional neurobehavioral outcomes, assessments were conducted at varying timepoints. For some studies, outcomes were assessed at fixed ages (11-16, 18-20, 23-25, 31, 33). Other studies assessed outcomes throughout follow-up or at the study termination date (17, 21, 27-29, 32, 34). For example, in Gustavson et al. (2019), the median age of children at the end of study completion ranged from seven to seventeen years of age (17) and in Saunders et al. (2019), it ranged from zero to ten years of age (34). For studies without fixed assessment times or those with assessments at very early ages, there may be censoring or right truncation of outcomes that may be found at, or more reliably assessed at, a later age. For example, the prevalence of ADHD

diagnoses may increase with age (44). One study that examined outcomes between 2008 and 2016 accounted for differential follow-up for ADHD by accounting for child age (captured in analysis with year of birth) (17). In the meta-analysis by Masarwa et al. (2018), child age at follow-up was a significant effect modifier (stronger associations with greater age) (36). However, a second meta-analysis found no effect modification by age at child assessment (35). Section 4.4 will discuss the implications of this data structure on the statistical analysis. Assessments of outcomes, like cognition, at young ages may be unreliable and be more reflective of educational interactions with parents (12).

Several studies used ICD codes/clinician diagnoses for ADHD, ASD, or other developmental disorders (11, 17, 18, 27, 28, 32, 34) or had parents report whether their child has been diagnosed with one of these disorders by a clinician (11, 29). Using clinician diagnoses improves the validity of outcome assessment (32). However, there may be inherent differences in those with greater healthcare utilization, compared to those with lower utilization, or those who seek out medical care to receive a diagnosis or verbalize concerns, versus those who do not. There also may be an under-ascertainment of outcomes (32). However, one meta-analysis found similar results between APAP and ADHD and APAP and ASD when restricting to those with diagnoses in a hospital register (35).

Differences in the age of diagnosis or the likelihood of a clinical diagnosis may be non-differential based on prenatal APAP exposure and bias findings towards the null. However, as noted previously, there are differences in health status among those with high prenatal APAP utilization, compared to those with none or lower utilization (3). If these underlying health differences impact how a mother seeks care for her child, it could increase or decrease the likelihood of a diagnosis and either bias findings away or toward the null, respectively.

The other studies assessed various neurobehavioral constructs with questionnaires and child assessments (12-16, 19, 20, 22-25, 31, 33). No one item was consistently used. Continuous overall scale and sub-scale scores were often calculated, standardized (e.g., T-scores), and dichotomized. Differences in scores on continuous scales may not be clinically meaningful and do not provide information on the number of children who would meet clinical, diagnostic criteria. Although there are often established, clinically meaningful thresholds for the scales used, these scores do not necessarily indicate a diagnosis of a specific disorder. These thresholds on screening tools sometimes simply suggest that further screening and assessment may be needed. Differences in scale validity would likely be non-differential and bias any results towards the null. However, if thresholds are too specific, it could result in outcome analyses being underpowered. For the questionnaires and assessments, the reporting source often varied (e.g., parent, teacher, child, trained assessor). Several studies examined differences in associations by reporting source (22-25, 33). Differences in associations were noted by reporting source for some studies (23, 24, 33), emphasizing the subjective nature of non-diagnostic outcome assessments. One meta-analysis found no difference in the association between APAP and ADHD by whether outcomes were obtained through parental report or standard assessments (37).

4.3 Confounding

Neurobehavioral disorders are complex. For example, the etiology of ADHD is multifactorial and risk factors may include genetics; male sex; low birthweight/prematurity; environmental toxins; nutrition; low SES; and in-utero exposures (e.g., maternal stress, cigarettes, alcohol, toxins, medications, illicit drugs) (45). As noted previously, there may be differences between those who use or do not use APAP during pregnancy (3). Consequently, there could be many potential confounders of the associations between APAP and neurobehavioral or urogenital outcomes. Potential confounders include genetic factors, environmental factors, familial characteristics, social characteristics, indications, and severity of indications.

As noted in Appendix C, the reviewed studies adjusted for a multitude of child, maternal, paternal, and household characteristics. There could still be unmeasured confounding in studies by unknown or unmeasured factors. This limitation was frequently acknowledged in the reviewed studies. Residual confounding, depending on the relationship between the confounder and prenatal APAP as well as the confounder and the outcome, could bias results either away or toward the null. It is also possible for residual confounding to switch the directionality of an association from a risk factor to a protective factor, or vice versa.

The meta-analysis by Masarwa et al. (2020) examined the association between APAP and ADHD; the study also assessed the impact of unmeasured confounding on study findings (38). The unmeasured confounding analysis examined parental ADHD; maternal fever, migraine, smoking and alcohol use; and air pollutants. Although the study found that the meta-analysis findings were robust to unmeasured confounding, individual studies were not. Unmeasured confounders both increased and decreased associations, depending on the study. Adjustment for parental ADHD, a proxy for genetic factors, and maternal migraine, an indication, resulted in the largest attenuation in measures of association (38).

A study by Leppert et al. (2019) found a significant association between maternal polygenic risk scores for ADHD and early and late pregnancy use of APAP (46). The same association was not found for ASD (46). This study suggests that genetic confounding, which is often not considered, could be important. However, genetic testing, as part of a large epidemiologic study, may be expensive and unfeasible. ADHD is highly heritable and familial factors, such as parental ADHD, may serve as reasonable proxies to account for genetic confounding (45). With respect to other familial factors, parental psychiatric disorders as well as parental educational level or socio-economic status may also be important confounders. Although studies may capture familial factors (e.g., neurobehavioral disorders, psychiatric disorders, education level, or socio-economic status), studies typically capture maternal (not paternal) characteristics (see Appendix C).

Potential confounding by indication is the major criticism of studies examining outcomes of prenatal APAP exposure. Several of the studies reviewed directly adjusted for indications for prenatal APAP use including fever (20, 21, 24-27, 33), pain (19-21, 33), headache/migraine (20, 22, 33), muscle or joint disease (24-26), and infections/inflammation (14, 15, 20-22, 24-27, 32, 33). One study used prenatal antibiotic use as a proxy for conditions likely to increase the use of APAP (13). Others incorporated confounding by indication in sensitivity, or supplemental, analyses. For example, Bornehag et al. (2018) adjusted for maternal number of colds in a

supplemental analysis and found no substantial changes in measures of association (16). Ji et al. (2020) conducted a subgroup analysis by maternal fever (28). Although studies may have adjusted for confounding by indication, it is not clear that the APAP use was used for a given indication. Four studies asked about use of APAP for specific indications (17, 20, 21, 33); however, only two studies analyzed associations stratified by concordant APAP use and indication (17, 21).

Use of other medications during pregnancy may impact neurodevelopment and may be used in addition to APAP. As a result, studies should consider adjustment for other medications. As discussed previously, migraine was identified as an important confounder. Polypharmacy, including the use of APAP, is seen among pregnant persons with migraine (47). There is mixed evidence on whether migraine medication exposure during pregnancy is associated with behavioral problems (48, 49). Maternal substance use (i.e., illicit drugs, tobacco, and alcohol) may also be an important confounder that merits adjustment. In Masarwa et al. (2020), adjustment for maternal smoking had a substantial impact on the measure of association in at least one study (unadjusted RR=1.44; bias corrected RR=1.05) (38).

Four studies used a NCE design to determine if there is substantial unmeasured confounding in studies of the association between prenatal APAP use and functional neurobehavioral outcomes (20, 21, 29, 32). Trønnes et al. (2020) and Chen et al. (2019) used pre-pregnancy APAP use (20, 32), Liew et al. (2019) used pre- and post-pregnancy APAP use (29), and Ystrom et al. (2017) used pre-pregnancy maternal and paternal APAP use (21) as NCEs. As discussed in the results section, Trønnes et al. (2020) found a significant association between pre-pregnancy APAP use and communication problems and activity levels (20); Ystrom et al. (2017) found an association between paternal APAP use and ADHD (21). These mixed findings suggest that there may be some unmeasured confounding in prenatal APAP studies.

The majority of studies included confounders as terms in adjusted regression models. Saunders et al. (2019) only examined the association between prenatal APAP and ASD using unadjusted, bivariate analysis (34). Only three studies used propensity scores and/or inverse probability treatment weight approaches (11, 20, 27). Overall, studies are mixed on whether adjustment for confounders and indications meaningfully changed study measures of association. One study conducted a bias analysis to address the impact of unmeasured confounding and found that residual confounding would only minimally impact study findings (26).

4.4 Statistical Approach

Studies primarily used adjusted linear and logistic regression to examine associations between APAP and outcomes. Three studies used Poisson or negative binomial regression (15, 23, 24) (20) and one used Cox proportional hazards models (21). Ernst et al. (2019) conducted regression analysis accounting for the data being interval-censored (26). Interval-censored data occurs when longitudinal data are collected at set intervals, despite the outcome possibly occurring in-between set intervals (50).

A methods paper by Stoltenberg et al. (2020) suggests that the analysis of prenatal APAP and outcomes like ADHD should be performed using cure models (51). Adverse neurobehavioral outcomes occur at birth; however, they are not observed until months or years after birth, which results in censoring of outcomes (51). Cure models address this challenging data structure. The paper demonstrated that logistic regression produces minimally biased estimates when the outcome is rare and occurs earlier in time (51). Cox models produce less biased estimates only when the outcome is rare (51).

There are important considerations in the application of cure models. Firstly, as the pointed out in Stoltenberg et al. (2020), to identify the fraction of susceptible individuals, semiparametric cure models assume that the distribution function of survival times in the susceptible individuals reaches unity before the distribution function of censoring times. This translates to assuming that in our data we have observed the largest diagnosis time for the susceptible individuals. It is important to consider if such an assumption is plausible. While parametric models do not need such an assumption, they make other unverifiable assumptions about the distribution of survival times. Another important consideration is the number of parameters to be estimated in the cure model. Cure models of the type discussed by Stoltenberg et al. (2020), have two component models. Namely, a model for the probability of being born susceptible and a model for the survival times of the susceptible individuals. The authors argue that in perinatal studies discussed in the paper, parameters related to the distribution of survival times are not of interest and should be treated as nuisance parameters. However, it is not clear that this approach would be acceptable for all scientific questions. The scientific question and clinical knowledge should determine if it is of interest to estimate the effect of an exposure or treatment on just one of or both the probability of being susceptible and the survival function of the susceptible fraction. Finally, as the authors point out, by using the cure model one assumes that there are no false positives, i.e., that the non-susceptible individuals are never diagnosed with having the condition (51). While there may be methods to try and quantify the bias stemming from such misclassification, it does not seem straightforward to do so.[39]

Within the reviewed studies, multiple analyses were often conducted operationalizing APAP and outcomes multiple different ways or considering many sub-scale scores. However, only one study corrected the p-value to account for multiple testing/comparisons in their adjusted analyses (23). For safety studies where signal detection is important, it may not be necessary to control for multiplicity. However, in studies attempting to evaluate causal associations, not accounting for multiple comparisons increases the risk of a type I error (incorrectly rejecting the null) due to chance alone. Only one study presented a power analysis calculation (30).

4.5 Selection Bias

Several studies noted differences between the characteristics of participants and non-participants (12, 13, 17, 20, 22, 25, 27, 33). For example, one study noted that non-participants typically were younger (12), were less educated (12), and had a lower income (12). One study found that

[39] Paragraph critique of cure models provided by Sai Dharmarajan of the Division of Biometrics VII (DB7). Additional review provided by Yueqin Zhao (team lead) from DB7.

non-participants were less likely to use APAP during pregnancy (12); however, several studies documented no difference in APAP use between participants and non-participants (12, 13, 17, 26). Loss to follow-up or missing data may also lead to selection bias. One study lost 15-20% of the sample, depending on the outcome being analyzed, due to missing data (22). The impact of non-participation, loss to follow-up, and missing data on measures of association depends on whether these factors differed by APAP exposure and outcome status.

The studies reviewed used multiple approaches to address and explore possible bias due to non-participation, missing data, and loss to follow-up. Approaches included multiple imputation (21) and inverse probability weighting (IPW) for missing data (12). Studies used IPW or IPTW to address missing outcome information (12, 13), stabilized IPW to account for study attrition, inverse probability of censoring weights to counter loss to follow-up (20), and inverse probability of selection weighting (24) or IPW (25) to factor in non-participation. Results are often similar when sensitivity analyses were conducted both with and without these weighting approaches (12, 13). The bias analysis by Masarwa et al. (2020) also demonstrated that selection bias due to greater or lower study participation among those with APAP exposure had minimal impacts on individual study findings on the association between APAP and ADHD (38). However, most studies did not incorporate imputation or weighting into their analyses and the impact of selection bias on study findings is unknown.

4.6 Generalizability

Several studies used population-based/national cohorts, which improves generalizability. The BBC is noted for being comprised of a mostly urban, low income, minority, U.S. population (27). Only three studies (12, 14, 15) included populations from outside of Europe and the U.S. One study included both a U.S. and Brazilian cohort (12). The two cohorts differed in terms of age, education level, income, smoking status, alcohol consumption, antibiotic use, and frequency of depression/anxiety (12). As noted previously, a review article found that the frequency of APAP use during pregnancy differs by geography (4). However, it is unclear if associations between APAP and outcomes would differ by sample characteristics or geography.

4.7 Publication Bias

Publication bias is a possibility in this field of research. However, we did not assess whether the association between prenatal APAP and neurobehavioral or urological outcomes have been assessed in other well-characterized, large, pregnancy cohorts.

4.8 Duration- or Dose-Response

Duration- or dose-response relationships between APAP and respective outcomes were explored in many of the reviewed studies (11-13, 16, 21, 24-28, 30-33, 52). Several found evidence of these types of relationships (11, 13, 16, 21, 24, 25, 27, 28, 52), but this finding was not universal (12, 25, 26, 30-33). The meta-analysis by Masarwa et al. (2018) found that duration of exposure

was an effect modifier with a stronger association noted for longer duration of exposure (interaction β coefficient per day of exposure=0.00; 95% CI: 0.00, 0.01) (36). The meta-analysis by Gou et al. (2019) also found duration-response relationships (37). However, the studies reviewed were typically unable to capture the number of APAP pills or the dosage of pills taken by pregnant women. Consequently, studies were unable to examine dose-response relationships in detail and often needed to extrapolate based on weeks or trimesters of use. Studies also captured and categorized duration or dose of use differently.

4.9 Biological Plausibility

The reviewed studies did not consistently find that APAP exposure during certain periods of pregnancy (e.g., first trimester exposure) are more important than others. Some studies found that certain trimesters appeared to be critical window periods; however, these critical window periods were often not replicated across studies. For example, findings in Chen et al. (2019) suggest that first and second trimester exposure to APAP may be more important for neurologic outcomes (32). The meta-analysis by Gou et al. found statistically significant associations[40] restricted to the first (RR=1.21; 95% CI: 1.01, 1.45) and third trimester (RR=1.26; 95% CI: 1.08, 1.47) (37). In general, across studies, APAP exposure in multiple trimesters was often more important to the development of negative health outcomes than exposure in any one trimester. Studies also suggest that postnatal APAP exposure may have no impact (24) or an additive impact (with prenatal exposure) (13) on behavioral difficulties. Postnatal exposure was found to have no impact on cognitive outcomes (12). Alemany et al. (2021) found no associations between postnatal APAP and ADHD or ASD in a meta-analysis (35).

It is plausible that the effect of APAP exposure on neurobehavioral outcomes may differ by sex. Sex was the most common effect measure modifier examined in the reviewed studies (12-16, 22, 24-29). Most studies found no difference (12, 13, 15, 22, 25, 28, 29) but others found associations in girls (but not boys) for neurobehavioral outcomes (12, 16, 27). Meanwhile, others found associations for neurobehavioral outcomes in boys but not girls (14) or stronger associations in boys (24). The meta-analysis by Alemany et al. found a slightly stronger association between APAP and ASD for boys (35). In the study of pubertal development, associations were noted in girls, but not boys (26).

In the study by Baker et al. (2020), magnetic resonance imaging (MRI) was conducted on a subset of individuals between the ages of 9 and 11 years (11). There were changes in brain connectivity noted on the MRI that were unique to those with ADHD and long-term APAP exposure (11). These findings suggest a mechanistic connection between prenatal APAP and ADHD (11). No other studies reviewed included a mediation analysis or other mechanistic component.

4.10 Studies Considered of Higher Quality

[40] Unclear if the pooled RRs were adjusted.

The *post hoc* list of study criteria that identified higher quality studies was developed considering the study features and limitations discussed in Sections 4.1 through 4.9. The criteria are listed in Appendix A. The comments column of Appendix B identifies which criteria were not met by each respective study.

4.10.1 Neurobehavioral Outcomes

Inoue et al. (2021) met all the study criteria in Appendix A (24). The study included 40,934 mother-child pairs from the Danish National Birth Cohort and assessed the association between prenatal APAP and behavioral problems. Pregnant women were initially asked about medication and supplement use in the four weeks prior to pregnancy through week of enrollment. At follow-up telephone interviews (gestational week 12, gestational week 30, 6-months postpartum), pregnant women were asked about weekly use of analgesics, specifically 44 medications that included APAP and APAP combination medications. Although some sub-scale scores do not have validated cut-offs, the study used the validated cut-off for SDQ total scores. Outcomes were reported by both parents and children when the child was eleven years of age. The study adjusted for multiple indications including self-reported muscle/joint disease, fever, and inflammation/infections during pregnancy. The study adjusted for other medications but only NSAIDs. The study adjusted for psychiatric illness before or during pregnancy, which could be a proxy for the use of other medications. The study adjusted for parent neurobehavioral problems in childhood (proxy for familial/genetic risk), maternal socio-economic status, maternal smoking, and maternal alcohol use. As discussed previously, the study found a weak, statistically significant association between prenatal APAP and total behavioral difficulties reported by parents (aRR=1.14; 95% CI: 1.01, 1.29) and children (aRR=1.40; 95% CI: 1.20, 1.63). When examining the highest durations of exposure, compared to those with no exposure, three trimesters of exposure was significantly associated with child-reported behavioral problems (aRR=1.53; 95% CI: 1.19, 1.96) but not parent-reported behavioral problems (aRR=1.20; 95% CI; 0.98, 1.96). Meanwhile, >10 weeks of exposure was significantly associated with both parent- (aRR=1.32; 95% CI: 1.06, 1.63) and child-reported (aRR=1.58; 95% CI; 1.22, 2.06) behavioral problems. Findings were robust to sensitivity analyses that assessed combined parent/child scores, adjusted for parent behavioral problems, treated SDQ as a continuous score, adjusted SDQ cut-offs, and used IPSW. Findings sometimes differed for sub-domain scores or composite scores, compared to total SDQ scores. Unfortunately, the study did not adjust for headache or migraine during pregnancy and did not adjust for additional medications that may have neurodevelopmental impacts. Lastly, although the study had parental behavioral problem information for a subset of the study sample, information was not available on both parents (24).

Three other studies met the majority of the aforementioned criteria with some more minor limitations (20, 21, 25). As a whole, these three studies suggest limited, weak associations between APAP and neurobehavioral outcomes. Liew et al. (2016) met all criteria except it was unclear if the dichotomized outcomes were based on clinical cut-offs (25). Additionally, adjustment for maternal intelligence quotient and parental education would only consider familial risk factors for executive function outcomes (not attention outcomes), although the study did adjust for maternal mental health. Focusing on executive function outcomes, the study only found a weak association between APAP and metacognition but not other measures of executive

function. Trønnes et al. (2019) conducted many sensitivity and bias analyses; however, it did not account for genetic/familial factors outside of maternal mental health (20). Only two of three outcomes were dichotomized at clinical thresholds. The study only found a statistically significant, weak association between APAP use in three trimesters and internalizing problems. There were no significant associations with other communication problems, behavioral problems, or temperament. The NCE analysis demonstrated an association between communication problems and lower activity levels. Ystrom et al. (2017) did not adjust for concomitant medications and ADHD diagnoses, based on ICD codes, were not assessed at a set year; however, some subjects had follow-up through 15 years of age (21). Ystrom et al. (2017) also conducted two NCE analyses, examined APAP use stratified by specific indications, and adjusted for parental symptoms of ADHD. The study examined ADHD diagnoses after age three. In this study, there was a weak, significant increased risk of ADHD among those with any prenatal exposure. This study was able to assess APAP use for specific indications; when days of exposure was examined stratified by indication, moderate to strong associations were noted for high durations of APAP exposure for all indications, fever and infections, and pain conditions. The study also found possible evidence of residual confounding (significant association with paternal APAP use but not maternal APAP preconception use) (21).

4.10.2 Urogenital Outcomes

Ernst et al. (2019) (26) assessed exposure at multiple time points; conducted follow-up at set ages; accounted for confounding by indication, socio-economic status, alcohol consumption, and smoking; and conducted several sensitivity analyses. The study did not adjust for other medications and the outcome (age of pubertal milestones) has unclear clinical implications. However, the major concern with the study is that it conducted an extensive number of analyses with the exposure and outcome operationalized in many different ways without correction for multiple testing. Despite the strengths of the study and multiple testing limitation, the study found almost no associations between APAP and age of pubertal milestones.

Navarro-Lafuente et al. (2021) (30) did examine the outcome (AGD) after birth using two examiners and performed a power calculation. However, the study was small (n=277), assessed prenatal APAP exposure at one point in time, only adjusted for two covariates in final models, did not adjust for other medications, and did not adjust for indications. The study only conducted one supplemental analysis that accounted for birth weight and gestational age and non-transformed AGD scores. This supplemental analysis only found an association between first trimester APAP and AGD (anus-penis) ($a\beta$=1.23; 95% CI: 0.06, 2.40).

Overall, both of these studies show no consistent effect of APAP on urogenital outcomes—AGD and pubertal milestones. However, outcomes in both studies are assessed as continuous outcomes with unclear clinical implications.

4.11 Comparison to Prior Reviews

Findings in this review are consistent with prior DEPI-I reviews of the association between APAP and neurobehavioral outcomes. Similar to the prior DEPI-I reviews, although not ubiquitous and limitations are still present, most studies find at least some associations between prenatal APAP and neurobehavioral outcomes, although associations are not particularly strong and are not adjusted for testing of multiple associations. For neurobehavioral outcomes, several studies more rigorously addressed confounding through sensitivity analyses, including NCE analyses. Studies in this review also incorporated direct measures of APAP, including measures from blood, urine, and meconium, that was not seen in previous DEPI-I reviews.

In the current review, there was only two studies of urogenital outcomes—AGD and pubertal milestones. Unlike what was seen in a prior DEPI-I review, there was no association found between prenatal APAP and AGD. No prior study reviewed by DEPI-I examined pubertal milestones. However, the study in the current review found very limited associations between prenatal APAP and pubertal milestones. As a whole, the association between prenatal APAP and urogenital outcomes may be more outcome specific and less persuasive than findings for neurobehavioral outcomes.

As a whole, it is still unclear, despite any associations noted, whether the totality of the evidence suggests that the association between prenatal APAP and neurobehavioral and urogenital outcomes is causal. It is unlikely that further observational studies will provide more clarity without more mechanistic data.

5 CONCLUSIONS

In general, the functional neurobehavioral outcome studies examined in this review along with the reviewed meta-analyses suggest a consistent association between APAP or long durations of prenatal APAP exposure and ADHD. However, findings for trimester-specific associations are not consistent. Findings for other diagnoses were mixed and sparse. Studies using questionnaires or scales to examine functional neurobehavioral constructs have more inconsistent findings across studies and within studies. However, the most methodologically rigorous study (Inoue et al. [2021]) found a weak association between APAP use overall and long durations of APAP use (>10 weeks of cumulative use) and behavioral problems (assessed with a scale) at 11 years of age (24). There were only two studies that examined urogenital outcomes and found no consistent impact of APAP on urogenital outcomes. Furthermore, studies reviewed in this review and prior DEPI-I reviews, suggest that prenatal APAP use may attenuate the impact of fever on childhood outcomes.

Studies are still limited by crude operationalizations of APAP exposure, unclear clinical meaning of findings, and the possibility of unmeasured confounding by factors such as indication, other medications, and genetic factors. Overall, as noted by Ji et al. (2016) (27), an ideal study would repeatedly assess APAP blood levels in a cohort of pregnant women. However, studies are limited to data collection approaches of established, pregnancy cohorts. Although the use of direct measures of APAP levels is a novel contribution to the literature in this area, the measurement methods in the studies reviewed do not reflect APAP exposure throughout pregnancy and may not be valid.

6 RECOMMENDATIONS

Overall, there are still study limitations and inconsistent study findings that prohibit causal interpretations of the association between APAP exposure and functional neurobehavioral outcomes as well as urogenital outcomes. Studies, including meta-analyses, consistently note associations between APAP and ADHD; however, associations are inconsistent by trimester of exposure. The most methodologically rigorous study found a weak association between overall APAP and long durations of APAP use and behavioral problems. The prior FDA communication focused on medication use for pain during pregnancy and did not discuss amount of use.[41] As a result, it may be prudent, as a precautionary measure, to (b) (5) [redacted]. Untreated fevers during pregnancy are associated with poor pregnancy outcomes (b) (5) [redacted]

(b) (5) [redacted]

[41] FDA Drug Safety Communication: FDA has reviewed possible risks of pain medicine use during pregnancy. January 9, 2015. Silver Spring (MD), U.S. Food and Drug Administration. Accessed January 6, 2022 at: https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-has-reviewed-possible-risks-pain-medicine-use-during-pregnancy.