Sarah Hjorth

Childhood outcomes after prenatal exposure to analgesics or antibiotics

Thesis submitted for the degree of Philosophiae Doctor

Department of Pharmacy
Faculty of Mathematics and Natural Sciences

PharmacoEpidemiology and Drug Safety Research Group
PharmaTox Strategic Research Initiative

2021
To my brother
Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor at the University of Oslo. The research presented here was conducted under the supervision of Professor Hedvig Nordeng, Dr Angela Lupattelli, and Professor Olav Spigset.

The thesis is a collection of five papers, presented in chronological order of writing. The papers focus on childhood outcomes after prenatal exposure to analgesics or antibiotics. All papers represent joint work with the main supervisor and other collaborators. The thesis synopsis, preceding the papers, provides background information and motivation for the research, demonstrates how the papers are related, and discusses implications of the work.

Acknowledgements

First, I would like to thank my supervisors Professor Hedvig Nordeng, Dr Angela Lupattelli, and Professor Olav Spigset for introducing me to the field of pharmacoepidemiology. It has been a privilege to be in such expert hands. Hedvig, thank you for entrusting me with this project in spite of my different background, for always believing in me, for cheering me on when I doubted myself, and for introducing me to your vast network of collaborators. If this acknowledgement section is uncommonly long, it is in part thanks to you. Angela, thank you for your everlasting patience, and day-to-day help and support. Olav, thank you for your attention to detail, and for not running away when you noticed that I had so little knowledge of the field that I wrote ACT code rather than ATC code throughout the first draft of my PhD protocol.

Thanks to my collaborators on specific projects. To Dr Rebecca Bromley for having me for a research stay at the University of Manchester, for introducing me to neurodevelopmental testing, and for making me feel so very welcome. To Professor Eivind Ystrøm for encouraging my interest in psychometrics. To Dr Mollie Wood for taking the time to discuss propensity scores, negative exposure controls etc. with me across the Atlantic. To Fatima Tauqeer for your optimism, and for our tea-talks about things professional and political. To Dr Marte Handal for valuable suggestions and encouragement. To the entire NorMedPregCCC for bearing with my limited project management skills, for sharing knowledge, and making a Nordic project possible. In particular, thanks to Professor Anton Pottegård for having me for a research stay at the University of Southern Denmark, and ensuring that I spent the time on career planning, as well as research. To Dr Marie Hargreave, Dr Maarit Leinonen, and Dr Ulrika Nörby for believing in the Nordic project from the beginning, and for your invaluable help through the legal labyrinth of ethical approvals, data protection considerations,
and applications for data. To Caroline H. Hemmingsen for spending way too much of your time running my scripts on Danish data before I got data access. I owe gratitude to the European Research Council, the Nordic Cancer Union, the PharmaTox Strategic Research Initiative, and the Norwegian PhD School of Pharmacy (NFIF) for financial support. I would also like to acknowledge NFIF and the National research school in population based epidemiology (EPINOR) for relevant courses.

And now for the acknowledgements of more personal character:

First, I want to thank the inspirational midwives and researchers who saw my research interest and encouraged me to pursue it: Dr Sara Kindberg, Associate Professor Dorte Hvidtjørn, and Professor Ellen Nøhr. To paraphrase Dorte: I am grateful to stand on your shoulders.

Thanks to Associate Professor Mette Bliddal for our regular talks.

Thanks to all my wonderful fellow PhD students for an atmosphere where it is legitimate to ask for advice and support. An extra thanks to Bich for the ‘almost done’-meetings during the write-up of our theses. I am so happy to have had your support on the final stretch.

Last, but definitely not least, thanks to my family and friends. You give me the strength to do more than I thought I was capable of. I will not write any more, because I know you find acknowledgements nauseatingly sentimental, and besides, I do not have adequate words.

Sarah Hjorth
Oslo, August 2021
Abstract

Background: The most frequently used medications in pregnancy are analgesics (between 51 and 70% of pregnancies), and antibiotics (between 27 and 42% of pregnancies). Little is known about the long-term child health after prenatal exposure to these medications. Especially for analgesics and child brain development, and antibiotics and childhood cancers, previous studies show conflicting results.

One reason for the conflicting results could be the challenges encountered when investigating childhood outcomes after prenatal medication exposure. The challenges relate to a lack of comparability between exposed and unexposed children, and to data validity.

Aim: This thesis aimed to assess the association between prenatal exposure to analgesics and child brain development, and the association between antibiotics and childhood cancers, including how to improve the validity when estimating these associations. Specifically, the aims were I: to describe the current literature, identify knowledge gaps and provide suggestions on how to improve the validity in future studies; II: to understand the patterns of analgesic and antibiotic use before, during, and after pregnancy, III: to investigate the association between prenatal exposure to non-steroidal anti-inflammatory drugs (NSAIDs) and child Attention-Deficit/Hyperactivity Disorder (ADHD); and IV: to investigate the association between prenatal exposure to nitrofurantoin and childhood leukaemia.

Methods: To address aim I, two systematic reviews were conducted (papers i and iv).

The remaining aims were addressed in cohort studies. Aim II was addressed by linking the Norwegian Prescription Database (NorPD) to the Medical Birth Registry of Norway (MBRN) (paper ii).

For aim III, data on self-reported NSAID use during pregnancy from the Norwegian Mother, Father and Child Cohort was linked to ADHD diagnoses from the Norwegian Patient Registry (paper iii). Marginal structural Cox models were applied to obtain hazard ratios (HR) with 95% confidence intervals (CI) while accounting for time-varying exposure and confounding.

Aim IV was addressed using data on prescription fills for nitrofurantoin from registries in Denmark, Finland, Norway, and Sweden, linked to data on cancer diagnoses from the national cancer registries. Comparators were children prenatally exposed to pivmecillinam. Poisson regression was used to obtain incidence rate ratios (IRR) with 95% CI.

Results: The systematic review on prenatal exposure to analgesics and child brain development (paper i) found that child brain development was measured in 31 different ways across 29 included studies. Long-term prenatal exposure to paracetamol was associated with child ADHD. Few studies had investigated
brain development after prenatal exposure to NSAIDs; none had investigated child ADHD.

Regarding antibiotics and childhood cancer, 24 of 36 studies in the systematic review (paper iv) relied on maternal recall of the exposure after the child was diagnosed with cancer. Prenatal exposure to antibiotics was not associated with other types of childhood cancers, but the newest studies identified an association with childhood leukaemia. All identified studies used unexposed comparators.

Paper ii included 172,585 pregnancies. Analgesic prescriptions were filled in 6.5% of pregnancies with the highest proportion in the first trimester. Antibiotic prescriptions were filled in 28.3% of pregnancies. The proportions did not differ much by trimester. The most commonly filled medications reflected current guidelines.

Among 56,340 children included in paper iii, we did not identify any association between prenatal exposure to NSAIDs and child ADHD, regardless of timing of exposure (first trimester: HR 1.12, 95% CI 0.86 to 1.45; second trimester: HR 0.98, 95% CI 0.69 to 1.38; third trimester: HR 0.68, 95% CI 0.31 to 1.46). We found no evidence of a dose-response relationship.

In paper v, we included 44,009 children prenatally exposed to nitrofurantoin and 247,306 children prenatally exposed to pivmecillinam. Any prenatal exposure to nitrofurantoin was associated with an IRR of 1.34 (95% CI 0.88 to 2.06). There was no evidence of a dose-response relationship.

Conclusions: This thesis did not find evidence of associations between short-term prenatal exposure to analgesics, and child adverse brain development. The systematic review showed an association between long-term use and child ADHD. However, it is unclear whether the association is causal. For NSAIDs, the cohort study did not identify a dose-response relationship for ADHD, but this needs to be confirmed in future studies.

For prenatal exposure to antibiotics, an association with childhood leukaemia cannot be ruled out based on data from the systematic review. Studies accounting for maternal infection are needed to guide causal inference. In the cohort study, prenatal nitrofurantoin exposure did not appear to be substantially associated with childhood leukaemia when compared to pivmecillinam, although this finding should be interpreted with caution due to limited statistical precision. Furthermore, the finding needs to be replicated.

Overall, the findings from this thesis support the appropriateness of existing guidelines on use of analgesics and antibiotics in pregnancy.

The thesis also formulated some considerations on how to improve the validity in future observational studies of childhood outcomes after prenatal exposure to medications. The considerations relate to the choice of outcome classification, use of evidence from in vitro and animal studies, and use of new methods in confounding control. The thesis provides examples of the use of some of these methods.
Contents

Abstract v
List of Papers ix
List of Figures xi
List of Tables xiii
Abbreviations xv

1 Introduction 1
  1.1 Medication use during pregnancy 2
  1.2 Challenges when investigating long-term medication safety in pregnancy 11

2 Thesis aims 19

3 Materials and methods 21
  3.1 Study design 21
  3.2 Data sources 21
  3.3 Study samples 24
  3.4 Measures 27
  3.5 Statistical analyses 30

4 Main findings 33
  4.1 Specific aim 1: Analgesics and child brain development 33
  4.2 Specific aim 2: Antibiotics and childhood cancer 37

5 Discussion 41
  5.1 Summary of main findings 41
  5.2 Strengths and limitations 41
  5.3 Interpretation 49
  5.4 Implications for practice 56
  5.5 Implications for research 57

6 Conclusion 59

7 Perspectives 61

Papers 64
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Use and validity of child neurodevelopment outcome measures in studies on prenatal exposure to psychotropic and analgesic medications – A systematic review</td>
<td>65</td>
</tr>
<tr>
<td>II</td>
<td>Fertility treatment and oral contraceptive discontinuation for identification of pregnancy planning in routinely collected health data – an application to analgesic and antibiotic utilisation</td>
<td>105</td>
</tr>
<tr>
<td>III</td>
<td>Prenatal Exposure to Non-Steroidal Anti-Inflammatory Drugs and Risk of Attention-Deficit/Hyperactivity Disorder – a Follow-Up Study in the Norwegian Mother, Father and Child Cohort</td>
<td>119</td>
</tr>
<tr>
<td>IV</td>
<td>Maternal Medication Use and Childhood Cancer in Offspring – Systematic Review and Considerations for Researchers</td>
<td>133</td>
</tr>
<tr>
<td>V</td>
<td>Prenatal exposure to nitrofurantoin and risk of childhood leukaemia – a registry-based cohort study in four Nordic countries</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>Appendices</td>
<td>205</td>
</tr>
<tr>
<td>A</td>
<td>Appendix to Chapter 1</td>
<td>207</td>
</tr>
<tr>
<td>B</td>
<td>Appendix to Chapter 4</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>239</td>
</tr>
</tbody>
</table>
List of Papers

Paper I

Paper II

Paper III

Paper IV

Paper V
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Vulnerable periods for development of various organ systems in the foetus.</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Graphical summary of the evidence regarding prenatal exposure to analgesics and child outcomes.</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>Graphical summary of the evidence regarding prenatal exposure to antibiotics and child outcomes.</td>
<td>10</td>
</tr>
<tr>
<td>1.4</td>
<td>Illustration of the research field on medication safety in pregnancy and this thesis’ place in the research field.</td>
<td>11</td>
</tr>
<tr>
<td>1.5</td>
<td>Terminology regarding the validity of dichotomous exposures and outcomes.</td>
<td>14</td>
</tr>
<tr>
<td>1.6</td>
<td>Illustration of the terminology used in directed acyclic graphs.</td>
<td>15</td>
</tr>
<tr>
<td>1.7</td>
<td>Illustration of this thesis’ place in the research field of medication safety in pregnancy, including the limitations associated with using existing data.</td>
<td>16</td>
</tr>
<tr>
<td>3.1</td>
<td>Study design diagram for the three cohort studies.</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>Flowcharts for the two systematic reviews</td>
<td>25</td>
</tr>
<tr>
<td>3.3</td>
<td>Flowcharts for the three cohort studies</td>
<td>26</td>
</tr>
<tr>
<td>3.4</td>
<td>Illustration of the exposure definition in the three cohort studies</td>
<td>28</td>
</tr>
<tr>
<td>4.1</td>
<td>Illustration of how the five papers in the thesis are connected.</td>
<td>33</td>
</tr>
</tbody>
</table>
## List of Tables

3.1 Overview of search strategies in the systematic reviews. 23
A.1 Overview of methods to account for confounding 208
A.2 Criteria for causal inference according to Bradford Hill, the target trial approach, and the pragmatic pluralism approach. 213
B.1 Most commonly filled analgesic and antibiotic substances in peripregnancy 216
B.2 Studies on the association between prenatal exposure to antibiotics and childhood cancer 217
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD</td>
<td>Attention-Deficit/Hyperactivity Disorder</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CPRS-R (S)</td>
<td>Conners’ Parent Rating Scale-Revised, Short Form</td>
</tr>
<tr>
<td>HR</td>
<td>Hazard ratio</td>
</tr>
<tr>
<td>ICCC</td>
<td>International Classification of Childhood Cancer</td>
</tr>
<tr>
<td>IPTW</td>
<td>Inverse probability of treatment weights</td>
</tr>
<tr>
<td>IRD</td>
<td>Incidence rate difference</td>
</tr>
<tr>
<td>IRR</td>
<td>Incidence rate ratio</td>
</tr>
<tr>
<td>MBRN</td>
<td>Medical Birth Registry of Norway</td>
</tr>
<tr>
<td>MoBa</td>
<td>Norwegian Mother, Father and Child Cohort</td>
</tr>
<tr>
<td>NorPD</td>
<td>Norwegian Prescription Database</td>
</tr>
<tr>
<td>NPR</td>
<td>Norwegian Patient Registry</td>
</tr>
<tr>
<td>NSAIDs</td>
<td>Non-steroidal anti-inflammatory drugs</td>
</tr>
<tr>
<td>USA</td>
<td>United States of America</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

‘Can I take this medication now that I am pregnant?’ is a question that many health care professionals have met during consultations with pregnant women. I too have frequently met the question in my practice as a midwife. The implied question was always ‘Is this medication safe for my baby?’

Between 60 and 96% of women use medications during pregnancy [1–5]. The most frequently used medications are analgesics (between 51 and 70% of pregnancies), and antibiotics (between 27 and 42% of pregnancies) [1–5].

With the majority of pregnant women reporting medication use during pregnancy, the need for knowledge about safety of medications in pregnancy is obvious. However, according to an American study, 89% of the medications approved since 1980 have no pregnancy safety data [6]. Safety data in pregnancy usually refer to information on outcomes that can be assessed during, or shortly after pregnancy, such as spontaneous abortions, or congenital malformations [7, 8]. It is mainly within the last decade that the associations between prenatal medication exposure and long-term child health have gained awareness [7]. Therefore, the proportion of medications with no long-term safety data may be even higher, and may extend to older medications as well.

One reason for the lack of long-term safety data could be the challenges related to investigating medication safety in pregnancy. Pregnant women are rarely included in randomised controlled trials [8, 9]. Furthermore, the follow-up required to investigate long-term safety is rarely feasible in the context of randomised controlled trials [10]. Therefore, studies on long-term medication safety in pregnancy rely on observational data [9]. In observational data, a major challenge is the lack of comparability between individuals who use the medication of interest and individuals who do not [8]. Moreover, in order to obtain a sufficient duration of follow-up, studies on long-term medication safety in pregnancy often need to use data that were collected years ago for a different purpose [11]. This poses the additional challenge of how to identify truly exposed individuals [11, 12], and accurately capture the outcomes of interest [8, 13].

The scope of this thesis is outcomes in childhood after prenatal exposure to analgesics and antibiotics, and methods that can be used to investigate long-term pregnancy safety when repurposing existing observational data.
1. Introduction

1.1 Medication use during pregnancy

1.1.1 Prevalence of medication use during pregnancy

Pregnancy is a period of many physiological changes in the maternal body [14]. Existing illnesses can improve or exacerbate, new acute or chronic illnesses can appear [14]. Medication use in pregnancy is common [1–5, 15–18]. In studies with data collection completed after 2010, the prevalence of use ranged from 60% of pregnancies [3] to 96% [5]. The lowest estimate, 60%, was reported in a study on prescription medications, where medications available over-the-counter were not captured [3]. However, in countries where prenatal vitamins and minerals are on prescription, up to 90% of pregnant women use prescription medications [1]. Use of over-the-counter medication ranges from 67% [4] to 93% [5].

In studies that include information on over-the-counter medications, the most commonly used medications are non-opioid analgesics (paracetamol and non-steroidal anti-inflammatory drugs [NSAIDs]) [4, 5, 16], used in 51% [4] to 68% of pregnancies [5]. In studies focusing on prescription medications, the most commonly used medications vary from analgesics (70% in a French cohort) [1], over anti-emetics (34% in the United States of America [USA]) [17] to antibiotics (51% in an Italian cohort, and 28% in a population-based study from Norway) [3, 18]. According to a review of observational studies, anti-infectives are widely prescribed in pregnancy regardless of country (27% to 42%), whereas the prevalence of use for other medication groups vary five- to ten-fold across countries [2]. Hence, analgesics and antibiotics are consistently widely used in pregnancy. Given their common use, any potential adverse long-term effects in children from prenatal exposure to these medications could have a large impact on a population level. Therefore, analgesics and antibiotics were chosen as focus for this thesis.

1.1.2 Risks associated with underlying maternal illness

Between 26% and 33% of pregnant women report that they have avoided the use of certain medications in pregnancy, including medications prescribed by their doctor [16, 19]. Reasons for avoidance of certain medications include fear of foetal harm, or a preference to cope with illness rather than exposing the foetus to medications [19].

While the pregnant woman provides the foetus with oxygen and nutrients, most medication exposures are also shared with the foetus [20]. Furthermore, maternal illnesses can affect the foetus [21, 22]. The timing of any potentially harmful exposures is important (Figure 1.1). The structures in the body are formed in early pregnancy, before the end of the first trimester, so only first trimester exposures have the potential to cause major congenital malformations [23]. However, some organ systems, such as the brain, are vulnerable throughout pregnancy and late exposures can cause minor malformations, or functional defects [23, 24].

2
This illustrates the dilemma that women are faced with, when it comes to medications in pregnancy. On the one hand, there may be a risk to maternal and foetal health from untreated illness; on the other hand, there may be risks, mainly to the foetus, from the use of medications. In this section, I will summarise the evidence concerning risks of underlying maternal illnesses for which analgesic or antibiotics treatment is indicated. Whereas antibiotics are used to treat bacterial infections, the indications for analgesic use are less uniform. Analgesics are indicated for pain-relief, but some also have anti-inflammatory, or anti-pyretic properties.

### 1.1.2.1 Bacterial infections

The severest consequence of infection in pregnancy is maternal sepsis resulting in death. Of all maternal deaths in high-income countries, 8% to 13% are attributable to infections [25–27]. Whereas some of these infections are viral, approximately half of the infection-related deaths are attributable to bacterial infections of the genito-urinary tract [26]. However, the absolute number of maternal deaths from sepsis is very low: 0.54 deaths in 100,000 maternities [26].

Intrauterine death of the foetus has also been described after infection in pregnancy [28, 29]. A systematic review including data on 756 stillbirths from high-income countries found that 40% of stillbirths were of infectious origin, and 1% attributable to group B streptococcus infection [29]. Again, the absolute risks are small, as the stillbirth rate ranges from 3 to 5 per 1000 live-births in high-income countries [30, 31]. A more common complication that is often attributed to infection is preterm birth [28, 32]. Preterm birth complicates 6 to 10% of pregnancies [31, 33]. According to a systematic review, antibiotic treatment of asymptomatic bacteriuria in pregnancy reduces the risk of preterm
birth (RR 0.34, 95% CI 0.13;0.88) [32].

As for long-term outcomes in the child, studies have shown associations between prenatal exposure to infection and childhood asthma, eczema [34], and autism spectrum disorder [35]. Maternal infection in pregnancy has also been suggested as a potential risk factor for childhood cancers, in particular leukaemia [36], but studies on bacterial infections are few and provide heterogeneous results [37]. Given the severity of the immediate outcomes associated with maternal infection, there is a consensus that all bacterial infections in pregnancy should be treated [32, 38].

1.1.2.2 Fever

It can be difficult to distinguish between the effects of the underlying infection and any effects of fever; however some attempts have been made [21, 39–42]. As opposed to infection, fever alone does not seem to be associated with a higher incidence of foetal death, or preterm delivery [21]. Some [21], but not all [41], studies indicate that fever increases the prevalence of specific congenital malformations, with odds ratios at or above 2 for neural tube defects, and oral clefts [21]. Most studies found that use of anti-pyretic medications reduced or eliminated the excess prevalence of malformations [21].

Long-term child outcomes that have been associated with prenatal exposure to maternal fever in more than one study include cerebral palsy [21], asthma [21], autism spectrum disorder [21, 40], and Attention-Deficit/Hyperactivity Disorder (ADHD) [39, 42]. Particularly first trimester exposure to fever seems implicated in brain development disorders [39, 40, 42]. Whereas anti-pyretic medication use appeared to mitigate the risk for autism spectrum disorders [21, 40], the risk of ADHD did not appear to decrease with anti-pyretic use [39, 42].

Though no clinical guidelines exist regarding fever in pregnancy, many pregnant women are encouraged by their health care providers to treat fever with anti-pyretic medication [43, 44].

1.1.2.3 Pain and inflammation

Pain conditions are among the predictors of low health-related quality of life during pregnancy [45]. Women with pain conditions, such as pelvic girdle pain, report that the pain negatively affects their sleep, and overall mood [46]. Some women attribute a post-partum depression to their pain condition [46]. Pain in itself does not appear to confer any risk to the foetus, but may be a symptom of an underlying illness that is associated with negative pregnancy outcomes [47, 48]. Examples include migraine that is associated with preterm birth [48], and rheumatoid arthritis that is associated with intrauterine growth restriction, and preterm prelabour rupture of the membranes among women with poorly controlled disease [47]. However, such underlying conditions make up a minority of all pain in pregnancy [49]. A study from the USA found that the most common reason for analgesic use in pregnancy is headache (reported in half the instances of analgesic use), whereas migraine was reported in 2.9% of instances,
Medication use during pregnancy

and arthritis in 0.2% of instances [49]. Unspecified pain was reported in 19.1% of instances of analgesic use [49].

Maternal inflammation seems to be implicated in several disorders of placental origin (e.g. pre-eclampsia, intrauterine growth restriction) [22]. Some inflammatory diseases improve during pregnancy (e.g. rheumatoid arthritis), whereas others follow a less predictable cause (e.g. systemic lupus erythematosus, where exacerbations increase the risk of spontaneous abortions, preterm labour, intrauterine growth restriction, pre-eclampsia, and foetal death) [22]. Maternal inflammatory diseases in pregnancy have also been implicated in child long-term health [50]. To date, limited data are available, but children born to women with inflammatory conditions during pregnancy appear to have a higher incidence of autism spectrum disorder [50], ADHD [51], lymphomas [50], and autoimmune diseases, such as arthritis, diabetes, and asthma [50].

1.1.3 Medication safety concerns

In this section, I will introduce the second perspective to women’s risk-benefit evaluation of medication use in pregnancy, namely the concerns about medication safety.

1.1.3.1 Brief historical perspectives

The focus on medication safety in pregnancy started with thalidomide [23, 52, 53]. Thalidomide was used as a treatment for nausea and vomiting of pregnancy [23, 53]. In 1961, a specific pattern of congenital malformations was reported in children prenatally exposed to thalidomide [52, 53]. The discovery opened a new research field in prenatal exposure to medications and negative immediate birth outcomes such as congenital malformations, stillbirth, spontaneous abortions, preterm birth, and birth weight [23]. To date, immediate birth outcomes remain the most frequently investigated potential adverse outcomes from prenatal medication exposures [7, 8].

A change in the field came about in 1971, when it was discovered that prenatal medication exposure could also have health consequences for the child that were not discovered until years after birth [54]. This time, the medication in focus was called diethylstilboestrol. Used mainly in the 1940’s and 1950’s to prevent spontaneous abortions, and support foetal growth in high risk pregnancies [55], diethylstilboestrol later proved to increase the risk of a specific cancer type, vaginal adenocarcinoma, in adolescence and early adulthood [56]. After this, the field of trans-placental carcinogenesis of medications started growing [57]. Moreover, other long-term effects of prenatal exposures became research topics. In the 1980’s, studies began to question the impact of prenatal medication exposure on child brain development; at first in children born with congenital malformations after prenatal medication exposure [58]. However, it was not until around 2010 that a medication was established as a cause for brain development disorders [58]. The anti-epileptic medication sodium valproate was shown to reduce offspring intelligence quotient by up to 11 points [59]. This sparked
an interest in the investigation of other neurally active medications, because
they, in common with sodium valproate, can cross both the placenta and the
blood-brain-barrier of the developing foetus.

The next two sections will be dedicated to the medication safety of analgesics
and antibiotics in pregnancy. In each section, the immediate birth outcomes will
be summarised first, followed by the long-term outcomes.

1.1.3.2 Analgesics

In this thesis, analgesics are defined according to the World Health Organization’s
medication classification system [60], as analgesic opioids (excluding opioid
maintenance treatment, and illicit drug use), anti-migraine medications, NSAIDs,
and other analgesics and antipyretics (a group that includes paracetamol, and
acetylsalicylic acid).

The most commonly used analgesic in pregnancy is paracetamol [1, 4, 16,
17]. Paracetamol is generally considered not to affect immediate birth outcomes
regardless of trimester of exposure [61–63], though recent studies have pointed
to a potential higher prevalence of malformations of the male genitals among
exposed [64]. Yet, in terms of immediate birth outcomes, paracetamol is still
considered the first choice for pain relief, and fever reduction in pregnancy [61,
64].

Analgesic opioids do not seem to increase the risk of adverse immediate birth
outcomes [65–68]. However, use in the third trimester is associated with neonatal
withdrawal symptoms, and use close to delivery increases the risk of respiratory
depression in the new-born [65, 66, 68, 69]. Therefore, analgesic opioids can be
used to treat severe pain in the first and second trimester, but should be avoided
in the third trimester if possible [61–63].

NSAIDs do not seem to increase the prevalence of congenital malformations
[64]. Associations with spontaneous abortion have been reported, but it is unclear
whether the associations are causal [64]. Use in the second half of pregnancy is
associated with increased risk of premature closure of the ductus arteriosus,
oligohydramnion, and delayed onset of delivery, or poor progress of labour [64,
70–72]. Hence, NSAIDs can be used in the first and second trimester, but should
be avoided in the third trimester of pregnancy according to Norwegian and
Great British experts [72, 73]. The United States Food and Drug Administration
recommends to avoid use after week 20 of pregnancy [74].

Acetylsalicylic acid in analgesic doses confers the same risks as NSAIDs when
used in late pregnancy [64, 75], and should be avoided in the third trimester [75].

The antimigraine medications triptans do not appear to increase the risk of
immediate birth outcomes when compared to untreated migraine [76]. However,
the studies investigating safety have mainly focused on first trimester exposure
to the most commonly used triptan [77–79]. In late pregnancy, some experts
recommend caution due to few and conflicting studies [77, 78], whereas others
describe that the medication can be used when clinically indicated [79].

For long-term outcomes after prenatal exposure to analgesics, the evidence
is less clear. Analgesics can cross the placenta and the blood-brain-barrier [80,
Medication use during pregnancy

81]. Both paracetamol and NSAIDs have demonstrated endocrine disruptive properties in animal studies [64]. Furthermore, both are thought to inhibit the prostaglandin synthesis [64]. Therefore, prenatal exposure to analgesics could potentially affect long-term outcomes across several organ systems, including asthma and allergy [82, 83], obesity [82], cancer [84], and brain development [82].

According to a review, several studies have reported associations between prenatal exposure to paracetamol and childhood asthma [82]. The same association has been identified for prenatal exposure to NSAIDs in a single study [85]. However, these associations were not replicated in sibling studies [86]. Single studies on paracetamol and celiac disease [83], or childhood overweight are reassuring [87]. Studies on the potential trans-placental carcinogenicity of analgesics are also mainly reassuring. For brain tumours, no association was seen for either paracetamol, NSAIDs, or opioids [88]. Two studies found associations between prenatal exposure to non-opioid analgesics and childhood leukaemia [89, 90], but two other studies did not [84, 91]. One of the studies even suggested that prenatal paracetamol exposure is associated with a lower incidence of childhood leukaemia [91]. Of all long-term outcomes after prenatal exposure to analgesics, most controversy seems to surround child brain development [82, 92–96]. For a systematic review of papers on prenatal exposure to analgesics and child brain development, please refer to Section 4.1. In brief, especially paracetamol has been associated with various adverse brain development outcomes, but the causality of the findings remains unclear [82, 92–96]. A graphical summary of the evidence regarding prenatal exposure to analgesics and child outcomes can be found in Figure 1.2.

Given the findings on sodium valproate as mentioned in Section 1.1.3.1, and the controversies regarding paracetamol, brain development has been highlighted in calls for research action from both European and American consortia on medication safety in pregnancy [97, 98]. Therefore, brain development after prenatal exposure to analgesics is one of the foci of this thesis.

1.1.3.3 Antibiotics

Most of the commonly used antibiotics are not considered to increase the prevalence of congenital malformations [99, 100]. Exceptions include trimethoprim due to its properties as a folic acid antagonist [100], and macrolides that some studies suggest are associated with a higher prevalence of heart malformations when compared to penicillins [101]. Some antibiotics should be avoided in late pregnancy if possible [100]. This includes tetracyclines that may affect the calcification of teeth and bones and cause discolouration of teeth in children if taken after the fourth month of pregnancy [100]. Some studies have suggested that sulfonamides and nitrofurantoin are associated with a higher incidence of icterus in the new-born if used in the third trimester [100]. It is recommended to avoid third trimester use among patient groups at increased risk of haemolytic anaemia in the new-born [102, 103]. Systemic treatment with amphenicols should be avoided close to delivery due to risk of circulatory collapse in the new-born (also known as grey baby syndrome) [100, 104].
Figure 1.2: Graphical summary of the evidence regarding prenatal exposure to analgesics and child outcomes.
White: Evidence against increased risk; Grey: Sparse or inconclusive evidence; Black: Evidence of increased risk; White with grey: Limited evidence suggests no increased risk; Black with grey: Limited evidence suggests increased risk. First line treatment marked with a thick frame.
Results regarding long-term outcomes of prenatal exposure to antibiotics are conflicting [83, 89, 105–127]. Prenatal exposure to antibiotics is thought to affect the developing immune system, and the gut microbiome [116]. Several potential adverse effects of these changes have been considered, including childhood asthma or allergy [116], obesity [116], brain development disorders [108], and cancer [117].

In a review on prenatal antibiotics exposure and childhood asthma, most studies found higher incidence among exposed children [116]. However, the two most recent studies included in the review accounted for maternal asthma and susceptibility for infections in sibling analyses, or using negative exposure controls, and concluded that any increased risk could be explained by confounding [116]. Fewer studies are available on allergies [116], but one study on prenatal exposure to antibiotics and childhood celiac disease did not find higher incidence among exposed [83]. Studies on childhood obesity after prenatal antibiotics exposure have been conflicting [116], but in a study that compared antibiotic treatment to untreated infection, and prophylactic antibiotics to no infection, no association with obesity was found [113]. Correspondingly, the studies on prenatal antibiotics exposure and childhood brain development disorders that accounted for genetic confounding found no increased risk [106, 107]. As bacterial infections in pregnancy should always be treated [32, 38], studies comparing different treatment regimens are important. A study comparing prenatal exposure to penicillins and macrolides found no difference in incidence of brain development disorders [101].

Several studies have investigated prenatal antibiotic exposure and childhood cancer outcomes (for a systematic review, please refer to Section 4.2). Notably, three newer studies have suggested higher incidences of childhood cancer (effect estimates ranging from 1.2 to 1.7) after first trimester exposure to antibiotics [127], any prenatal exposure to antibiotics [110], or any prenatal exposure to penoxymethylpenicillin, pivampicillin, ciprofloxacin, or nitrofurantoin [117]. These findings are known from older studies [109, 115, 118, 122], but there are also several studies that have not identified any association [105, 111, 112, 114, 119–121, 123–126]. A graphical summary of the evidence regarding prenatal exposure to antibiotics and child outcomes can be found in Figure 1.3.

To date, studies on antibiotics and childhood cancer have failed to account for the underlying maternal infection. Therefore, antibiotics and childhood cancer is the second focus of this thesis.

Please see Figure 1.4 for an overview of the research field on medication safety in pregnancy, and how this thesis places itself within the field.
### 1. Introduction

Selected child outcomes:
- Tetracyclines
- Amphenicols
- Penicillins
- Sulfonamides
- Trimethoprim
- Macrolides
- Nitrofurantoin

Immediate birth outcomes:

#### Growth/obesity

#### Brain development

#### Asthma/allergy

#### Cancers

---

**Figure 1.3:** Graphical summary of the evidence regarding prenatal exposure to antibiotics and child outcomes.

White: Evidence against increased risk; Grey: Sparse or inconclusive evidence; Black: Evidence of increased risk; White with grey: Limited evidence suggests no increased risk; Black with grey: Limited evidence suggests increased risk. First line treatment marked with a thick frame.
1.2 Challenges when investigating long-term medication safety in pregnancy

1.2.1 Data sources

Randomised controlled trials are considered the preferred study design for research on causal effects [128]. However, pregnant women are rarely included in randomised controlled trials [8, 9]. Even if they were, inclusion and long-term follow-up of the numbers needed to investigate the risk of rare outcomes in childhood, such as cancers or brain development disorders, would probably not be feasible [10]. Therefore, observational studies are necessary. Moreover, to gain access to an appropriate duration of follow-up, studies on long-term medication safety in pregnancy often use data that were collected previously for a different purpose. These data sources include electronic health data, and birth cohort studies [8]. Electronic health data are recorded routinely when a patient has contact with the health care system [8]. Sometimes the data are stored for the purpose of insurance claims, sometimes as part of national health registries [8]. Birth cohort studies that prospectively recruit pregnant women for participation have typically been established to investigate how prenatal exposures impact child health [129, 130]. However, the cohorts used today to study long-term outcomes after prenatal exposure to medications will necessarily have been established in the past, so the data collection cannot be tailored to each new research question. Therefore, I use the term existing data to cover both birth cohorts and administrative databases, though the term is often reserved for administrative databases [8, 13]. In the following sections, the most important limitations when repurposing existing observational data will be introduced. These include data validity [13], and comparability between individuals who use...
the medication of interest and individuals who do not (exchangeability) [8]. The final section of the introduction will deal with the challenge of understanding whether identified associations from observational data can be considered causal.

1.2.2 Data validity

In observational data, we may be unable to accurately capture information on key variables, be that exposure, outcome, or covariates. The information may be completely unavailable (missing data), or available, but of questionable validity (resulting in potential misclassification). The following sections will deal with both scenarios.

1.2.2.1 Missing data

Data can be missing in different ways. They may be completely missing, because the data source is a registry or claims database where the information on the variable is not registered (for example whether the pregnancy was planned). Data may also be sporadically missing, for example if not all participants answered a survey-question, or if registration is not mandatory (for instance it is possible for Norwegian pregnant women to opt out of having smoking information reported to the Medical Birth Registry [131]). If a variable is missing for everyone, it can be considered as missing completely at random [132]. Sporadically missing variables may also be missing completely at random, but will often be missing at random, or missing not at random [132]. Missing at random is a rather unfortunate name for variables that are missing at random if we condition on other measured variables. The term missing not at random refers to variables that are systematically missing in a way that we cannot account for using available data [132]. It can be tested whether data are missing completely at random, but whether data are missing at random, or missing not at random can only be assumed [132].

Typically, participants with completely missing information on the exposure or the outcome are not included in analyses. To account for potential selection bias from excluding persons lost to follow-up, inverse probability of censoring weights can be used [133]. The weights are based on the estimated probability of having data available for follow-up given the measured baseline covariates [133]. When calculating the estimates of association, individuals with a lower probability of being followed up are given a larger weight, or importance, than individuals with a higher probability of being followed up [133].

Studies may have completely missing information on important covariates. In many cases, this will lead to unmeasured confounding (a concept that will be expanded in Section 1.2.3). However, if the covariates of interest are measured in a sample that is similar in confounding structure to the study sample, propensity score calibration can be used to correct the estimate of association in the study sample [134].

When information on covariates is sporadically missing, and the data are missing at random, multiple imputation can be used to replace missing values
Challenges when investigating long-term medication safety in pregnancy

in a way that accounts for the uncertainty of the missing values [132]. When data are missing at random, multiple imputation has been shown to reduce bias and improve statistical precision compared to only analysing participants with complete information on the covariates [132]. If data are missing not at random, both multiple imputation, and analyses restricted to participants with complete data will be biased [132].

1.2.2.2 Measurement and (mis)classification

Regarding the exposures in this thesis, it is debated how medication use during pregnancy should ideally be measured [12, 135]. Whereas prescribed medications can often be detected in registries or claims databases, medications that are available over-the-counter can only be captured by self-report among the pregnant women [8]. Both prescription records and self-report may be limited by misclassification. Not all prescribed medications that are filled by pregnant women will actually be taken [136]. Self-report may be limited by poor recall [137], or unwillingness to disclose medication use that is associated with social stigma – though a Norwegian study suggests that the latter is a small problem [138]. Exposure data from prescription registries will be gathered prospectively and typically independent of the outcome. Therefore, any misclassification can be assumed to be non-differential with any bias generally being towards the null [139]. However, maternal self-report may not always result in non-differential misclassification, even if gathered prospectively [95, 139]. This particularly applies to studies where the outcome is also parent-reported. In such cases, the reporting of the exposure and the outcome are not independent, and it has been suggested that some underlying factors could affect how both exposure and outcome are reported [95], yielding unpredictable direction of the resulting bias [95, 139].

Regarding the outcomes in this thesis, a number of the controversies surrounding prenatal exposure to analgesics and child brain development, mentioned in Section 1.1.3.2, have been concerned with the validity of the outcome measures [95, 96, 140]. In brief, some studies use medical diagnoses [141–144], whereas others use psychometric instruments (questionnaires or tests) completed by health care professionals [145–147], teachers [145, 147, 148], or parents [142, 149, 150]. It has been questioned whether medical diagnoses are sensitive enough to detect small changes in brain development following medication exposure [151]. For instance, prenatal exposure to sodium valproate increases the risk of autism, but the clinical manifestation of the cases is atypical and may not always meet the diagnostic criteria [152]. Other researchers have questioned the clinical relevance of parent-reported symptoms [95, 140]. These considerations relate to content validity (the degree to which the method used to measure the outcome captures all aspects of the outcome, without capturing aspects that are not relevant to the outcome [153]). However, concerns about misclassification also apply to the outcome, and are relevant for both brain development and cancer. Such misclassification can be differential if children prenatally exposed to the medication of interest are monitored more closely than
unexposed children [9].

Both exposure and outcome misclassification can be handled through probabilistic sensitivity analyses [154]. The theoretical underpinning for this is that there exists a gold standard against which the exposure or outcome classifications can be compared. Furthermore, validation studies are necessary that compare the exposure or outcome classifications to the gold standard. Based on the findings from validation studies, bias corrected estimates can be calculated accounting for the validity of the exposure and/or outcome classifications [154]. Important terminology when investigating the validity of dichotomous exposures and outcomes is introduced in Figure 1.5.

Though very useful when a gold standard exists or can be assumed, probabilistic sensitivity analyses do not provide guidance on which outcomes are the most valid, or clinically relevant. This is an aspect that will be reviewed further in the present thesis.

<table>
<thead>
<tr>
<th>Gold standard exposed</th>
<th>Gold standard unexposed</th>
<th>Positive predictive value: $\frac{TP}{TP + FP}$</th>
<th>Negative predictive value: $\frac{TN}{FN + TN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recorded exposed</td>
<td>True positive (TP)</td>
<td>False positive (FP)</td>
<td></td>
</tr>
<tr>
<td>Recorded unexposed</td>
<td>False negative (FN)</td>
<td>True negative (TN)</td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity: $\frac{TP}{TP + FN}$

Specificity: $\frac{TN}{FP + TN}$

Figure 1.5: Terminology regarding the validity of dichotomous exposures. The same terminology applies to dichotomous outcomes.

Finally, information on important potential confounders may be misclassified, resulting in residual confounding. The topic of confounding will be covered in the following section.

### 1.2.3 Comparability

Another limitation of observational studies is illustrated in the original case-control study linking prenatal diethylstilboestrol exposure to vaginal
Challenges when investigating long-term medication safety in pregnancy

adenocarcinoma in early adulthood [56]. In that study, not only was use of diethylstilboestrol during pregnancy more common among case mothers than among control mothers, maternal bleeding during pregnancy, and previous pregnancy loss were also more common among cases than controls [56]. These maternal conditions are markers of risk of spontaneous abortion, one of the indications for diethylstilboestrol treatment [56].

In randomised controlled trials, it is expected that background characteristics will be randomly distributed between exposed and unexposed participants [128], but in observational studies on medication exposure, the indication for medication use will be more prevalent among medication users than among population controls [155]. Several other baseline characteristics may also differ systematically between exposed and unexposed [9, 134]. If a baseline factor also influences the risk of having the outcome, confounding may be present [134]. To make assumptions about potential confounders explicit, a graphical depiction of the relationship between variables of interest using directed acyclic graphs has been suggested [156]. Directed acyclic graphs make it possible to distinguish between potential confounders, intermediates, and colliders, as illustrated in Figure 1.6.

![Directed Acyclic Graph Illustration](image)

**Figure 1.6**: Illustration of the terminology used in directed acyclic graphs. The variable that an arrow points at is affected by the variable that the same arrow origins from.

In order to obtain a valid effect estimate, it is necessary to account for all confounding factors, whereas it is inappropriate to condition on colliders, or intermediate factors [156]. However, as we saw in Section 1.2.2, data on important confounders may be unavailable, or misclassified. It is therefore common to distinguish between strategies to account for measured and unmeasured confounding [134]. The former include confounder summary scores, such as the propensity score, marginal structural models [134], and disease comparators [9]. Strategies to account for unmeasured confounding include sibling designs,
1. Introduction

instrumental variables [134], and active comparators (if the underlying indication is unmeasured but the medications of interest are used for fairly specific indications, or if there are unmeasured factors associated with willingness to seek and use treatment) [155]. Methods that estimate the impact of potential unmeasured confounding include negative controls [10], probabilistic sensitivity analysis [9], and calculation of the e-value [157]. A brief overview of the listed methods, also specifying which methods will be used in this thesis, is available in Table A.1.

In Figure 1.7, the Figure 1.4 of this thesis’ place within the research field of medication safety in pregnancy has been expanded to include the limitations associated with using existing data.

Limitations in existing data
- Data validity
- Exchangeability

Figure 1.7: Illustration of this thesis’ place in the research field of medication safety in pregnancy, including the limitations associated with using existing data.

1.2.4 Causal inference

Causal inference is made difficult by the lack of a precise definition of causation [158]. Several attempts have been made in defining useful criteria for causal inference [158]. Among the most famous are the Bradford Hill criteria [159], and the newer potential outcomes approach, especially the target trial approach advocated by Hernán and Robins [128, 160] (see Table A.2 for a brief overview). The target trial approach lends itself well to the study of medication use in pregnancy and long-term child outcomes, because medication exposure during pregnancy can be considered as a well-defined intervention, which is a requirement in the target trial approach [128, 161]. Some methods used in this thesis, such
as directed acyclic graphs and marginal structural models, are based on the potential outcomes framework [158, 160].

However, the target trial approach has been criticised for too narrow a focus on exposures that are currently feasible to change using interventions [158]. For example, biological sex cannot be investigated as an exposure in the target trial approach [158]. Furthermore, the critics state that the target trial approach puts too much emphasis on individual studies and not enough on the integration of evidence from several studies using different approaches, or from different methods used within the same study [158]. Vandenbroucke et al. have therefore suggested a pragmatic pluralism approach to causal inference [158]. This approach focuses on integration of evidence from several studies, evidence from other disciplines (what they refer to as ‘integrating evidence’), and the use of triangulation (evidence from different methods with diverging bias structures [10]) [158]. The pragmatic pluralism approach thus also welcomes methods based on the potential outcomes framework [158]. Hence, this approach seems well suited for the present thesis, given that the thesis has synthesised evidence from several studies, and used several different methods, including negative controls, a method for triangulation that is not covered by the target trial approach [158].
Chapter 2

Thesis aims

This thesis aimed to investigate the associations between prenatal exposure to analgesics or antibiotics and long-term child outcomes, including how to improve the validity when estimating these associations in existing data. Specifically, the thesis focused on analgesics and childhood brain development, and antibiotics and childhood cancer. Hence, the five papers that make up this thesis can be organised according to two specific aims:

Aim 1: Assessing the association between prenatal exposure to analgesics and child brain development

- Paper i: Describe how child brain development has been measured in medication safety studies on prenatal exposure to analgesics, identify relevant directions for future research, and formulate considerations for how to improve measurement and reporting.
- Paper ii: Identify patterns of prescription fills for analgesics before, during, and after pregnancy.
- Paper iii: Investigate the association between prenatal exposure to NSAIDs and ADHD in childhood.

Aim 2: Assessing the association between prenatal exposure to antibiotics and childhood cancer

- Paper iv: Describe how childhood cancer has been investigated in medication safety studies on prenatal exposure to antibiotics, and provide considerations for how to improve reporting and limit potential biases.
- Paper ii: Identify patterns of prescription fills for antibiotics before, during, and after pregnancy.
- Paper v: Investigate the association between prenatal exposure to nitrofurantoin and childhood leukaemia.
Chapter 3
Materials and methods

3.1 Study design

For both specific aims of this thesis, the first step was to conduct a systematic literature review (papers i and iv) that could inform the study design and conduct of the following studies. The second step was a descriptive cohort study to understand the patterns of medication use in pregnancy (paper ii). The third step was an analytic cohort study (papers iii and iv). Figure 3.1 illustrates the study designs in the three cohort studies.

3.2 Data sources

This thesis relied solely on existing data sources. For two of the three cohort studies, sufficient sample sizes could be achieved using data from Norway, but because childhood cancer is fortunately very rare (156 per million person-years [162]), the third study included data from four Nordic countries: Denmark, Finland, Norway, and Sweden. Below, the data sources used in this thesis will be described.

3.2.1 Systematic reviews

For both systematic reviews, we searched several databases (PsycInfo [paper i only], PubMed, Embase, Scopus, Cochrane, and Web of Science) to ensure coverage of the published literature. In addition, we screened the reference lists of any relevant reviews, and of all included studies. The search was structured according to the Participants, Interventions, Comparators, Outcomes framework as described in the Preferred reporting items for systematic review and meta-analysis protocols [163]. Please see Table 3.1 for an overview of the concepts included in the search strings along with relevant synonyms. No date restrictions were applied, but only published original studies with a comparator group were included.

3.2.2 Medical birth registries

The Medical Birth Registry of Norway (MBRN) was established in 1967 [131]. Today MBRN has mandatory registration of all births after week 12 [164], and covers a wide range of information on perinatal factors [131]. Of importance for research purposes, mothers, partners, and children are deterministically linked using unique personal identification numbers [164] given to all residents in Norway. Furthermore, information is available on the date of birth for the
3. Materials and methods

Figure 3.1: Study design diagram including data sources for the three cohort studies. ADHD: Attention-Deficit/Hyperactivity Disorder; MoBa: Norwegian Mother, Father and Child Cohort.
Data sources

Table 3.1: Overview of search strategies in the systematic reviews.

<table>
<thead>
<tr>
<th></th>
<th>Participants</th>
<th>Intervention</th>
<th>Comparators</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper i</td>
<td>Prenatal exposure</td>
<td>Analgesics</td>
<td>Intentionally left blank</td>
<td>Brain development</td>
</tr>
<tr>
<td>Paper iv</td>
<td>Child</td>
<td>Prenatal exposure</td>
<td>Intentionally left blank</td>
<td>Cancer</td>
</tr>
</tbody>
</table>

children, and the gestational age at birth [131], so the duration of pregnancy can be accurately determined. The main limitation of MBRN from a research perspective is that data on several important life-style factors (for example maternal smoking, body mass index, and alcohol use) are partially or completely unavailable because their registration is not mandatory [131].

The medical birth registries in the other Nordic countries have similar contents as the Norwegian and were established in the 1970s and 1980s [165–167].

3.2.3 Prescription registries

The Norwegian Prescription Database (NorPD) was established in 2004 and is considered complete from 2005 [168]. NorPD has information on all prescriptions filled at pharmacies by patients in ambulatory care [168]. The registry is linkable via personal identification numbers. The main strength of NorPD for research is the national coverage, whereas the main limitation is that over-the-counter medications, and medications used during hospital admissions are not captured [169]. Furthermore, it is unknown whether or when the dispensed medications were used [169].

In the other Nordic countries, the prescription registries were established in the 1990s and 2000s [170–172]. The contents of the registries differ slightly, as the Finnish Prescription Registry only contains data on medications that are potentially eligible for reimbursement [171].

3.2.4 Patient registries

In Norway, a linkable patient registry was established in 2008 [173]. The Norwegian Patient Registry (NPR) contains information on diagnoses and procedures from government-owned hospitals and outpatient clinics, and private health clinics that receive governmental reimbursement [173]. A limitation of the NPR is that somatic outpatient care tends to be underreported [173].

Though some of the Nordic countries established their patient registries earlier, linkable data on both inpatient and outpatient contacts only became available in the 1990s and 2000s [174–176].
3. Materials and methods

3.2.5 Cancer registries

The Nordic cancer registries have recorded incident cases of cancer since the 1940s or 1950s, depending on country [177–180]. Countrywide coverage and mandatory reporting ensures that the registries are almost 100% complete [177–180].

3.2.6 Registry-based information on death and migration

The Nordic countries have civic registries that record dates of death and migration for all residents [181, 182]. Cause of death registries also record dates of death [183, 184]. In Norway, information on dates of death and migration in children is available from the MBRN throughout childhood. Based on our experience in obtaining data for the present thesis, it is difficult to obtain permissions to share the data from the civic registries across countries, whereas other registries are more open to data sharing.

3.2.7 The Norwegian Mother, Father and Child Cohort (MoBa)

MoBa is a pregnancy cohort that included pregnant women from all over Norway between 1999 and 2008 [185]. In 41% of pregnancies, the women consented to participation in surveys, and linkage of their data from the Norwegian health registries, resulting in 114,500 included children [185]. Women were invited to participate when they booked their routine pregnancy-week 17 ultrasound scan [185]. Pregnant women and their partners each were asked to complete a questionnaire around week 17 of pregnancy, and the women completed two additional questionnaires during pregnancy (weeks 22 and 30), and a questionnaire six months after delivery [185]. Several questionnaires have been sent throughout the childhood of the index child, and follow-up is ongoing.

The main strength of MoBa as a data source is the detailed information about important potential confounders and pregnancy exposures, including exposure to over-the-counter medications [8]. The main limitation of MoBa is the selected sample, and an increasing loss to follow-up as the children grow older [185]. Compared to the general birthing population of Norway, participants were less likely to be young parents, more likely to be married or cohabiting, and had a healthier lifestyle during pregnancy [186]. Women who were still participating in the cohort when the index child was 3 years old had higher educational levels than other women who gave birth during the same period, and the index child was more likely to be the firstborn child [187].

3.3 Study samples

The flowcharts for the two systematic reviews have been combined in Figure 3.2. The study samples for the three cohort studies are illustrated in Figure 3.3. The difference in study sizes reflect the difference in outcome prevalences.
Study samples

Figure 3.2: Flowcharts for the two systematic reviews.
Figure 3.3: Flowcharts for the three cohort studies. ADHD: Attention-Deficit/Hyperactivity Disorder; MBRN: Medical Birth Registry of Norway; MoBa: Norwegian Mother, Father and Child Cohort; NorPD: Norwegian Prescription Database; NSAID: Non-steroidal anti-inflammatory drug.
3.3.1 Ethics and data protection

MoBa was approved by the Norwegian Data Protection Agency, and The Regional Committee for Medical Research Ethics. All adult participants in the cohort provided written, informed consent to participation, and to the use of their data from the Norwegian health registries. All cohort studies in this thesis were approved by The Regional Committee for Medical Research Ethics in South-Eastern Norway; approval numbers 2015/2137/REK Sør-Øst, 2018/140/REK Sør-Øst, and 2018/142/REK Sør-Øst. In addition, the studies were approved by the data protection officer at the University of Oslo. The Nordic study was further approved by the Swedish Ethical Review Authority (approval number: 2018/2604-31/1 2019-00268 (2019-02311)). In Denmark and Finland, registry-based studies are exempt from ethical review, but the study was approved by local data protection officers (approval number, Denmark: 2019-DCRC-0096, approval numbers, Finland: THL/2297/5.05.00/2018, Kela 120/522/2019, TK-53-1405-19). Data were handled in accordance with the European Union General Data Protection Regulation (2016/79).

3.4 Measures

3.4.1 Medication in pregnancy

Given the importance of timing of prenatal medication exposure (cf. Section 1.1.2), all three cohort studies estimated not only whether any prenatal exposure had taken place, but also the trimester of exposure. Figure 3.4 illustrates how the exposure was classified in the three cohort studies. In the prescription registries, a filled prescription at any point during a trimester was considered as exposure in said trimester for papers ii and v. In MoBa, women were asked to recall if they had experienced any of a list of health problems and whether the health problem had been treated with any medication. They also provided the timing of the health problem by ticking one or more boxes representing four-week intervals (e.g. pregnancy week 9-12). The four-week intervals were categorised to pregnancy trimesters for the purpose of paper iii.

In both analytic studies (papers iii and v), we also attempted to identify measures of cumulative dose or duration of treatment. In paper iii, based on the MoBa, the best approximation we could make was number of four-week intervals with reported exposure. Number of prescription fills was the approximation chosen for paper v. In both papers, we a priori assumed that several treatment instances would be the exception rather than the norm, so we chose to categorise the number of treatment instances.

3.4.1.1 Choosing an appropriate comparator

For the two analytic studies, an important consideration was choosing an appropriate comparator. In paper iii on analgesics, we took into account that many women forego analgesic treatment during pregnancy [19, 188]. Therefore,
Papers ii & v: Prescription fills

<table>
<thead>
<tr>
<th>≤ 1 year BP</th>
<th>D 0-89</th>
<th>D 90-179</th>
<th>D 180-birth</th>
</tr>
</thead>
</table>

Timing of exposure: 1st trimester 2nd trimester 3rd trimester

Paper iii: Maternal report in MoBa

<table>
<thead>
<tr>
<th>Q1 coverage</th>
<th>Q3 coverage</th>
<th>Q4 coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>13-16</td>
<td>≥29</td>
</tr>
<tr>
<td>5-8</td>
<td>17-20</td>
<td>Last part of pregnancy</td>
</tr>
<tr>
<td>9-12</td>
<td>21-24</td>
<td></td>
</tr>
<tr>
<td>4-week intervals:</td>
<td>25-28</td>
<td>≥29</td>
</tr>
<tr>
<td>≤6 months BP</td>
<td></td>
<td>End of pregnancy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Q1 completed</th>
<th>Q3 completed</th>
<th>Q4 completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>≅ W 18</td>
<td>≅ W 31</td>
<td>≅ 6 months AP</td>
</tr>
</tbody>
</table>

Timing of exposure: 1st trimester 2nd trimester 3rd trimester

Figure 3.4: Illustration of the exposure definition in the three cohort studies. AP: After pregnancy; BP: Before pregnancy; D: gestational days; MoBa: Norwegian Mother, Father and Child Cohort; Q: questionnaire; W: gestational weeks.
conditions of fever, infection, or pain that were not treated with the medication of interest (and hence could be either untreated or treated with a different medication) were considered relevant comparators for paper iii. In paper v on antibiotics, a different approach was chosen. Based on lessons learned from paper iii, and the fact that bacterial infections during pregnancy should always be treated [32, 38], we decided to use an active comparator design. Therefore, the study compared children prenatally exposed to two different antibiotics commonly used to treat the same underlying type of infection.

### 3.4.2 Child brain development

As mentioned in Section 1.1.3.2, there have been debates about how to measure brain development. Therefore, the first paper of this thesis aimed at identifying how brain development has been measured in previous studies in order to provide suggestions on how to quantify brain development. Based on the suggestions and research gaps identified in paper i, we decided to investigate child attention problems in two ways for paper iii. Primarily, as the presence or absence of a clinical diagnosis of hyperkinetic disorder, F90 according to the International Classification of Diseases, 10th revision [189], and/or a prescription fill for a medication used in the treatment of ADHD. This outcome classification was chosen in order to also capture children who were diagnosed before the establishment of the NPR in 2008. In Norway, general practitioners are not allowed to initiate treatment with ADHD medications [190], so children with prescription fills must be diagnosed by a specialist. Children were followed from birth until the first diagnosis of ADHD, prescription fill for ADHD medication, or end of 2016, whichever came first.

As a secondary outcome, we investigated parent-reported symptoms of ADHD on the 12-item ADHD index from the Conners’ Parent Rating Scale-Revised, Short Form (CPRS-R (S)). Higher scores on the CPRS-R (S) are indicative of more symptoms [191]. To facilitate an assessment of the clinical relevance of the findings regarding symptoms of ADHD, the CPRS-R (S) was standardised to a z-score, where the mean is zero and the standard deviation is one. Z-scores of two or more are then considered indicative of clinically important problems with attention and/or hyperactivity [191].

### 3.4.3 Childhood cancer

Based on the previous literature, and our findings from paper iv, we decided to investigate childhood leukaemia according to the International Classification of Childhood Cancer, 3rd revision (ICCC-3) [192] in paper v. Childhood leukaemia was defined as an incident diagnosis from ICCC-3 site group 1 (codes 011-015). In a secondary analysis, we investigated the two most common childhood leukaemia subtypes: lymphoid leukaemia (ICCC-3 code 011) and acute myeloid leukaemia (ICCC-3 code 012). Children with more than one cancer diagnosis were only considered at risk of the outcome until their first diagnosis, and subsequent diagnoses were disregarded, in order to capture incident primary
3. Materials and methods

cancers. Follow-up started at birth and continued until the first cancer diagnosis, death, migration, the child’s 20th birthday, or the end of 2017, whichever came first.

3.4.4 Covariates

The relevant covariates differed by study, as they were selected based on topic knowledge and assumptions about the underlying causal relationships as depicted in directed acyclic graphs [156]. Please refer to the individual papers for more detailed descriptions of the included covariates.

3.5 Statistical analyses

In both systematic reviews, the aim was a qualitative synthesis of the previous literature to inform decisions regarding materials and methods in future studies. Therefore, no statistical methods were used. In paper ii only descriptive statistics were used, so this section will focus on papers iii and v. All statistical analyses were performed in Stata (versions 14.2 to 16.1; StataCorpLP).

3.5.1 Missing data and multiple imputation

In both papers, data on some important covariates were sporadically missing. We assumed data to be missing at random, and therefore used multiple imputation to replace the missing values. To allow for categorical variables, we used multiple imputation by chained equations [193]. As recommended [132], we included exposures, outcomes, and potential confounders in the imputation models. In paper v, where few potential confounders were identified, some auxiliary variables were included in the imputation model as well. This should in theory increase the plausibility of the assumption that data are missing at random [132]. Both papers used outcome models that took different length of follow-up between participants into consideration. To account for this, the imputation models included the cumulative Nelson Aalen hazard function for the outcomes [194, 195].

3.5.2 Modelling

Paper iii used Cox regression to estimate the crude hazard ratios (HRs) of ADHD diagnosis with 95% confidence intervals (CI) comparing children prenatally exposed to NSAIDs and children unexposed to NSAIDs but exposed to maternal fever, infection, or pain. Generalised linear models were used to estimate crude mean differences in ADHD symptom z-scores with 95% CIs for the same comparison. To account for time-varying confounding by the severity of the indication (proxied by analgesic and psychotropic co-medication use, and physical exercise), we used marginal structural models. Separate propensity scores were estimated at baseline, and for each trimester of pregnancy including NSAID treatment history, time-varying covariates, and baseline covariates (please
Statistical analyses

Statistical analyses refer to paper iii for a full description of the included variables). The propensity scores were converted to inverse probability of treatment weights (IPTWs), and the resulting four IPTWs were multiplied to obtain a total weight. The total weight was used in marginal structural Cox models [133] with robust standard errors to obtain weighted HRs with 95% CIs. The robust standard errors were used to account for the weighting and for some dependence between observations, as a mother could have more than one child in the cohort.

In the analysis on ADHD symptoms, there was loss to follow-up. Therefore, we estimated inverse probability of censoring weights based on baseline covariates in all eligible pregnancies and multiplied these with the IPTWs [133]. The resulting total weight was used in generalised linear marginal structural models with robust standard errors to obtain weighted standardised mean differences with 95% CIs.

Analyses on both ADHD diagnosis and symptoms were repeated after stratification by indication for NSAIDs use to further account for confounding by indication.

**Paper v** used Poisson regression to estimate crude incidence rate ratios (IRRs) and incidence rate differences (IRDs) of childhood leukaemia with 95% CIs comparing children prenatally exposed to nitrofurantoin, and children prenatally exposed to pivmecillinam. As the exposure was expected to be much more prevalent than the outcome, we used propensity scores to account for confounding. The propensity scores included potential confounders and predictors of the outcome as recommended [196]. In active comparator studies, it is preferable to use the propensity scores in IPTWs [197], so we did. Generalised linear models with log link and robust standard errors were used in the weighted dataset to obtain weighted IRRs and IRDs with 95% CIs.

Analyses were conducted separately for each country using a common data model. To combine the findings, we used fixed-effects meta-analysis, assuming a common treatment effect across the Nordic countries [198]. Heterogeneity was examined using $I^2$, where values above 25% suggest non-negligible heterogeneity [199]. For analyses with zero exposed or unexposed cases in one or more countries, results were combined using multilevel mixed-effects Poisson regression with random effect terms for the variance components [200]. This method has been shown to yield less biased results than standard meta-analysis techniques when combining incidence rate data in the presence of zero counts [200].

### 3.5.3 Sensitivity analyses

Pre-planned sensitivity analyses were used in both papers iii and v to test the robustness of the findings.

To test the robustness of the statistical methods, both papers did complete case analyses to compare with the analyses in the imputed data. Paper iii also tested several different specifications of the IPTWs.

To account for potential exposure misclassification, we used probabilistic sensitivity analysis in paper iii as a validation study was available, comparing
3. Materials and methods

Maternal report of NSAID use in medication diaries with report in a questionnaire with approximately the same recall period as the MoBa questionnaires [201]. For paper v, we were unable to identify an appropriate validation study that could substantiate any assumptions for a probabilistic sensitivity analysis on exposure misclassification.

To evaluate the outcomes in paper iii, we assessed the correspondence of the CPRS-R (S) z-scores and the ADHD diagnoses, and we restricted the classification of ADHD to only include diagnoses of F90, not ADHD medication. In paper v, we excluded diagnoses of infant leukaemia (diagnoses during the first year of life), as the aetiology of infant leukaemia and childhood leukaemia are thought to be different [36].

In both papers, disease severity was unmeasured. To account for unmeasured confounding by disease severity, we used unexposed children born to pre-pregnancy NSAID-users as a comparator group in paper iii. In paper v, we restricted the sample to children who were not exposed to other antibiotics than the medications of interest. We did not use these methods in the main analyses, as they come at the cost of conditioning on potential intermediate factors.

To estimate the extent of unmeasured confounding, we used negative exposure controls in paper iii. Children born to mothers using NSAIDs before pregnancy, but not during, were compared to children whose mothers reported indications for NSAID use, but did not use NSAIDs before pregnancy. In paper v, we calculated the e-value to estimate how strong confounding would be required to reduce the estimate of association to the null if the confounding was removed. Furthermore, in a post hoc analysis, we used an unexposed comparator group to estimate the extent of any confounding by maternal indication for nitrofurantoin treatment.
Chapter 4

Main findings

The main findings will be presented separately for the two specific aims of the thesis. For each of papers i-iv, it will be described how the findings contributed to the design of later papers. Figure 4.1 displays the connection between the papers.

![Figure 4.1: Illustration of how the five papers in the thesis are connected.](image)

4.1 Specific aim 1: Analgesics and child brain development

**Paper i** In the systematic review, we identified 29 studies on prenatal exposure to analgesics and child brain development. The most important finding from the review was the high heterogeneity between studies when it came to choice of outcome measure. Across the 29 studies, we identified 12 different psychometric instruments completed by health care professionals, nine different psychometric instruments completed by parents, six different psychometric instruments completed by teachers, and four different diagnostic categories. Box 4.1 summarises five points for consideration in future studies that we made based on our findings.
4. Main findings

Box 4.1: Five points for consideration in future studies on prenatal exposure to analgesics and child brain development

A wide variety of outcomes must be assessed to establish whether a medication affects child brain development. Child brain development consists of several domains, including cognition, behaviour, and emotionality. All domains of brain development are important for children’s daily living. Therefore, it is not reasonable to select one domain of brain development as the priority in medication safety research. However, to facilitate meta-analyses, it would be helpful if studies assessing the same domain of brain development and using the same assessor (e.g. parent, or teacher) would also use the same outcome measure.

Previous literature, both animal and human, should inform choice of outcomes

Biological plausibility, as well as signals from previous studies, should guide the choice of outcomes prioritised for each medication. Therefore, collaboration with pre-clinical researchers, or knowledge of the literature on in vitro or animal studies, is necessary.

Different sources of outcome ascertainment have different strengths and weaknesses and should complement each other

A study might use both data on child diagnosis and parental assessment to see if the results are in agreement. However, in meta-analyses the use of both diagnoses and psychometric instruments can be challenging. In-depth knowledge of the various outcome measures is necessary when deciding which outcomes can be combined.

The outcome measure should be age appropriate

Many tests or questionnaires are developed for a certain age group and should not be used for older or younger children. Similarly, some diagnoses cannot be given to children below a certain age. Brain development continues into early adulthood, and some problems are not detected until the teenage years, when more is expected of the child.

Reliability and validity of the outcome measure should be reported

The use of several different outcome measures across studies makes it difficult for readers to be familiar with all the different measures. This increases the responsibility of authors to provide information on validity and reliability. For the types of validity and reliability where statistical tests can be done, quantitative measures of validity and reliability should be reported.

Regarding the identification of directions for future research, we found that most research had been done on paracetamol. Of the 29 studies, 18 investigated paracetamol. The longest follow-up available was 18 years. Prenatal exposure to paracetamol did not appear to be associated with cognitive problems, nor
emotional problems (although only two studies investigated the latter). The evidence about behavioural problems was more heterogeneous, although the majority of studies found a higher incidence of ADHD or other behavioural problems among exposed; more consistently so in studies investigating diagnostic outcomes.

Analgesic opioids were only investigated in two studies. One investigated language skills in 3-year-olds and did not identify more language problems among exposed. The other investigated diagnoses of autism or developmental delay in pre-schoolers and found elevated estimates of association, but with confidence intervals including the null.

Only two papers investigated NSAIDs other than indomethacin for the treatment of premature contractions. One paper investigated children at the age of 3 years, the other followed children until the age of 6 years. Both studies were reassuring with no evidence of more cognitive or behavioural problems among exposed. None of the studies used diagnostic outcomes.

Acetylsalicylic acid was investigated in five studies with the longest follow-up being 11 years. Findings on cognitive and behavioural outcomes were reassuring. None of the studies used diagnostic outcomes. Both NSAIDs and acetylsalicylic acid were used as negative controls in studies on prenatal paracetamol exposure and child brain development.

Five studies investigated prenatal triptan exposure. The longest follow-up available was 18 years. The investigations included cognition, behaviour, and emotionality, and none of the studies found any evidence of association with adverse brain developmental outcomes.

In a general way, the five points for consideration, mentioned in Box 4.1, all informed the design and reporting of paper iii. More particularly, the findings from paper i motivated the design of paper iii in the following manner:

- We chose to focus on NSAID use, because it was used as a negative exposure control in studies on paracetamol and child behaviour despite the limited data available on brain development after prenatal exposure to NSAIDs. Furthermore, in studies on prenatal exposure to paracetamol, it was mainly the studies with diagnostic outcomes that identified increased risks, but no studies had investigated the association between ADHD diagnosis and prenatal exposure to NSAIDs.

- We chose to investigate the outcome of interest using both diagnosis and parent report.

**Paper ii** Among 172,585 pregnancies included in the descriptive cohort study, 9.8% filled a prescription for an analgesic in the three months before pregnancy. The most commonly dispensed analgesic medication was an NSAID, diclofenac (Table B.1). During pregnancy, 6.5% filled a prescription for an analgesic. The highest proportion was seen in the first trimester, 4.2%, and the most commonly dispensed medication switched from diclofenac to paracetamol with codeine, and remained so throughout pregnancy. The proportion with filled analgesic
prescriptions was lower in the second (1.8%), and third (1.7%) trimesters. In the three months after pregnancy, the proportion rose to 5.7%, and diclofenac was again the most commonly dispensed. First trimester NSAID-fills were seen in 2.0% of pregnancies, but second and third trimester fills were only seen in 0.23% and 0.14% of pregnancies.

NSAID prescription fills seemed in accordance with the recommendation to avoid use in the second half of pregnancy if possible. Hence, women with second, and especially third trimester use of NSAIDs probably differ substantially from the general population. This contributed to the design of paper iii in the following ways:

- We chose to use women with indications for NSAID treatment as comparators rather than using the general population as comparators.
- We decided to use marginal structural models to better account for time-varying exposure and confounding.
- We chose to investigate trimester-specific NSAID exposure and not report a composite exposure of NSAIDs at any point during pregnancy. The rationale for this was that the disease severity probably differed between first and third trimester exposed, in addition to the foetal brain having different susceptibility in different trimesters.

Another objective in paper ii, although slightly off topic for the present thesis, was to investigate whether it was possible to identify pregnancy planning in routinely collected health data. This did not appear to be the case, given the proxies that we had selected (hormonal contraceptive discontinuation, fertility treatment, and folic acid use). Pregnancy planning was considered an important potential confounder for paper iii, but fortunately information on pregnancy planning was available in MoBa, so we were not obliged to rely on registry data.

**Paper iii** In an analytic cohort study, we included 56,340 children for the analysis on ADHD diagnosis. For ADHD symptoms, 34,961 children were included. NSAID use was reported in 6.2% of pregnancies, and the prevalence of child ADHD diagnosis after an average of 9.8 years of follow-up (range 8 to 12 years) was 2.2%. Regarding parent-reported ADHD symptoms at age 5 years, the proportion of children who had a z-score of two or more standard deviations from the mean was 4.5%.

Prenatal exposure to NSAIDs was not associated with higher incidence of ADHD diagnosis in any trimester or duration category. In the crude analysis, first trimester exposure to NSAIDs was associated with a higher incidence of ADHD diagnosis (HR 1.32, 95% CI 1.03 to 1.68). However, after weighting, the association was no longer seen (HR 1.12, 95% CI 0.86 to 1.45), pointing to the importance of accounting for confounding. We noted an elevated point estimate for NSAID exposure in two to three four-week intervals, but with wide CIs including the null (HR 1.32, 95% CI 0.89 to 1.96). For a single four-week interval, and for four or more intervals, the HR was at and below the null, respectively.
In the secondary analysis on ADHD symptoms, we did not identify any evidence of higher symptom scores associated with any timing of exposure. In the analysis on duration, we found slightly higher symptom scores in children exposed in a single four-week interval (weighted standardised mean difference 0.11, 95% CI 0.05 to 0.17) compared to unexposed children. The same point estimate, but with wider confidence intervals, was seen in children exposed for two to three intervals (weighted standardised mean difference 0.10, 95% CI -0.00 to 0.20), but not in children exposed for four or more intervals (weighted standardised mean difference -0.02, 95% CI -0.16 to 0.12).

The results did not appear to differ by strata of maternal indication, but the statistical precision was limited in the stratified analyses.

The findings from our sensitivity analyses showed that non-differential exposure misclassification could have masked a clinically relevant difference in ADHD risk between first trimester exposed and unexposed (HR around 1.5). However, we identified similar point estimates for the negative exposure controls, which spoke against a causal association. The results from the analysis on duration also spoke against a causal association, although if only first trimester exposure is associated with an increased risk, the duration-response analysis should have been limited to the first trimester. Unfortunately, we did not have an adequate sample size for such investigations.

In paper iii we saw that estimates still appeared to be influenced by confounding even though we used disease comparators. This illustrated the need for careful consideration regarding the choice of a comparator that could be more similar in terms of disease severity. These considerations were carried on to paper v, where we decided to use an active comparator design. This seemed promising, because bacterial infections in pregnancy should always be treated [32, 38].

### 4.2 Specific aim 2: Antibiotics and childhood cancer

**Paper iv** In the systematic review, we identified 36 studies on prenatal exposure to antibiotics and childhood cancer. Most studies (33 of 36) reported on a specific type of childhood cancer, rather than a composite endpoint of any childhood cancer. However, half of the studies that reported on a specific type of cancer (17 of 33 studies) did not specify what classification system they had used to classify the cancers. The source of exposure information was maternal retrospective recall in 24 of 36 studies; six studies investigated more than one exposure window (e.g. any time during pregnancy, and any time during the first trimester); three studied dose or duration of exposure. In five of the studies, adjustment was made for at least one potential intermediate factor in the main analysis. All studies used population comparators. Based on our findings, we provided three points for consideration in future studies, as summarised in Box 4.2.
4. Main findings

**Box 4.2: Three points for consideration in future studies on prenatal exposure to antibiotics and childhood cancer**

**Reporting of cancer types according to the ICCC when possible**
No known substance increases the risk of every type of cancer. Therefore, a composite outcome of *any childhood cancer* should be avoided. The International Classification of Childhood Cancer (ICCC) takes into account that childhood cancers do not respect the traditional sites used in the classification of adult cancer, so the ICCC should be used when possible.

**Biological plausibility should guide the exposure definition, and exposure windows**
Findings from *in vitro* or animal studies, and signals from previous human studies could aid in determining which exposures could be combined. The relevant exposure window for trans-placental carcinogenesis is unknown, so studies should consider using several exposure windows. Analyses by dose or duration of exposure are also helpful, as there may be threshold levels for effect.

**Use the new developments in confounder assessment and control**
Use of directed acyclic graphs could help ensure that adjustments are not made for intermediate factors. The measured confounders could be made into propensity scores, as the exposure will typically be more common than the very rare childhood cancer outcomes. To account for confounding by indication, studies could consider using disease comparators or active comparators.

The findings from studies on prenatal exposure to antibiotics are displayed in Table B.2. In 19 studies, the outcome was childhood leukaemia, and 11 studies found no difference in incidence between exposed and unexposed [105, 111, 112, 114, 119–121, 123–126], whereas eight studies found a higher incidence among exposed [89, 109, 110, 115, 117, 118, 122, 127]. Notably, three newer studies with large samples have suggested higher incidences of childhood leukaemia (estimates of association ranging from 1.2 to 1.6) after first trimester exposure to antibiotics [127], any prenatal exposure to penicillin [89], or any prenatal exposure to phenoxyethylpenicillin, pivampicillin, or nitrofurantoin [117]. For all other cancer types that have been investigated (central nervous system tumours in seven studies, neuroblastoma in five studies, renal tumours in four studies, lymphomas in three studies, liver tumours, eye tumours, and sarcomas each in two studies), at most a single study has identified a higher incidence among exposed.

In a general way, the three points for consideration all informed the design and reporting of paper v. More particularly, the findings from paper iv motivated the design of paper v in the following manner:
Specific aim 2: Antibiotics and childhood cancer

- We decided to focus on a specific type of cancer, rather than any cancer as a composite endpoint. Based on the findings from previous studies, we decided to focus on leukaemia.

- We decided to investigate the antibiotic substance nitrofurantoin, as it was the antibiotic substance that showed the strongest association with childhood leukaemia in newer studies (a HR of 1.6 [117]). Furthermore, nitrofurantoin is used for a very specific indication (urinary tract infection) that is common during pregnancy. Treatment guidelines in Scandinavia have pivmecillinam as the first line treatment and nitrofurantoin as an equivalent or second line treatment depending on country [202–204]. The one study that investigated leukaemia in children prenatally exposed to pivmecillinam did not find the incidence to be higher than the incidence in unexposed children [117]. Therefore, paper v became an active comparator study, comparing the incidence of childhood leukaemia among children prenatally exposed to nitrofurantoin and children prenatally exposed to pivmecillinam.

Paper ii In the descriptive cohort study including 172,585 pregnancies, 12.0% filled a prescription for an antibiotic in the three months before pregnancy. During pregnancy, 28.3% filled a prescription for an antibiotic. The proportion was quite similar in the first (11.5%), second (12.2%), and third trimester (13.6%). Throughout pregnancy, the five most commonly dispensed antibiotics in descending order of frequency were pivmecillinam, phenoxymethylpenicillin, amoxicillin, nitrofurantoin, and erythromycin (first and second trimester), or trimethoprim (third trimester) (Table B.1). In the three months after pregnancy, the proportion of prescription fills for an antibiotic rose to 18.4%.

Antibiotic prescription fills seemed in accordance with the national guidelines, and filling patterns appeared similar in the three trimesters. This contributed to the design of paper v in the following ways:

- We decided that an investigation of any antibiotic exposure throughout pregnancy would be meaningful to present along with the trimester-specific estimates.

- The adherence to guidelines and frequency of use supported our choice of nitrofurantoin and pivmecillinam as medications of interest in the active comparator design.

Paper v The analytic cohort study included 44,091 children prenatally exposed to nitrofurantoin and 247,306 children prenatally exposed to pivmecillinam. The children were followed for 9.3 years on average (standard deviation 4.1). During follow-up, 161 children were diagnosed with leukaemia (134 with lymphoid leukaemia).
4. Main findings

Any prenatal exposure to nitrofurantoin as compared to pivmecillinam was associated with a weighted IRR of 1.34 (95% CI 0.88 to 2.06), corresponding to an IRD of 15 cases per million person-years. The IRR did not appear higher when children exposed to two or more nitrofurantoin treatments were compared to children exposed to two or more pivmecillinam treatments (weighted IRR 1.57, 95% CI 0.54 to 4.55). However, the analysis was based on six cases exposed to nitrofurantoin. Regarding timing of exposure, first trimester (weighted IRR 1.92, 95% CI 0.84 to 4.41) and third trimester (weighted IRR 1.73, 95% CI 1.00 to 2.98), but not second trimester (weighted IRR 0.53, 95% CI 0.19 to 1.47), nitrofurantoin exposure was associated with higher incidence of leukaemia, albeit with confidence intervals including or overlapping the null.

The secondary analysis on leukaemia subtypes was only possible for lymphoid leukaemia, as there were too few cases of acute myeloid leukaemia. The findings were slightly attenuated for the trimester specific estimates, but the statistical precision was low.

Sensitivity analyses were generally in accordance with the main analysis. By calculating the e-value, we saw that the removal of unmeasured confounding of the magnitude 2.02 would be able to reduce the IRR to the null, but weaker confounding would not. Findings from the post hoc analysis were similar to the main analysis (weighted IRR 1.23, 95% CI 0.87 to 1.76), suggesting against important confounding by maternal indication for nitrofurantoin treatment.
Chapter 5

Discussion

5.1 Summary of main findings

In the first systematic review of this thesis, the most investigated analgesic in terms of child brain development was paracetamol (paper i). Findings were reassuring for most domains of brain development, but three studies from different cohorts found an association between prenatal exposure to paracetamol and child ADHD (paper i). The few studies on other analgesics were reassuring, but follow-up was short and diagnostic outcomes lacking, in particular for NSAIDs (paper i). Prescription fills for NSAIDs declined substantially over the course of pregnancy (paper ii). We therefore investigated the association between prenatal exposure to NSAIDs and child ADHD in a cohort study of 56,340 children, accounting for time-varying exposure, and confounding. We did not find any association between prenatal exposure to NSAIDs and child ADHD, regardless of timing or duration of exposure, when compared to children whose mothers had indications for NSAID use, but did not use NSAIDs (paper iii).

According to the second systematic review of this thesis, previous studies did not find any consistent associations between prenatal exposure to antibiotics and childhood cancers of any specific types, except for leukaemia (paper iv). Here the evidence was conflicting, but all studies had used population comparators (paper iv). Antibiotic prescription fills during pregnancy reflected national guidelines, with both pivmecillinam and nitrofurantoin in the top five most commonly filled antibiotics (paper ii). We therefore investigated the association between prenatal exposure to nitrofurantoin and childhood leukaemia in an active comparator study, comparing 44,091 prenatally nitrofurantoin exposed children to 247,306 prenatally pivmecillinam exposed children. We did not find any substantial association between prenatal exposure to nitrofurantoin and childhood leukaemia, although we observed a slightly elevated IRR with wide confidence intervals overlapping the null (weighted IRR 1.34, 95% CI 0.88;2.06, IRD 15 cases per million person-years). There was no evidence of a dose-response relationship (paper v).

5.2 Strengths and limitations

Strengths and limitations of the five studies in this thesis will be discussed separately for the two systematic reviews and the three cohort studies, before the generalisability of the findings will be discussed for all five studies.
5. Discussion

5.2.1 Systematic reviews

Strengths of the two reviews include comprehensive search strategies developed with the help from research librarians, searches completed in several databases supplemented with screening of the reference lists of included studies, and screening for eligibility by two independent reviewers.

Limitations include restriction according to publication language. In one of the reviews (paper i), no studies were excluded due to language restrictions, but in the other review (paper iv), five studies in Spanish or Chinese were excluded. For these studies, we were unable to assess eligibility in a full-text reading, thus we do not know whether they fulfilled the inclusion criteria for the review. Authors who do not have English as their first language are more likely to publish studies with negative findings in non-English journals [205]. However, it is unknown whether methods or reporting differ systematically between studies published in English and non-English journals.

Another limitation is the risk of publication bias. According to the Grading of Recommendations, Assessment, Development and Evaluations framework, the risk of publication bias is higher for observational studies than for randomised controlled trials [205]. In particular, reviews including small studies or studies based on routinely collected health data (which is the case for both reviews in this thesis) should consider the risk of publication bias as substantial [205].

Though not a limitation in the traditional sense of the word, the shear amount of material in both reviews, especially in the supplementary tables, could negatively affect readability. A narrower scope of the reviews could perhaps have made them more reader-friendly.

5.2.2 Cohort studies

Below, the strengths of the three cohort studies will be presented, followed by considerations regarding selection, exposure and outcome classification, confounding, and statistical methods. Some strengths and limitations apply to all the three cohort studies (strengths, exposure classification, selection). The sections on outcome classification, confounding, and statistical methods only apply to the two analytic studies. Selection problems have almost not been covered in this thesis so far, wherefore the section on selection problems will be comparatively long.

5.2.2.1 Strengths

The main strengths of the cohort studies include the population-based design (papers ii and v), complete follow-up for diagnostic outcomes (papers iii and v), and the focus on outcome classification (more on this in Section 5.2.2.4). Another strength is that we employed several methods to account for confounding (papers iii and v), some of which have not previously been used in studies on prenatal exposure to analgesics and child brain development (marginal structural models to account for time-varying confounding), or prenatal exposure to antibiotics and childhood cancer (active comparator design).
Strengths and limitations

5.2.2.2 Selection

The inclusion criteria differed across the three cohort studies, so the potential for selection problems also differed. Because an aim of paper ii was to investigate whether it was possible to identify pregnancy planning in routinely collected health data, the sample was restricted to pregnancies with information on the selected proxies for pregnancy planning (hormonal contraceptive use or fertility treatment). This meant that the study sample was younger and more likely to be nulliparous than the general birthing population of Norway, but the prevalence of chronic illness recorded in the MBRN was comparable. Given the selected sample, the prevalence of medication use that we identified in paper ii may not be generalisable to the entire pregnant population. However, our findings were largely similar to estimates from a population-based study in Norway [3].

Paper iii was affected by the selective participation in MoBa. All pregnant women in Norway were invited to participate during the study period, but only 41% consented to participation. The participants were less likely to be young parents, more likely to be married or cohabiting, and had a healthier lifestyle during pregnancy compared to the general birthing population [186]. Previous studies have shown that the selection into cohort studies does not appear to bias the estimates of association [186, 206].

Paper v was population-based.

Live-birth bias Both paper iii and paper v studied outcomes that are only observable in live-born children. Therefore, we restricted our samples to live-born children. However, there are differing opinions as to whether this can introduce bias [207–209]. The key difference of opinion is whether the long-term outcomes are conceptualised as determined during foetal life, or occurring after birth (note the distinction between occurrence and detection – it is universally agreed that the outcomes are only observable after birth). If the outcomes are determined during foetal life, stillbirths and abortions are competing risks and should be considered [207, 209]. If the outcomes do not occur until after birth, non-live births are not competing risks and should not introduce bias [208]. In addition to the outcome being determined during foetal life, the exposure also has to be a risk factor for non-live births for bias to occur [207]. Some studies have reported associations between NSAID use during pregnancy and spontaneous abortions, but it is unclear whether the associations are causal [64]. If NSAIDs are associated with non-live births, we cannot rule out that conditioning on live births could introduce bias from unmeasured factors associated with both foetal death and ADHD. The direction of such bias would be unpredictable. Probabilistic sensitivity analysis has been suggested to estimate the impact of potential live-birth bias [207]. However, one could argue that too many assumptions would be required to perform such analysis. Instead, the e-value [157] could be used to estimate how strong any unmeasured factors would have to be in order for the bias corrected estimate to reach the null (if bias is away from the null), or conversely to reach a pre-specified magnitude (if the bias is towards the null).
Though antibiotic use may lower the risk of foetal death compared to untreated infection, there is no evidence to suggest that the risk of foetal death will differ between nitrofurantoin and pivmecillinam exposed. Therefore, live-birth bias is unlikely to have had an impact on the findings from paper v.

**Potential unintended consequences of inclusion criteria** During the writing of this thesis, I have become aware of some additional selection factors in the two analytic studies. In paper iii, we required participants to have answered both of the MoBa pregnancy questionnaires with information on medication exposures. We considered it a necessity in order to know whether the child was truly exposed or unexposed. However, the second pregnancy questionnaire on medication exposures was sent to women around week 30 of pregnancy. The questionnaire did not give women an opportunity to reply that they have already given birth, and all questions related to pregnancy exposures, so it seems unlikely that women who gave birth before pregnancy week 30 would fill in the questionnaire. I have verified this assumption in the MoBa dataset combined with information from the MBRN. Indeed, only 1% of the women who answered the week-30-questionnaire gave birth before week 33 of pregnancy, whereas 10% of the women who did not answer the questionnaire gave birth before week 33. This means that paper iii, and several other papers using MoBa data to study medication safety in pregnancy [210–215], have unintentionally conditioned on extremely preterm birth (birth before week 28 [216]), and maybe also very preterm birth (birth between week 28 and 32 [216]). The incidence of ADHD appears to increase with decreasing gestational age [217, 218]. As NSAID use is associated with prolonged pregnancy [64, 70–72], preterm birth could be an intermediate factor. The pitfalls of adjusting for preterm birth are well-described [219]. It would be possible to perform a probabilistic sensitivity analysis to obtain estimates corrected for the selection under different assumptions [154]. However, we could also have limited the extent of the bias, at least for some of our analyses, by making different samples for the analyses on first trimester exposure and the analyses on second and third trimester exposure. For first trimester exposure, the requirement to answer the week 30 questionnaire was unnecessary as all baseline factors, and first trimester exposures were reported in the first pregnancy questionnaire.

In paper v, we excluded children who were exposed to both pivmecillinam and nitrofurantoin. Active comparator studies are recommended not to exclude individuals who switch medications, but rather assign the time at risk for the outcome to the medication that is currently being used [155]. However, this approach requires that the outcome can at least theoretically occur during the period with exposure. Another study applying active comparators to investigate medication safety in pregnancy also excluded children exposed to both medications of interest [101]. They did not discuss how this could have affected study findings [101]. Exposure to more than one type of antibiotic might be an indicator of disease severity, thus restricting our study to children exposed to less severe maternal infection. This would only be an advantage, if it made
the two groups more comparable in terms of disease severity. However, it cannot be ruled out that polytherapy is in fact an intermediate factor and that bias was introduced by conditioning on polytherapy.

**Loss to follow-up in MoBa** Whereas follow-up was complete for the diagnostic outcomes in papers iii and v, ADHD symptoms in paper iii were ascertained from parental report in a questionnaire filled in when the children were 5 years old. Approximately half of the eligible parents completed the questionnaire [185]. A study from MoBa showed that the loss to follow-up appeared to bias estimates of association for long-term outcomes such as ADHD [187]. The study also showed that inverse probability of censoring weights was a robust method to handle such bias [187]. Therefore, we used inverse probability of censoring weights in paper iii to reduce bias from loss to follow-up.

### 5.2.2.3 Exposure classification

Potential exposure misclassification is a limitation in all three cohort studies. As mentioned in Section 1.2.2.2, women may choose not to use filled prescriptions [136], or they may inaccurately recall and/or report use [137]. In one paper on analgesic use, we relied on prescription records (paper ii). In that paper, we only captured prescribed analgesics, which was in accordance with the aim of the study: to investigate patterns of prescription fills throughout pregnancy. However, as a measure of actual use of analgesics, prescription fills seem to be poor. In our paper, we found a prevalence of analgesic prescription fills in pregnancy of 6.5%, compared to a survey where 62.9% of participants from Northern Europe reported use of over-the-counter analgesics in pregnancy [4].

In the other paper on analgesic use (paper iii), exposure was reported by the mothers during and shortly after pregnancy. Here 6.2% of pregnancies were exposed to NSAIDs according to maternal self-report, and more than 95% of users reported using a medication that is available over-the-counter (mainly ibuprofen). To assess how prescription fills correspond to maternal self-report, I have linked the ADHD-diagnosis sample from paper iii to NorPD in an analysis completed specifically for this thesis synopsis. I identified all prescription fills for NSAIDs at any point during pregnancy. There were 953 women (1.7%) who filled a prescription for an NSAID during pregnancy. The majority of these (698 women) did not report NSAID use during pregnancy in the MoBa questionnaires. Among the women reporting NSAID use to the MoBa questionnaires, 7.2% had filled a prescription for an NSAID during pregnancy. Using MoBa as the gold standard, this yields a specificity of 98.7%, and a sensitivity of 7.2%.

Though better than prescription records, maternal self-report of medication use also has limitations. The longer the recall period for medication use, the higher the risk for misclassification [137]. A Dutch validation study compared maternal self-report on web-based questionnaires to medication diaries. Women were randomised to different timings of the web-based questionnaire relative to the completion of the medication diary. For analgesics, the study concluded that recall on the web-based questionnaires was sufficient if there were two
months or less between questionnaires [137]. In MoBa, this is not the case, so inaccurate recall is to be expected. In paper iii, we used probabilistic sensitivity analysis to obtain estimates of association corrected for exposure misclassification. We assumed non-differential misclassification, as 1) exposure and outcome information came from independent sources, 2) there were two exposure categories per trimester (exposed or unexposed), and 3) maternal report took place during pregnancy, and six months after birth, where child symptoms of ADHD are unlikely. According to the findings from the probabilistic sensitivity analysis, misclassification could not have masked an increased risk of ADHD after second or third trimester NSAID exposure. For first trimester exposure, the HR could have been as high as 1.5 in the absence of exposure misclassification. However, the misclassification corrected HR associated with the negative control, maternal NSAID use six months before pregnancy, would have been of the same magnitude.

In the Nordic countries, antibiotic prescription fills may be a good approximation of antibiotic use during pregnancy, because systemic antibiotics are not available over-the-counter. In addition, the Nordic prescription registries collect information on dispensed medications, which is a better approximation of medication use than information on prescribed medications [220]. However, women may still choose not to use the filled prescriptions [136]. In a Danish study of 170 filled prescriptions for antibiotics during pregnancy, only 108 were also reported as used in prospective biweekly questionnaires [136], resulting in a sensitivity of 93% and a specificity of 88%. Assessing the potential impact of exposure misclassification in an active comparator design requires more assumptions than in a traditional exposed/unexposed scenario. Even in the simplest scenario, assuming non-differential misclassification of exposure, we would have three exposure categories rather than two: nitrofurantoin exposed, pivmecillinam exposed, and children exposed to untreated maternal infection. This would give bias towards the null if we assume that 1) women are equally likely to avoid nitrofurantoin and pivmecillinam use during pregnancy, and 2) the risk of leukaemia is the same in children exposed to untreated maternal infection that should have been treated with nitrofurantoin, and children exposed to untreated maternal infection that should have been treated with pivmecillinam. This not only requires that the severity of infection is the same between the two groups, but also that the different pathogens that are sensitive to different types of antibiotics pose the same risk. If assumptions 1 and 2 are not met, bias towards the null cannot be assumed. The direction of the bias would depend on the strength of association between maternal untreated infection and childhood leukaemia, relative to the strengths of association between nitrofurantoin and childhood leukaemia, and pivmecillinam and childhood leukaemia, respectively. Hence, the direction of the resulting bias could be both towards and away from the null.
5.2.2.4 Outcome classification

As mentioned in Section 1.2.2.2, and further developed in papers i and iv, outcome classifications have been heterogeneous in the previous literature. In both analytic studies in this thesis, we focused on applying some lessons learned about outcome classification from the systematic reviews. Both studies used specific, rather than composite, diagnostic outcomes from routinely collected health data.

The validity of an ADHD diagnosis in the Norwegian Patient Registry has not been investigated, but in Denmark, where the health care system is comparable to the Norwegian, a recorded diagnosis of ADHD has a positive predictive value of 0.87 [221]. The validity of childhood leukaemia diagnoses has been investigated in the Finnish Cancer Registry, where the completeness was 96% [178]. To further improve the comparability of childhood leukaemia diagnoses between the countries, we recoded all cancer cases to the ICCC-3 based on the recorded information on cancer site, and type and behaviour of the cancer cells.

Given the findings from the Finnish validation study, the severity of childhood leukaemia, and the mandatory reporting to the Nordic cancer registries, including from autopsies, it seems fair to assume that any detection bias for childhood leukaemia would be of negligible magnitude. For child ADHD, a complete detection seems less likely. Detection bias could occur if women with a higher health care utilisation are more likely to both use NSAIDs and have their children examined for ADHD. We therefore used a second source for ADHD outcome assessment, namely parental report of child ADHD-symptoms according to a well-validated psychometric instrument. Here differential misclassification of the outcome could again occur if underlying factors affect maternal report of both exposure and outcome [95]. However, this bias structure could be slightly different from detection bias, so comparing the two different sources was a strength of our paper. Furthermore, we found a good correspondence between ADHD diagnosis, and the parent-reported symptom score.

5.2.2.5 Confounding

We used several methods to reduce confounding in papers iii and v. In both papers, potential covariates were entered into directed acyclic graphs to make our assumptions explicit and avoid adjustment for potential intermediate factors. The measured confounders were combined into propensity scores in both papers, as we a priori expected the exposures to be more common than the outcomes. In paper iii, we had the opportunity to adjust for time-varying confounding using marginal structural models. Furthermore, we used disease comparators. In paper v, we used active comparators to account for the underlying maternal condition. To estimate the impact of potential unmeasured confounding, we used negative exposure controls in paper iii. In paper v, we calculated the e-value, and estimated the associations when using a population comparator instead of the active comparator. The negative exposure controls in paper iii indicated the presence of unmeasured confounding. In paper v, the post hoc analysis using
unexposed comparators suggested against important confounding by maternal indication for nitrofurantoin treatment. However, residual confounding of the study findings cannot be ruled out. The e-value was 2.02, meaning that an unmeasured confounder with an association of more than 2 with the outcome would be required to explain the association. This seems unlikely given the few known strong risk factors for childhood leukaemia. However, the combined effects of several smaller confounders should be considered. This points to a limitation of the e-value. It does not account for combined effects of several unmeasured confounders, and could therefore provide false security about research findings. A limitation of the negative controls that we used in paper iii is that we required NSAID use before pregnancy and no NSAID use during pregnancy. This involves conditioning on a variable that is affected by the exposure (in this analysis NSAID use before pregnancy) which could have introduced bias. A better negative control in this scenario might have been NSAID use by the mother’s partner, while the mother was pregnant. We did not have access to this information. Instead we had information on the partner’s use of NSAIDs before pregnancy, which coincides with the period of spermatogenesis for male partners. Therefore, we did not consider it as a truly negative control. In paper v, we were unable to identify an appropriate negative exposure control, as treatment guidelines differ for pregnant and non-pregnant women [202–204], so the underlying bias structures might also differ. Furthermore, it was unclear whether pivmecillinam and nitrofurantoin would also be appropriate active comparators outside of pregnancy. Use of inappropriate active comparators may introduce more bias than they remove [222].

It would have strengthened the inference from our studies, if we had been able to add analyses with different bias structures, such as sibling analysis or cross-context comparisons [10]. Unfortunately, this was not feasible.

### 5.2.2.6 Statistical methods

In both analytic cohort studies, children could participate with different lengths of follow-up. Therefore, we used statistical models that can account for varying follow-up, but at the cost of assuming that time to diagnosis will be affected by the exposure. New developments within biostatistics suggest that future studies should perhaps model associations between prenatal exposures and childhood outcomes differently [223]. The rationale is that the exposure no longer has an effect after the child is born, and therefore does not necessarily affect the time to diagnosis [223]. Rather, if the exposure has an effect, it will be to change the child’s susceptibility to the outcome of interest already at birth, even though the outcome is not observable until later in life [223]. The proposed statistical model for this situation as called a cure model [223]. In the cure model, there is a logistic part that models child susceptibility, and a survival part that models time to diagnosis [223]. This allows for the time to diagnosis to be independent of exposure [223].

However, from a pragmatic point of view, the outcomes of interest in both papers iii and v are so rare that that the estimates of association are unlikely to
have differed substantially between Cox and cure models [223], or even between Poisson and logistic models [224]. Yet, the low incidence of the outcomes also meant that both papers iii and v encountered problems with the statistical precision, especially in the analyses on dose or duration of treatment. This happened even though paper v started out with a larger cohort than any of the identified previous studies of prenatal exposure to antibiotics and childhood cancer. Experiences from papers iii and v highlight the need for meta-analyses, which again requires that individual studies are conducted and reported in a more uniform manner, as suggested in papers i and iv.

5.2.3 Generalisability
Due to the systematic reviews, the findings from this thesis are based on data from many different countries and data sources. Nonetheless, most research was done in high-income countries, and no studies from African countries were identified. Both the genetic composition of the included children, and the confounding structures may differ between countries where research has been done, and countries where research has not been done. For example, analgesic use was associated with lower maternal education in the MoBa (paper iii), but with higher maternal education in a Brazilian cohort [225, 226]. Therefore, findings from this thesis may not be generalisable to other contexts.

An additional problem specific to long-term outcomes is that the longer the follow-up, the further the included pregnancies will be removed from contemporary practice [11]. The oldest children included in papers iii and v were born in the late 1990’s. Both guidelines for medication use and actual patterns of use may have changed since then.

5.3 Interpretation
In the following sections, I will use the pragmatic pluralism approach to causal inference, as mentioned in Section 1.2.4, to give an interpretation of the thesis findings. This will include looking at the thesis findings in light of evidence from other studies (published after the completion of the systematic reviews), evidence from in vitro and animal models (interlocking evidence), and evidence from negative controls, sibling studies, and different cultural contexts (triangulation). The interpretation will be presented separately for the two specific aims of the thesis.

5.3.1 Specific aim 1: Analgesics and child brain development
Since the completion of the systematic review on analgesics and brain development, several studies have been published [214, 225–233]. With the exception of one study on analgesic opioids [234], all of the studies investigated prenatal exposure to paracetamol. Two studies investigated prenatal exposure to NSAIDs in addition [231, 232].
5. Discussion

5.3.1.1 Paracetamol

Prenatal exposure to paracetamol does not appear to affect child cognitive abilities [141, 145, 214, 225, 226, 229, 233, 235–238], though three studies from the same cohort identified more communication problems among exposed children at preschool age [149, 212, 215]. However, the association between prenatal paracetamol exposure and communication problems was not found among older children in the same cohort [214].

Fewer studies have investigated emotional problems [149, 214, 215, 229, 230]. One study using both parent and child report identified more emotional problems among exposed children [230], but other studies using parent report [149, 214, 215], or parent, teacher, and child report [229] did not. To date, studies using diagnostic outcomes are lacking.

The majority of studies have focused on behavioural problems, in particular ADHD. In the current body of evidence, all studies using diagnostic outcomes have identified higher incidence of ADHD after prenatal paracetamol exposure, in particular long-term exposure [142, 144, 227, 228, 239]. Among the studies using parent report, half found more behavioural problems in exposed children [149, 150, 214, 229, 230, 232], whereas the other half did not [146, 148, 215, 226, 231, 240, 241]. Notably, both studies using child report identified more behavioural problems among exposed [148, 230], whereas the majority of studies using teacher report did not find more behavioural problems among exposed [145, 146, 229, 231]. A difference between the studies that found associations between prenatal paracetamol exposure and behavioural problems, and studies that did not, lies in the presence of a measure of long-term or high-dose paracetamol use. Of 13 estimates pointing to more problems among exposed, nine were from studies that had this information [142, 144, 149, 214, 227, 228, 230, 232, 239], whereas only four of nine estimates pointing to no difference between exposed and unexposed were from studies with information on dose or duration [145, 146, 215, 231].

In general, the studies on prenatal exposure to paracetamol and child behavioural problems conditioned on several important confounders. However, few studies have had access to information on maternal ADHD diagnosis or symptoms. In a systematic review on prenatal exposure to paracetamol and child ADHD diagnosis or symptoms, probabilistic sensitivity analysis was applied to all identified studies [242]. The findings illustrated how adjustment for parental ADHD alone, or in combination with maternal migraine could explain away all associations between prenatal paracetamol exposure and child ADHD [242].

Contrarily, animal studies support that prenatal exposure to paracetamol, especially in high therapeutic doses, could affect offspring brain development [82, 243]. In addition, an epigenetic study in human cord-blood found that short-term prenatal exposure to paracetamol was not associated with epigenetic changes [244]. Long-term exposure was associated with changes in the gene expression of genes linked to ADHD among children who were later diagnosed with ADHD, but not among long-term exposed children who remained undiagnosed [244]. This could suggest that some individuals are more vulnerable to long-term prenatal
paracetamol exposure. Hence, paracetamol exposure alone is not sufficient to cause ADHD. A study with brain imaging diagnostics in children at 9 to 11 years of age found a difference in brain connectivity between children prenatally exposed and unexposed to paracetamol, and linked the brain connectivity to ADHD symptoms [227]. Both the epigenetic study and the brain imaging study were limited by small sample sizes, and could not adjust for parental ADHD [227, 244].

From the point of view of triangulation, the different results from different observers (e.g. parents and teachers) are interesting, as the methods have different strengths and weaknesses (paper i). It has been argued that studies using maternal report of both exposure and outcome could be affected by misclassification bias away from the null because personality traits may affect reporting of both exposure and outcome [95]. Some authors argue that this could also be the case for studies using diagnostic outcomes, if maternal personality traits affect the likelihood to seek medical care for child behavioural problems [231]. Whereas child report could in theory also be affected by inherited personality traits, teacher report should be independent. However, teachers observe children in one setting only, whereas problems in more than one setting is among the diagnostic criteria for ADHD [245, 246].

Other methods for triangulation used in the literature on paracetamol and child behaviour include negative controls, used by several studies [144, 149, 150, 214, 228, 231, 239, 241]. In some of the studies, the negative controls were associated with more behavioural problems or higher incidence of ADHD [144, 214, 228]. This could suggest that the main findings in these studies are biased away from the null. However, as mentioned in Section 5.2.2.5, some negative controls are perhaps not ideal. The negative controls in the studies that found associations between the negative control and the outcome were maternal paracetamol use before but not during pregnancy, and paternal paracetamol use during spermatogenesis [144, 214, 228].

Only one study used sibling design [149], and this study still found an association between prenatal exposure to paracetamol and child behavioural problems.

Most studies were conducted in high-income countries, but two studies were from the Brazilian Pelotas cohort [226, 241] that has previously been used as a cross-cultural comparison [225, 247]. Neither of the studies from this cohort with a potentially different confounding structure found more behavioural problems in children prenatally exposed to paracetamol than in unexposed children [226, 241]. However, the studies did not have information on duration of treatment.

To sum up the evidence regarding paracetamol, short-term or low cumulative dose prenatal exposure does not seem to affect child brain development. Long-term exposure is associated with child behavioural problems, mainly ADHD. However, more research is needed in order to understand whether the associations are causal. To strengthen the causal inference, future studies on prenatal exposure to paracetamol and child behaviour should provide analyses by dose or duration of treatment. Further, use of sibling design, preferably in studies using diagnostic outcomes, and cross-cultural comparisons would be beneficial. Finally,
5. Discussion

epigenetic studies and studies using brain imaging techniques could improve our understanding, provided that they attempt to adjust for genetic confounding.

5.3.1.2 Analgesic opioids

The evidence regarding analgesic opioids is very scarce. Two studies from the same cohort found no evidence of more problems with language and communication among exposed children [212, 234]. One study investigated diagnosis of autism and found a higher incidence in exposed although with confidence intervals overlapping the null [143]. The estimate of association appeared higher for maternal opioid use before pregnancy, a negative exposure control, but the statistical precision was limited [143]. Given this evidence, it seems premature to make any conclusions regarding prenatal exposure to analgesic opioids and child brain development. Though most animal studies have investigated medicines used in opioid maintenance treatment rather than analgesic opioids [248], animal studies may still point to areas of interest for future investigation. These areas include memory [248, 249], anxiety or depression [249], and hyperactivity [248]. Future epidemiologic studies should also carefully consider the choice of comparators, seeing that only pregnant women with severe pain will be prescribed analgesic opioids. Population comparators may therefore be inappropriate.

5.3.1.3 NSAIDs

Based on the few studies available, prenatal exposure to NSAIDs does not appear to affect child cognitive abilities [147, 149]. No studies have investigated child emotionality. As for behaviour, one study using parent and teacher report found a small increase in problems with executive functioning, corresponding to one fifth of a standard deviation, and overall problem behaviour, corresponding to one fourth of a standard deviation [232]. Two other studies using parent and teacher report did not identify more behavioural problems among exposed [147, 231], and the only study to investigate child ADHD diagnosis did not identify a higher incidence among exposed either (paper iii). The study that found an association between prenatal NSAID exposure and child behavioural problems accounted for many important potential confounders, but did not have access to other information on indication for NSAID use than co-medication with antibiotics as a proxy for infection [232]. Furthermore, the study accounted for maternal depressive symptoms and use of antidepressants, but not maternal behavioural problems or ADHD. Other studies accounted for indication (paper iii), [231], or adjusted for all maternal psychotropic co-medication, and for maternal ADHD (paper iii), or other maternal mental health disorders including behavioural problems [147]. Information on dose or duration of treatment was available in two of the studies that did not identify more behavioural problems among the exposed (paper iii), [147], as well as in the study that did find more behavioural problems among the exposed [232].
A study in mice supports that prenatal exposure to NSAID, even in high therapeutic doses, does not affect brain development [250].

From a triangulation point of view, two studies used negative controls (paper iii), [147]. In one study the negative exposure control was in fact associated with child ADHD (paper iii), whereas it was not associated with behavioural problems in the other study [147]. One of the studies also used a negative outcome control and this was associated with prenatal exposure to NSAIDs [147], so bias away from the null seemed likely in both studies. One study used sibling design and did not identify more behavioural problems among exposed [149].

In summary, based on current evidence, prenatal NSAID exposure does not seem to affect child cognition and behaviour. We do not have information on emotionality. Studies from other cultural contexts would add to our understanding, and the findings regarding ADHD diagnosis should be replicated, preferably using sibling design where the bias structure is expected to differ from the one in our study [10]. However, seeing that the previous estimates of association have been close to the null, sources for bias towards the null should be considered in future studies. This could for instance be through probabilistic sensitivity analysis to correct for misclassification of exposure, as we did in paper iii. More focus on analyses by dose or duration could also potentially reduce misclassification, because women with a single episode of NSAID use during pregnancy may be more likely to forget the exposure than women with long-term use [136, 137].

5.3.1.4 Acetylsalicylic acid

The findings regarding acetylsalicylic acid either stem from studies that are more than 30 years old [251–253], or studies where acetylsalicylic acid was considered a negative exposure control to the main exposure [148, 240]. Although the findings were reassuring, the relevance to current clinical practice is questionable. Today, analgesic aspirin is rarely used during pregnancy, whereas low-dose aspirin use is being explored for the prevention of preeclampsia, and gestational hypertension [64].

5.3.1.5 Triptans

All studies that investigated prenatal exposure to triptans compared to prenatal exposure to untreated migraine were reassuring [254–257]. However, a limitation in the current body of literature is that all four studies using disease comparators were from the same cohort [254–257]. Triptans are serotonin receptor agonists, and a review of evidence from in vitro, animal, and human studies mentions the potential implication in child autism, ADHD, and depression based on theoretical knowledge of the role of serotonin [258]. However, the review did not include any animal studies on triptans [258], nor have I been able to identify any for this thesis. Yet future studies may still want to start with an investigation of the mentioned outcomes because of their biological plausibility to be affected by
5. Discussion

changes in serotonin levels. Furthermore, the use of disease comparators should be considered.

5.3.2 Specific aim 2: Antibiotics and childhood cancer

5.3.2.1 Any antibiotics and childhood cancer

No new studies on prenatal exposure to antibiotics and childhood cancer were identified when updating the literature search for this thesis synopsis.

The previous literature has identified several signals of carcinogenicity for prenatal exposure to antibiotics when compared to unexposed children. However, only the signals for leukaemia have been replicated, although in particular central nervous system tumours and neuroblastoma have also been investigated in several studies. Among the studies that did identify a higher incidence of leukaemia after prenatal exposure to antibiotics, three studies had prospective exposure ascertainment, adjustment for important potential confounders, and no adjustment for potential intermediate factors [117, 122, 127]. None of the studies that did not identify a higher incidence of leukaemia after prenatal antibiotics exposure fulfilled all these criteria. The main limitation in all the studies is that they compared children exposed to maternal infection and antibiotic treatment to children unexposed to both. Therefore, it cannot be concluded whether the infection, the antibiotic, the genetic susceptibility to infections, or all these in combination are responsible for the higher incidence of childhood leukaemia among exposed.

A study in mice concluded that early life antibiotic treatment appeared to induce leukaemia in genetically predisposed mice [259]. The mice were exposed to several different antibiotic medications at once (ampicillin, metronidazole, vancomycin, ciprofloxacin, and imipenem) over a period of 8 weeks. This is hardly comparable to the scenario of pregnancy exposure, as most women are only treated once during pregnancy (paper v), corresponding to one week of exposure [202–204].

None of the studies on prenatal exposure to antibiotics and childhood leukaemia used any form of triangulation.

To sum up the evidence, it cannot be ruled out that prenatal exposure to maternal bacterial infection and antibiotic treatment increases the risk of childhood leukaemia. One way to attempt to disentangle the effect of antibiotics from the effect of the underlying infection in future studies, could be to investigate the risk associated with antibiotic treatment in women without infection. This approach has previously been used in a study on prenatal exposure to antibiotics and childhood obesity [113]. In some countries, pregnant women are routinely screened for colonisation with group B streptococcus during pregnancy and offered treatment with prophylactic antibiotics [113]. In this setting, comparing treated and untreated women will remove any effect of infection.

If antibiotic use during pregnancy does in fact induce childhood leukaemia among genetically susceptible individuals, a pertinent question would be: which antibiotic is the safest? To answer that question, the active comparator design
that we used in paper v is ideal. If the aim is a broader screening of which antibiotics are the safest to use when it comes to risk of specific types of childhood cancer in addition to leukaemia, it may be useful to mimic a design from malformations research: a predefined number of the most commonly prescribed antibiotics could be compared to penicillins that are not considered to be associated with increased risk [99]. The study could have a hypothesis generating approach [260], and investigate for example the ten most common ICC-3-diagnoses. Signals would then have to be followed up in other cohorts.

5.3.2.2 Nitrofurantoin and childhood leukaemia

The literature review did not identify any other studies using active comparator design, so we do not have much wherewith to compare our findings from paper v. Sometimes, as in the case of diethylstilboestrol, one study is sufficient for causal inference. However, in that case, there was a pattern of a specific rare type of cancer that was strongly associated with the exposure [56]. The results from paper v are far from equally conclusive.

From studies on immediate birth outcomes, we have seen indications that nitrofurantoin can cross the placenta, and affect the developing foetus [100]. However, our findings that nitrofurantoin is not substantially associated with childhood leukaemia are supported by animal studies. Though few studies have investigated the trans-placental carcinogenicity of nitrofurantoin in animals, there is no evidence that exposure to nitrofurantoin at any point during the animal’s lifespan causes leukaemia, even if the dosage is high [261].

The limited statistical precision, absence of a dose-response relationship, and lack of a clear biological mechanism of action, suggests against a causal interpretation of the elevated IRR from our study. To come closer to an understanding about whether prenatal exposure to nitrofurantoin can be ruled out as a cause of childhood leukaemia, our findings need replication. In particular, it would be useful if future studies could provide additional data on childhood leukaemia incidence among children with prenatal exposure to more than one nitrofurantoin treatment, as our findings were based on only six exposed cases. An validity concern in our study that should be remedied in future studies was exposure misclassification (most likely non-differential, but the direction of the resulting bias could be unpredictable, as discussed in Section 5.2.2.3). As a potential source of bias towards the null, this is important to take into account before concluding that there is no association between prenatal nitrofurantoin exposure and childhood leukaemia. Substantially reducing exposure misclassification will be difficult in registry-based studies, as information on actual medication use is typically not available. Probabilistic sensitivity analyses could be applied in future studies, but this would require several assumptions if applied to an active comparator design, as it appears unfeasible to assess the risk associated with untreated infection. This is because all bacterial infections during pregnancy should be treated [32, 38], so the amount of untreated women should be very low. However, findings from our
study suggested against important associations between the maternal indication for nitrofurantoin treatment and childhood leukaemia.

Birth cohort studies could perhaps collect more accurate information on medication use, but a sufficient sample size could hardly be obtained, even by combining data from different birth cohorts. In paper v, we ended up with very imprecise estimates, even though our study sample started out with more than three million live-born children. In comparison, a study combining data from the two largest birth cohorts, MoBa and the Danish National Birth Cohort, included 185,617 live-born children with information on medication exposure [262].

5.4 Implications for practice

Overall, the findings from this thesis support the appropriateness of existing guidelines on use of analgesics and antibiotics in pregnancy.

It is recommended that analgesics should be used in the lowest effective dose for the shortest possible time with paracetamol as first line treatment [263, 264]. When long-term treatment is needed, the pregnant woman should consult a physician [264]. For some types of pain, there are non-pharmacologic methods of pain relief that should be attempted before resorting to analgesics [263]. The non-pharmacologic methods include rest, hot or cold compresses, massage, and physiotherapy [263].

Specifically regarding NSAIDs, the findings from this thesis (papers i and iii) are reassuring with regards to child brain development for women who need to use NSAIDs during pregnancy.

Concerning antibiotics, guidelines recommend that infections in pregnancy or asymptomatic bacteriuria should be treated with antibiotics [32, 38, 202]. If the bacteria are sensitive to penicillin, this is the first line treatment [38, 202].

Even if prenatal exposure to antibiotics does increase the risk of childhood leukaemia (which cannot be ruled out based on the findings from paper iv), this should not lead to a change in the guidelines that bacterial infection during pregnancy should always be treated due to the immediate risks to mother and foetus from an untreated infection [32, 38]. It would, however, be yet another reason to restrict antibiotic use to situations with an indication. For example, pregnant women can be tested and treated for group B streptococcus intrapartum rather than receiving prophylactic treatment during pregnancy as they do in some countries outside of Scandinavia. The rationale for intrapartum screening is that up to 25% of women who are screened positive for group B streptococcus during the third trimester of pregnancy will no longer have group B streptococcus colonisation at the time of delivery if left untreated [265].

Regarding nitrofurantoin, our findings from paper v are reassuring, but need replication. There are several reasons for this, including low statistical precision, and risk of bias. In the meantime, it is possible that the elevated estimates of association from our study or previous studies may cause worry among pregnant women. Pregnant women stating concerns about nitrofurantoin use during pregnancy should be reassured that it is unlikely that nitrofurantoin
causes childhood leukaemia. Normally, if a medication does increase the risk of cancer, we would expect that higher doses of the medication would give higher cancer risk, but the studies did not find this for nitrofurantoin. In addition, the medication has been tested in animals and even very high doses did not cause leukaemia. Pregnant women should be encouraged to take antibiotics as prescribed, seeing that antibiotic treatment of bacteria in the urine prevents 1 preterm birth for every 9 pregnant women treated, and 1 case of maternal kidney infection for every 7 pregnant women treated [32]. A detailed discussion on how to communicate risks to pregnant women is outside the scope of this thesis. For a short introduction to the topic and several relevant references, please refer to Widnes and Schjøtt [266].

5.5 Implications for research

Sections 5.3.1 and 5.3.2 touched upon several specific implications for research. In this section, I will consider some more general implications.

Papers i and iv provided a number of suggestions for future studies on prenatal exposure to medications and child brain development and cancer, respectively. These suggestions were motivated by the current difficulty in performing meta-analyses for both child brain development and childhood cancer due to the heterogeneity in outcome measurement, and for childhood cancer also exposure definitions. The general suggestions for future research based on the present thesis can be synthesised to the following:

- **Brain development is not just brain development; cancer is not just cancer.** Child brain development consists of several domains that all need to be investigated before we can conclude that a medication does not affect child brain development. Childhood cancer can be considered as several different diseases, so a composite outcome of any childhood cancer should be avoided. For both brain development and childhood cancer, the reporting of outcome classification and validity could be improved.

- **Different sources of outcome ascertainment should complement each other.** This suggestion pertains to child brain development only, where the use of several different sources for outcome ascertainment (diagnoses, clinical tests, parent -, teacher -, and child report) can be used as a method for triangulation.

- **The outcome measure should be age appropriate.** For child brain development, some psychometric instruments are only validated for use in children of specific age groups. For both brain development and childhood cancer, studies should have adequate follow-up for the child to develop the disease in question. For example, the mean age at first ADHD diagnosis in Norway is 8.6 years [245] and malignant bone tumours typically occur among teenagers [267], so follow-up for these two conditions in preschool children would be inadequate.
5. Discussion

- **Biological plausibility should guide the choice of outcomes and exposure definitions.** For both child brain development and childhood cancer, findings from animal or *in vitro* studies, as well as signals from previous studies in humans, could point to potential signals worth investigating. For both outcomes, the relevant exposure windows are uncertain, so it may be beneficial to investigate several exposure windows. Dose-response or duration-response analyses should be done when possible, as they may help in the interpretation of the biological plausibility of the findings.

- **Use the new methods for confounder assessment and control.** Adjustment for potential intermediate factors was seen in both reviews (papers i and iv). Use of directed acyclic graphs may help avoid this potential source of bias. Studies should consider whether sibling comparators (particularly for studies on analgesic use and risk of hereditary conditions such as ADHD or autism), disease comparators, or active comparators may be relevant to use instead of, or in addition to, population comparators, both in order to account for potential confounding and in order to increase the clinical relevance.
Chapter 6

Conclusion

In two systematic reviews and three cohort studies, this thesis has investigated selected outcomes in childhood after prenatal exposure to the two most commonly used medication groups in pregnancy: analgesics and antibiotics. For analgesics, the focus was on child brain development; for antibiotics, the focus was on childhood cancers. We did not find evidence to suggest any associations between short-term prenatal exposure to paracetamol, NSAIDs, analgesic opioids, or triptans, and child adverse brain development. For paracetamol, the systematic review suggested an association between long-term use and child behavioural problems, in particular ADHD. However, more research is needed in order to understand whether the association is causal. For NSAIDs, based on findings from our cohort study, there did not appear to be a dose-response relationship for ADHD, but this needs to be confirmed in future studies.

For prenatal exposure to antibiotics, the systematic review did not reveal any evidence of associations with other types of childhood cancer, but an association with childhood leukaemia cannot be ruled out. Studies accounting for maternal infection are needed to guide causal inference. In our cohort study, prenatal nitrofurantoin exposure did not appear to be substantially associated with childhood leukaemia when compared to pivmecillinam, although we identified a slightly higher incidence among nitrofurantoin exposed. However, the confidence intervals overlapped the null, and there was no evidence of a dose-response relationship. Taken together with the lack of a biological mechanism to explain how prenatal exposure to nitrofurantoin should increase the incidence of childhood leukaemia, this suggests against a causal interpretation. Yet, further studies are needed to provide additional data, as the statistical precision was limited.

Overall, the findings from this thesis support the appropriateness of the existing guidelines on use of analgesics and antibiotics in pregnancy.

The thesis also investigated how to limit potential biases when investigating the above-mentioned associations using existing data. Here suggestions relate to the choice of outcome classification, use of evidence from in vitro and animal studies, and use of new methods in confounding control. This thesis provides examples of the use of several new methods in confounding control, including marginal structural models, active comparators, negative controls, and calculation of the e-value.

Both when designing studies, and when discussing medication use with pregnant women in the clinical setting, it is necessary to consider the risk of untreated illness, as well as the risk associated with medication use.
Chapter 7

Perspectives

This thesis has focused almost exclusively on child outcomes after medication use during pregnancy. However, in order to perform an adequate risk-benefit evaluation, the risks and benefits for the mother from taking medications during pregnancy also needs investigation. For many medications, little evidence exists about the effectiveness when used during pregnancy [268], where the drug metabolism and blood volume differs from non-pregnant women [14]. Especially for analgesics, there is little focus on the consequences for the mother associated with untreated pain conditions in pregnancy [46].

Another perspective that has not been covered in this thesis, is the development towards including pregnant women in randomised controlled trials [268]. If pregnant women were to be increasingly included in randomised controlled trials, this could prove a valuable source of evidence, relatively unbiased by confounding. Because loss to follow-up is very likely over the course of the number of years required to study long-term child outcomes, it would be highly beneficial to conduct the trials in countries with nationwide registries and obtain participant consent for themselves and their children to be followed throughout childhood using registry linkage.
Papers
Paper I

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

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

*PloS One*, 2019, volume 14, issue 7, pp. e0219778. DOI: 10.1371/journal.pone.0219778.
Use and validity of child neurodevelopment outcome measures in studies on prenatal exposure to psychotropic and analgesic medications – A systematic review

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

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

Abstract

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

Review registration number: CRD42018086101 in the PROSPERO database.

Introduction

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

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

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

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

Independent of which outcome measures are used, these should be reliable and valid in order for research data to be of value. Evaluations of reliability and validity are important for both medical diagnoses [17] and psychometric instruments [18,19], as both methods to some extent rely on subjective assessments [20]. Reliability is the degree to which the assessment is free from measurement error [18] (see also S1 File for definitions of different types of validity and reliability). Low reliability in an outcome measure mainly introduces random error and can therefore affect the precision of the results. This is particularly problematic in studies with
smaller sample sizes. Validity is defined as the extent to which the outcome measure truly measures the construct (e.g. shyness) it is intended to measure [19]. An outcome measure that is not valid in the context where it is used, will give systematically erroneous results, which is problematic regardless of sample size.

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

Methods

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

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

- Motor skills: Including ICD-10 code F82, specific developmental disorder of motor skills [24]
- Cognition: Including ICD-10 codes F70-79, mental retardation, F80, specific developmental disorder of speech and language, F81, specific developmental disorder of scholastic skills, and F84, pervasive developmental disorders (ASD)
- Behaviour: Including ICD-10 codes F90, hyperkinetic disorders, and F91, conduct disorders
- Emotionality: Including ICD-10 codes F30-39, mood disorders, F40-49, neurotic, stress-related and somatoform disorders (including anxiety), and F93, emotional disorders with onset specific to childhood
• Other: Including sleep disorders (ICD-10 code F51), and tic disorders (ICD-10 code F95)

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

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

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

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

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

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

Antidepressants

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

Anxiolytics

Twenty papers studied neurodevelopment after prenatal exposure to anxiolytics [35,38,40,56,69,79–81,87,92,94,104–112], with ten of the papers published ten or more years ago. Follow-up was available from two months to 15 years (S2 Table). Most papers had used assessment from a health care professional, and only one paper had investigated risk of psychiatric diagnosis [36]. Most work has been done on the cognitive domain; the least has been done on emotionality (Fig 2).

Antipsychotics

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

Six papers studied hypnotics and sedatives [94,104,109,115–117], and three of these papers looked at overdoses taken for suicide attempts. Follow-up was available from 1 year to 5 years.
Three papers did not specify age at follow-up, but other papers from the same study have assessed toddlers. Most papers, four of six, had used psychometric instruments completed by health care professionals (Fig 2). Cognition was studied in four, and behaviour in five papers. One paper investigated motor skills [108], none studied emotionality.

**Paracetamol**

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

**NSAIDs**

Five papers investigated NSAIDs [120,133–136]. Three of the papers were more than ten years old. Children were followed from 6 months to 6 years. Three papers studied exposure to indomethacin for treatment of preterm contractions, so the populations consisted of a higher proportion of children born prematurely than the general population. Four papers reported...
assessments by health care professionals using psychometric instruments, two papers used parent reporting, and one teacher reporting [135]. The papers covered motor skills, cognition and behaviour domains of neurodevelopment, but not emotionality (Fig 3).

**Acetylsalicylic acid**

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

**Triptans**

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

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

Validity and reliability of neurodevelopmental outcome measures

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

Antidepressants

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

Anxiolytics

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

Antipsychotics

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

Hypnotics and sedatives

Half of the six papers on hypnotics and sedative commented on validity and/or reliability of at least one of their outcome measures. No type of validity or reliability was mentioned in more than one paper. None of the papers mentioned content validity, inter-rater–or test-retest reliability.
### Table 1. Validity and reliability of outcome measures as reported by the authors of the study, papers on antidepressants.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Outcome</th>
<th>Psychometric properties of outcome measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assessment using psychometric instruments</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Infant (&lt;2 years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suri 2011 [99]</td>
<td>BNIBSAS</td>
<td>Assessors were trained and certified to 0.90 reliability</td>
</tr>
<tr>
<td>Mortensen 2003 [81]</td>
<td>Boel test</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Batton 2013 [44]</td>
<td>Bayley Infant Neurodevelopment Screener</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Weikum 2013a [101]</td>
<td>BSID, unspecified edition</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Gustafsson 2018 [61]</td>
<td>BSID-II</td>
<td>Administrators had completed several practice administrations with repeated feedback prior to engaging in study administration of the BSID</td>
</tr>
<tr>
<td>Oberlander 2004 [67]</td>
<td>BSID-II</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Reeyeby 2002 [92]</td>
<td>BSID-II</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Reeyeby 2012 [38]</td>
<td>BSID-II</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Santucci 2014 [93]</td>
<td>BSID-II</td>
<td>Good reliability for infants from 1 to 42 months</td>
</tr>
<tr>
<td>Austin 2013 [43]</td>
<td>BSID-III</td>
<td>BSID is considered gold standard. Good reliability and internal consistency</td>
</tr>
<tr>
<td>Hanley 2013 [64]</td>
<td>BSID-III</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Heikkinen 2002 [67]</td>
<td>Gesell Development scales</td>
<td>Psychometric properties not mentioned (Comment from review authors: Not mentioned which psychometric instrument is used, we have taken this information from the 2003 paper)</td>
</tr>
<tr>
<td>Heikkinen 2003 [16]</td>
<td>Gesell Development scales</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Johnson 2012 [73]</td>
<td>Infant Neurological International Battery</td>
<td>Interrater reliability 0.97, test-retest 0.95</td>
</tr>
<tr>
<td>de Vries 2013 [53]</td>
<td>Psychomotor assessment according to Prechtl</td>
<td>Interrater reliability of general movement assessment is high (89% to 93%)</td>
</tr>
<tr>
<td><strong>Preschool (2–5 years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nulman 2002 [84]</td>
<td>BSID-II, Reynell developmental language scale, MSCA</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Nulman 1997 [83]</td>
<td>BSID-II</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Casper 2003 [49]</td>
<td>BSID-II</td>
<td>The two psychologist assessors were reliability certified on the BSID-II annually</td>
</tr>
<tr>
<td>Batton 2013 [44]</td>
<td>BSID-III</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Galbally 2011 [57]</td>
<td>BSID-III</td>
<td>Test-retest reliability for total score is 0.9 at 12 months</td>
</tr>
</tbody>
</table>

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

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

**School child (6–12 years)**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Outcome</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hutchison 2019 [79]</td>
<td>BRIEF</td>
<td>Not mentioned</td>
<td>Maternal perception of her child may have been negatively affected by depression, yet study using performance based measures of executive function found similar results</td>
</tr>
<tr>
<td>Hermansen 2016 [68]</td>
<td>CBCL</td>
<td>See validity</td>
<td>Widely used and standardised. Excellent reliability and validity</td>
</tr>
<tr>
<td>Nulman 1997 [83]</td>
<td>CBCL</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Nulman 2012 [85]</td>
<td>CBCL, CPRS</td>
<td>Not mentioned</td>
<td>Maternal perception of her child may have been negatively affected by depression and its associated anxiety and stress</td>
</tr>
</tbody>
</table>

(Continued)
### Table 1. (Continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Outcome</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nulman 2015</td>
<td>CBCL, CPRS-R</td>
<td>Not mentioned</td>
<td>Severity of maternal depression may influence responses when filling out questionnaires. Bias minimised as mothers evaluate both her exposed and unexposed children simultaneously. (Comment from review authors: the mother still knows that one child was exposed and the other was not)</td>
</tr>
<tr>
<td>El Marroun 2014</td>
<td>CBCL and Social responsiveness scale (SRS)</td>
<td>Correlations between the CBCL measurements at different ages fell in the expected range, based on a mean correlation ($r = 0.60$)/Cronbach's $\alpha$ indicated high inter-item reliability for the SRS ($\alpha = 0.79$)</td>
<td>CBCL: Validated in Dutch. Good predictive validity to identify preschoolers at risk of autism spectrum disorder. Scales could not be normalised, so scores were dichotomised. The SRS correlated well with the pervasive developmental problems scale of the CBCL, $r = 0.59$</td>
</tr>
<tr>
<td>Hanley 2015</td>
<td>HBQ-P</td>
<td>See validity</td>
<td>Maternal report poses some risk of differential outcome reporting. The HBQ-P has strong psychometric properties</td>
</tr>
<tr>
<td>Weikum 2013b</td>
<td>HBQ-P</td>
<td>See validity</td>
<td>The HBQ-P has strong psychometric properties. Mental health scales discriminate groups of children with and without signs of early psychopathology</td>
</tr>
<tr>
<td>Grzeskowiak 2016</td>
<td>SDQ</td>
<td>Not mentioned</td>
<td>Validated in Danish. Excellent discrimination for the identification of emotional (AUC 0.80) and behavioural (AUC 0.89) disorders</td>
</tr>
</tbody>
</table>

**iii. Assessment by teachers/others**

**Preschool (2–5 years)**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Outcome</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson 2016</td>
<td>CBCL (other caregiver)</td>
<td>Not mentioned</td>
<td>CBCL is not a diagnostic tool, but children scoring high on the PDD subscale are more likely to evidence behaviours commonly associated with ASDs</td>
</tr>
<tr>
<td>Misri 2006</td>
<td>CBCL (teacher)</td>
<td>Test-retest reliability $r = 0.85$, intrarater reliability $r = 0.61$</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Oberlander 2007</td>
<td>CBCL (teacher)</td>
<td>Not mentioned</td>
<td>Widely used and well validated. Close relationship between maternal depression and maternal report of child behaviour. Increased level of aggression could represent lower maternal thresholds for tolerating preschool behaviour</td>
</tr>
</tbody>
</table>

**Assessment using medical diagnosis**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Outcome</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boukhris 2017</td>
<td>ADHD</td>
<td>Not mentioned</td>
<td>Diagnosis defined from hospital diagnosis or redemption of a prescription for ADHD medication. Diagnoses of ADHD in the cohort were not validated. Sensitivity analysis on children with a diagnosis confirmed by neurologists and psychiatrists was consistent with those of the main analyses</td>
</tr>
<tr>
<td>Figueroa 2010</td>
<td>ADHD</td>
<td>Not mentioned</td>
<td>ADHD identified from diagnoses or treatment from practice, not from any formalised test or direct observation. This could lead to false-positive and false-negative errors. The young age when ADHD was diagnosed represents a limitation. Possible that only the most severe cases of ADHD were identified or that other behavioural problems that resemble ADHD are included</td>
</tr>
<tr>
<td>Laugesen 2013</td>
<td>ADHD</td>
<td>Not mentioned</td>
<td>Detection bias could have led to an overestimation of the association. Children with ADHD were identified based on hospital diagnoses and drug prescriptions. Patients with ADHD diagnosed by private psychiatrists or general practitioners and not prescribed drug treatment would be misclassified as not having ADHD</td>
</tr>
<tr>
<td>Man 2017</td>
<td>ADHD</td>
<td>Not mentioned</td>
<td>The registry contains information from publicly funded healthcare medical records. Does not include data from private medical practitioners or hospitals</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Reference</th>
<th>Outcome</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Castro 2016 [50]</td>
<td>ASD and ADHD</td>
<td>Not mentioned</td>
<td>ICD-9 codes have a high sensitivity and specificity for ASD and ADHD versus generally healthy control. Case definition was previously validated and included scores from autism diagnostic observation scale</td>
</tr>
<tr>
<td>Clements 2015 [51]</td>
<td>ASD and ADHD</td>
<td>Not mentioned</td>
<td>For ASD diagnosis in registry, sensitivity was 1.00, specificity 0.94. For ADHD, sensitivity was 0.84, specificity 0.90</td>
</tr>
<tr>
<td>Sujan 2017 [98]</td>
<td>ASD and ADHD</td>
<td>Not mentioned</td>
<td>Previous research has validated the diagnoses in used registries. Associations were estimated excluding offspring with diagnoses before age 2 years to address concerns about validity of early neurodevelopmental diagnoses. This did not markedly alter results</td>
</tr>
<tr>
<td>Wibroe 2017 [103]</td>
<td>ASD and ADHD</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Malm 2016 [77]</td>
<td>ASD, depression, anxiety, ADHD</td>
<td>Not mentioned</td>
<td>Quality of the registry has been validated and is good for psychiatric diagnoses</td>
</tr>
<tr>
<td>Liu 2017 [75]</td>
<td>ASD, F30-39, F40-49, F70-79, F90-99</td>
<td>Not mentioned</td>
<td>Cannot rule out detection bias, but similar associations were observed for all disorders, irrespective of age at onset. Cases were redefined as at least two hospital contacts for psychiatric disorders, but similar results were obtained</td>
</tr>
<tr>
<td>Boukhris 2016 [45]</td>
<td>ASD</td>
<td>Not mentioned</td>
<td>Diagnoses of ASD in the cohort were not validated. Sensitivity analysis on children with a diagnosis of ASD confirmed by neurologists and psychiatrists was consistent with those of the main analyses</td>
</tr>
<tr>
<td>Brown 2017 [48]</td>
<td>ASD</td>
<td>Not mentioned</td>
<td>ASD was defined as 2 or more outpatient diagnoses by either a paediatrician or psychiatrist, 1 or more diagnoses in hospital databases, after the age of 2 years. A similar definition using US insurance data had a positive predictive value of 87.4%</td>
</tr>
<tr>
<td>Croen 2011 [52]</td>
<td>ASD</td>
<td>Not mentioned</td>
<td>Validation study of diagnoses in the registry against the Autism Diagnostic Interview–Revised and the Autism Diagnostic Observation Schedule–Generic; 94% of cases met criteria for ASD on both instruments, and 100% on at least one. A full review of diagnostic information recorded in cohort medical records demonstrated that at least 90% of ASD cases in the cohort meet DSM-IV criteria</td>
</tr>
<tr>
<td>Gidaya 2014 [59]</td>
<td>ASD</td>
<td>Not mentioned</td>
<td>Register diagnosis of childhood autism was confirmed 94% of the time with an additional 3% classified with another ASD in validation study</td>
</tr>
<tr>
<td>Hvid 2013 [71]</td>
<td>ASD</td>
<td>Not mentioned</td>
<td>A previous study showed that 94% of children registered with autism spectrum disorder diagnoses met diagnostic criteria in chart review. Not all children in the study have been followed throughout childhood. Some may receive a diagnosis of ASD at older ages. Detection bias cannot be ruled out</td>
</tr>
<tr>
<td>Janecka 2018 [72]</td>
<td>ASD</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Rai 2017 [91]</td>
<td>ASD</td>
<td>Not mentioned</td>
<td>Previous validation studies found high validity of the diagnoses recorded in the registers</td>
</tr>
<tr>
<td>Sorensen 2013 [97]</td>
<td>ASD, infantile autism</td>
<td>Not mentioned</td>
<td>The quality of the infantile autism diagnosis in the registry has been validated. 94% met the criteria for correct diagnosis</td>
</tr>
<tr>
<td>Viktorin 2017b [41]</td>
<td>ASD</td>
<td>Not mentioned</td>
<td>The ASD diagnoses in the register have previously been validated</td>
</tr>
</tbody>
</table>

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


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

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

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

Triptans
Of five papers on triptans, four had comments on validity and/or reliability of at least one of their outcome measures. Most frequently mentioned was predictive validity and cross-cultural validity. None of the papers mentioned structural validity or content validity.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Psychometric properties of outcome measure</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platt 1989 [114]</td>
<td>Antipsychotics</td>
<td>BSID</td>
<td>Not mentioned</td>
<td>Measures obtained at different ages may differ in sensitivity. The categorical measure reported was not used in original examination, but derived by combining relevant motor items which showed some variance in study population, and dichotomising resulting scores to maximise drug effects</td>
</tr>
<tr>
<td>Peng 2013 [113]</td>
<td>Antipsychotics</td>
<td>BSID-III</td>
<td>Not mentioned</td>
<td>The scale is widely used and has potential to provide clinically relevant information on early neurodevelopment</td>
</tr>
<tr>
<td>Johnson 2012 [73]</td>
<td>Antipsychotics</td>
<td>Infant Neurological International Battery</td>
<td>Interrater reliability 0.97, test-retest 0.95</td>
<td>Sensitivity 90%, specificity 83%, PPV 79%, NPV 93%</td>
</tr>
<tr>
<td>Mortensen 2003 [81]</td>
<td>Antipsychotics, anxiolytics</td>
<td>Boel test</td>
<td>Not mentioned</td>
<td>Test was performed at children’s home, not under standard settings in which the test was developed. Surroundings could be associated with both exposure and outcome</td>
</tr>
<tr>
<td>Oberlander 2004 [87]</td>
<td>Anxiolytics (clonazepam combined with SSRI)</td>
<td>BSID-II</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Reebye 2002 [92]</td>
<td>Anxiolytics (clonazepam combined with SSRI)</td>
<td>BSID-II</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Reebye 2012 [38]</td>
<td>Anxiolytics (clonazepam combined with SSRI)</td>
<td>BSID-II</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Viggedal 1993 [40]</td>
<td>Anxiolytics</td>
<td>Griffiths’ mental development scale I, Neuropsychological assessment</td>
<td>Griffiths’: Not mentioned. Neuropsychological assessment: Deviations from normal activity and attention was diagnosed when present at two independent observations</td>
<td>Griffiths’: The high IQ in the reference group is considered normal as the average IQ nowadays is about 110. Neuropsychological assessment: Not mentioned</td>
</tr>
<tr>
<td>Gidai 2008a [105]</td>
<td>Anxiolytics (alprazolam)</td>
<td>Hungarian development test, Behavioural style questionnaire</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Gidai 2008b [106]</td>
<td>Anxiolytics (medazepam)</td>
<td>Hungarian development test, Behavioural style questionnaire</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Gidai 2008c [107]</td>
<td>Anxiolytics (chlordiazepoxide)</td>
<td>Hungarian development test, Behavioural style questionnaire</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Timmermann 2008b [113]</td>
<td>Anxiolytics (meprobamate)</td>
<td>Hungarian development test, Behavioural style questionnaire</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Laegreid 1992 [108]</td>
<td>Anxiolytics</td>
<td>Touwen Neurologic Assessment, Clinical neurologic assessment</td>
<td>Not mentioned</td>
<td>The test showed a fair differentiation and an evident developmental sequence</td>
</tr>
<tr>
<td>Petik 2008a [115]</td>
<td>Hypnotics (glutethimide)</td>
<td>Hungarian development test, Behavioural style questionnaire</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Petik 2008b [116]</td>
<td>Hypnotics (Amobarbital)</td>
<td>Hungarian development test, Behavioural style questionnaire</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
</tbody>
</table>

(Continued)
Table 2. (Continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preschool (2–5 years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hurault-Delarue 2016 [62]</td>
<td>Antipsychotics, anxiolytics and hypnotics</td>
<td>Compulsory medical exam</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Schechter 2017 [94]</td>
<td>Antipsychotics, anxiolytics, hypnotics</td>
<td>DAS</td>
<td>Fidelity checks were done approximately every 6-months by a licensed clinical psychologist</td>
<td>The DAS is a standardised, age-normed, well-validated measure of cognitive ability</td>
</tr>
<tr>
<td>Hartz 1975 [35]</td>
<td>Anxiolytics (meprobamate, chloridiazepoxide)</td>
<td>Stanford Binet Intelligence Scale</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Mattson 2002 [79]</td>
<td>Anxiolytics</td>
<td>WPPSI-R</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td><strong>School child (6–12 years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platt 1989 [114]</td>
<td>Antipsychotics</td>
<td>Paediatric neurologic assessment</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td><strong>Adolescent (13–18 years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mattson 2002 [79]</td>
<td>Anxiolytics</td>
<td>WISC-III</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td><strong>ii. Assessment by parents</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Infant (&lt;2 years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reebye 2002 [92]</td>
<td>Anxiolytics (clonazepam combined with SSRI)</td>
<td>Early Infancy Temperament Questionnaire</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td><strong>Preschool (2–5 years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lupattelli 2019 [109]</td>
<td>Anxiolytics, hypnotics</td>
<td>ASQ, CPRS-R</td>
<td>ASQ: Internal consistency was 0.6 to 0.7. CPRS-R: Internal consistency 0.9</td>
<td>Widely used and validated. In the supplement, the authors present associations between ASQ and diagnosis of motor delay (Beta 1.96 to 4.48) or language impairment (Beta 2.06 to 2.45), and between CPRS-R and parental ADHD symptoms (Beta 0.12 to 0.30 for paternal symptoms and 0.41 to 0.76 for maternal symptoms)</td>
</tr>
<tr>
<td>Brandlistuen 2017 [104]</td>
<td>Anxiolytics, hypnotics</td>
<td>CBCL (shortened)</td>
<td>Not mentioned</td>
<td>Validated, representative of full scale, the domains predict later psychopathology. High factor loadings</td>
</tr>
<tr>
<td>Misri 2006 [80]</td>
<td>Anxiolytics (clonazepam combined with SSRI)</td>
<td>CBCL</td>
<td>Test-retest reliability $r = 0.85$, intrarater reliability $r = 0.61$</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Odusu 2015 [110]</td>
<td>Anxiolytics</td>
<td>Intelligibility/Complexity of 3-year-old Children’s Utterances</td>
<td>Not mentioned</td>
<td>Parental self-report is a good measure of early expressive vocabulary, especially for severe language delay. Validity of the language grammar rating scale in the cohort has been described previously</td>
</tr>
<tr>
<td><strong>School child (6–12 years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radojić 2017 [111]</td>
<td>Anxiolytics</td>
<td>CBCL</td>
<td>Cronbach’s alphas for all scales were the same in 6 year-old children and in older children, indicating that problems were reliably measured in children older than 6 years of age</td>
<td>Validated in the Netherlands</td>
</tr>
</tbody>
</table>

iii. Assessment by teachers/others

**Preschool (2–5 years)**
Both papers on opioids commented on the validity of at least one of their outcome measures, one on concurrent validity, the other did not specify type of validity. None of the papers commented on reliability, construct validity or content validity.

Discussion

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

Strengths and limitations

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

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

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misri 2006 [80]</td>
<td>Anxiolytics (clonazepam combined with SSRI)</td>
<td>CBCL</td>
<td>Test-retest reliability r = 0.85, interrater reliability r = 0.61</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>School child (6–12 years)</td>
<td>Radojčić 2017 [111]</td>
<td>Anxiolytics</td>
<td>CBCL</td>
<td>Cronbach’s alphas for all scales were the same in 6 year-old children and in older children, indicating that problems were reliably measured in children older than 6 years of age</td>
</tr>
<tr>
<td>Assessment using medical diagnosis</td>
<td>Figueroa 2010 [56]</td>
<td>Anxiolytics</td>
<td>ADHD</td>
<td>Not mentioned</td>
</tr>
<tr>
<td></td>
<td>Janecka 2018 [72]</td>
<td>Lithium</td>
<td>ASD</td>
<td>Not mentioned</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Psychometric properties of outcome measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assessment using psychometric instruments</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Infants (&lt;2 years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salokorpi 1996 [136]</td>
<td>Indomethacin</td>
<td>Autti-Rämö neurodevelopmental test battery</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Amin 2008 [134]</td>
<td>Indomethacin</td>
<td>BSID-II, mental development index</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Al-Alaiyan 1996 [133]</td>
<td>Indomethacin</td>
<td>Gesell development scales, revised</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Avella-Garcia 2016 [118]</td>
<td>Paracetamol</td>
<td>BSID, unspecified ed.</td>
<td>Not mentioned in the paper, but in supplement. Cronbach’s α 0.70 (good to moderate)</td>
</tr>
<tr>
<td><strong>Preschool (2–5 years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barr 1990 [137]</td>
<td>ASA</td>
<td>Items from Gross Motor Scale (University of Oregon Medical School), Gesell- and Bayley Scales, Wisconsin Fine Motor Stadiness Battery, and Halstead Reitan Neuropsychological Battery</td>
<td>Monthly reliability checks revealed good interrater reliability</td>
</tr>
<tr>
<td>Klebanoff 1988 [138]</td>
<td>ASA</td>
<td>Stanford Binet Intelligence Scale</td>
<td>3 months test-retest, and in addition inter-rater, reliability was 0.83</td>
</tr>
<tr>
<td>Streissguth 1987 [139]</td>
<td>ASA</td>
<td>WPPSI</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Avella-Garcia 2016 [118]</td>
<td>Paracetamol</td>
<td>McCarthy Scales of Children’s Abilities, CAST†</td>
<td>McCarthy Scale of Children’s Abilities: Cronbach’s α 0.90. CAST: Cronbach’s α 0.64 (good to moderate)</td>
</tr>
<tr>
<td>Bornhag 2018 [119]</td>
<td>Paracetamol</td>
<td>Swedish language development scale†</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Liew 2016b [121]</td>
<td>Paracetamol</td>
<td>TEACCh-5</td>
<td>The authors do not specify the reliability or validity but refers readers interested in psychometric properties to two papers on the psychometrics of the instrument</td>
</tr>
<tr>
<td>Liew 2016c [125]</td>
<td>Paracetamol</td>
<td>WPPSI-R, shortened</td>
<td>Not mentioned</td>
</tr>
<tr>
<td><strong>Child (6–12 years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markovic 2019 [135]</td>
<td>NSAIDs</td>
<td>SON-R</td>
<td>Reliability of the tests used in the study was 0.73 and 0.71</td>
</tr>
<tr>
<td>Laue 2019 [121]</td>
<td>Paracetamol</td>
<td>WISC-JV</td>
<td>Not mentioned</td>
</tr>
<tr>
<td><strong>ii. Assessment by parents</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Infants (&lt;2 years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vlenterie 2016 [42]</td>
<td>Paracetamol</td>
<td>Motor milestone questionnaire, ASQ, CBCL, EAS</td>
<td>Motor milestone is believed to be objective and therefore a reliable maternal report on motor development. ASQ: Not mentioned. CBCL: Not mentioned. EAS: The short form was as reliable and precise as the full scale</td>
</tr>
</tbody>
</table>

(Continued)
### Table 3. (Continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood 2016a [140]</td>
<td>Triptans</td>
<td>ASQ, CBCL, EAS</td>
<td>ASQ: The questions had excellent test-retest reliability and agreement between parents and professional examiners. CBCL: Not mentioned. EAS: In a Norwegian sample, internal consistency (α) within each scale ranged from 0.48 to 0.79</td>
<td>The ASQ is predictive of school performance. The short version has been validated in Norway and in young children. CBCL: The shortened CBCL has been validated in Norway and in young children. Parents are better reporters of externalising symptoms, whereas children are better reporters of internalising symptoms. EAS has been validated in children as young as those studied</td>
</tr>
</tbody>
</table>

#### Preschool (2–5 years)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Markovic 2019 [135]</td>
<td>NSAIDs</td>
<td>CBCL</td>
<td>Internal consistency (α) 0.68</td>
<td>Good validity, validated in Dutch. The subscales had good fit in international studies in diverse societies</td>
</tr>
<tr>
<td>Brandlistuen 2013 [120]</td>
<td>NSAIDs and paracetamol</td>
<td>Motor milestone questionnaire, ASQ, CBCL, EAS</td>
<td>Maternal reports of gross motor milestone attainment have been reported to be highly reliable. ASQ: Not mentioned. CBCL: Not mentioned. EAS: The short form was as reliable and precise as the full scale</td>
<td>Motor milestones: Not mentioned. ASQ: Validated in Norway. Average factor loading 0.61 for fine motor and 0.75 for gross motor items, adequate reliability. Average factor loading 0.82 for communication, good reliability. (Comment from review authors: According to COSMIN principles, factor loadings are considered part of construct validity). CBCL: Correlation to CBCL full scale was 0.92. Average factor loading 0.58 for externalising and 0.52 for internalising behaviour scales, adequate reliability. EAS: Factor loading for emotionality 0.71, activity 0.68, for sociability 0.58, for shyness 0.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liew 2014 [122]</td>
<td>Paracetamol</td>
<td>SDQ</td>
<td>Reliable screening tool</td>
<td>With the cut-off used, the scale has high specificity for ADHD-like behaviours and 17% of children with problems on the SDQ have received a diagnosis of HKD</td>
</tr>
<tr>
<td>Liew 2016b [124]</td>
<td>Paracetamol</td>
<td>BRIEF</td>
<td>The authors do not specify the reliability or validity but refers readers interested in psychometric properties to two papers on the psychometrics of the instrument</td>
<td></td>
</tr>
<tr>
<td>Skovlund 2017 [129]</td>
<td>Paracetamol and opioids</td>
<td>Intelligibility/Complexity of 3-year-old Children’s Utterances &amp; ASQ</td>
<td>Not mentioned</td>
<td>Parental report is considered a valid measure and validation against clinical assessment has been described for the instrument in the cohort</td>
</tr>
<tr>
<td>Wood 2016b [141]</td>
<td>Triptans</td>
<td>CBCL</td>
<td>Not mentioned</td>
<td>The shortened CBCL has been validated in Norway. The domains used, have been shown to predict later psychopathology in children and adolescents</td>
</tr>
<tr>
<td>Wood 2016c [142]</td>
<td>Triptans</td>
<td>ASQ, EAS</td>
<td>Not mentioned</td>
<td>The ASQ is predictive of school performance. This short version has been validated in Norway. EAS: Not mentioned</td>
</tr>
<tr>
<td>Harris 2018 [139]</td>
<td>Triptans</td>
<td>ASQ, CBCL, EAS</td>
<td>Not mentioned in the article, but in the supplement internal consistency is reported quantitatively for each subscale</td>
<td>Not mentioned</td>
</tr>
</tbody>
</table>

#### Child (6–12 years)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stergioukoli 2016 [129]</td>
<td>Paracetamol</td>
<td>SDQ</td>
<td>SDQ is a validated and reliable screening instrument</td>
<td></td>
</tr>
<tr>
<td>Tovo-Rodrigues 2018 [131]</td>
<td>Paracetamol</td>
<td>SDQ</td>
<td>Not mentioned</td>
<td>Validated for a Brazilian population and for the studied age group</td>
</tr>
</tbody>
</table>

(Continued)
## Table 3. (Continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Psychometric properties of outcome measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thompson 2014 [130]</td>
<td>Paracetamol and ASA</td>
<td>SDQ, CPRS-R-L (only for paracetamol)</td>
<td>SDQ has a test-retest reliability of 0.62 after 4 to 6 months. Internal consistencies of the subscales range from 0.62 to 0.75. Self-reported problem behaviour has been shown to be a more valid indicator of mental and physical health than parent-reported problems.</td>
</tr>
<tr>
<td>Ruisch 2018 [127]</td>
<td>Paracetamol and ASA</td>
<td>Development and Well-Being Assessment</td>
<td>Inter-rater differences between maternal and teacher assessments could reflect different behaviours in different settings or bias in the assessment. Good validity.</td>
</tr>
</tbody>
</table>

### iii. Assessment by teachers/others

#### Preschool (2–5 years)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Psychometric properties of outcome measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avella-Garcia 2016 [118]</td>
<td>Paracetamol</td>
<td>California Preschool Social Competence Scale, ADHD, DSM-IV form list</td>
<td>California Preschool Social Competence Scale: Cronbach’s α 0.89. ADHD, DSM-IV form list: Cronbach’s α 0.90. Not mentioned.</td>
</tr>
<tr>
<td>Liew 2016b [124]</td>
<td>Paracetamol</td>
<td>BRIEF</td>
<td>The authors do not specify the reliability or validity but refers readers interested in psychometric properties to two papers on the psychometrics of the instrument.</td>
</tr>
</tbody>
</table>

#### Child (6–12 years)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Psychometric properties of outcome measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thompson 2014 [130]</td>
<td>Paracetamol and ASA</td>
<td>SDQ (children)</td>
<td>SDQ has a test-retest reliability of 0.62 after 4 to 6 months. Internal consistencies of the subscales range from 0.62 to 0.75. Self-reported problem behaviour has been shown to be a more valid indicator of mental and physical health than parent-reported problems.</td>
</tr>
<tr>
<td>Ruisch 2018 [127]</td>
<td>Paracetamol and ASA</td>
<td>Development and Well-Being Assessment</td>
<td>Inter-rater differences between maternal and teacher assessments could reflect different behaviours in different settings or bias in the assessment. Good validity.</td>
</tr>
</tbody>
</table>

### Assessment by medical diagnosis

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Psychometric properties of outcome measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rubenstein 2019 [143]</td>
<td>Opioids</td>
<td>Developmental delay and ASD</td>
<td>Not mentioned. All children in the study were screened for ASD and had general developmental evaluations by a clinician. Children with positive screenings had comprehensive ASD evaluations.</td>
</tr>
<tr>
<td>Liew 2016a [123]</td>
<td>Paracetamol</td>
<td>Infantile autism and ASD</td>
<td>Not mentioned. Diagnoses of ASD were ascertained from the general and psychiatric hospital registries in Denmark using standardised diagnostic criteria. Diagnoses of infantile autism in the psychiatric registry have previously been shown to have high validity.</td>
</tr>
<tr>
<td>Liew 2014 [122]</td>
<td>Paracetamol</td>
<td>HKD</td>
<td>Not mentioned. Children who received diagnoses solely prior to 5 years of age were not considered as having HKD due to higher diagnostic uncertainty at younger ages. 79% of all children diagnosed with HKD had redeemed medications at least twice.</td>
</tr>
<tr>
<td>Liew 2019 [126]</td>
<td>Paracetamol</td>
<td>ADHD</td>
<td>Diagnoses of ADHD were ascertained through maternal report. This method is reliable. In a validation study, all girls with maternal report of ADHD scored above 90% on ADHD Rating Scale-IV, as did 64% of boys. ADHD prevalence in the study was comparable to estimates by Centers for Disease Control and Prevention.</td>
</tr>
</tbody>
</table>

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

**Points for consideration**

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

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

Table 3. (Continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome measure</th>
<th>Psychometric properties of outcome measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ystrom 2017 [132]</td>
<td>Paracetamol</td>
<td>ADHD</td>
<td>Not mentioned</td>
</tr>
</tbody>
</table>

*Structured interview of parents by the health care professional
†Mixture of parental questionnaires and nurse observation

https://doi.org/10.1371/journal.pone.0219778.s003
so studies assessing the same domain of neurodevelopment and using the same data source also would use compatible outcome measures and report results in a uniform manner.

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

---

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

https://doi.org/10.1371/journal.pone.0219778.g004
neurodevelopment following valproate exposure with the same gravitas as major congenital malformations. As early as 1996 an association between valproate and ASD like behaviours was noted [146], whilst ASD in exposed children was not the focus of a prospective investigation until much later [22].

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

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

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

Teacher reporting may be blinded to exposure status. However, the expectations of children in a classroom setting may be different from what is expected from children elsewhere. One review concluded that teacher reporting results in a higher prevalence of ADHD than if the disease is classified according to diagnostic criteria [151]. In addition, not all domains of neurodevelopment may be assessed with equal ease in a classroom setting. Hence in a meta-analysis of 119 studies, parent reporting of emotional problems correlated better than teacher reporting with children’s own assessments [152]. When parent and teacher reporting only
show moderate correlations, it is important to consider that they represent assessment of the child in very different settings. Some children have problems at school that are not present in the home-environment. Hence one assessment is not necessarily more correct than the other. In both parent and teacher reporting, the concurrent and predictive validity should be considered, as well as content validity to enable clinicians and researchers to determine whether the symptoms or traits evaluated by the scales are in fact clinically relevant problems.

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

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

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

The outcome measure should be age appropriate. Standardised tests and questionnaires are often validated for a certain age group. If the outcome measure is to be used in a different age group, it should first be validated for use in that age group [19]. Researchers using
diagnoses as outcomes should be aware that certain diagnoses are not valid below a certain age. For example, the American Academy of Pediatrics recommends that children should be 4 years old before DSM-IV diagnostic criteria of ADHD can be used [144]. Children should not be considered as having a diagnosis, if they only have a diagnosis recorded at an age where it is considered implausible that a correct diagnosis can be made. Length of follow-up should be guided by the average age at diagnosis for the specific disorder in the country where the research is carried out.

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

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

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

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

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

If the content or criterion validity is low or unknown, this should be reflected in the language used in the paper. For example, if a study uses a psychometric instrument to assess the
presence of ADHD, and the instrument has not shown acceptable validity when tested against diagnostic interviews for ADHD, authors should write that they have assessed “symptoms of ADHD” and not “ADHD” as such.

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

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

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

Conclusions

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

Supporting information

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

S2 File. Search terms and search strategy, example for the MEDLINE database.
(PDF)
S3 File. PRISMA checklist.
(PDF)

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

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

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

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

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

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

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

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

Data curation: Sarah Hjorth.

Formal analysis: Sarah Hjorth.

Funding acquisition: Hedvig Nordeng.

Investigation: Sarah Hjorth.

Supervision: Hedvig Nordeng.

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

Visualization: Sarah Hjorth.

Writing – original draft: Sarah Hjorth, Rebecca Bromley.

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

References


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


Child neurodevelopment outcomes in studies on prenatal exposure to psychotropics and analgesics


Paper II

Fertility treatment and oral contraceptive discontinuation for identification of pregnancy planning in routinely collected health data – an application to analgesic and antibiotic utilisation

Sarah Hjorth, Mollie Wood, Fatima Tauqeer, Hedvig Nordeng

Fertility treatment and oral contraceptive discontinuation for identification of pregnancy planning in routinely collected health data – an application to analgesic and antibiotic utilisation

Sarah Hjorth 1*, Mollie Wood 1,2, Fatima Tauqeer 1 and Hedvig Nordeng 1,3

Abstract

Background: Women with unplanned pregnancies use folic acid less frequently, and more often use potentially teratogenic medications in the first trimester. Yet most studies based on routinely collected data lack information on pregnancy planning. Further, only pregnancies proceeding beyond a certain gestational age appear in routinely collected data, creating the possibility for collider-stratification bias. If pregnancy intention could be identified, pregnancies could be ascertained earlier. This study aimed to investigate fertility treatment and discontinuation of oral contraception (OC) as proxies for pregnancy planning by describing variations in patterns of prescription fills for antibiotics and analgesics during the peri-pregnancy period by these proxies of pregnancy intention.

Methods: Fertility treatment with clomiphene and discontinuation of OC were identified in the Norwegian Prescription Database (NorPD) and linked with data from the Medical Birth Registry of Norway for the years 2006 to 2017. Filled prescriptions for antibiotics and analgesics from NorPD were displayed for women on fertility treatment, women who discontinued OC before pregnancy, and women who discontinued during pregnancy.

Results: Of 172,585 included pregnancies, fertility treatment was identified in 19,449, and OC discontinuation before or during pregnancy in 153,136. Women who discontinued OC during pregnancy were less likely to use preconception folic acid (25.4%) than women who discontinued before pregnancy (32.9%), and women on fertility treatment (51.0%). Proportions of first trimester prescription fills were 4.9% (analgesics) and 12.8% (antibiotics) for women who discontinued OC during pregnancy, compared to 4.0 and 11.4% in women who discontinued OC before pregnancy, and 4.7 and 11.0% in women on fertility treatment.

(Continued on next page)
Conclusions: There were no substantial differences in patterns of prescription fills for analgesics and antibiotics before or during pregnancy by fertility treatment and OC discontinuation. This suggests that there were few differences in medication use between women with planned and unplanned pregnancies, or that fertility treatment and timing of OC discontinuation from routinely collected health data cannot stand alone in the identification of unplanned pregnancies. A narrower definition of OC discontinuation during pregnancy seemed to be a better proxy, but this should be confirmed in other studies.

Keywords: “Analgesics, Non-narcotic”, “Analgesics, Opioids”, “Anti-bacterial agents”, “Drug Utilization”, “Pregnancy, Unplanned”, “Registries”

Background
An estimated 45% of pregnancies in high-income countries and 29% of pregnancies in Northern Europe are unplanned [1]. Women with unplanned pregnancies are less likely to use folic acid before pregnancy [2, 3], more likely to smoke [2, 4], and more often use medications in the first trimester, including potentially teratogenic medications [5]. Most studies on pregnancy intention and lifestyle in pregnancy have used self-reported data [3]. In routinely collected health data, where study samples are more representative of the underlying population, information on pregnancy intention is typically unavailable. Therefore, studies of prescription drug utilisation or safety usually lack information on pregnancy intention [6–8]. Further, for pregnancies to appear in routinely collected data and be identified in studies, they need to proceed beyond a certain point, for example week 12 [7], or end in live births [6]. Hence, early pregnancy losses cannot be identified, and collider stratification bias may be introduced [9]. If pregnancy intention could be identified in routinely collected data, it would be possible to identify pregnancies prospectively from an earlier time.

A way to study pregnancy intention in routinely collected data could be to identify women on fertility treatment. A previous study investigated prescription drug utilisation in women on fertility treatment compared to women with spontaneous pregnancies [10]. However, to make inferences about women with planned pregnancies who conceive spontaneously, a different approach is needed. A few studies on pregnancy intention and lifestyle in pregnancy have used oral contraceptive discontinuation as a marker of pregnancy intention [3]. To our knowledge, this approach has not been used in drug utilisation studies. Folic acid use may also be a marker of pregnancy planning in data sources that record such use, as most public health authorities recommend 400 micrograms folic acid daily from at least 1 month prior to conception and throughout the first 12 weeks of pregnancy [11]. We therefore aimed to investigate fertility treatment and discontinuation of oral contraception as proxies for pregnancy planning by describing variations in patterns of prescription fills for antibiotics and analgesics during the peri-pregnancy period by these proxies of pregnancy intention, while accounting for folic acid use.

Antibiotic and analgesics were chosen for two reasons. First, they are commonly used classes of prescription medications in pregnancy [6–8]. In a US population, prescription antibiotics were filled in 50%, and analgesics in 30% of pregnancies [6]. Second, we expected that the two classes would show different patterns of use in the peri-pregnancy period. Previous studies have found that prescription analgesic use declines over the course of pregnancy, whereas prescription antibiotic use does not [7]. In studies based on self-reported surveys, women with unplanned pregnancies had 29% higher odds of reporting use of prescription medications for chronic or long-term illnesses [12], and 54% higher likelihood of reporting frequent paracetamol use in the first trimester [13]. There was no difference in the use of prescription medications for acute indications, such as urinary tract infections [12]. However, no studies have replicated these results in a population-based sample. Furthermore, medication use was studied either as any use during pregnancy [12], or in the first trimester only [13], and it is unknown whether pregnancy intention will affect medication use beyond the first trimester.

We sought to answer these lingering questions, using a population-based cohort of Norwegian women. Specifically, we evaluated fertility treatment or discontinuation of oral contraceptives (OC) as proxies for pregnancy planning, with the hypothesis that 1) women on fertility treatment, and women who discontinue OC before pregnancy, would be less likely to use prescription analgesics before pregnancy and in the first trimester than women who discontinue OC during pregnancy, and 2) there would be no difference in the use of prescription antibiotics. Additionally, we expected that in the second and third trimester, and after pregnancy, there would be no difference in the use of prescription analgesics and antibiotics by proxies for pregnancy planning.
Methods
Study population and data sources
The study was based on data from the Norwegian Prescription Database (NorPD) and the Medical Birth Registry of Norway (MBRN), which were linked using unique personal identification numbers. NorPD was established in 2004 and contains information on all prescriptions filled at pharmacies [14]. Prescriptions are classified using the World Health Organization’s Anatomical Therapeutic Chemical (ATC) Classification System [15]. The MBRN was established in 1967 [16] and contains data on all pregnancies with a duration beyond week 12 [17]. The MBRN has information on the date of birth of the child and the gestational age at birth, allowing calculation of the start of pregnancy.

We included all registered pregnancies from 2006 to 2017. To make inferences about pregnancy intention, we included women with either filled prescriptions for medications for fertility treatment in the year before pregnancy, or OC dispensed that covered the date 1 year before the start of pregnancy. We focused on OC to be able to approximate the timing of contraceptive discontinuation; women using alternate forms of long-acting birth control (e.g., intrauterine devices) were excluded from our sample. Figure 1 shows the exclusion criteria applied to achieve the study sample.

Variables
Proxies for pregnancy intention
Fertility treatment was identified in NorPD as filled prescriptions for clomiphene (ATC-code G03GB02) in the year before pregnancy. Clomiphene is the first-line medication in fertility treatment in Norway [18]. OC use was assessed by filled prescriptions for hormonal contraception for systemic use (ATC-code G03A, excluding contraceptive patches, G03AA13, injections, G03AC06, implants, G03AC08, and emergency contraceptives, G03AD). Discontinuation status was divided into three categories: Early OC discontinuers, late OC discontinuers, and within-pregnancy OC discontinuers (Fig. 2).
Early OC discontinuers had their last day covered by OCs between 1 year before pregnancy and three cycles before pregnancy, where each cycle was assumed to last 28 days. Late OC discontinuers had their last day covered between two cycles and 27 days before pregnancy and 1 day before pregnancy. Within-pregnancy OC discontinuers had coverage overlapping with pregnancy.

To account for folic acid use, we stratified the groups by timing of folic acid initiation: initiation before pregnancy, during pregnancy, or no folic acid. Information on folic acid initiation was obtained from MBRN (recorded by health care professionals) and/or NorPD (filled prescriptions of ATC-codes B03BB, A11EA or A11EB).

The validity of data on fertility treatment with clomiphene and OC discontinuation has not been investigated for NorPD. A Finnish study found that 10% of pregnancies were misclassified as spontaneous if fertility treatment was identified by prescription fills [19]. A US study comparing claims data to self-report of OC use found moderate agreement, Cohen’s kappa 0.46 [20].

A study investigating the validity of preconception folic acid registration in MBRN found underreporting: 45% of self-reported users before pregnancy were registered as non-users. The validity was not investigated for registration of folic acid use during pregnancy [21].

**Drug utilisation**
We studied five peri-pregnancy periods: 90 days before pregnancy, first trimester, second trimester, third trimester, and 90 days after pregnancy. For each period, we identified the proportion of women who filled a prescription for an analgesic (ATC-code N02 or M01A) or an antibiotic (ATC-code J01) from NorPD. For antibiotics, we included prescriptions where the defined daily doses overlapped with the period, because antibiotics are used continuously. In addition, we studied prescription fills for non-steroidal anti-inflammatory drugs (NSAIDs, ATC-code M01A) and tetracyclines (ATC-code J01A) separately, as these medications are contraindicated in the second half of pregnancy in Norway [22, 23].

**Participant characteristics**
We included information from MBRN on sociodemographics, lifestyle, and obstetric factors including obstetric comorbidity index, adapted from Bateman et al. [24].

**Statistical analysis**
Participant characteristics by proxies for pregnancy planning were compared using descriptive statistics. For prescription fills, we compared proportions with 95% confidence intervals (CIs). The proportion of missingness was compared for variables with missing values.

In a sensitivity analysis to account for uncertainty in date of conception, the start of pregnancy was considered as start of pregnancy 1) plus 14 days, and 2) minus 14 days. In another sensitivity analysis to improve specificity, we considered only women with filled prescriptions for OC one cycle before pregnancy as within-pregnancy discontinuers. A third sensitivity analysis included women with filled prescriptions of vaginal rings with hormones (ATC-code G02BB) and/or contraceptive patches. A fourth sensitivity analysis was restricted to live births, as claims data often only is reliable for live births [25]. All analyses were performed using Stata (version16;StataCorpLP).

**Results**
Of 172,585 included pregnancies, 19,449 occurred after fertility treatment with clomiphene. Among 153,136
women with OC coverage 1 year before pregnancy, 50.8% were early, 27.8% late, and 21.4% within-pregnancy discontinuers. Of women on fertility treatment, half (51.0%) initiated folic acid use before pregnancy. The proportion of women who initiated folic acid before pregnancy was lower for early (35.3%), late (32.9%), and within-pregnancy OC discontinuers (25.4%). Early OC discontinuers were more likely to have experienced previous pregnancy loss than within-pregnancy OC discontinuers (Table 1). Women on fertility treatment were older, had more obstetric comorbidities, and were more likely to be nulliparous and to have experienced previous pregnancy loss. Women with no folic acid use were more likely to smoke in early and late pregnancy (Additional file 1).

### Analgesics

In total, 9.8% filled a prescription for an analgesic before pregnancy. In all groups, proportions declined in the first trimester, continued to decline throughout pregnancy, and rose after pregnancy, but not to pre-pregnancy levels. Within-pregnancy OC discontinuers had similar proportions of analgesic prescription fills before pregnancy (10.5, 95% CI 10.2 to 10.9%) to early (9.3, 95% CI 9.1 to 9.5%) and late discontinuers (9.4, 95% CI 9.1 to 9.7%) (Fig. 3, panel a). Women on fertility treatment filled the most prescriptions for analgesics (11.4, 95% CI 11.0 to 11.9%). During pregnancy, there were no substantial differences by group. After pregnancy, there were no differences in analgesic prescription fills by OC group, but women in the fertility treatment group had higher

### Table 1 Characteristics of the included pregnancies by proxies of pregnancy intention

<table>
<thead>
<tr>
<th></th>
<th>Fertility treatment (n = 19, 449)</th>
<th>Timing of oral contraceptive discontinuation</th>
<th>Within-pregnancy (n = 32,780)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age</td>
<td>31.8 (5.0)</td>
<td>28.9 (4.7)</td>
<td>28.4 (4.6)</td>
</tr>
<tr>
<td>Married/cohabiting</td>
<td>95.3</td>
<td>94.4</td>
<td>94.9</td>
</tr>
<tr>
<td>Employed</td>
<td>73.2</td>
<td>74.1</td>
<td>75.5</td>
</tr>
<tr>
<td>Nulliparous</td>
<td>58.2</td>
<td>55.3</td>
<td>57.6</td>
</tr>
<tr>
<td>Previous pregnancy loss</td>
<td>27.5</td>
<td>19.9</td>
<td>10.8</td>
</tr>
<tr>
<td>Obstetric comorbidity index</td>
<td>0.77 (1.2)</td>
<td>0.39 (0.8)</td>
<td>0.36 (0.8)</td>
</tr>
<tr>
<td><strong>Components of the index</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>5.3</td>
<td>5.2</td>
<td>5.3</td>
</tr>
<tr>
<td>Diabetes, pre-gestational</td>
<td>1.2</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Hypertension, chronic</td>
<td>0.9</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Hypertension, gestational</td>
<td>2.5</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Kidney disease</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Multiple gestation</td>
<td>5.6</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Preeclampsia, mild</td>
<td>3.0</td>
<td>2.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Preeclampsia, severe</td>
<td>2.1</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>Previous caesarean section</td>
<td>5.5</td>
<td>5.3</td>
<td>4.6</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>0.7</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Folic acid use</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiation before pregnancy</td>
<td>51.0</td>
<td>35.3</td>
<td>32.9</td>
</tr>
<tr>
<td>Initiation during pregnancy</td>
<td>31.4</td>
<td>43.4</td>
<td>44.9</td>
</tr>
<tr>
<td>No folic acid</td>
<td>17.7</td>
<td>21.6</td>
<td>22.2</td>
</tr>
<tr>
<td>Smoking in early pregnancy</td>
<td>5.4</td>
<td>8.1</td>
<td>7.9</td>
</tr>
<tr>
<td>Smoking at the end of pregnancy</td>
<td>3.3</td>
<td>4.6</td>
<td>4.5</td>
</tr>
<tr>
<td>Weight gain in pregnancy</td>
<td>13.8 (7.7)</td>
<td>14.6 (8.5)</td>
<td>14.6 (7.3)</td>
</tr>
</tbody>
</table>

* Table stratified by folic acid use is shown in Additional file 1

**Figures shown are percent of non-missing values with the exception of maternal age, calendar year, obstetric comorbidity index, and weight gain in pregnancy, presented as mean (standard deviation). Missing values ranged from 0% (maternal age, calendar year, parity) to 17.6% (maternal employment). Women could choose not to have smoking and weight reported to the MBRN. For smoking, 12.4 to 16.3% chose not to report. For weight, 75.9 to 77.7% chose not to report

**Adapted from Bateman et al. [24], using the variables available in MBRN (age, asthma, pre-gestational diabetes, chronic hypertension, kidney disease, previous caesarean section, multiple gestation, severe preeclampsia, mild preeclampsia, gestational hypertension) and weighting the variables as done by Bateman et al. [24].
proportions of analgesic prescription fills (8.0, 95% CI 7.7 to 8.4%) than OC discontinuers (5.2 to 5.5%, 95% CIs 5.0 to 5.4% and 5.4 to 5.7%, respectively).

Prescription fills for NSAIDs followed the same patterns across peri-pregnancy periods and proxies for pregnancy planning as analgesics overall (Fig. 3, panel b).

Antibiotics
For all groups, the proportion of filled antibiotic prescriptions was similar before (12.0%), and during (12.4%) pregnancy. After pregnancy, the proportion rose (18.4%).

Within-pregnancy OC discontinuers had the highest proportion of antibiotic prescription fills before pregnancy (13.7, 95% CI 13.3 to 14.0%), but the proportions were quite similar for early (11.5, 95% CI 11.3 to 11.7%) and late OC discontinuers (11.9, 95% CI 11.6 to 12.2%), and women on fertility treatment (11.4, 95% CI 11.0 to 11.9%) (Fig. 4, panel a), with comparable patterns in the first trimester. Antibiotic prescription fills did not differ by proxies for pregnancy intention in the second and third trimester. After pregnancy, women in the fertility treatment group filled more prescriptions for antibiotics (20.7, 95% CI 20.2 to 21.3%) than OC discontinuers (18.0 to 18.1%, 95% CIs 17.7 to 18.4% and 17.7 to 18.6%, respectively).

Prescription fills for tetracyclines decreased throughout pregnancy, and was never higher than 0.8% in any of the groups. The proportion rose slightly after pregnancy, but not to pre-pregnancy levels (Fig. 4, panel b). There were no differences in proportion of tetracycline fills by proxies for pregnancy intention.

Sensitivity analyses
In different sensitivity analyses, we redefined pregnancy start as 14 days earlier or later, included women with
filled prescriptions of vaginal rings with hormones and/or contraceptive patches, and restricted to live births. Participant characteristics by group and patterns of prescription fillings were largely unchanged (Additional files 4, 5, 6 and 7).

In women who filled prescriptions of OC within one cycle before pregnancy, 14.7% used folic acid before pregnancy. The women were slightly younger, less likely to be married, and more likely to smoke during pregnancy compared to within-pregnancy OC discontinuers in the primary analysis (Additional file 8). They also tended to have filled more analgesic and antibiotic prescriptions before pregnancy and in the first trimester than early and late OC discontinuers.

Discussion
In this population-based cohort of 172,585 Norwegian pregnancies, we found no substantial differences in filled prescriptions for analgesics and antibiotics in the peri-pregnancy period by OC group. There were no substantial differences in prescription fills for the potentially harmful medication groups NSAIDs and tetracyclines either. This suggests that there were few differences in medication use between women with planned and unplanned pregnancies, or that fertility treatment and timing of OC discontinuation are not good proxies for pregnancy planning.

Women in the fertility treatment group were more likely to fill prescriptions for analgesics and antibiotics after pregnancy. In all groups, we found similar patterns
across strata of folic acid use. Women who used folic acid prior to pregnancy had the lowest proportion of analgesic, and antibiotic prescription fills before pregnancy and in the first trimester. However, the proportions were not substantially different from those in women who used folic acid during pregnancy only, and women who did not use folic acid.

**Interpretation**

A previous study found that women on fertility treatment were twice as likely to fill prescriptions for potentially hazardous medications in pregnancy as women with spontaneous pregnancies [10]. We did not find differences in prescription fills during pregnancy, but after pregnancy women in the fertility treatment group had the highest proportion of antibiotic and analgesic prescription fills. An explanation could be that women on fertility treatment are older and have more comorbidities.

The high proportions of filled prescriptions for antibiotics before pregnancy and in the first trimester in within-pregnancy OC discontinuers, especially those who filled a prescription within one cycle before pregnancy, are unlikely to be explained by reduced OC efficacy among women on antibiotics. One study found an association between any antibiotic prescription fills and reduced OC efficacy [26]. However, most studies found no association for systemic antibacterials (ATC group J01) [27, 28]. Differences in participant characteristics between our study and the previous volunteer-based studies that we based our hypotheses on, could explain the differences in findings. Participants in volunteer-based studies are often healthier and have a higher socioeconomic position than the general population. In our sample, included women were similar to the general birthing population in measured health parameters (Additional file 9).

We found that women with no folic acid use were more likely to smoke in early pregnancy (an expected marker of unplanned pregnancy) and in late pregnancy (not expected to indicate pregnancy intention), and less likely to be employed. A study from Norway found that self-reported pregnancy planning predicted folic acid use, but the strongest predictor was high education level [29]. Folic acid might, therefore, be a marker of both pregnancy intention and socioeconomic position.

**Limitations**

A limitation of our chosen proxies for pregnancy planning is the inability to assess pregnancy planning in women who did not use contraception, or who used other types of contraception than short-acting hormonal. This meant that only 24.2% of the population could be classified by proxies for pregnancy intention. We did not include women without short-acting hormonal contraceptive use as a control group in our study, as they are likely a heterogeneous group of women who plan pregnancies, and women who do not plan pregnancies, but use other types of contraception. Furthermore, we ascertained that women were under observation for 1 year before the start of pregnancy by identifying prescription fills in NorPD. For women without contraceptive or fertility treatment use, 1 year of look-back would only be available if they had filled other prescriptions, thus introducing different selection criteria among the groups. Some conditions, including migraine with aura, hypertension, or a history of thromboembolic disorders, are considered contraindications for using hormonal contraceptives [30], and so women with these diagnoses are likely under-represented in our sample. We did not have information on migraine or thromboembolic disorders, but we found a similar prevalence of chronic hypertension in women on OC and women who were excluded from our study sample (Additional file 9). The main differences between included and excluded women were that included women were younger and more often nulliparous. Hence, our findings on patterns of prescription fills in peri-pregnancy may not be generalisable to the entire pregnant population. However, our findings regarding the selected proxies for pregnancy planning should be applicable to other settings that have the same availability of fertility treatment and similar use contraceptive medications.

We had no information on women who did not become pregnant, had early miscarriages, or terminations of pregnancy occurring before gestational week 12. Women with unplanned pregnancies may be more likely to have early terminations, leading to an underrepresentation of unplanned pregnancies in our sample. Women who have early terminations may differ systematically from women who continue pregnancy in lifestyle and medication use, so the underrepresentation could have attenuated group differences in our study. However, the absolute number of pregnancies missing from this study due to early terminations should be small. There are 13 early terminations per 1000 women in Norway, and in 19% of these, women report the use of short-acting hormonal contraceptives at pregnancy start [31].

Early miscarriages may be more likely among women with low fertility. One of our proposed markers for pregnancy planning - early OC discontinuation - may be a proxy for long time-to-pregnancy and therefore low fertility. We found that women on fertility treatment and early OC discontinuers were more likely to have experienced previous pregnancy loss. Women with pregnancy intention and long time-to-pregnancy may, therefore, be underrepresented in our study sample. Women with longer time-to-pregnancy will have had more time to
make lifestyle changes before pregnancy; hence the underrepresentation could have attenuated group differences in our study. In cohorts that only have access to live births, underrepresentation of women with low fertility may be larger, as we found that women on fertility treatment were more often excluded in the sensitivity analysis restricted to live births. Repeating our study in a cohort that is not selected on achieving a pregnancy of at least 12 weeks’ duration may provide results that are more representative.

Misclassification of the proxies for pregnancy intention cannot be ruled out. We found that 21.4% of pregnancies between 2006 and 2017 were to women who filled a prescription for an OC in the year before pregnancy. A previous study showed that around 20% of Norwegian women of fertile age use OC [32], supporting the sensitivity of our classification. Women could have discontinued earlier than their last day covered, reducing specificity. This seems likely, as 21.4% of OC users were classified as within-pregnancy discontinuers. When we considered only women with filled OC prescriptions one cycle before pregnancy as within-pregnancy discontinuers to increase the specificity of within-pregnancy OC discontinuation, the proportion decreased from 21.4 to 4.7%. In addition, we saw lower proportions of folic acid use before pregnancy, and higher proportions of smoking and filled prescriptions in the first trimester, as expected. This classification might, therefore, warrant further investigation.

We identified fertility treatment with clomiphene in 2.7% of pregnancies between 2006 and 2017. Norwegian health authorities reported fertility treatment in 3 to 4% of pregnancies from 2003 to 2013 [33, 34]. It is possible that some women on fertility treatment were not treated with clomiphene. Since they almost certainly were not also taking an OC, they were likely excluded. We would expect these women to exhibit behaviours similar to women undergoing clomiphene treatment, but were unable to test this assumption.

Conclusions
Peri-pregnancy patterns of prescription fills for analgesics and antibiotics did not vary substantially by the studied proxies for pregnancy planning. This suggests that there were few differences in medication use between women with planned and unplanned pregnancies, or that fertility treatment and timing of OC discontinuation from routinely collected health data, even when combined with data on folic acid use, cannot stand alone in the identification of unplanned pregnancies. A narrower definition of within-pregnancy OC discontinuation seemed a better proxy, but this should be confirmed in other studies. It should also be considered that women on fertility treatment or OC are not representative of all pregnancy planners. Future studies should examine how proxies for pregnancy intention perform, and whether pregnancies can be identified prospectively, in a cohort that is not selected based on achieving and maintaining a pregnancy beyond the 12th gestational week. To enable identification of pregnancy intention in a larger proportion of the population, future studies could consider more data-driven methods making use of a wider range of health indicators as proxies for pregnancy intention, such as age, smoking cessation, removal of long-acting hormonal contraceptives, and discontinuation of other medications.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12884-020-03435-4.

Additional file 1. Characteristics of the included pregnancies by proxies of pregnancy intention, stratified on folic acid use.

Additional file 2. Analgesic prescription fills by proxies of pregnancy intention, stratified on folic acid use. Proportion of pregnancies with analgesic prescription fills by peri-pregnancy period and proxies of pregnancy intention, stratified on folic acid use.

Additional file 3. Antibiotic prescription fills by proxies of pregnancy intention, stratified on folic acid use. Proportion of pregnancies with antibiotic prescription fills by peri-pregnancy period and proxies of pregnancy intention, stratified on folic acid use.

Additional file 4. Results from sensitivity analysis redefining pregnancy start as 14 days earlier. Characteristics of included pregnancies, proportion of pregnancies with analgesic prescription fills, proportion of pregnancies with antibiotic prescription fills by proxies of pregnancy intention, sensitivity analysis redefining pregnancy start as 14 days earlier.

Additional file 5. Results from sensitivity analysis redefining pregnancy start as 14 days later. Characteristics of included pregnancies, proportion of pregnancies with analgesic prescription fills, proportion of pregnancies with antibiotic prescription fills by proxies of pregnancy intention, sensitivity analysis redefining pregnancy start as 14 days later.

Additional file 6. Results from sensitivity analysis including women with filled prescriptions for other short-acting hormonal contraceptives. Characteristics of included pregnancies, proportion of pregnancies with analgesic prescription fills, proportion of pregnancies with antibiotic prescription fills by proxies of pregnancy intention, sensitivity analysis including women with filled prescriptions of vaginal rings with hormones and/or contraceptive patches.

Additional file 7. Results from sensitivity analysis restricting to live births. Characteristics of included pregnancies, proportion of pregnancies with analgesic prescription fills, proportion of pregnancies with antibiotic prescription fills by proxies of pregnancy intention, sensitivity analysis restricting to live births.

Additional file 8. Results from sensitivity analysis with a more narrow definition of oral contraceptive discontinuation during pregnancy. Characteristics of included pregnancies, proportion of pregnancies with analgesic prescription fills, proportion of pregnancies with antibiotic prescription fills by proxies of pregnancy intention, sensitivity analysis redefining oral contraceptive discontinuation during pregnancy as fillings within one cycle before pregnancy.

Additional file 9. Characteristics of pregnancies by inclusion and exclusion from the study sample.

Abbreviations
ATC: Anatomical Therapeutic Chemical; MBRN: Medical Birth Registry of Norway; NorPD: Norwegian Prescription Database; OC: Oral contraceptives
Acknowledgements
Data was stored at the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (UST).

Authors’ contributions
All authors contributed to design of the study and interpretation of the results. SH analysed the data and drafted the manuscript. MW, FT and HN critically revised the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the European Research Council Starting Grant “DrugInPregnancy” [grant number 639377 to HN and SH], the National Heart Lung and Blood Institute [grant number 1T32HL080448 NHLBI to MW] and The PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Norway. The funding sources had no role in the design of the study, collection, analysis, or interpretation of data, in writing the manuscript, or in the decision to publish.

Availability of data and materials
The data that support the findings of this study are available from the Norwegian Institute of Public Health but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of a Norwegian Regional Committee for Research Ethics, a Data Protection Officer at a Norwegian research institution, and the Norwegian Institute of Public Health.

Ethics approval and consent to participate
The study was approved by the Regional Committee for Research Ethics in South Eastern Norway (approval number 2018/140/REK sør øst) and by the Data Protection Officer at the University of Oslo (approval number 58033). The Regional Committee for Research Ethics in South Eastern Norway has granted exemption from the obligation to seek consent from participants.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, and PharmaTox Strategic Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Postboks 1068 Blindern, 0316 Oslo, Norway. 2Department of Epidemiology, T.H. Chan School of Public Health, Harvard University, Boston, USA. 3Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway.

Received: 16 July 2020 Accepted: 17 November 2020

Published online: 25 November 2020

References

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Paper III

Prenatal Exposure to Non-Steroidal Anti-Inflammatory Drugs and Risk of Attention-Deficit/Hyperactivity Disorder – a Follow-Up Study in the Norwegian Mother, Father and Child Cohort

Sarah Hjorth, Angela Lupattelli, Marte Handal, Olav Spigset, Eivind Ystrom, Hedvig Nordeng

*Pharmacoepidemiology and Drug Safety.* Online ahead of print. DOI: 10.1002/pds.5250.
Prenatal exposure to non-steroidal anti-inflammatory drugs and risk of attention-deficit/hyperactivity disorder: A follow-up study in the Norwegian mother, father and child cohort

Sarah Hjorth¹ | Angela Lupattelli¹ | Marte Handal² | Olav Spigset³,4 | Eivind Ystrom¹,2,5 | Hedvig Nordeng¹,6

¹PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, and PharmaTox Strategic Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway
²Department of Mental Disorders, Norwegian Institute of Public Health, Oslo, Norway
³Department of Clinical Pharmacology, St. Olav's University Hospital, Trondheim, Norway
⁴Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway
⁵PROMENTA Research Center, Department of Psychology, University of Oslo, Oslo, Norway
⁶Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway

Correspondence
Sarah Hjorth, PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, and PharmaTox Strategic Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Postbox 1068 Blindern, 0316 Oslo, Norway. Email: s.h.andersen@farmasi.uio.no

Funding information
European Research Council, Grant/Award Number: 639377; PharmaTox Strategic Research Initiative

Abstract

Purpose: To estimate the association between Attention-Deficit/Hyperactivity Disorder (ADHD) in children in preschool and primary school, and prenatal exposure to non-steroidal anti-inflammatory drugs (NSAIDs) by timing and duration.

Methods: This study was based on the Norwegian Mother, Father and Child Cohort Study linked to the Medical Birth Registry of Norway, the Norwegian Patient Registry (NPR) and the Norwegian Prescription Database (NorPD). NSAID exposure was identified by maternal self-report in pregnancy. Child diagnosis of ADHD was identified by maternal self-report in pregnancy. Child diagnosis of ADHD was obtained from NPR and NorPD. Symptoms of ADHD at age 5 years were measured using Conners’ Parent Rating Scale-Revised, where higher scores correspond to more symptoms. To account for time-varying exposure and confounders, marginal structural models were fitted to estimate hazard ratios and mean difference in z-scores.

Results: The analyses on ADHD diagnosis and ADHD symptoms included 56 340 and 34 961 children respectively. Children exposed to NSAIDs prenatally had no increased risk of ADHD diagnosis (first trimester: HR 1.12, 95% CI 0.86;1.45, second trimester: HR 0.98, 95% CI 0.69;1.38, third trimester: HR 0.68, 95% CI 0.31; 1.46) or ADHD symptoms (first trimester: standardized mean difference 0.03, 95% CI 0.03;0.09, second trimester: standardized mean difference 0.03, 95% CI 0.04;0.11, third trimester: standardized mean difference 0.11, 95% CI 0.03; 0.25). There was no duration-response relationship for either outcome.

Conclusion: Though non-differential misclassification of the exposure may have attenuated results, these findings are reassuring and suggest no substantially increased risk of ADHD diagnosis or symptoms in children prenatally exposed to NSAIDs, regardless of timing or duration.

KEYWORDS
anti-inflammatory agents, attention deficit disorder with hyperactivity, Medical Birth Registry of Norway, Norwegian mother, prenatal exposure delayed effects, father and child cohort study, non-steroidal

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2021 The Authors. Pharmacoepidemiology and Drug Safety published by John Wiley & Sons Ltd.
1  |  INTRODUCTION

Non-steroidal anti-inflammatory drugs (NSAIDs) are used in 5–15% of pregnancies, can cross the placenta, and the blood–brain-barrier. NSAIDs inhibit cyclooxygenase-1 and -2, and first and third trimester use is associated with an increased risk of negative birth outcomes. Both cyclooxygenase-1 and -2 are expressed in the brain. Prenatal exposure to NSAIDs might therefore influence child neurodevelopment.

Findings from previous studies on child neurodevelopment after prenatal NSAID exposure are in general reassuring with no associations. However, only two studies had follow-up beyond 3 years of age, and one found slightly poorer executive function in exposed children. Brain development continues into early adulthood, and some functions cannot be assessed until children have reached an age where more complex tasks are demanded. Among these tasks are behavioral inhibition and sustained attention; tasks that are problematic for children with Attention-Deficit/Hyperactivity Disorder (ADHD).

ADHD is among the most common behavioral disorders in childhood. The worldwide prevalence is approximately 7% using DSM-IV criteria, and approximately 3% using ICD-10 criteria. The etiology of ADHD is unclear, but thought to be highly genetic. Many environmental factors have also been proposed as influencing ADHD risk. Among them maternal inflammation in pregnancy, one of the indications for NSAID use.

Several studies have investigated prenatal exposure to acetaminophen and risk of ADHD, but to our knowledge, only one previous study on prenatal NSAID exposure had information on symptoms of ADHD, and none had information on ADHD diagnosis. Further, NSAIDs are often used intermittently, so any time-varying effect of prenatal exposure would be important to guide clinical decisions.

In this study, the primary objective was to investigate associations between timing and duration of prenatal exposure to NSAIDs and risk of ADHD diagnosis and symptoms. A secondary objective was to investigate whether associations differed by maternal indication for NSAID use.

2  |  MATERIALS AND METHODS

2.1  |  Data sources and study population

This study was based on data from the Norwegian Mother, Father and Child Cohort (MoBa). MoBa is a population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health. Participants were recruited from all over Norway from 1999 to 2008. The women consented to participate in 41% of the pregnancies. The cohort now includes 114,500 children, 95,200 mothers and 75,200 fathers. The present study is based on version 9 of the quality-assured data files, which was released for research in 2016. The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from The Regional Committee for Medical Research Ethics. All mothers and fathers in the cohort provided written, informed consent to participation and to the use of their data from the Norwegian Health Registries. MoBa is currently regulated by the Norwegian Health Registry Act. The present study was approved by The Regional Committee for Medical Research Ethics in South-Eastern Norway; approval number: 2015/2137/REK Sør-Ost. Data were handled in accordance with the General Data Protection Regulation.

In addition to MoBa data, we used data from the Medical Birth Registry of Norway (MBRN), the Norwegian Prescription Database (NorPD), and the Norwegian Patient Registry (NPR). MBRN is a national health registry containing information about all births in Norway since 1967. NorPD has stored data on prescriptions redeemed at pharmacies by patients in ambulatory care since 2004. NPR has stored individual level data on diagnoses in secondary and tertiary health care settings since 2008. Data was linked using the unique personal identification number given to all residents in Norway. Only live born singletons were included. In an attempt to meet the assumption of positivity and account for confounding by indication, the study sample was restricted to women reporting indications for NSAID use during pregnancy (fever, infection, pain, or headache/migraine). The indications are specified in the Supporting Information. Further inclusion and exclusion criteria are presented in Figure 1.

2.2  |  Exposure

Exposure was defined as prenatal exposure to NSAIDs (M01A in the World Health Organization’s Anatomical Therapeutic Chemical [ATC] Classification System, except glucosamine, M01AX05), as reported by the mother in any of two prenatal and one post-partum self-administered questionnaires (Figure 2). The mothers were presented with a list of symptoms and asked to check the ones that they had experienced. For each checked item on the list, the mothers were also asked to note any medications taken and specify the timing of use by checking one or more boxes that each represented a four-week interval (e.g., week 5–8 of pregnancy). To investigate the first objective of

Key Points

- In a large Norwegian cohort, there was no substantially increased risk of ADHD diagnosis or ADHD symptoms in children prenatally exposed to NSAIDs, when accounting for time-varying exposure and confounding.
- There was no duration-response relationship for either outcome, and the results were stable in sensitivity analyses.
- If these findings are corroborated in other populations, prenatal exposure to NSAIDs may be used as a negative exposure control in studies on prenatal exposure to acetaminophen and risk of ADHD in children.

...
In this study, we evaluated (a) timing of NSAID use (first trimester [0–12 weeks of gestation], second trimester [13–28 weeks, or to delivery if born before week 28], or third trimester [29 weeks to delivery]), (b) duration of use, defined as number of 4-week intervals during pregnancy with exposure (grouped as “0,” “1,” “2–3,” or “4 or more”), and (c) substance-level analysis on ibuprofen (ATC code M01AE01), the most commonly used NSAID in Norway. To investigate the second objective of the study, we stratified the analyses by five

**FIGURE 1** Flowchart of the study samples. Conditions of exclusion can overlap. ADHD, attention-deficit/hyperactivity disorder; CPRS-R (S), Conners’ Parent Rating Scale-Revised, Short Form; NSAID, non-steroidal anti-inflammatory drug; Q, questionnaire.

**FIGURE 2** Timing and coverage of the questionnaires in the Norwegian mother, father and child cohort. BP, before pregnancy; GW, gestational week; PP, post partum; Q, questionnaire; t, time point in the marginal structural model.
indications for NSAID use: musculoskeletal pain, headache/migraine, fever, infection, and other.

No studies have investigated the validity of self-reported NSAID use in pregnancy in Norway. In a US cohort, data from maternal medication diaries in early pregnancy and recall of NSAID use during a first trimester interview showed moderate agreement (Cohen's kappa 0.41, sensitivity 0.79 and specificity 0.62).23 We did not use data from NorPD on filled prescriptions for NSAIDs, as (a) some NSAIDs (ibuprofen, diclofenac, and naproxen) are available over-the-counter, and (b) the filled prescriptions may be used in a later trimester than it was filled, or not at all. The first point is supported by a Danish study comparing self-reported use of acetaminophen or NSAIDs (investigated as one group) to prescription fills.32 In 95% of 348 self-reports of use, the medication was purchased over-the-counter.32 The number of identified prescription fills was correspondingly low (20 fills in total).32

2.3 | Outcome

The primary outcome was child ADHD, defined as a diagnosis of ADHD (ICD-10 code F90)16 recorded in NPR by a specialist in the Norwegian health care system, and/or a filled prescription for ADHD medication (ATC codes N06BA01, N06BA02, N06BA04, N06BA09, and N06BA12)30 recorded in NorPD. In Norway, these were the only ADHD medications available during the study period. Children who had filled prescriptions for the indication of narcolepsy (ICD-10 code G47.4),16 and who had no diagnosis of ADHD in NPR, were considered as not having ADHD. Children were followed from birth to incident ADHD or ultimo 2016, whichever came first. We had information about whether children died or migrated during follow-up, but not about dates of deaths or migrations. We restricted the sample to children born in 2004 or later to ensure that outcome data was available from birth in at least one registry. This was done as there is no lower age limit for receiving a F90-diagnosis, though it is rare in children under the age of 5 years.

The validity of ADHD diagnoses has not been investigated for Norwegian registries, but the documentation of diagnostic procedures in medical charts was found to be poor.33 In Denmark, where the health care system is similar to the Norwegian, a registered diagnosis of ADHD has a positive predictive value of 0.87.24

To identify children who had difficulties, but did not necessarily meet the diagnostic criteria for ADHD, the secondary outcome was child hyperkinetic/inattentive symptoms at age 5 years, as reported by parents on the 12-item ADHD index from the Conners’ Parent Rating Scale-Revised, Short Form (CPRS-R [S]). Children were excluded if less than eight items had been completed.

The CPRS-R (S) items show high internal consistency (Cronbach’s alpha 0.88), and ability to predict later ADHD diagnosis in the MoBa population.35 As to content validity, the items are based on DSM-IV ADHD criteria.24

2.4 | Covariates

Potential confounders were identified a priori using subject knowledge and directed acyclic graphs (Figures S1 and S2).27,28 The sufficient adjustment set contained socioeconomic position, maternal ADHD, unplanned pregnancy, and disease severity. Data on covariates were obtained from MoBa questionnaires, MBRN, and NorPD. We did not have information on disease severity, but used proxies for disease severity (co-medication with other analgesics and psychotropics,29,40 exercise in pregnancy, and severity of depressive symptoms, which may affect pain perception).

2.5 | Statistical analysis

For ADHD diagnosis, crude hazard ratios with 95% confidence intervals (HRs with 95% CIs) were obtained using Cox proportional hazards models. For ADHD symptoms, an average standardized score (z-score) was calculated. The z-score has a mean of zero and a standard deviation (SD) of one. For the CPRS-R (S), higher z-scores indicate more ADHD symptoms, with a score two SD above the mean usually considered indicative of clinically important problems with attention and/or hyperactivity.24 Crude mean differences in z-score with 95% CIs were identified using generalized linear models.

To account for time-varying confounding, propensity scores were estimated and used as inverse probability of treatment weights (IPTWs).41 In the analyses on timing, IPTWs were estimated at three points in addition to baseline to account for time-varying exposure to NSAIDs, time-varying confounding (by co-medication and exercise in pregnancy), and confounding by baseline covariates (socioeconomic position, maternal health and lifestyle). The resulting four IPTWs were multiplied to obtain a total weight that was used in marginal structural models. In the analysis on duration, a single IPTW was estimated for any pregnancy exposure, using baseline covariates and baseline values of the time-varying covariates. For ADHD diagnosis, the weight was used in Cox proportional hazards models with robust standard errors to obtain weighted HRs with 95% CIs. For ADHD symptoms, we additionally accounted for loss to follow-up by estimating inverse probability of censoring weights (IPCWs) in eligible pregnancies.41 The IPCWs were multiplied by the IPTWs in the sample that had answered CPRS-R (S). The IPTW*IPCW was used in generalized linear models with robust standard errors to obtain weighted standardized mean differences with 95% CIs. Details on the variables included in the weights are presented in the Supporting Information.

To answer the second objective of the study, the analyses were repeated in the five strata of maternal indications for medication use.

2.6 | Missing data

Up to 27% of included pregnancies had missing values on at least one variable used to generate IPTWs. The variables with the highest proportions of missing values were alcohol intake in pregnancy (up to 12.7%), maternal education (up to 5.2%), and maternal depressive symptoms (up to 4.8%).
## Table 1  Characteristics of pregnancies exposed and unexposed to NSAIDs in the Norwegian mother, father and child cohort

<table>
<thead>
<tr>
<th>Characteristicsa</th>
<th>ADHD diagnosis sample N = 56,340</th>
<th>ADHD symptoms sample N = 34,961</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exposed N = 3542</td>
<td>Unexposed N = 52,798</td>
</tr>
<tr>
<td>Maternal sociodemographics and lifestyle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25 years</td>
<td>401 (11.3)</td>
<td>5084 (9.6)</td>
</tr>
<tr>
<td>25–29 years</td>
<td>1154 (32.6)</td>
<td>17,390 (32.9)</td>
</tr>
<tr>
<td>30–34 years</td>
<td>1325 (37.4)</td>
<td>20,862 (39.5)</td>
</tr>
<tr>
<td>35–39 years</td>
<td>589 (16.6)</td>
<td>8379 (15.9)</td>
</tr>
<tr>
<td>≥40 yearsb</td>
<td>73 (2.1)</td>
<td>1083 (2.1)</td>
</tr>
<tr>
<td>Married/cohabiting, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3377 (95.3)</td>
<td>50,881 (96.4)</td>
<td>2090 (96.1)</td>
</tr>
<tr>
<td>College/university educationc, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2134 (60.3)</td>
<td>35,002 (66.3)</td>
<td>1435 (66.0)</td>
</tr>
<tr>
<td>Gross yearly income, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>2088 (59.0)</td>
<td>31,632 (59.9)</td>
</tr>
<tr>
<td>Low</td>
<td>944 (26.7)</td>
<td>12,694 (24.0)</td>
</tr>
<tr>
<td>High</td>
<td>417 (11.8)</td>
<td>6944 (13.2)</td>
</tr>
<tr>
<td>Primiparous, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1750 (49.4)</td>
<td>24,898 (47.2)</td>
<td>1098 (50.5)</td>
</tr>
<tr>
<td>Planned pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2653 (74.9)</td>
<td>43,450 (82.3)</td>
<td>1663 (76.5)</td>
</tr>
<tr>
<td>Pre-pregnancy BMI, kg/m²; mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.6 (4.7)</td>
<td>24.0 (4.2)</td>
<td>24.5 (4.6)</td>
</tr>
<tr>
<td>Leisure time physical activity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than weekly</td>
<td>1325 (37.4)</td>
<td>20,454 (38.7)</td>
</tr>
<tr>
<td>Once or twice a week</td>
<td>1472 (41.6)</td>
<td>21,421 (40.6)</td>
</tr>
<tr>
<td>More than twice a week</td>
<td>647 (18.3)</td>
<td>9322 (17.7)</td>
</tr>
<tr>
<td>Folic acid supplementation, any, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2989 (84.4)</td>
<td>46,381 (87.9)</td>
<td>1863 (85.7)</td>
</tr>
<tr>
<td>Smoking in pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>362 (10.2)</td>
<td>3055 (5.8)</td>
<td>187 (8.6)</td>
</tr>
<tr>
<td>Alcohol intake in pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>451 (12.7)</td>
<td>4891 (9.3)</td>
<td>286 (13.2)</td>
</tr>
<tr>
<td>I illicit drug use in pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 (0.4)</td>
<td>94 (0.2)</td>
<td>8 (0.4)</td>
</tr>
<tr>
<td>Maternal health</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic disease registered in MBRN, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>319 (9.0)</td>
<td>4639 (8.8)</td>
<td>206 (9.5)</td>
</tr>
<tr>
<td>Obstetric comorbidity, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6 (1.1)</td>
<td>0.5 (1.0)</td>
<td>0.6 (1.1)</td>
</tr>
<tr>
<td>Comedication in pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opioid analgesics (ATC code N02A)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>248 (7.0)</td>
<td>925 (1.8)</td>
<td>139 (6.4)</td>
</tr>
<tr>
<td>Acetaminophen (ATC code N02BE01)</td>
<td>2694 (76.1)</td>
<td>25,041 (47.4)</td>
</tr>
<tr>
<td>Migraine medications (ATC code N02C)</td>
<td>129 (3.6)</td>
<td>495 (0.9)</td>
</tr>
<tr>
<td>Antipsychotics/anxiolytics/hypnotics (ATC code N05)</td>
<td>122 (3.4)</td>
<td>809 (1.5)</td>
</tr>
<tr>
<td>Antidepressants (ATC code N06)</td>
<td>79 (2.2)</td>
<td>580 (1.1)</td>
</tr>
<tr>
<td>Depressive symptoms in pregnancy, mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average SCL-5 at GW 17 &amp; 29</td>
<td>1.3 (0.4)</td>
<td>1.2 (0.3)</td>
</tr>
<tr>
<td>ADHD symptom level (ASRS), mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130 (3.6)</td>
<td>125 (3.4)</td>
<td>12.8 (3.5)</td>
</tr>
<tr>
<td>Use of ADHD medications, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 (1.4)</td>
<td>478 (0.9)</td>
<td>23 (1.1)</td>
</tr>
<tr>
<td>Paternal characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25 years</td>
<td>170 (4.8)</td>
<td>2174 (4.1)</td>
</tr>
<tr>
<td>25–29 years</td>
<td>825 (23.3)</td>
<td>11,663 (22.1)</td>
</tr>
<tr>
<td>30–34 years</td>
<td>1326 (37.4)</td>
<td>20,768 (39.3)</td>
</tr>
<tr>
<td>35–39 years</td>
<td>850 (24.0)</td>
<td>12,799 (24.2)</td>
</tr>
<tr>
<td>40–44 years</td>
<td>260 (7.3)</td>
<td>3912 (7.4)</td>
</tr>
<tr>
<td>≥45 years</td>
<td>104 (2.9)</td>
<td>1371 (2.6)</td>
</tr>
<tr>
<td>College/university education, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1538 (43.4)</td>
<td>27,059 (51.3)</td>
<td>1021 (46.9)</td>
</tr>
<tr>
<td>Depressive symptoms (SCL-8), mean (SD)</td>
<td>1.2 (0.3)</td>
<td>1.1 (0.3)</td>
</tr>
</tbody>
</table>

(Continues)
exposure misclassification, we used probabilistic bias analysis.45,46 To 
comparing them to children exposed to NSAIDs in utero. To handle potential 
congenital malformations were used as a disease comparator by com-
in utero, but whose mothers used NSAIDs in the 6 months prior to preg-
address potential residual confounding, we used negative exposure con-
additional model specifications. Among these were models including 
paternal characteristics (age, education, depressive symptoms, and use of 
NSAID medications) and parental symptoms of ADHD. In an attempt to 
address potential residual confounding, we used negative exposure con-
trols by comparing unexposed children to children unexposed to NSAIDs 
in utero, but whose mothers used NSAIDs in the 6 months prior to preg-
nancy. The latter group was also used as a disease comparator by com-
paring them to children exposed to NSAIDs in utero. To handle potential 
exposure misclassification, we used probabilistic bias analysis.45,46 To 
assess the validity of the outcome measures, we investigated the corre-
respondence between CPRS-R (S) score and ADHD diagnosis, and the 
association between in utero exposure to NSAIDs and risk of ADHD, 
based only on diagnostic data from NPR. For ADHD diagnosis, we 
excluded children who died or emigrated during follow-up to investigate 
the impact of misclassified time at risk. All statistical analyses were per-
formed using Stata (version15; StataCorpLP).

3 | RESULTS

For ADHD diagnosis, 56,340 children of 50,572 mothers were 
included. For ADHD symptoms, 34,961 children of 31,696 mothers 
were included. A child could be included in one or both samples. 
NSAID use was reported in 6.2% of pregnancies with medications 
available over-the-counter, mainly ibuprofen, accounting for more 
than 95% of users. A majority of mothers had a college or university 
education, but mothers of exposed children were less likely to have 
high education, and more likely to report unplanned pregnancy, 
smoking, and alcohol use in pregnancy (Table 1).

3.1 | ADHD diagnosis

The children were followed for 9.8 years on average (SD 1.5, range 8– 
12 years). The prevalence of child ADHD diagnosis was 2.2%, and 
the average age at first diagnosis was 8.2 years (SD 1.7). In the crude anal-
ysis, first trimester exposure to NSAIDs was associated with a higher 
risk of ADHD (HR 1.32, 95% CI 1.03; 1.68; Table 2). After weighting, 
the association was no longer seen (HR 1.12, 95% CI 0.86; 1.45), and 
prenatal exposure to NSAIDs was not associated with higher risk of 
ADHD in any trimester or duration category. 

Results did not differ substantially in the substance-level analysis 
on ibuprofen (Table S1), nor by maternal indication for NSAID use 
(Figure 3).

3.2 | ADHD symptoms

The mean average CPRS-R (S) score at age 5 years was 1.37 
(SD 0.38). The proportion of children who had a z-score of two or 
more SD from the mean was 4.5%. In the analysis on timing of expo-
sure, we observed no association with CPRS-R (S) score.

In the analysis on duration, we found slightly higher CPRS-R 
(S) scores in children exposed in 1 four-week interval (weighted
standardized mean difference 0.11, 95% CI 0.05;0.17) compared to unexposed. In children exposed for 2–3 intervals and 4 or more intervals, the weighted standardized mean differences were 0.10 (95% CI −0.00:0.20) and −0.02 (95% CI −0.16:0.12), respectively.

Results did not differ substantially in the substance-level analysis on ibuprofen (Table S1), nor by maternal indication for NSAID use (Figure 3).

3.3 | Sensitivity analyses

The inclusion of paternal characteristics and parental ADHD symptoms in alternative model specifications did not alter the estimates of association substantially (Figure S4).

In the negative exposure controls, we identified similar estimates of association between pre-pregnancy use of NSAIDs and child

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>ADHD diagnosis and symptoms by timing and duration of prenatal NSAID exposure in the Norwegian mother, father and child cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Incidence (%)</td>
</tr>
<tr>
<td>Timing</td>
<td></td>
</tr>
<tr>
<td>Exposed before pregnancy only, negative control</td>
<td>4067</td>
</tr>
<tr>
<td>Unexposed before pregnancy</td>
<td>48 731</td>
</tr>
<tr>
<td>Exposed in 1st trimester</td>
<td>2354</td>
</tr>
<tr>
<td>Unexposed in 1st trimester</td>
<td>53 986</td>
</tr>
<tr>
<td>Exposed in 2nd trimester</td>
<td>1524</td>
</tr>
<tr>
<td>Unexposed in 2nd trimester</td>
<td>54 816</td>
</tr>
<tr>
<td>Exposed in 3rd trimester</td>
<td>612</td>
</tr>
<tr>
<td>Unexposed in 3rd trimester</td>
<td>55 728</td>
</tr>
<tr>
<td>Duration</td>
<td></td>
</tr>
<tr>
<td>0 periods (unexposed)</td>
<td>52 798</td>
</tr>
<tr>
<td>1 period</td>
<td>2297</td>
</tr>
<tr>
<td>2–3 periods</td>
<td>899</td>
</tr>
<tr>
<td>4 or more periods</td>
<td>346</td>
</tr>
</tbody>
</table>

| ADHD symptoms (CPRS-R(S)) sample, N = 34 961 |
| n         | Mean z-score (SD) | Crude mean difference (95% CI) | Weighted mean difference (95% CI) |
| Timing    |                                                           |                  |                      |
| Exposed before pregnancy only, negative control    | 2627             | 0.15 (1.0)        | 0.17 (0.13:0.20)     | 0.14 (0.10:0.19)  |
| Unexposed before pregnancy                         | 30 159           | −0.02 (1.0)       | -                    | -                  |
| Exposed in 1st trimester                           | 1441             | 0.10 (1.1)        | 0.10 (0.05:0.15)     | 0.03 (−0.03:0.09) |
| Unexposed in 1st trimester                         | 33 520           | −0.00 (1.0)       | -                    | -                  |
| Exposed in 2nd trimester                           | 922              | 0.08 (1.0)        | 0.08 (0.01:0.14)     | 0.03 (−0.04:0.11) |
| Unexposed in 2nd trimester                         | 34 039           | 0.00 (1.0)        | -                    | -                  |
| Exposed in 3rd trimester                           | 369              | 0.18 (1.2)        | 0.18 (0.08:0.28)     | 0.11 (−0.03:0.25) |
| Unexposed in 3rd trimester                         | 34 592           | 0.00 (1.0)        | -                    | -                  |
| Duration                                           |                 |                   |                      |
| 0 periods (unexposed)                              | 32 798           | −0.01 (1.0)       | -                    | -                  |
| 1 period                                           | 1416             | 0.13 (1.1)        | 0.14 (0.09:0.19)     | 0.11 (0.05:0.17)  |
| 2–3 periods                                        | 549              | 0.11 (1.1)        | 0.11 (0.03:0.20)     | 0.10 (−0.00:0.20) |
| 4 or more periods                                   | 210              | −0.00 (0.9)       | 0.00 (−0.13:0.14)    | −0.02 (−0.16:0.12) |

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; CI, confidence interval; CPRS-R (S), Conners’ parent rating scale-revised, short form; HR, hazard ratio; NSAID, non-steroidal anti-inflammatory drug.

*aInverse probability of treatment weights, model additionally adjusted for co-medication with acetaminophen as the weights failed to balance that covariate.

*bInverse probability of treatment weights, model additionally adjusted for co-medication with acetaminophen, as the weights failed to balance that covariate. For third trimester exposure, the weights also failed to balance illicit drug use, which was added to the regression model.
ADHD diagnosis, as between first trimester exposure and ADHD diagnosis (weighted HR 1.14, 95% CI 0.92; 1.40). We also identified an association between pre-pregnancy use of NSAIDs and child ADHD symptoms (weighted mean difference 0.14, 95% CI 0.10; 0.19; Table 2).

In the probabilistic bias analysis, we found that failure to account for non-differential exposure misclassification could have biased the findings toward the null by about 26–37% according to trimester.

Children with an ADHD diagnosis had a mean CPRS-R (S) z-score of 1.85 (SD 1.9) which was almost two standard deviations from the mean in children without a diagnosis (−0.03, SD 0.9).

Results from the remaining sensitivity analyses showed that the estimates of association were generally robust (Supporting Information).

4 | DISCUSSION

In this Norwegian birth cohort with 9.8 years of follow-up on average, we found no substantially increased risk of ADHD diagnosis among children prenatally exposed to NSAIDs. For ADHD symptoms in 5-year-olds, we observed no associations by timing of NSAID exposure, but we found higher symptom scores in children exposed for one 4-week interval of pregnancy. Children exposed for four or more intervals did not have higher symptom scores, suggesting that associations are not causal, albeit number of cases exposed for four or more intervals was low.

To our knowledge, this is the first study to investigate prenatal NSAID exposure and risk of ADHD diagnosis. Our results on ADHD symptoms are in line with findings from Markovic et al, who used the Child Behavior Checklist to identify attention problems in children aged 5 years, and found no difference in adjusted mean score.11 In spite of the current debate on acetaminophen and risk of ADHD,22-24 NSAIDs are not an alternative to acetaminophen as first line analgesic in pregnancy. First trimester NSAID use is associated with increased risk of early miscarriage, and third trimester use with increased risks of oligohydramnios, premature closure of ductus arteriosus, prolonged pregnancy, and has the biological plausibility to increase blood loss at delivery.2,3 The identification of safe and effective medications for the treatment of pain during pregnancy should be a priority in future research, especially given the questions surrounding the safety of long-term use of acetaminophen during pregnancy.22-24 In the meantime, our findings, if corroborated in other populations, are reassuring regarding child ADHD for
the women who need to use NSAIDs during pregnancy to manage pain conditions.

An implication for research is that if our findings are corroborated, exposure to NSAIDs could be used as a negative exposure control in studies on prenatal exposure to acetaminophen and child ADHD, as the structure of bias is probably similar for NSAIDs and acetaminophen, albeit the contraindications are different.

Based on previous findings on maternal inflammation and child ADHD,20,21 we expected to find a lower risk of ADHD in children exposed to anti-inflammatory treatment than in children exposed to untreated inflammation. We did not find such an association, when comparing exposed children to unexposed children whose mothers reported similar symptoms. This could be because of our pre-specified categories of indications grouped heterogeneous diseases together. It is also possible that women with inflammatory conditions, who did not use NSAIDs were treated with other anti-inflammatory drugs.

Strengths of the present study include a large sample size with long follow-up and access to both diagnostic outcomes, and well-validated parent-reported outcomes. The study also has several limitations. Findings from the negative exposure controls suggest that some residual confounding, such as confounding by genetics or severity of indication, is present.

Exposure misclassification cannot be ruled out. First and second trimester exposure was reported during pregnancy, third trimester exposure 6 months after birth, where child symptoms of ADHD are unlikely, so any exposure misclassification is likely non-differential. This could have biased results toward the null. In the probabilistic bias analysis, we estimated the magnitude of such bias around 26–37%. In first trimester NSAID exposed, we might have found a higher risk of ADHD (HR around 1.5) in the absence of misclassification. However, the HR in the negative exposure controls would have been similarly higher, cautioning against a causal interpretation.

The prevalence of ADHD diagnoses was 2.2% in our sample after an average 9.8 years of follow-up, whereas the prevalence among Norwegian 12-year-olds is 3.4%.23 This could reflect systematic differences between the study sample and the general population, and/or a shorter period to observe the outcomes in the present study. If exposure is in any way associated with earlier or later detection of ADHD, this could have affected our results. The validity of the ADHD diagnoses is supported by a correspondence between ADHD diagnosis and a higher score on the well-validated CPRS-R (S).

Participation rate in MoBa was 41%. Compared to the general birthing population of Norway, participants were less likely to be young parents, more likely to be married or cohabiting, and had a healthier lifestyle during pregnancy.47 A study found that selection into the cohort and loss to follow-up appeared to affect estimates of association for longer-term outcomes such as child ADHD, but that IPCW was a robust method to handle such bias.65 Still, the selected sample may affect the prevalence of ADHD diagnoses, and the generalizability of our findings.

5 CONCLUSION

In this large cohort study with follow-up of 9.8 years on average, we found no substantially increased risk of ADHD diagnosis in children exposed to NSAIDs in utero, regardless of timing or duration of exposure. We identified a slightly higher ADHD symptom score at age 5 years in children exposed to NSAIDs for one 4-week interval in pregnancy. Exposure for more intervals was not associated with higher symptom scores, suggesting that the finding is an artifact. Our findings are reassuring for women who need to use NSAIDs in pregnancy.

ACKNOWLEDGMENTS

This work was supported by the European Research Council Starting Grant “DrugsInPregnancy” [grant number 639377 to SH] and The PharmaTox Strategic Research Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway. The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. Data was stored at the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT).

CONFLICT OF INTEREST

The authors declare there is no conflict of interest.

ORCID
Sarah Hjorth https://orcid.org/0000-0003-2841-5868
Angela Lupattelli https://orcid.org/0000-0002-8787-3183
Marte Handal https://orcid.org/0000-0003-1773-0184
Olav Spigset https://orcid.org/0000-0001-7902-9014
Eivind Ystrom https://orcid.org/0000-0003-4390-6171
Hedvig Nordeng https://orcid.org/0000-0001-6361-2918

REFERENCES


**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

Paper IV

Maternal Medication Use and Childhood Cancer in Offspring – Systematic Review and Considerations for Researchers

Sarah Hjorth, Caroline H. Hemmingsen, Justine Bénévent, Anne Broe, Anton Pottegård, Lina S. Mørch, Maarit K. Leinonen, Susanne K. Kjaer, Marie Hargreave, Hedvig Nordeng

Maternal Medication Use and Childhood Cancer in Offspring – Systematic Review and Considerations for Researchers

Sarah Hjorth, Caroline H. Hemmingsen, Justine Bénévent, Anne Broe, Anton Pottegaard, Lina S. Mørch, Maarit K. Leinonen, Susanne K. Kjaer, Marie Hargreave*, and Hedvig Nordeng*

*Shared last authorship

Correspondence to Sarah Hjorth, PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, University of Oslo, Postboks 1068 Blindern 0316 Oslo, Norway (e-mail: s.h.andersen@farmasi.uio.no)
Author affiliations: PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, and PharmaTox Strategic Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway (Sarah Hjorth, Justine Bénévent, and Hedvig Nordeng); Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark (Caroline H. Hemmingsen, Susanne K. Kjaer, and Marie Hargreave); Department of Medical and Clinical Pharmacology, Toulouse Faculty of Medicine, Centre Hospitalier Universitaire de Toulouse, Toulouse, France (Justine Bénévent); Clinical Pharmacology, Pharmacy and Environmental Medicine, Department of Public Health, University of Southern Denmark, Odense, Denmark (Anne Broe, and Anton Pottegaard); Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark (Anne Broe); Cancer Surveillance and Pharmacoepidemiology, Danish Cancer Society Research Center, Copenhagen, Denmark (Lina S. Mørch); Data and Analytics, Information Services Department, Finnish Institute for Health and Welfare, Helsinki, Finland (Maarit K. Leinonen); Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway (Hedvig Nordeng).

This work was supported by PharmaTox QLS Visiting Scholarship (to JB); and European Research Council Starting Grant (grant number 639377 to HN).

Conflicts of interest: none declared

Running head: Maternal Medication Use and Childhood Cancer
Abstract

Cancer is an important cause of childhood mortality, yet the etiology is largely unknown. A combination of pre- and postnatal factors is thought to be implicated, including maternal medication use. We aimed to provide 1) a systematic review of peer-reviewed publications on associations between maternal medication use and childhood cancer with a focus on study design and methodology, and 2) suggestions for how to increase transparency, limit potential biases, and improve comparability in studies on maternal medication use and childhood cancer. We conducted a systematic search in the PubMed, Embase, Scopus, Cochrane, and Web of Science databases until June 8th 2020. Altogether, 112 studies were identified. The reviewed studies were heterogeneous in study design, exposure, and outcome classification. In 21 studies (19%), the outcome was any childhood cancer. Of the 91 papers that reported on specific types of cancer, 62% did not report the cancer classification system. The most frequently investigated medication groups were sex hormones (46 studies, excluding fertility medications), and anti-infectives (37 studies). Suggestions for strengthening future pharmacoepidemiological studies on maternal medication use and childhood cancer relate to choice of cancer classification system, exposure windows, and methods for identification of -, and control for, potential confounders.

Key words: Cancer, Child, Medications, Pharmacoepidemiology, “Prenatal exposure, Delayed Effects”

Abbreviations

ICCC: International Classification of Childhood Cancers

ICD: International Classification of Diseases
The worldwide incidence of childhood cancer is estimated at 140 per million person-years, and is increasing (1–3). The etiology of childhood cancer is largely unknown, but thought to be explained by both pre- and postnatal factors (4–6). The increasing incidence of childhood cancer could point to environmental risk factors that have changed over time (2–4).

The only fully established trans-placental chemical carcinogen is diethylstilbestrol (7). The research on diethylstilbestrol sparked an interest in the investigation of maternal medication use and risk of childhood cancers (8). In recent years, some studies (9–12), but not all (13,14), pointed to an association between maternal medication use and childhood cancer. A review of the literature before 1997 illustrated the heterogeneity in the literature in particular with regard to the applied cancer classification (8). Suggestions for methods have been provided for pharmacoepidemiological studies of medication-cancer associations in adults (15), but to our knowledge, guidance has not been provided for maternal medication-childhood cancer associations. Therefore, we aimed to provide 1) a review of peer-reviewed publications on associations between maternal medication use and childhood cancer with a focus on study design and methodology, and 2) suggestions on how to increase transparency, limit potential biases, and improve comparability in studies on maternal medication use and childhood cancer.

METHODS

Search strategy

We conducted a systematic search in the PubMed, Embase, Scopus, Cochrane, and Web of Science databases from database inception (1966 or earlier depending on database) until June 8th 2020 to address the following specific questions:

1. Classification of childhood cancer: How was the outcome classified (as any cancer, or according to specific diagnoses; specified by the International Classification of Diseases (ICD), the International Classification of Childhood Cancers (ICCC), or other)?
2. **Exposure windows**: What timing of maternal medication use was investigated (during pregnancy only, or including periods before pregnancy, or during breastfeeding)?

3. **Study design**: What study designs were used?

4. **Methods and statistics**: What statistical and epidemiological methods were used?

5. **Follow-up/age at case ascertainment**: What was the maximum age at follow-up (cohort studies)/age at case ascertainment (case-control studies)?

Reference lists of relevant reviews and included studies were screened to ensure complete coverage of the published literature. Our initial search used the search terms “child” AND “prenatal” AND “medication” AND “cancer” including relevant synonyms. However, this search proved too narrow, as more studies were identified from reference lists than from the search itself. Many such studies identified via reference lists did not include the term “medication” in their titles or abstracts. The search was therefore repeated using only “child” AND “prenatal” AND “cancer”, and relevant synonyms. An example of search terms and search strategy for the PubMed database can be found in Web Table 1.

References were imported into the reference management program Endnote (16), where duplicates were removed. The remaining references were imported to Rayyan QCRI (17), a platform for management of systematic review data. Title and abstract screening, as well as full text screening were performed independently by two reviewers (SH and CH). Any disagreement was solved by a discussion among all authors.

**Inclusion criteria**

Studies were considered eligible for inclusion if they fulfilled the following criteria for participants, exposures, comparators, outcomes, and study design. Participants were required to be children, defined as individuals under the age of 20 years. The exposure was restricted to maternal pre-pregnancy or pregnancy use of prescription or over-the-counter medication, as
identified in prescription data or from self-reported data. To maintain the study focus on therapeutic medications, studies on supplements (vitamins/minerals), and studies on use of illegal substances were not included. Studies that classified exposure as “any medication” were also excluded. Comparators were children born to mothers who did not use the specified medications. This included children born to healthy mothers (population comparators), children born to mothers with illnesses, but not receiving treatment (disease comparators), and children born to mothers who used other specified medications than the medication of interest for the study (active comparators). The outcome was childhood cancer. Studies that used children with other types of cancer as controls were excluded. Randomized controlled trials, cohort studies, and case-control studies were eligible for inclusion. Non-original studies (e.g. reviews and editorials), studies without a comparison group, cross-sectional studies, ecological studies, and animal studies were excluded, as were conference abstracts, study protocols, and pilot studies. No restrictions were applied as to study date or setting, but for resource reasons, the search was limited to peer-reviewed publications in English, French, or one of the Scandinavian languages.

Data extraction

Data items extracted from the included studies were decided a priori as follows: study design, setting, sample size, number of exposed, number of cases, age at end of follow-up/case ascertainment, exposure and outcome classification, statistical analysis, and adjustment variables. Data was extracted by SH and JB.

No systematic tool like GRADE (18) or ROBINS-I (19) was used to assess risk of bias in the individual studies. Instead, all eligible studies were assessed according to the pre-specified questions mentioned above, and discussed in the author group to identify suggestions for future research.
Data was grouped by medication exposure according to indication for use. Given the study aims, no synthesis of study findings was planned.

Post hoc sensitivity analyses

We did two sensitivity analyses post hoc to assess the robustness of the findings. The first was restricted to the most recent studies published between 2011 and 2020, as methodological developments over the years might have had an impact on the quality of the included studies. The second was a restriction to studies where childhood cancer was the main outcome, as analyses of main and secondary outcomes might differ systematically.

RESULTS

The literature search yielded 10,033 studies. After removal of duplicate records, 6,879 studies were left for title and abstract screening. Of these, 160 were relevant for full text assessment, and 78 were eligible for inclusion (7,9–14,20–90). An additional 34 studies were identified from reference lists of included studies and relevant reviews (91–124). See Figure 1 for flowchart according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (125).

Data was extracted from 112 studies performed between 1958 and 2020 originating from Europe (50 studies), North America (48 studies), Asia (six studies), South America (four studies), and Australia (two studies). In addition, there were two trans-continental studies (Table 1).

Classification of childhood cancer

In 21 studies (19%), the outcome was a composite endpoint of ‘any childhood cancer’ (Table 1). An additional 13 studies (12%) reported results for any childhood cancer, in addition to specific types of cancer. Of the 91 papers that reported on specific types of cancer, 62% did
not report the cancer classification system. In the remaining papers, the most commonly reported classification system was the ICD. Some studies had their classification system labelled as “other”, for instance because they used more detailed classifications by disease staging (74). Several studies (32%) reported more than one group or specific type of cancer (range 1 to 16). The most commonly reported main diagnostic groups were leukemias (25 studies) and central nervous system tumors (14 studies). The most commonly investigated subgroups of childhood cancer were acute lymphoblastic leukemias (26 studies), acute myeloid leukemias (15 studies), and neuroblastomas (14 studies).

Exposure

Several studies (40%) investigated more than one medication group (range 1 to 13). The most frequently investigated medication groups were sex hormones (46 studies, excluding fertility medications), anti-infectives (37 studies), and fertility medications (30 studies) (Web Table 2).

In 15% of studies, more than one exposure window was investigated. The majority (56%), investigated maternal medication use anytime during pregnancy, whereas only 6% investigated use in specific trimesters of pregnancy. Many studies (37%) investigated use prior to pregnancy, mainly use of fertility medications or sex hormones. Though most studies distinguished between use before, during, and after pregnancy, 14% of studies investigated an exposure window including more than one peri-pregnancy period (e.g. pre-pregnancy and pregnancy, or pregnancy and breastfeeding). Analyses by dose or duration were available in 10 studies (9%).

Exposure was ascertained by maternal retrospective self-report in 70 studies (63%), from routinely collected health data in 43 studies (38%), and from maternal prospective self-report in four studies (4%). Five studies (4%) used two sources of exposure ascertainment. A total of
27 studies (24%) accounted for potential exposure misclassification; 24 of these had ascertained exposure by maternal retrospective recall.

Study design

The majority of studies were case-control studies (81 studies, 72%), 21% were cohort studies, and the remainder were case-cohort studies (Table 1). Two studies had less than 100 participants, and 21% of studies had more than 10,000 participants. In contrast to the high number of participants, 61% of the studies had less than 100 exposed, and 29% of studies had less than 100 cases in their main analysis. In nine studies, there were either no cases, or no exposed cases or controls, for the main analysis. Almost all studies used population comparators (90%), but disease comparators were used in eight studies, and active comparators were used in two studies. Some studies used more than one type of comparator.

Statistical and epidemiological methods

The most commonly used statistical analysis was logistic regression (58% of studies). Cox regression was used in 15% of studies, and two studies used Poisson regression. Purely descriptive analysis was used in 11% of the studies, mainly when the sample size was limited, or when cancer was not the main outcome of the study. The studies using other statistical methods (14%) were predominantly studies performed in the 1970’s and 1980’s. An example of a statistical method used in these studies is the Mantel-Haenszel method for stratified analysis (44,54,103). Whereas 73% of studies included matching or adjustment for one or more potential confounding factors, 7% had no adjustment, and 17% adjusted for at least one potential intermediate factor in their main analysis. The most frequently included intermediate factors were gestational age at birth (11 studies) and birth weight (10 studies).

Follow-up/age at case ascertainment
The mean upper limit of follow-up/age at case ascertainment for the children was 14 years. In 35% of studies, the age range was 0–14 years, and in 18% of studies, the age range was 0–19 years. The remaining studies used other age ranges, either determined by the peak incidence of the type of cancer investigated (e.g. less than 5 years of age for hepatoblastoma (104)), or by data availability (e.g. less than 9 years in a study using data from an existing birth cohort (82)).

Post hoc sensitivity analyses

In the first sensitivity analysis of 38 studies published between 2011 and 2020, ‘any childhood cancer’ was used as the outcome by a larger proportion of studies than in the studies published before 2011 (26% compared to 15%), and more studies on specific cancer types reported on the classification system used (14 of 28 studies on specific cancer types, 50% compared to 33% in the studies published before 2011). A lower proportion of studies ascertained exposure by maternal retrospective report (39% compared to 74%), but a larger proportion adjusted for at least one intermediate factor in a main analysis (29% compared to 11%).

The second sensitivity analysis excluded seven studies that did not have childhood cancer as their main outcome. The results were largely similar to the main analysis, except that a lower proportion of studies (14 of 105 studies, 13%) had investigated childhood cancer as a composite outcome.

DISCUSSION

Main findings

In this review of 112 studies on maternal use of medication before or during pregnancy and childhood cancer in offspring, 19% used “any cancer” as outcome. A majority of the studies that investigated specific types of cancer did not report the cancer classification system
In most studies (56%), the exposure window was anytime during pregnancy, but 15% of studies investigated more than one exposure window. A majority of studies were case-control studies (72%). Most studies (73%) accounted for potential confounding by matching and/or adjustment, but 17% of studies adjusted for at least one intermediate factor in a main analysis. The mean upper limit of follow-up/age at case ascertainment was 14 years.

Limitations

Of 112 included studies, 34 were identified from reference lists. Half of these investigated fertility medications, whereas the other half did not have any pattern of common characteristics. This could indicate that the search terms included in the literature search were not exhaustive for fertility medications. As it takes some time from a study is published until it can be cited, we may have an incomplete coverage of the literature from late 2019 and early 2020. In addition, we had to exclude five studies due to language restrictions. For these studies, we have not been able to assess eligibility in a full-text reading, thus we do not know whether they fulfilled the inclusion criteria for the present review. Authors who do not have English as their first language, are more likely to publish studies with negative findings in local, non-English journals (126). It is not known whether methods or reporting differ systematically between studies published in English and studies published in local journals.

Considerations for future research

Reporting of cancer types according to the ICCC when possible. It has been argued that “any cancer” should not be used as an outcome in studies of medication-cancer associations in adults (15). In brief, this is because cancer is a heterogeneous disease, and no known carcinogens increase the risk of every type of cancer (15). This also applies to childhood cancer, arguing for investigations of specific types of cancer when possible. Though limited sample sizes can render a detailed outcome classification impracticable, it should be noted that the
grouping of outcomes might in fact reduce study precision. This seemingly contra-intuitive claim stems from the fact that investigating several cancer types in one group will introduce heterogeneity leading to increased variance and therefore wider confidence intervals (127). However, for new and rarely used treatments, e.g. biologics, studies with any cancer as the outcome may be the only option, and might still provide important reassurance or serve to flag initial safety signals. Yet, even if the risk estimate for overall cancer is not increased, this does not rule out that the studied medication increases the risk of a specific type of cancer. Therefore, results indicating null-associations from studies using any cancer as the outcome should be reported and interpreted with caution. On the other hand, if studies using imprecise outcomes do identify a signal, this could warrant further investigation. To prioritize among multiple signals for further investigation, it can be useful to employ methods such as empirical Bayes shrinkage, that adjusts observed estimates of association for random variation, and is thought to reduce the number of false positive findings (128). In studies with only a few exposed cases, it might be beneficial to apply a lesson learned from teratology (129,130), and report on any patterns of specific cancer types, even if statistical analysis is only feasible for a combination of all cancers. It was the presence of a specific type of cancer that first suggested the trans-placental carcinogenicity of diethylstilbestrol (7), just as the specific patterns of syndromes or malformations flagged the potential teratogenicity of thalidomide and valproic acid (129).

For studies that have the statistical power to report risk estimates for specific types of childhood cancer, the next question is what classification scheme to use. To facilitate pooling of data in meta-analyses, a standardization of the outcome classification would be helpful (8). The need for standardization led to the development of the ICCC in 1987 (131). Childhood cancers differ from cancers in adults by being embryonal in type and arising in organ systems that do not respect the traditional sites that are used in classification of adult cancers (131).
Therefore, ICCC is primarily based on the type and behavior of the cancer cells, with some site-based groupings to facilitate comparisons to ICD. In addition, the most common childhood cancers have individual codes (131). This makes the ICCC preferable to ICD for classification of childhood cancers. The International Classification of Diseases for Oncology, while not adapted specifically to childhood cancers, is still preferable to the ICD, as it takes morphology into account. From a pharmacological point of view, it is plausible that morphologically similar cancers will react to medication exposures in a similar fashion, whereas cancers of the same site may not (15). Hence, researchers should consider using the ICCC classification system when possible. The International Classification of Diseases for Oncology recommended for use in cancer registries could be an alternative in situations where mapping to the ICCC is not feasible.

*Biological plausibility should guide the exposure definition, and exposure windows.* Both preclinical findings (e.g. from in vitro or animal studies) and signals from previous human studies could be helpful to inform the exposure definition. If, for instance, the hypothesis is that hormonal disruption from exposure to oral contraceptives plays a causal role in the development of childhood non-lymphoid leukemias (9), it may be insufficient to study oral contraceptives as a group. Most oral contraceptives contain an estrogen analogue and progestin, whereas others contain progestin only. Findings would then be attenuated for “any oral contraceptives”, if estrogen was the causally important substance. Further, analyses by dose or duration of exposure are helpful when feasible, as there may be threshold levels for effect (15).

Once the exposure has been defined, the relevant exposure window should be chosen. As opposed to teratology, where the relevant exposure window for most malformations is the first trimester (129), relevant exposure window(s) for childhood cancers are largely unknown. One proposed relevant exposure window is immediately after fertilization, where the epigenome is thought to be highly sensitive to environmental factors (132). In the present review, different
approaches were seen across studies, mainly ranging from a year before conception until the end of pregnancy. One study considered maternal exposure when the mother was a fetus (84). Seeing that the ovaries and egg cells for future offspring are formed in fetal life (133), it is possible that medications taken throughout the life-course could affect the egg cells. Some studies investigated maternal medication use during pregnancy and/or breastfeeding (30,63,66,89,100). In studies where the exposure window extends beyond the prenatal period, the outcome can happen during the exposure window, thus potentially introducing immortal time bias (134). Immortal time bias can be avoided; for example, by moving the start of follow-up or case ascertainment to the end of the exposure window (134). The use of study design diagrams as advocated by Schneeweiss et al. (135), could also help to clarify the timing of eligibility -, exposure -, covariate -, and outcome assessment in the study. In Web Figure 1, we provide an example of a study design diagram. Because the relevant exposure window is uncertain, it could be beneficial to investigate several exposure definitions within one study (i.e. ever before pregnancy, 1 year before pregnancy, during pregnancy, during the first trimester etc.). Another argument for the use of several different exposure definitions is specific to registry- or claims based studies. In these studies, exposure in a given time window, for example the first trimester, is not solely driven by prescriptions filled in the time window. Prescriptions filled immediately before the time window may also contribute, if the dispensed medications cover part of the time window (136).

Regardless of the data source, authors should consider the risk of exposure misclassification. With prospective exposure ascertainment, the misclassification is often (but not always) non-differential, whereas this cannot be assumed for retrospective maternal recall (137). Probabilistic sensitivity analysis is a useful tool to assess the potential impact of misclassification (138).
Use the new developments in confounder assessment and control. Most studies in the present review addressed potential confounding through matching, adjustment, or both. Some, mainly newer, studies adjusted for potential intermediate factors such as gestational age and birth weight. In the best case scenario, adjusting for intermediate factors can preclude an estimation of the total effect of the exposure (139). However, it may potentially introduce collider-stratification bias from unmeasured variables associated with both the intermediate factor and the outcome (139). Use of Directed Acyclic Graphs could help ensure that adjustments are not made for intermediate factors (140). In Web Figure 2, we present an example of a Directed Acyclic Graph. In many studies, the limited sample size can pose additional challenges for confounder adjustment. To avoid overfitting the models, serial change-in-estimate approaches for variable selection (141), or confounder summary scores can be used (142). Propensity score methods may be used if the exposure is more prevalent than the disease, because propensity scores are constructed from regression models with exposure as the dependent variable and covariates as independent variables (142).

Another method to estimate confounding in etiological epidemiology is by introducing a negative control (142,143). A negative exposure control that has been suggested in perinatal pharmacoepidemiology is maternal medication use before pregnancy (142). However, as stated above, it cannot be ruled out that maternal pre-pregnancy exposure can affect the risk of childhood cancers, and so other negative controls should probably be preferred. Maternal medication use after birth can be used as a negative control in non-communicable diseases, especially if maternal medication use is assessed when the breastfeeding has ceased. Paternal medication use while the mother is pregnant has been proposed to assess confounding by unmeasured genetic factors, or other unknown factors associated with medication use (144). If the main concern is confounding by the underlying maternal illness, the choice of disease- or active comparators could be considered.
A majority of the studies in the present review used population comparators. This is a good choice in situations where the indication for medication use is not thought to influence cancer risk (e.g. contraceptives). However, if the underlying maternal disease has been linked to an increased risk of childhood cancer (e.g. autoimmune disease, HIV-infection (145)), use of population comparators may be inadequate. Here a disease comparator could instead be used, comparing illness treated with medications to the same illness managed without medication.

An active comparator (i.e. comparing different substances used for the same indication, or comparing mono- and polytherapy) should be considered if a suitable comparator exists (15). Sibling comparators are primarily useful if the main source of confounding is thought to be inherited genetic risk (142). Provided that life-course maternal exposure to medications can affect risk of childhood cancers, carryover effects between siblings should be considered (146).

Compared to the choice of epidemiological methods, the choice of statistical methods is less important. That is because childhood cancer, in particular specific types of cancer, occur so rarely that estimates of rate ratios and risk ratios will be virtually identical (147).

Conclusion

Studying associations between maternal medication use and childhood cancer is methodologically challenging. This systematic literature review showed that such studies are largely heterogeneous in their study design, exposure, and outcome classification. To improve the transparency, limit potential biases, and improve comparability of future studies, we propose three points of consideration bridging the fields of prenatal pharmacoepidemiology and cancer epidemiology. The points include 1) investigating specific types of childhood cancer according to the ICCC-classification when possible, or as a minimum stating the classification system used, 2) carefully considering relevant exposure windows, including whether several exposure
windows should be investigated, and 3) using appropriate methods for identification of potential covariates (i.e. Directed Acyclic Graphs), and control of confounding (e.g. disease comparators, active comparators, negative controls).

**Web materials**

Web Table 1. Example of Search Terms and Search Strategy for the PubMed Database.

Web Table 2. Characteristics of the Included Studies by Medication Group.

Web Figure 1. Example of Study Design Diagram.

Web Figure 2. Example of Directed Acyclic Graph.

**ACKNOWLEDGEMENTS**

**Author contributions**

Hedvig Nordeng conceptionalised the study and all authors contributed to the study design. Sarah Hjorth, Hedvig Nordeng, and Caroline H. Hemmingsen developed the search strategy. Sarah Hjorth and Caroline H. Hemmingsen performed the literature search and screening for eligibility. Sarah Hjorth and Justine Bénévent extracted the data from the studies. Sarah Hjorth prepared the original draft and all other authors critically revised the work.

**Funding**

This work was supported by PharmaTox QLS Visiting Scholarship (to JB); and European Research Council Starting Grant (grant number 639377 to HN).
Data availability

All relevant data are available within the article and Web material.

Conflicts of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

References


29. Chaparro M, Verreth A, Lobaton T, et al. Long-Term Safety of In Utero Exposure to Anti-TNFα Drugs for the Treatment of Inflammatory Bowel Disease: Results from the Multicenter European TEDDY Study: *Am. J. Gastroenterol.* 2018;113(3):396–403.


Table 1. Summary Characteristics of 112 Included Studies on Associations Between Maternal Medication Use and Childhood Cancer.

<table>
<thead>
<tr>
<th>Study characteristic</th>
<th>n studies (N=112)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of publication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1971</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1971–1980</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>1981–1990</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>1991–2000</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>2001–2010</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>2011–2020</td>
<td>38</td>
<td>34</td>
</tr>
<tr>
<td>Continent of study conduct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Asia</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Australia</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Europe</td>
<td>50</td>
<td>45</td>
</tr>
<tr>
<td>North America</td>
<td>48</td>
<td>43</td>
</tr>
<tr>
<td>South America</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Trans-continental</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Number of medication groups investigated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>67</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>4+</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>Cancer classification system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any cancer</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>International Classification of Childhood Cancers</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>International Classification of Diseases</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>International Classification of Diseases for Oncology</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Not specified</td>
<td>56</td>
<td>50</td>
</tr>
<tr>
<td>Study design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-control</td>
<td>81</td>
<td>72</td>
</tr>
<tr>
<td>Case-cohort</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Cohort</td>
<td>24</td>
<td>21</td>
</tr>
<tr>
<td>Randomized controlled trial</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Exposure ascertainment^b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal retrospective self-report</td>
<td>70</td>
<td>63</td>
</tr>
<tr>
<td>Maternal prospective self-report</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Routinely collected health data</td>
<td>43</td>
<td>38</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis Type</td>
<td>Count 1</td>
<td>Count 2</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Cox regression</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>Descriptive analysis</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>65</td>
<td>58</td>
</tr>
<tr>
<td>Poisson regression</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Maximum follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5 years</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>5–9 years</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>10–14 years</td>
<td>51</td>
<td>46</td>
</tr>
<tr>
<td>≥15 years</td>
<td>33</td>
<td>29</td>
</tr>
<tr>
<td>Not specified</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Comparator group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population comparator</td>
<td>101</td>
<td>90</td>
</tr>
<tr>
<td>Disease comparator</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Active comparator</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;100</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>100–499</td>
<td>34</td>
<td>30</td>
</tr>
<tr>
<td>500–999</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>1000–9999</td>
<td>29</td>
<td>26</td>
</tr>
<tr>
<td>10,000+</td>
<td>24</td>
<td>21</td>
</tr>
<tr>
<td>Number of cases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;100</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>100–499</td>
<td>48</td>
<td>43</td>
</tr>
<tr>
<td>500–999</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>1000+</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Number of exposed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;100</td>
<td>68</td>
<td>61</td>
</tr>
<tr>
<td>100–499</td>
<td>42</td>
<td>38</td>
</tr>
<tr>
<td>500–999</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1000+</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>Not stated</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Adjustment or matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No adjustment or matching</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Only confounders</td>
<td>82</td>
<td>73</td>
</tr>
<tr>
<td>At least one intermediate factor</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>Not specified</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

*The numbers do not add up, as one study presented results from two countries. In one country, the design was case-cohort, in the other case-control.
The numbers do not add up, as five studies used two sources of exposure ascertainment.

The numbers do not add up, as two studies had more than one type of comparators.

In one study the comparator group included cousins of exposed, in another study it included children of unexposed mothers and exposed fathers, and in the third study it included children of women who used the medication prior to pregnancy only.

Numbers reported for the main analysis. If main exposure or outcome were not stated, all exposures and outcomes were considered equal, and the study could end up in more than one category.
Figure legends

Figure 1. Flowchart according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (125)
Paper V

Prenatal exposure to nitrofurantoin and risk of childhood leukaemia – a registry-based cohort study in four Nordic countries

Sarah Hjorth, Anton Pottegård, Anne Broe, Caroline H. Hemmingsen, Maarit K. Leinonen, Marie Hargreave, Ulrika Nörby, Hedvig Nordeng

Submitted for publication.
Prenatal exposure to nitrofurantoin and risk of childhood leukaemia – a registry-based cohort study in four Nordic countries

Sarah Hjorth (ORCID 0000-0003-2841-5868)1, Anton Pottegård (ORCID 0000-0001-9314-5679)2, Anne Broe (ORCID 0000-0002-4149-8808)2,3, Caroline H. Hemmingsen (ORCID 0000-0001-5420-1981)4, Maarit K. Leinonen (ORCID 0000-0002-7631-4749)5, Marie Hargreave (ORCID 0000-0001-6821-9242)4, Ulrika Nörby (ORCID 0000-0003-0396-5986)6, Hedvig Nordeng (ORCID 0000-0001-6361-2918)1,7

1PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, and PharmaTox Strategic Initiative, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway

2Clinical Pharmacology, Pharmacy and Environmental Medicine, Department of Public Health, University of Southern Denmark, Odense, Denmark

3Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark

4Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark

5Data and Analytics, Information Services Department, Finnish Institute for Health and Welfare, Helsinki, Finland

6Health and Medical Care Administration, Region Stockholm, Stockholm, Sweden

7Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway

Correspondence

Sarah Hjorth
Department of Pharmacy, University of Oslo
Postboks 1068 Blindern
0316 Oslo, Norway
E-mail: s.h.andersen@farmasi.uio.no

Word count: 2925
Abstract

Background: Studies have suggested increased risks of childhood leukaemia after prenatal exposure to antibiotics, particularly nitrofurantoin. However, these findings may be related to the underlying maternal infection. This multinational study aimed to investigate the association between prenatal nitrofurantoin exposure and childhood leukaemia while accounting for maternal infection.

Methods: In a population-based cohort study of children born in Denmark, Finland, Norway or Sweden from 1997 to 2013, prenatal exposure to nitrofurantoin or pivmecillinam (active comparator) was ascertained from national prescription registries. Childhood leukaemia was identified by linkage to national cancer registries. Poisson regression was used to estimate incidence rate ratios (IRR), and incidence rate differences (IRD) with inverse probability of treatment weights applied to account for confounding.

Results: We included 44,091 children prenatally exposed to nitrofurantoin and 247,306 children prenatally exposed to pivmecillinam. The children were followed for 9.3 years on average (standard deviation 4.1). There were 161 cases of childhood leukaemia. The weighted IRR for prenatal nitrofurantoin exposure when compared to pivmecillinam was 1.34 (95% CI 0.88;2.06), corresponding to an IRD of 15 per million person-years. Higher point estimates were seen for first and third trimester exposure. There was no evidence of a dose-response relationship.

Conclusions: Prenatal exposure to nitrofurantoin was not substantially associated with childhood leukaemia, although a slightly elevated IRR with confidence intervals including the null was observed, corresponding to a small absolute risk. The lack of a dose-response relationship, and a clear biological mechanism to explain the findings, suggest against a causal association.

Keywords: leukaemia, nitrofurantoin, “prenatal exposure, delayed effects”

Key messages:

- Previous studies on prenatal exposure to antibiotics and childhood cancer have used unexposed comparators, which may introduce confounding from underlying maternal infection. Furthermore, findings based on a non-user comparator are of limited clinical value seeing that infections during pregnancy should always be treated.
• In a multinational cohort study, we investigated the association between prenatal exposure to nitrofurantoin and childhood leukaemia when compared to pivmecillinam, another antibiotic used for the same indication.

• There was no substantial association between prenatal exposure to nitrofurantoin and childhood leukaemia, albeit a slightly elevated incidence rate ratio with confidence intervals overlapping the null. The current body of evidence, including the lack of clear biological mechanism of action, speaks against interpreting the association as causal.

Background

Cancer is the second most common cause of death in children in affluent countries with leukaemia as the most common type (1). Childhood cancer incidence in Scandinavia is 160 per million per year, and leukaemia accounts for a third of cases (1). The aetiology of childhood leukaemia is largely unknown, but is thought to involve a combination of pre- and postnatal factors (2,3). A number of studies have investigated potential associations between prenatal exposure to antibiotics and risk of childhood leukaemia when compared to no exposure with heterogeneous findings (4–8). The only large study to investigate at the individual antibiotic substance level and specific types of childhood cancers, found associations with leukaemia for exposure to some commonly used antibiotics, including nitrofurantoin (HR 1.56, 95% CI 1.02-2.37), when compared to unexposed children (7). However, maternal infection has also been proposed to affect risk of childhood leukaemia (2). Therefore, it will be important to also compare to children who were exposed to maternal infection, but received another antibiotic treatment, i.e. use an active comparator (9). This is particularly relevant for urinary tract infections in pregnancy, as treatment is always indicated, even in asymptomatic cases (10,11). Clinicians and pregnant women therefore need to choose between different antibiotic treatment regimens, rather than between antibiotics or no treatment. Up to 10% of pregnancies are affected by bacteriuria, making it one of the most common pregnancy complications (10). The Nordic countries have similar clinical guidelines for treatment of bacteriuria in pregnancy with pivmecillinam as the first line treatment and nitrofurantoin as an equivalent or second line treatment depending on country (Supplementary information). In the above-mentioned study, based on Danish and Swedish data, pivmecillinam was not associated with childhood leukaemias (HR 1.11, 95% CI 0.83;1.48) (7). Therefore, pivmecillinam could be used as an active comparator to nitrofurantoin
in a Nordic setting. The objective of the present study was to investigate the risk of childhood leukaemia after prenatal exposure to nitrofurantoin compared to pivmecillinam.

**Methods**

The design was an active comparator cohort study using individual level data from national registries in Denmark, Finland, Norway, and Sweden. We estimated incidence rate ratios (IRRs) and incidence rate differences (IRDs) for childhood leukaemia by comparing the incidence among children who were prenatally exposed to nitrofurantoin and children who were prenatally exposed to pivmecillinam.

**Data sources**

We used data from Nordic Cancer Registries, Prescription Registries, Medical Birth Registries, and Patient Registries. In addition, we used data from the Danish Civil Registration System, and from the Cause of Death Registries in Finland, and Sweden. In the Nordic countries, the registries cover all residents. Linkage between registries is made possible by unique personal identification numbers.

The Nordic Cancer Registries have recorded incident cases of cancer since 1942-1958, depending on country, with a coverage of close to 100% (12–15). See Supplementary information for an overview of the periods covered by the registries in each country.

The Nordic Prescription Registries provide information on all prescriptions redeemed at pharmacies by patients in ambulatory care (16). Medications are categorised according to the Anatomical Therapeutic Chemical (ATC) Classification System (17). The Prescription Registries were established in 1993-2005 (16).

The Nordic Medical Birth Registries cover a wide range of information on ante- and perinatal factors, as well as some background information on mother, father, and infant (18). Notification to the Medical Birth Registries has been mandatory for live births since 1967-1987, depending on country (18).

The Nordic Patient Registries record all diagnoses and procedures in government-owned hospitals and outpatient clinics. The Patient Registries have contained individual level data since 1977-2008, depending on country (19–22).
The Nordic Cause of Death Registries record dates of death (23,24); the Civil Registries record dates of death and migration for all citizens (25).

**Study sample**

The initial study population was live-born singletons registered in the Nordic Medical Birth Registries. We included all children born between 1997 and 2013 in Denmark and Finland, and between 2007 and 2013 in Norway and Sweden, who were prenatally exposed to either nitrofurantoin or pivmecillinam. Exclusion criteria was prenatal exposure to both pivmecillinam and nitrofurantoin, or inability to determine exposure to medications in pregnancy because of 1) missing maternal or child ID number, or 2) missing or unrealistic (>45 weeks) gestational age at birth, and thus missing information about the start of pregnancy. Follow-up continued until the first cancer diagnosis, death (handled as competing risk), emigration, the child’s 20th birthday, or December 31st 2017, whichever came first.

**Exposures**

Exposure was maternal filling of one or more prescriptions of the medication of interest during pregnancy, as recorded in the Nordic Prescription Registries. Pivmecillinam was identified by ATC code J01CA08, nitrofurantoin by ATC code J01XE01.

Analysis was also performed for number of prescriptions filled in pregnancy (1 and 2+), and for timing of prescription fills (first-, second-, or third trimester, defined as days 0-89; 90-179 or birth, if the child was born during the second trimester; and 180 to birth).

In the Nordic countries, antibiotics are only available on prescription, so the sensitivity of the exposure classification is expected to be high.

**Outcomes**

The primary outcome was any leukaemia as recorded in the Nordic Cancer Registries (International Classification of Childhood Cancer, third edition [ICCC-3] site group 1, codes 011-015) (26). As secondary outcomes, leukaemia was specified according to the most common types of childhood leukaemia, namely lymphoid leukaemia (ICCC-3 code 011), and acute myeloid leukaemia (ICCC-3 code 012) (26).

A validation study in the Finnish Cancer Registry showed 95.7% completeness of childhood leukaemia registrations (13).
Covariates

Covariates were chosen a priori using subject knowledge and Directed Acyclic Graphs (Supplementary Figure 3) (27). As the outcomes were expected to be very rare, a propensity score-based approach to confounder adjustment was chosen. The propensity score was estimated using logistic regression (28). As recommended, the propensity score included both potential confounders and predictors of the outcome (28). The following covariates were included: Calendar year at birth (numerical), maternal age (numerical), parity at start of pregnancy (0, 1, 2, 3, 4+), maternal history of cancer before pregnancy (yes/no), prescription fills for immunosuppressants, systemic corticosteroids, and systemic antibiotics (apart from pivmecillinam and nitrofurantoin) in the year before start of pregnancy (in Finland three months before the start of pregnancy) (yes/no, used as a proxy for susceptibility to infections), maternal smoking status during first trimester (yes/no), and child sex. Information on covariates was obtained from the Medical Birth Registries, the Prescription Registries, and the Cancer Registries. Smoking in pregnancy is highly correlated to socio-economic position in the Nordic countries (29–31), and was considered as a proxy for socio-economic position.

Statistical analysis

Baseline characteristics and antibiotic utilisation (number of fills in pregnancy, trimester of fills, and number of fills for other antibiotics in pregnancy) were compared between pivmecillinam and nitrofurantoin exposed. Analyses were conducted separately for each country. To allow for different lengths of follow-up, we used Poisson regression to obtain IRRs and IRDs. Crude estimates with 95% confidence intervals (95% CIs) were obtained using generalised linear models with a log link. To account for confounding, propensity scores were estimated using logistic regression, and used as inverse probability of treatment weights (IPTW). IPTW is recommended when using active comparators (32), and will (under the assumption of no residual confounding) answer the question: what would the incidence rate have been if everyone had been treated with nitrofurantoin as opposed to if everyone had been treated with pivmecillinam? The non-overlapping regions of the propensity score were trimmed (59 children excluded), with the baseline characteristics of the mothers to the excluded children found to be similar to the remaining cohort (data not shown for reasons of confidentiality due to the small numbers). The balance of the weights was checked using standardised mean differences (33). Covariates were considered balanced if the standardised mean differences were below 0.1 (33). Generalised linear models with robust standard errors were used in the weighted dataset to obtain weighted
estimates with 95% CIs. Robust standard errors were chosen to account for weighting and clustering, as more than one child of the same mother could be included in the cohort.

Missing data on covariates (smoking and parity), was seen for 6.9% of the study sample. Under the assumption that data were missing at random, missing data was imputed using multiple imputation by chained equations (34) with 50 datasets created. As done in previous studies using Poisson regression, the imputation model included exposure, outcome, all covariates, and the cumulative Nelson Aalen hazard function for leukaemia (35).

Fixed-effects meta-analysis was used to pool the results from Danish, Finnish, Norwegian, and Swedish data, assuming a common treatment effect across the Nordic countries (36). In a Nordic context, fixed-effects meta-analysis has been shown to yield results similar to those obtained from pooling individual data (37). Heterogeneity was examined using \( I^2 \). For analyses were there were zero exposed or unexposed cases in one or more countries, results were combined using multilevel mixed-effects Poisson regression with random effect terms for the variance components (38). This method has been shown to yield results with less bias than standard meta-analysis techniques in meta-analyses of incidence rate data with zero counts (38).

**Supplementary analyses**

We performed several pre-planned supplementary analyses to assess the robustness of our findings.

First, to estimate the impact of potential unmeasured confounding, for instance by severity of infection, we calculated the e-value. The e-value was calculated to examine how strong any unmeasured confounding should be to explain the observed effect to the extent it reduces the observed point estimate to the null (39).

Second, to compare with the results from the imputed dataset, we did a complete case analysis.

Third, we started follow-up at 1 year of age, as infant leukaemia may have a different aetiology from later onset childhood leukaemia (2).

Fourth, to strengthen comparability between nitrofurantoin exposed and pivmecillinam exposed, we restricted the sample to children who were unexposed to other systemic antibiotics *in utero* than pivmecillinam or nitrofurantoin.
Fifth, we restricted the sample to women who had a contact with the health care system (diagnosis or prescription fill) before pregnancy. This was expected to ensure that women were present in the country throughout pregnancy. However, this analysis was also expected to restrict the sample to women with more comorbidities than the general population.

Results

Among 3,135,376 live-born children, we identified 312,026 (10.0%) children prenatally exposed to nitrofurantoin and/or pivmecillinam (Figure 1). Of these, 20,570 children, corresponding to 7.7% of pivmecillinam exposed, and 31.8% of nitrofurantoin exposed, were excluded due to exposure to both medications. We included 44,091 (1.4%) prenatally nitrofurantoin exposed, and 247,306 (7.9%) pivmecillinam exposed. Pivmecillinam exposed children were more often exposed to maternal smoking in pregnancy and to more than one treatment with the antibiotic medication of interest (Table 1). The prevalence of prenatal exposure to other systemic antibiotics was similar between nitrofurantoin exposed and pivmecillinam exposed. All covariates included in the IPTW had a standardised mean difference below 0.1 after weighting. An exception was the model for Finland, where birth year was not balanced and hence added to the outcome model. In all countries, the weights had a mean of 1.00. The highest weight in any country was 3.64.

The included children were followed for 984,784,867 person-years in total, with each child followed for 9.3 years on average (standard deviation 4.1). During follow-up there were 161 cases of leukaemia, 134 of which were lymphoid leukaemia. Compared to pivmecillinam, prenatal nitrofurantoin exposure was associated with a slightly elevated IRR for leukaemia (fixed effects IRR 1.34, 95% CI 0.88 to 2.06, $I^2=0$), albeit with wide confidence intervals that included the null (Table 2). This corresponds to an IRD of 15 per million person-years. For almost all analyses, there were zero exposed or unexposed cases in at least one country. Therefore, remaining analyses were combined using mixed-effects Poisson models. The statistical precision was limited, but we found no evidence of a larger increase in leukaemia incidence after two or more prenatal nitrofurantoin exposures when compared to two or more pivmecillinam exposures. Trimester specific analyses pointed to increased incidences of leukaemia after first and third trimester nitrofurantoin exposure, but not after second trimester exposure.
In the secondary analyses on lymphoid leukaemia, nitrofurantoin exposure was associated with an IRR of 1.34 (95% CI 0.83 to 2.17). Analyses on acute myeloid leukaemia were not feasible due to the small number of cases.

**Supplementary analyses**

In general, the statistical precision was low in the supplementary analyses, but results corresponded with results from the main analysis (Supplementary Tables 1-4).

For the association between any prenatal nitrofurantoin exposure and childhood leukaemia, the e-value was 2.02, meaning that any unmeasured confounder would have to have an association of 2.02 with both the exposure and the outcome to fully explain the IRR.

**Post hoc analysis**

In a *post hoc* analysis to estimate the extent of confounding by indication, we compared children prenatally exposed to nitrofurantoin to 2,254,684 children who were unexposed to antibiotics in pregnancy. The same variables were included in the propensity scores as in the main analysis, but the propensity scores were used as standardised mortality ratio weights, as recommended for population comparators (32). Findings were similar to the results from the active comparator design, IRR 1.23 (95% CI 0.87;1.76), IRD 14 per million person-years (Supplementary Table 5).

**Discussion**

In this active comparator study of 291,397 children from the Nordic countries, we found no substantial association between prenatal exposure to nitrofurantoin and childhood leukaemia, although we observed a slightly elevated IRR with wide confidence intervals overlapping the null. There was no evidence of a dose-response relationship. The results were stable in supplementary analyses.

Our findings are in accordance with the previous study on Danish and Swedish data that used population comparators (7), although point estimates in the present study are lower. Our study sample partially overlaps with the sample from that study by three of the included years from Sweden and 11 of the included years from Denmark.

We could only estimate associations between trimester specific exposure and childhood leukaemia with imprecision, but first and third trimester exposure were associated with the largest
estimates of association. Previous studies did not investigate nitrofurantoin exposure by trimester, but one Canadian study has investigated exposure to any antibiotic by trimester (8). That study found an increased risk of childhood acute lymphoid leukaemia after first trimester exposure to antibiotics (HR 1.5, 95% CI 0.9; 2.5), but not after second or third trimester exposure (both HR 0.8) (23).

A biological mechanism that could explain the findings is lacking, as there has not been evidence to suggest that even high doses of nitrofurantoin can cause leukaemia in animals (40).

Strengths of this study include the population-based design with information on all exposed pregnancies across four Nordic countries and follow-up in nationwide registries of high completeness. Another strength is the use of an active comparator design to account for confounding by maternal infection and aid clinical decision-making.

However, residual confounding must be considered. It is possible that there are unmeasured systematic differences between pregnancies that are exposed to different antibiotic treatment regimens for urinary tract infection. Whereas maternal infection in pregnancy has been proposed as a risk factor for childhood leukaemia (2), few studies have investigated this for maternal urinary tract infections (41). A systematic review from 2020 identified two studies on prenatal exposure to maternal urinary tract infection and childhood lymphoid leukaemia (41), where one found an odds ratio of 0.7 (95% CI 0.4; 1.2) and the other an odds ratio of 1.9 (95% CI 1.0; 3.9). The study that found a point estimate above 1 used information from medical records, whereas the other study relied on retrospective maternal report where recall bias could occur. Therefore, an association between maternal urinary tract infection and childhood leukaemia cannot be ruled out. The biological mechanism of action is speculative, but maternal urinary tract infections can affect the foetal environment, as they are implicated in preterm births and low birth weight (10). Though findings from the post hoc analysis suggested against important confounding by maternal urinary tract infection, confounding by other unmeasured factors cannot be ruled out. Calculation of the e-value showed that a confounder with an association of 2.02 with both the exposure and the outcome could fully explain the IRR in the present study.

Another limitation is potential misclassification of the exposures, as non-adherence to prescribed antibiotics in pregnancy has been reported (42). A Danish study found a sensitivity of 93% and a specificity of 88% for antibiotic prescription fills when compared to self-report in prospective biweekly questionnaires (42). It is unknown whether such non-adherence differs
by antibiotic substance. If so, the direction of the resulting bias would be unpredictable, as there would be three exposure groups in our study: pivmecillinam exposed, nitrofurantoin exposed, and children exposed to untreated maternal infection. However, we had access to information on dispensed antibiotics which is a better approximation of medication use than prescription records (43).

We did not have information about the ethnicity of the children included in the study sample. With a majority of Caucasian ethnicity in the Nordic countries, and differences in disease susceptibility and drug responses between different ethnic groups, the findings from the present study may not be generalisable to populations with a different genetic composition.

In conclusion, we found no substantial association between childhood leukaemia and prenatal exposure to nitrofurantoin, albeit a slightly elevated IRR with confidence intervals including the null, and corresponding to a small absolute risk. There was no dose-response relationship, and a biologically plausible mechanism of action is lacking. Jointly, this suggests against a causal interpretation of the elevated IRR.

**Ethics approval**

The study was approved by The Regional Committee for Medical Research Ethics in South-Eastern Norway (approval number: 2018/142/REK Sør-Øst), and by the Swedish Ethical Review Authority (approval number: 2018/2604-31/1 2019-00268 (2019-02311)). In Denmark and Finland, registry-based studies are exempt from ethical review. The study was approved by local data protection officers (approval number, Denmark: 2019-DCRC-0096; approval numbers, Finland: THL/2297/5.05.00/2018, Kela 120/522/2019, TK-53-1405-19; approval number, Norway: 233835). Data were handled in accordance with the research approvals and the applied legal norms including European Union General Data Protection Regulation (2016/79).

**Author contributions**

All authors contributed to the design of the study. Data management was performed by CHH, MKL and SH. SH analysed the data. All authors contributed to the interpretation of the findings. SH prepared the original draft and all other authors critically revised the work.

**Acknowledgements**
Data from Finland, Norway, and Sweden was stored at the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT).

The authors would like to thank Charlotte Wessel Skovlund for help with the data management.

**Funding**

The study was financed with support from the Nordic Cancer Union R275-A15824. Sarah Hjorth and Hedvig Nordeng are supported by Professor Nordeng’s European Research Council Starting Grant “DrugsInPregnancy” grant number 639377.

**Conflicts of interest**

The authors have no conflicts of interest to declare that are relevant to the content of this article.

**References**


Figure 1: Flowchart of the study population. *Individuals with missing ID or gestational age were excluded by registry holders, so we do not have information on how many individuals were excluded for these reasons.
**Tables**

**Table 1: Characteristics of included pregnancies exposed to nitrofurantoin or pivmecillinam according to the Nordic Prescription Registries.**

<table>
<thead>
<tr>
<th></th>
<th>Exposed to nitrofurantoin (n=44,091)</th>
<th>Exposed to pivmecillinam (n=247,306)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age, mean (sd)</td>
<td>29.6 (5.5)</td>
<td>29.0 (5.4)</td>
</tr>
<tr>
<td>Maternal nulliparity, n (%)</td>
<td>20,607 (46.7)</td>
<td>119,069 (48.1)</td>
</tr>
<tr>
<td>Maternal history of cancer before pregnancy, n (%)</td>
<td>276 (0.6)</td>
<td>1356 (0.5)</td>
</tr>
<tr>
<td>Maternal smoking in early pregnancy, n (%)</td>
<td>5279 (12.0)</td>
<td>40,923 (16.5)</td>
</tr>
<tr>
<td>Child country of birth, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>8794 (19.9)</td>
<td>118,303 (47.8)</td>
</tr>
<tr>
<td>Finland</td>
<td>1210 (2.7)</td>
<td>44,422 (18.0)</td>
</tr>
<tr>
<td>Norway</td>
<td>6908 (15.7)</td>
<td>50,899 (20.6)</td>
</tr>
<tr>
<td>Sweden</td>
<td>27,179 (61.6)</td>
<td>33,682 (13.6)</td>
</tr>
<tr>
<td>Child male sex, n (%)</td>
<td>22,611 (51.3)</td>
<td>127,091 (51.4)</td>
</tr>
<tr>
<td>Number of prescription fills in the year before pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>41,243 (93.5)</td>
<td>242,714 (98.1)</td>
</tr>
<tr>
<td>1</td>
<td>2229 (5.1)</td>
<td>3887 (1.6)</td>
</tr>
<tr>
<td>2+</td>
<td>619 (1.4)</td>
<td>705 (0.3)</td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>39,591 (89.8)</td>
<td>219,623 (88.8)</td>
</tr>
<tr>
<td>1</td>
<td>3664 (8.3)</td>
<td>21,045 (8.5)</td>
</tr>
<tr>
<td>2+</td>
<td>836 (1.9)</td>
<td>6638 (2.7)</td>
</tr>
<tr>
<td>Number of prescription fills for study medication in pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>38,448 (87.2)</td>
<td>199,598 (80.7)</td>
</tr>
<tr>
<td>2+</td>
<td>5643 (12.8)</td>
<td>47,708 (19.3)</td>
</tr>
<tr>
<td>Trimester of exposure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>11,688 (26.5)</td>
<td>75,160 (30.4)</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>15,657 (35.5)</td>
<td>88,231 (35.7)</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>20,460 (46.4)</td>
<td>116,817 (47.2)</td>
</tr>
<tr>
<td>Number of prescription fills for other antibiotics in pregnancy, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>29,554 (67.0)</td>
<td>166,635 (67.4)</td>
</tr>
<tr>
<td>1</td>
<td>9358 (21.2)</td>
<td>53,404 (21.6)</td>
</tr>
<tr>
<td>2+</td>
<td>5179 (11.7)</td>
<td>27,267 (11.0)</td>
</tr>
</tbody>
</table>
### Table 2: Incidence rate ratios and differences for leukaemia comparing children prenatally exposed to nitrofurantoin and pivmecillinam.

<table>
<thead>
<tr>
<th></th>
<th>Any leukaemia cases</th>
<th>IRR (95% CI)</th>
<th>IRD pr 100,000 person-years (95% CI)</th>
<th>wIRD&lt;sup&gt;b&lt;/sup&gt; pr 100,000 person-years (95% CI)</th>
<th>Lymphoid leukaemia cases</th>
<th>IRR (95% CI)</th>
<th>IRD pr 100,000 person-years (95% CI)</th>
<th>wIRD&lt;sup&gt;b&lt;/sup&gt; pr 100,000 person-years (95% CI)</th>
<th>AML cases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pivmecillinam</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 or more</td>
<td>47 708</td>
<td>23</td>
<td>reference</td>
<td>2.15 (-&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>reference</td>
<td>17</td>
<td>reference</td>
<td>-</td>
<td>&lt;5</td>
</tr>
<tr>
<td>prescription fills</td>
<td></td>
<td></td>
<td>(0.88;5.28)</td>
<td>(5.92;12.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>5643</td>
<td>6</td>
<td>reference</td>
<td>1.57 (-&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>reference</td>
<td>5</td>
<td>reference</td>
<td>-</td>
<td>&lt;5</td>
</tr>
<tr>
<td><strong>1st trimester</strong></td>
<td></td>
<td></td>
<td>(0.54;4.55)</td>
<td>(4.24;16.74)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>75 160</td>
<td>26</td>
<td>reference</td>
<td>1.83 (-&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>1.42</td>
<td>24</td>
<td>reference</td>
<td>-</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>11 688</td>
<td>7</td>
<td>(0.79;4.21)</td>
<td>2.61;8.82</td>
<td>(0.54;3.71)</td>
<td></td>
<td>(0.57;3.87)</td>
<td>3.28;6.15</td>
<td></td>
</tr>
<tr>
<td><strong>2nd trimester</strong></td>
<td></td>
<td></td>
<td>(0.84;4.41)</td>
<td>2.60;9.65</td>
<td>(0.57;3.87)</td>
<td></td>
<td>(0.57;3.87)</td>
<td>3.8;6.76</td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>88 231</td>
<td>48</td>
<td>reference</td>
<td>0.74 (-&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>0.67</td>
<td>5</td>
<td>0.58</td>
<td>-1.81 (-&lt;sup&gt;b&lt;/sup&gt;)</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>15 657</td>
<td>6</td>
<td>(0.32;1.72)</td>
<td>5.31;2.16</td>
<td>(0.27;1.69)</td>
<td></td>
<td>(0.21;1.61)</td>
<td>5.20;158</td>
<td></td>
</tr>
<tr>
<td><strong>3rd trimester</strong></td>
<td></td>
<td></td>
<td>(0.19;1.47)</td>
<td>5.97;0.43</td>
<td>(0.27;1.69)</td>
<td></td>
<td>(0.21;1.61)</td>
<td>5.49;0.93</td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>116 817</td>
<td>67</td>
<td>reference</td>
<td>1.73 (-&lt;sup&gt;a&lt;/sup&gt;)</td>
<td>1.63</td>
<td>51</td>
<td>reference</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>20 460</td>
<td>21</td>
<td>(1.03;2.92)</td>
<td>1.37;10.90</td>
<td>(0.89;3.00)</td>
<td></td>
<td>(0.88;3.11)</td>
<td>1.90;8.52</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Results from fixed effects meta-analysis, $I^2 = 0.0%$. Findings from mixed effects Poisson models were comparable: IRR 1.44 (0.96;2.17), wIRR 1.40 (0.91;2.15).

<sup>b</sup>Inverse probability of treatment weights including calendar year at birth, maternal age, parity, maternal history of cancer before pregnancy, prescription fills for immunosuppressants, systemic corticosteroids, and systemic antibiotics before start of pregnancy, maternal smoking status during first trimester, and child sex. In Finland, birth year was not balanced after weighting and added to the outcome model.

IRD: Incidence rate difference; IRR: Incidence rate ratio; wIRD: weighted incidence rate difference; wIRR: weighted incidence rate ratio.
Supplementary information

Guidelines for the treatment of urinary tract infections in pregnancy:

<table>
<thead>
<tr>
<th>Country</th>
<th>Asymptomatic bacteriuria or cystitis:</th>
<th>Recurrent cystitis:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark (1)</td>
<td>1. Pivmecillinam</td>
<td>1. Pivmecillinam or nitrofurantoin</td>
</tr>
<tr>
<td></td>
<td>2. Nitrofurantoin (do not use in last weeks of pregnancy)</td>
<td></td>
</tr>
<tr>
<td>Norway (2)</td>
<td>1. Pivmecillinam</td>
<td>1. Pivmecillinam</td>
</tr>
<tr>
<td></td>
<td>2. Nitrofurantoin</td>
<td>2. Nitrofurantoin</td>
</tr>
<tr>
<td>Sweden (3)</td>
<td>1. Pivmecillinam or nitrofurantoin</td>
<td>1. Nitrofurantoin or cefadroxil</td>
</tr>
<tr>
<td></td>
<td>2. Cefadroxil</td>
<td></td>
</tr>
<tr>
<td>Finland (4)</td>
<td>1. Pivmecillinam, nitrofurantoin (do not use in pregnancy weeks 38-42), amoxicillin, or first generation cephalosporins.</td>
<td>1. Nitrofurantoin or methenamine hippurate</td>
</tr>
</tbody>
</table>

Sources:

Periods of data inclusion for the different national registries:

<table>
<thead>
<tr>
<th>Country</th>
<th>Years of data inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cancer registries</strong></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>1943-2017</td>
</tr>
<tr>
<td>Finland</td>
<td>1953-2017</td>
</tr>
<tr>
<td>Norway</td>
<td>1953-2017</td>
</tr>
<tr>
<td>Sweden</td>
<td>1958-2017</td>
</tr>
<tr>
<td><strong>Prescription registries</strong></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>1995-2013</td>
</tr>
<tr>
<td>Finland</td>
<td>1993-2013</td>
</tr>
<tr>
<td>Norway</td>
<td>2004-2013</td>
</tr>
<tr>
<td>Sweden</td>
<td>2005-2013</td>
</tr>
<tr>
<td><strong>Medical birth registries</strong></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>1997-2013</td>
</tr>
<tr>
<td>Finland</td>
<td>1997-2013</td>
</tr>
<tr>
<td>Norway</td>
<td>2007-2013</td>
</tr>
<tr>
<td>Sweden</td>
<td>2007-2013</td>
</tr>
<tr>
<td><strong>Patient registries</strong></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>1995-2013</td>
</tr>
<tr>
<td>Finland</td>
<td>1994-2013</td>
</tr>
<tr>
<td>Norway</td>
<td>2008-2013</td>
</tr>
<tr>
<td>Sweden</td>
<td>2001-2013</td>
</tr>
<tr>
<td><strong>Civil Registration System</strong></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>1996-2017</td>
</tr>
<tr>
<td><strong>Cause of death registries</strong></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>1997-2017</td>
</tr>
<tr>
<td>Sweden</td>
<td>2007-2017</td>
</tr>
</tbody>
</table>

*In Denmark, data on death and migration were obtained from the Civil Registration System, in Finland and Sweden, data on death were obtained from the cause of death registries. In Norway, data on death and migration were obtained from the Medical Birth Registry.*
**Supplementary Table 1: Incidence rate ratios and differences for leukaemia comparing children prenatally exposed to nitrofurantoin and pivmecillinam. Complete case analysis.**

<table>
<thead>
<tr>
<th></th>
<th>Any leukaemia</th>
<th>Lymphoid leukaemia</th>
<th>AML</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>cases</td>
<td>IRR (95% CI)</td>
<td>IRD pr 100,000 person-years (95% CI)</td>
<td>IRD pr 100,000 person-years (95% CI)</td>
<td>IRD pr 100,000 person-years (95% CI)</td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>230,627</td>
<td>118</td>
<td>reference</td>
<td>11.50 (1.01-2.42)</td>
<td>2.78 (-0.55-6.11)</td>
<td>2.47 (-1.21-6.15)</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>41,066</td>
<td>30</td>
<td>1.50 (0.91-2.27)</td>
<td>1.90 (0.64-5.57)</td>
<td>6.96 (-4.38-16.30)</td>
<td>4.31 (-5.52-14.13)</td>
</tr>
<tr>
<td>2 or more prescription fills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>44,482</td>
<td>22</td>
<td>reference</td>
<td>2.24 (0.91-5.53)</td>
<td>6.96 (-0.55-6.11)</td>
<td>2.47 (-1.21-6.15)</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>5290</td>
<td>6</td>
<td>1.90 (0.64-5.57)</td>
<td>4.31 (-5.52-14.13)</td>
<td>5.80 (-1.80-16.30)</td>
<td>4.31 (-5.52-14.13)</td>
</tr>
<tr>
<td>Trimester of exposure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>70,109</td>
<td>24</td>
<td>reference</td>
<td>1.71 (0.70-4.17)</td>
<td>2.63 (-3.06-8.33)</td>
<td>1.83 (-3.61-7.27)</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>10,907</td>
<td>6</td>
<td>1.48 (0.57-3.86)</td>
<td>3.06 (8.33)</td>
<td>1.83 (-3.61-7.27)</td>
<td>1.83 (-3.61-7.27)</td>
</tr>
<tr>
<td>2nd trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>82,377</td>
<td>43</td>
<td>reference</td>
<td>0.69 (0.27-1.74)</td>
<td>-1.80 (-2.14-7.52)</td>
<td>-2.33 (-5.79-1.13)</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>14,632</td>
<td>5</td>
<td>0.59 (0.21-1.65)</td>
<td>3.11 (5.72)</td>
<td>-2.33 (-5.79-1.13)</td>
<td>-2.33 (-5.79-1.13)</td>
</tr>
<tr>
<td>3rd trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>108,904</td>
<td>62</td>
<td>reference</td>
<td>1.86 (1.09-3.18)</td>
<td>5.80 (-0.90-12.51)</td>
<td>5.04 (-1.65-11.73)</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>19,041</td>
<td>21</td>
<td>1.76 (0.99-3.13)</td>
<td>9.00 (12.51)</td>
<td>1.65 (5.79-11.73)</td>
<td>1.65 (5.79-11.73)</td>
</tr>
</tbody>
</table>

*Results from mixed effects poisson models.

bInverse probability of treatment weights including calendar year at birth, maternal age, parity, maternal history of cancer before pregnancy, prescription fills for immunosuppressants, systemic corticosteroids, and systemic antibiotics before start of pregnancy, maternal smoking status during first trimester, and child sex. In Finland, birth year was not balanced after weighting and added to the outcome model.

IRD: Incidence rate difference; IRR: Incidence rate ratio; wIRD: weighted incidence rate difference; wIRR: weighted incidence rate ratio.
Supplementary Table 2: Incidence rate ratios and differences for leukaemia comparing children prenatally exposed to nitrofurantoin and pivmecillinam. Infant leukaemia excluded. a

<table>
<thead>
<tr>
<th></th>
<th>Any leukaemia cases</th>
<th>Lymphoid leukaemia cases</th>
<th>AML cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>IRR (95% CI)</td>
<td>IRD pr 100,000 person-years (95% CI)</td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 or more prescription fills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>245,555</td>
<td>122</td>
<td>reference</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>43,882</td>
<td>28</td>
<td>1.34 (0.87;2.08)</td>
</tr>
<tr>
<td>Trimester of exposure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st trimester</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>74,648</td>
<td>25</td>
<td>reference</td>
</tr>
<tr>
<td></td>
<td>11,628</td>
<td>6</td>
<td>1.64 (0.67;4.00)</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>87,584</td>
<td>47</td>
<td>reference</td>
</tr>
<tr>
<td>2nd trimester</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>115,996</td>
<td>62</td>
<td>reference</td>
</tr>
<tr>
<td></td>
<td>20,362</td>
<td>18</td>
<td>1.63 (0.93;2.85)</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd trimester</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>115,996</td>
<td>62</td>
<td>reference</td>
</tr>
<tr>
<td></td>
<td>20,362</td>
<td>18</td>
<td>1.63 (0.93;2.85)</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aResults from mixed effects poisson models.

bInverse probability of treatment weights including calendar year at birth, maternal age, parity, maternal history of cancer before pregnancy, prescription fills for immunosuppressants, systemic corticosteroids, and systemic antibiotics before start of pregnancy, maternal smoking status during first trimester, and child sex. In Finland, birth year was not balanced after weighting and added to the outcome model.

IRD: Incidence rate difference; IRR: Incidence rate ratio; wIRD: weighted incidence rate difference; wIRR: weighted incidence rate ratio.
Supplementary Table 3: Incidence rate ratios and differences for leukaemia comparing children prenatally exposed to nitrofurantoin and pivmecillinam. Children prenatally exposed to other antibiotics excluded.\(^a\)

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Any leukaemia cases</th>
<th>IRR (95% CI)</th>
<th>wIRR(^b) (95% CI)</th>
<th>IRD pr 100,000 person-years (95% CI)</th>
<th>wIRD(^b) pr 100,000 person-years (95% CI)</th>
<th>Lymphoid leukaemia cases</th>
<th>IRR (95% CI)</th>
<th>wIRR(^b) (95% CI)</th>
<th>IRD pr 100,000 person-years (95% CI)</th>
<th>wIRD(^b) pr 100,000 person-years (95% CI)</th>
<th>AML cases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pivmecillinam</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 or more prescription fills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>28,084</td>
<td>11</td>
<td>reference</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>3097</td>
<td>&lt;5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>&lt;5</td>
</tr>
<tr>
<td><strong>Trimester of exposure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1st trimester</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>50,440</td>
<td>12</td>
<td>reference</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>8163</td>
<td>&lt;5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>&lt;5</td>
</tr>
<tr>
<td><strong>2nd trimester</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>57,652</td>
<td>29</td>
<td>reference</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>10,361</td>
<td>&lt;5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>&lt;5</td>
</tr>
<tr>
<td><strong>3rd trimester</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>77,173</td>
<td>41</td>
<td>reference</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>13,031</td>
<td>15</td>
<td>2.08 (-1.08;3.98)</td>
<td>2.03 (1.04;3.98)</td>
<td>6.79 (1.74;15.32)</td>
<td>6.48 (2.11;15.07)</td>
<td>2.15 (1.05;4.39)</td>
<td>2.04 (0.98;4.29)</td>
<td>5.99 (2.01;13.98)</td>
<td>5.44 (2.44;13.33)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Results from mixed effects poisson models.

\(^b\)Inverse probability of treatment weights including calendar year at birth, maternal age, parity, maternal history of cancer before pregnancy, prescription fills for immunosuppressants, systemic corticosteroids, and systemic antibiotics before start of pregnancy, maternal smoking status during first trimester, and child sex. In Finland, birth year was not balanced after weighting and added to the outcome model.

IRD: Incidence rate difference; IRR: Incidence rate ratio; wIRD: weighted incidence rate difference; wIRR: weighted incidence rate ratio.
Supplementary Table 4: Incidence rate ratios and differences for leukaemia comparing children prenatally exposed to nitrofurantoin and pivmecillinam. Restricted to children who are followed for the entire pregnancy.*

<table>
<thead>
<tr>
<th></th>
<th>Any leukaemia cases</th>
<th>Lymphoid leukaemia cases</th>
<th>AML cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>IRR (95% CI)</td>
<td>IRD pr 100,000 person-years (95% CI)</td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>220,648</td>
<td>115</td>
<td>reference</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>40,181</td>
<td>28</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.89;2.11)</td>
</tr>
<tr>
<td>2 or more prescription fills</td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>44,403</td>
<td>21</td>
<td>reference</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>5210</td>
<td>6</td>
<td>2.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.96;5.87)</td>
</tr>
<tr>
<td>Trimester of exposure</td>
<td>1st trimester</td>
<td></td>
<td>65,472</td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>10,702</td>
<td>7</td>
<td>1.84</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td></td>
<td></td>
<td>(0.79;4.27)</td>
</tr>
<tr>
<td>2nd trimester</td>
<td></td>
<td></td>
<td>79,020</td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>14,303</td>
<td>6</td>
<td>0.88</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td></td>
<td></td>
<td>(0.37;2.07)</td>
</tr>
<tr>
<td>3rd trimester</td>
<td></td>
<td></td>
<td>106,779</td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>18,599</td>
<td>17</td>
<td>1.51</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td></td>
<td></td>
<td>(0.85;2.67)</td>
</tr>
</tbody>
</table>

*aResults from mixed effects poisson models.

*bInverse probability of treatment weights including calendar year at birth, maternal age, parity, maternal history of cancer before pregnancy, prescription fills for immunosuppressants, systemic corticosteroids, and systemic antibiotics before start of pregnancy, maternal smoking status during first trimester, and child sex. In Finland, birth year was not balanced after weighting and added to the outcome model.

IRD: Incidence rate difference; IRR: Incidence rate ratio; wIRD: weighted incidence rate difference; wIRR: weighted incidence rate ratio.
Supplementary Table 5: Post hoc analysis of incidence rate ratios and differences for leukaemia comparing children prenatally exposed and unexposed to nitrofurantoin.

<table>
<thead>
<tr>
<th></th>
<th>Any leukaemia cases</th>
<th>Lymphoid leukaemia cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>IRR (95% CI)</td>
</tr>
<tr>
<td>Unexposed</td>
<td>2,254,684</td>
<td>1435</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>44091</td>
<td>32</td>
</tr>
<tr>
<td>2 or more prescription fills</td>
<td>Unexposed</td>
<td>2,254,684</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>5643</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trimester of exposure</th>
<th>Unexposed</th>
<th>Nitrofurantoin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st trimester</td>
<td>Unexposed</td>
<td>2,287,087</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>11,688</td>
<td>7</td>
</tr>
<tr>
<td>2nd trimester</td>
<td>Unexposed</td>
<td>2,283,118</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>15,657</td>
<td>6</td>
</tr>
<tr>
<td>3rd trimester</td>
<td>Unexposed</td>
<td>2,275,852</td>
</tr>
<tr>
<td>Nitrofurantoin</td>
<td>20,460</td>
<td>21</td>
</tr>
</tbody>
</table>

| aResults from mixed effects poisson models. |
| bInverse probability of treatment weights including calendar year at birth, maternal age, parity, maternal history of cancer before pregnancy, prescription fills for immunosuppressants, systemic corticosteroids, and systemic antibiotics before start of pregnancy, maternal smoking status during first trimester, and child sex. In Finland, birth year was not balanced after weighting and added to the outcome model. |

IRD: Incidence rate difference; IRR: Incidence rate ratio; wIRD: weighted incidence rate difference; wIRR: weighted incidence rate ratio.
Supplementary Figure 1: Data included in the study
Supplementary Figure 2: Study design diagram.
Supplementary Figure 3: Directed Acyclic Graph. SEP: socio-economic position.
Appendices
Appendix A

Appendix to Chapter 1
Table A.1: Overview of methods to account for confounding [9, 10, 134].

<table>
<thead>
<tr>
<th>Method – attributed to</th>
<th>Characteristic</th>
<th>Strengths</th>
<th>Limitations</th>
<th>Examples of medication safety studies using this method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active comparators –</strong>&lt;br&gt;Kramer et al. 1987 [269]</td>
<td>Compares exposure to one medication with exposure to another medication for the same indication.</td>
<td>When untreated illness is not an option, this design answers the clinically relevant question. Accounts for the underlying indication for use, and for unmeasured factors associated with willingness to seek and use treatment.</td>
<td>If both exposures increase the risk of the outcome compared to no treatment, this increased risk will be missed. The method requires that two fairly equivalent treatment options exist.</td>
<td>Hleyhel et al. 2016 [270] compared the risk of cancer in children prenatally exposed to two different HIV treatments. Fan et al. 2020 [101] compared the risk of adverse birth outcomes and brain development disorders between children prenatally exposed to penicillin and children prenatally exposed to macrolides.</td>
</tr>
<tr>
<td><strong>Confounder summary scores, e.g. propensity scores –</strong>&lt;br&gt;Miettinen 1976 [271], Rosenbaum &amp; Rubin 1983 [272]</td>
<td>Reduces a large set of confounders to one summary score. The propensity score models the probability of exposure given the measured confounders and predictors of the outcome.</td>
<td>Can prevent over-fitting of the statistical models. The propensity score is useful when the exposure is more common than the outcome.</td>
<td>Does not account for time-varying, measured confounding, nor for unmeasured confounding.</td>
<td>Fan et al. 2020 [101] used stratification on the propensity score, Trønnes et al. 2019 [214] used weighting on the propensity score.</td>
</tr>
</tbody>
</table>
Table A.1: Overview of methods to account for confounding [9, 10, 134], continued.

<table>
<thead>
<tr>
<th>Method – attributed to</th>
<th>Characteristic</th>
<th>Strengths</th>
<th>Limitations</th>
<th>Examples of medication safety studies using this method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-context comparisons – Brion et al. 2011 [247]</td>
<td>Compares findings from different contexts where the underlying confounding structure is thought to be different.</td>
<td>Can account for unmeasured confounding. Is also important for generalisability of findings.</td>
<td>Requires that the underlying confounding structure truly differs between contexts. Will rarely account for confounding by indication as this rarely differs between contexts, but the characteristics of medication users might.</td>
<td>Bertoldi et al. 2020 [225] compared associations between prenatal exposure to paracetamol and child brain development in one cohort from the US, and one from Brazil.</td>
</tr>
<tr>
<td>Disease comparators – unidentified</td>
<td>Compares exposure to a medication with exposure to untreated illness</td>
<td>When untreated illness is an option, this answers a clinically important question. Accounts for the underlying indication for medication use.</td>
<td>Does not necessarily account for the severity of the underlying disease. Does not account for unmeasured factors associated with the willingness to seek and use treatment.</td>
<td>Harris et al. 2018 [257] compared brain development in children born to women with migraine during pregnancy who were treated with triptans, and children born to women with migraine during pregnancy who were not treated with triptans. Li et al. 2020 [113] compared the risk of childhood obesity between children prenatally exposed to antibiotics and children prenatally exposed to untreated infection.</td>
</tr>
<tr>
<td>Method – attributed to</td>
<td>Characteristic</td>
<td>Strengths</td>
<td>Limitations</td>
<td>Examples of medication safety studies using this method</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------</td>
<td>-----------</td>
<td>-------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td><strong>E-value (bounding factor) – Ding &amp; VanderWeele 2016 [157]</strong></td>
<td>Estimates how strong unmeasured confounding is required to reduce the effect estimate and lower limit of the confidence interval to the null if the confounding is removed.</td>
<td>Requires no assumptions about the unmeasured confounder(s). Can also be used to assess how strong a confounding would be required to move a small effect size to a clinically meaningful effect size.</td>
<td>It is up to the researchers to determine whether it is plausible that any unmeasured confounders exist that could yield as strong confounding as suggested by the e-value.</td>
<td>Masarwa et al. 2020 [242] calculated the e-value to estimate the extent of confounding that would be necessary to explain away the finding from their meta-analysis on prenatal paracetamol and child risk of attention problems.</td>
</tr>
<tr>
<td><strong>Instrumental variables – method from economics, described systematically for epidemiology by Greenland 2000 [273]</strong></td>
<td>Compares the association between the outcome and an instrument, adjusted for the instrument-exposure association. The instrument should be strongly associated with the exposure, but not associated with the confounders of the exposure-outcome relationship, and should only be associated with the outcome through the exposure.</td>
<td>Can account for unmeasured confounding and misclassification.</td>
<td>Good instruments are difficult to find for medication use in pregnancy.</td>
<td>Swanson et al. 2015 [274] used prescriber preference and calendar years as instruments but found that these performed poorly.</td>
</tr>
</tbody>
</table>
Table A.1: Overview of methods to account for confounding [9, 10, 134], continued.

<table>
<thead>
<tr>
<th>Method – attributed to</th>
<th>Characteristic</th>
<th>Strengths</th>
<th>Limitations</th>
<th>Examples of medication safety studies using this method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal structural models – Robins et al. 2000 [133]</td>
<td>Builds on inverse probability of treatment weights. At each new treatment assignment, a new probability of receiving treatment is calculated based on covariates and previous treatment. The weights are multiplied to yield a total weight.</td>
<td>Useful in scenarios with time-varying exposure. Can account for measured, time-varying confounding.</td>
<td>Does not account for unmeasured confounding. May yield very wide confidence intervals if the treatment and the confounders are strongly associated.</td>
<td>Wood et al. 2016 [255] used marginal structural models analyse associations between the timing of prenatal triptan exposure and child brain development.</td>
</tr>
<tr>
<td>Negative controls – method from experimental biology, described systematically for epidemiology by Lipsitch et al. 2010 [275]</td>
<td>Negative exposure controls compare the association between the outcome and an exposure that cannot plausibly be causally related to the outcome, but has similar underlying confounding structures as the association of interest. Negative outcome controls compare the association between the exposure and an outcome that cannot plausibly be causally related to the exposure, but has similar underlying confounding structures as the association of interest.</td>
<td>Estimates the extent of bias from unmeasured confounding, but also from misclassification.</td>
<td>Does not identify the source of the bias. If the controls are not truly negative, the findings from the negative control analysis will not be useful.</td>
<td>Shaheen et al. 2019 [86] used paternal analgesic use while the mother was pregnant with the index child as a negative control to maternal analgesic use, Hamad et al. 2020 [107] used maternal antibiotic use before pregnancy as a negative control to maternal antibiotic use during pregnancy.</td>
</tr>
<tr>
<td>Method – attributed to</td>
<td>Characteristic</td>
<td>Strengths</td>
<td>Limitations</td>
<td>Examples of medication safety studies using this method</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------</td>
<td>-----------</td>
<td>-------------</td>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Probabilistic sensitivity analysis – Hoffman &amp; Hammonds 1994</strong> [276]</td>
<td>The researcher specifies priors – assumptions regarding the distribution of the bias parameters. Probabilistic bias analysis provides bias-corrected effect estimates.</td>
<td>Can be specified for unmeasured confounders, selection, and misclassification bias. Several types of bias can be accounted for simultaneously.</td>
<td>The result depends on the used priors and their distribution, hence assumptions must be made about the prevalence of the confounder among exposed and unexposed, and the confounder-disease relationship. The existing tools for implementation of this method typically only allow one unmeasured confounder at a time.</td>
<td>Masarwa et al. 2020 [242] used probabilistic sensitivity analysis to estimate the impact of unmeasured confounding by parental ADHD in the individual studies included in their meta-analysis on prenatal paracetamol and child risk of attention problems.</td>
</tr>
<tr>
<td><strong>Sibling design – method from economics, described systematically for epidemiology by Susser et al. 2010</strong> [277]</td>
<td>Compares exposed children with their unexposed siblings.</td>
<td>Accounts for measured and unmeasured confounding that is shared between siblings (genetic confounding and shared environment).</td>
<td>Does not account for confounders that are not shared between siblings. More vulnerable to exposure misclassification than non-sibling designs.</td>
<td>Brandlistuen et al. 2013 [149], Hamad et al. 2020 [107].</td>
</tr>
</tbody>
</table>

Methods used in this thesis are marked in boldface.
Table A.2: Criteria for causal inference according to Bradford Hill [159], the target trial approach [128, 160], and the pragmatic pluralism approach [158].

<table>
<thead>
<tr>
<th>Bradford Hill</th>
<th>Target trial approach</th>
<th>Pragmatic pluralism approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength of the association (effect size)</td>
<td>Exchangeability (no residual confounding)</td>
<td>Evidence from several studies</td>
</tr>
<tr>
<td>Consistency (replication across time, different sites and in different ways/methods)</td>
<td>Positivity (in all strata of confounding, there has to be at least one person with each exposure level)</td>
<td>Interlocking evidence (using evidence from other disciplines, such as \textit{in vitro} or animal studies)</td>
</tr>
<tr>
<td>Analogy (seeing that it is known that some medications can cause congenital malformations, it is easier to accept that other medications could do the same)</td>
<td>Consistency (there should be a well-defined difference between exposed and unexposed, meaning that the exposure should be possible to make into an intervention)</td>
<td>Triangulation (evidence from different methods, each with their own bias structure)</td>
</tr>
<tr>
<td>Temporality (the cause must come before the effect)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biological gradient (dose-response)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plausibility (it is helpful if there is a known biological mechanism to explain the findings, but not an absolute requirement)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coherence (findings should not be opposed to the known facts about disease biology)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment (strong evidence for causality when feasible)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity (the association is limited to a particular type of disease)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Appendix to Chapter 4
Table B.1: Most commonly filled analgesic and antibiotic substances in peri-pregnancy among 172,585 Norwegian pregnancies, descending order of frequency.

<table>
<thead>
<tr>
<th></th>
<th>Before pregnancy</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; trimester</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; trimester</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; trimester</th>
<th>After pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analgesics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diclofenac</td>
<td></td>
<td>Paracetamol with codeine</td>
<td>Paracetamol with codeine</td>
<td>Paracetamol with codeine</td>
<td>Diclofenac</td>
</tr>
<tr>
<td>Paracetamol with codeine</td>
<td></td>
<td>Diclofenac</td>
<td>Paracetamol</td>
<td>Paracetamol</td>
<td>Paracetamol</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>Paracetamol</td>
<td>Sumatriptan</td>
<td>Sumatriptan</td>
<td>Paracetamol with codeine</td>
<td></td>
</tr>
<tr>
<td>Paracetamol</td>
<td>Ibuprofen</td>
<td>Diclofenac</td>
<td>Tramadol</td>
<td>Ibuprofen</td>
<td></td>
</tr>
<tr>
<td>Sumatriptan</td>
<td>Sumatriptan</td>
<td>Ibuprofen</td>
<td>Ibuprofen</td>
<td>Naproxen</td>
<td></td>
</tr>
<tr>
<td><strong>Antibiotics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenoxymethylpenicillin</td>
<td>Pivmecillinam</td>
<td>Pivmecillinam</td>
<td>Pivmecillinam</td>
<td>Didoxacillin</td>
<td></td>
</tr>
<tr>
<td>Pivmecillinam</td>
<td>Phenoxymethylpenicillin</td>
<td>Phenoxymethylpenicillin</td>
<td>Phenoxymethylpenicillin</td>
<td>Phenoxymethylpenicillin</td>
<td></td>
</tr>
<tr>
<td>Erythromycin</td>
<td>Amoxicillin</td>
<td>Amoxicillin</td>
<td>Amoxicillin</td>
<td>Pivmecillinam</td>
<td></td>
</tr>
<tr>
<td>Doxycycline</td>
<td>Nitrofurantoin</td>
<td>Nitrofurantoin</td>
<td>Nitrofurantoin</td>
<td>Erythromycin</td>
<td></td>
</tr>
<tr>
<td>Azithromycin</td>
<td>Erythromycin</td>
<td>Erythromycin</td>
<td>Trimethoprim</td>
<td>Amoxicillin</td>
<td></td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Results&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonaventure et al.</td>
<td>Penicillins (846).</td>
<td>ALL (725), AML (128),</td>
<td>Match on sex,</td>
<td>ALL 1.0 (0.8:1.2), AML 1.6 (1.1:2.4), NHL 1.1 (0.6:1.9), HL 1.2 (0.5:3.0), Astrocytoma 0.8 (0.4:1.4), Medulloblastoma 1.6 (0.8:3.1), Rhabdomyosarcoma 1.7 (1.0:3.2), Ewings sarcoma 1.7 (0.6:4.6), Neuroblastoma 1.1 (0.6:1.9), Nephroblastoma 0.9 (0.4:1.7).</td>
</tr>
<tr>
<td>2015 [89]</td>
<td>Routinely collected health data</td>
<td>NHL (83), HL (31), astrocytoma (100), medulloblastoma (48), rhabdomyosarcoma (54), Ewings sarcoma (20), neuroblastoma (78), nephroblastosma (60)</td>
<td>Case-control (4122). UK, UKCCS, mean age 5.8 (max age 15 years). Unconditional logistic regression</td>
<td>Matching on sex, month and year of birth, and region of residence, adjustment for year of birth</td>
</tr>
<tr>
<td>Reference</td>
<td>Exposure</td>
<td>Outcome</td>
<td>Materials and methods</td>
<td>Results$^a$</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Buckley et al. 1989</td>
<td>Antibiotics in the year before pregnancy or during pregnancy (31). Maternal retrospective report.</td>
<td>Hepatoblastoma (75)</td>
<td>Case-control (150). US and Canada, CCSG, upper age limit not stated, but the majority of children were under 2 years of age. Conditional logistic regression</td>
<td>Matching on age 0.8, CI not reported.</td>
</tr>
<tr>
<td>Bunin et al. 1994</td>
<td>Medication against vaginal infection (59). Maternal retrospective report.</td>
<td>Astrocytoma (155), PNET (166)</td>
<td>Case-control (332). US and Canada, Children’s Cancer Group, ages 0-5 years. Conditional logistic regression</td>
<td>Astrocytoma 0.4 (0.2;0.9), PNET 1.2 (0.6;2.3)</td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure Medication (number of exposed). *Ascertained. Exposure window: during pregnancy, unless stated otherwise</th>
<th>Outcome Type of cancer (number of cases)</th>
<th>Materials and methods Design (sample size). *Study characteristics. Statistical analysis</th>
<th>Results* Estimate of association (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carozza et al. 1995 [280]</td>
<td>Ampicillin (13), Ampicillin trihydrate (1), Doxycycline (2), Erythromycin (4), Oxytetracycline (2), Penicillin V potassium (7), Tetracycline HCL (1). *Maternal retrospective report.</td>
<td>Any brain tumours (361)</td>
<td>Case-control (1444). *US, SEER, upper age limit 18 years. Descriptive analysis</td>
<td>Ampicillin: 4 of 361 cases and 9 of 1083 controls, Ampicillin trihydrate: 1 of 361 cases and 0 of 1083 controls, Doxycycline: 1 of 361 cases and 1 of 1083 controls, Erythromycin: 0 of 361 cases and 4 of 1083 controls, Oxytetracycline: 1 of 361 cases and 1 of 1083 controls, Penicillin V potassium: 1 of 361 cases and 6 of 1083 controls, Tetracycline HCL: 0 of 361 cases and 1 of 1083 controls</td>
</tr>
<tr>
<td>Reference</td>
<td>Exposure</td>
<td>Outcome</td>
<td>Materials and methods</td>
<td>Resultsa</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------</td>
<td>---------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td><strong>Cook et al. 2004</strong> [281]</td>
<td>Amoxicillin (102), Ampicillin (21), Cephalexin (17), Erythromycin (27), Nitrofurantoin (5), Penicillin (23), Sulfamethoxazole (14), Trimethoprim (15) – each during pregnancy or breastfeeding. <em>Maternal retrospective report.</em></td>
<td>Neuroblastoma (504)</td>
<td>Case-control (1008). <em>US and Canada, Children’s Cancer Group, upper age limit 18 years.</em> Conditional logistic regression</td>
<td>Amoxicillin: 0.8 (0.5;1.3), Ampicillin: 1.5 (0.6;3.7), Cephalexin: 0.7 (0.3;1.9), Erythromycin: 1.9 (0.8;4.3), Nitrofurantoin (not reported due to few exposed), Penicillin: 0.9 (0.4;2.1), Sulfamethoxazole: 1.1 (0.4;3.2), Trimethoprim: 1.2 (0.4;3.4)</td>
</tr>
<tr>
<td><strong>Gilman et al. 1989</strong> [282]</td>
<td>Antibiotics (631), Sulphonamides (116). <em>Maternal retrospective report and medical records. Children are considered exposed if medication use is present in either source.</em></td>
<td>Any cancer (8059)</td>
<td>Case-control (16,118). <em>UK, OSCC, age not specified, only that cases are childhood deaths from cancer.</em> Conditional logistic regression</td>
<td>Antibiotics: 1.6, CI not reported, Sulphonamides: 1.6, CI not reported</td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Results&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradel &amp; Kaerlev 2015 [105]</td>
<td>Antibiotics (1201). &lt;br&gt; <em>Routinely collected health data</em></td>
<td>Main outcome: Any leukaemias (360), secondary outcome: ALL (284)</td>
<td>Case-control (3869). &lt;br&gt; <em>Denmark, ages 0-15 years</em>. Conditional logistic regression</td>
<td>Any leukaemia: 1.1 (0.9;1.4), ALL: 1.1 (0.8;1.4)</td>
</tr>
<tr>
<td>Grufferman et al. 1982 [283]</td>
<td>Antibiotics 1 year prior to or during pregnancy (32). &lt;br&gt; <em>Maternal retrospective report.</em></td>
<td>Rhabdomyosarcoma (33)</td>
<td>Case-control (132). &lt;br&gt; <em>North Carolina, ages 0-14 years</em>. 'Unmatched methods', not specified further</td>
<td>2.7 (1.1;6.5)</td>
</tr>
<tr>
<td>Hartley et al. 1988 [284]</td>
<td>Urinary anti-infectives (not reported). &lt;br&gt; <em>Maternal retrospective report, validated in prescription data</em></td>
<td>Bone or soft tissue sarcomas (73)</td>
<td>Case-control (219). &lt;br&gt; <em>UK, IRESCC, ages 0-15 years</em>. Mantel-Haenszel, 95% CI by Cornfield’s formula</td>
<td>Matching. Matching factors not reported, but they refer to a paper where methods are described (age and sex)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Estimate of association (95% CI)
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Resultsa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medication (number of exposed). Ascerta-</td>
<td>Type of cancer</td>
<td>Method to account for confounding in the main analysis</td>
<td>Estimate of associ-</td>
</tr>
<tr>
<td></td>
<td>tainment. Exposure window: during pregnancy,</td>
<td>(number of cases)</td>
<td></td>
<td>ation (95% CI)</td>
</tr>
<tr>
<td></td>
<td>unless stated otherwise</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heck et al. 2015 [285]</td>
<td>Antibiotics (53). Maternal retrospective</td>
<td>Sporadic retinoblastoma, unilateral</td>
<td>Matching on year of birth, adjustment for maternal ethnicity, education, income, maternal age, and smoking in pregnancy</td>
<td>Unilateral: 0.8</td>
</tr>
<tr>
<td></td>
<td>report</td>
<td>(187), Sporadic retinoblastoma, bilateral</td>
<td></td>
<td>(0.3;1.8), Bilateral:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5 (0.2;1.4)</td>
</tr>
<tr>
<td>Infante-Rivard et al. 2000</td>
<td>Antibiotics (118). Maternal retrospective</td>
<td>ALL (491)</td>
<td>Matching on sex, age, and area of residence, adjustment for maternal age and education</td>
<td>1.5 (1.0;2.2)</td>
</tr>
<tr>
<td>[109]</td>
<td>report</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaatsch et al. 2010 [110]</td>
<td>Antibiotics (224). Maternal retrospective</td>
<td>Main outcome: Any cancer (1867), secondary</td>
<td>Matching on community, sex, and year of birth, adjustment for sex, age, degree of urbanisation, and socioeconomic position</td>
<td>ALL: 1.2 (0.8;1.8),</td>
</tr>
<tr>
<td></td>
<td>report</td>
<td>outcomes: ALL (650), AML (105), Burkitt</td>
<td></td>
<td>AML: 3.2 (1.7;6.0),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lymphoma (12),</td>
<td></td>
<td>Burkitt lymphoma: 7.4 (1.8;30.1), NHL: 1.3 (0.6;3.0), Medulloblastoma: 2.1 (1.0;4.2),</td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Resultsa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kumar et al. 2014</td>
<td>Chloramphenicol (not reported). Maternal retrospective report</td>
<td>NHL (141), medulloblastoma (94), astrocytoma (92), ependymoma (44), neuroblastoma (160), nephroblastoma (147), bone tumour (97), soft tissue sarcoma (137)</td>
<td>Case-control (264). Study characteristics. Statistical analysis Matching on age, sex, and residency</td>
<td>Astrocytoma: 2.3 (1.1;4.7), Ependymoma: 1.2 (0.4;4.1), Neuroblastoma: 1.3 (0.7;2.5), Nephroblastoma: 1.1 (0.6;2.3), Bone tumour: 1.4 (0.5;4.1), Soft tissue sarcoma: 0.8 (0.3;2.0)</td>
</tr>
<tr>
<td></td>
<td>Medication (number of exposed). Ascertainment. Exposure window: during pregnancy, unless stated otherwise</td>
<td>ALL or AML (132)</td>
<td>Matching on age, sex, and residency</td>
<td>Not reported. Stated that more cases than controls were exposed, but that p=0.695</td>
</tr>
<tr>
<td>Reference</td>
<td>Exposure</td>
<td>Outcome</td>
<td>Materials and methods</td>
<td>Results^a</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------------------</td>
<td>----------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Kwan et al. 2006 [112]</td>
<td>Antibiotics three months before and/or during pregnancy (128). Maternal retrospective report</td>
<td>Main outcome: Leukaemia (365), secondary outcome: ALL (311)</td>
<td>Case-control (825). US, NCCLS, ages 0-14 years. Conditional logistic regression</td>
<td>Leukaemia: 1.0 (0.7;1.5), ALL: 1.0 (0.6;1.5)</td>
</tr>
<tr>
<td>McKinney et al. 1987 [114]</td>
<td>Antibiotics (not reported). Maternal retrospective report and medical records, findings from the two data sources are reported separately</td>
<td>Leukaemias or lymphomas (234)</td>
<td>Case-control (702). UK, IRESCC, ages 0-15 years. Mantel-Haenszel</td>
<td>Matching on age and sex</td>
</tr>
<tr>
<td>Reference</td>
<td>Medication (number of exposed). Ascertainment. Exposure window: during pregnancy, unless stated otherwise</td>
<td>Type of cancer (number of cases)</td>
<td>Design (sample size). Study characteristics. Statistical analysis</td>
<td>Method to account for confounding in the main analysis</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>McKinney et al. 1999 [115]</td>
<td>Antibiotics (60). Routinely collected health data</td>
<td>Leukaemia (144), secondary outcome: ALL (124), lymphomas (45), CNS tumours (75), other solid tumours (126)</td>
<td>Case-control (415). Scotland, ages 0-14 years, some of this data overlaps with the UKCCS, but the amount of overlap is unclear. Conditional logistic regression</td>
<td>Matching on age, area of residence, and sex</td>
</tr>
<tr>
<td>Michalek et al. 1996 [286]</td>
<td>Medications for urinary tract infection (26), Medications for bladder infection (12). Maternal retrospective report</td>
<td>Neuroblastoma (183)</td>
<td>Case-control (555). New York, ages 0-14 years. Unconditional logistic regression</td>
<td>Matching on year of birth</td>
</tr>
<tr>
<td>Momen et al. 2015 [117]</td>
<td>Main exposure: Antibiotics (506,194), secondary exposures: Antibiotics on 3rd ATC-level,</td>
<td>Any cancer (1479), secondary outcomes: Leukaemias (591), ALL (444)</td>
<td>Cohort (1,442,114). Denmark and Sweden, ages 0-14 years. Cox regression</td>
<td>Adjustment for maternal age, parity, education, smoking during pregnancy, and country of birth</td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Results&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medication (number of exposed). Ascertainment. Exposure window: during pregnancy, unless stated otherwise</td>
<td>AML (74), HL (23), NHL (20), CNS or SNS tumours (249), renal tumours (121), hepatic tumours (29), endocrine tumours (93), testis cancer (21), eye cancers (77), bone cancers (44)</td>
<td>Design (sample size). Study characteristics. Statistical analysis</td>
<td>CNS/SNS tumours: 1.2 (0.9:1.7), Eye: 0.7 (0.4:1.5), Renal: 1.1 (0.7:1.8), Hepatic: 1.0 (0.4:2.7), Bone: 1.2 (0.6:2.4), Endocrine: 2.1 (1.3:3.3); Pivmecillinam: Leukaemias: 1.1 (0.8:1.5), CNS/SNS tumours: 0.7 (0.4:1.2), Eye: 1.6 (0.8:3.1), Renal: 1.2 (0.7:2.3), Endocrine: 1.0 (0.5:2.1); Pivampicillin: Leukaemias: 1.6 (1.1:2.2), CNS/SNS tumours: 1.2 (0.6:2.5), Renal: 1.5 (0.6:3.6), Hepatic: 8.3 (2.9:24.0);</td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Results(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medication (number of exposed). Ascertainty. Exposure window: during pregnancy, unless stated otherwise</td>
<td>Type of cancer (number of cases)</td>
<td>Design (sample size). Study characteristics. Statistical analysis</td>
<td>Method to account for confounding in the main analysis</td>
</tr>
<tr>
<td></td>
<td>Amoxicillin:</td>
<td>Leukaemias: 1.2 (0.8;1.8), CNS/SNS tumours: 1.3 (0.7;2.4), Eye: 2.4 (1.1;5.6).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sulfamethizole:</td>
<td>Leukaemias: 1.1 (0.8;1.5), CNS/SNS tumours: 0.8 (0.4;1.6), Eye: 1.1 (0.4;2.7), Endocrine: 1.3 (0.6;2.6);</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nitrofurantoin:</td>
<td>Leukaemias: 1.6 (1.0;2.4), CNS/SNS tumours: 1.4 (0.7;2.5); Cefadroxil:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leukaemias: 1.4 (0.6;3.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Resultsa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naumburg et al. 2002 [118]</td>
<td>Local treatment for vaginal infections (88), Oral antibiotics (76)</td>
<td>Main outcome: Leukaemia (652)</td>
<td>Case-cohort (1304). Sweden, age 0-16 years. Conditional logistic regression</td>
<td>Local treatment for vaginal infections: 1.7 (1.0;2.6), Oral antibiotics: 1.2 (0.7;1.8)</td>
</tr>
<tr>
<td>Pombo de Oliveira et al. 2006[119]</td>
<td>Amoxicillin (87), Ciprofloxacin (16) – each 3 months before pregnancy,</td>
<td>Infant acute leukemias (202)</td>
<td>Case-control (642). Brazil, ages 0-21 months. Unconditional logistic regression</td>
<td>Amoxicillin: 0.9 (0.6;1.3), Ciprofloxacin: 0.9 (0.3;2.8)</td>
</tr>
<tr>
<td></td>
<td>during pregnancy, or during breastfeeding. Maternal retrospective report</td>
<td></td>
<td>Matching on age and region, adjustment for region, sex, income, maternal age, and birth weight</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Exposure</td>
<td>Outcome</td>
<td>Materials and methods</td>
<td>Results</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>---------</td>
<td>-----------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Robison et al. 1989 [120]</td>
<td>Antibiotics in the year before, through pregnancy and breastfeeding (97). <em>Maternal retrospective report, validated by comparisons to associations in four other independent case-control studies where cases of different cancers did not report more medication use than controls</em></td>
<td>ANLL (204)</td>
<td>Case-control (408). US, ages 0-18 years. Conditional logistic regression</td>
<td>1.3 (0.8;2.0)</td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Results¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roman et al. 1997</td>
<td>Antibiotics (27).</td>
<td>Main outcomes: Leukaemias (143), NHL (34), secondary outcomes: ALL (113), AML (15)</td>
<td>Case-control (429). England, ages 3 months to 29 years. Conditional logistic regression</td>
<td>Total leukaemias: 0.8 (0.4;2.0), ALL: 0.7 (0.2;1.8), AML: ∞(0.3;∞), NHL : 1.7(0.1;21.1)</td>
</tr>
<tr>
<td></td>
<td>Routinely collected health data</td>
<td>Main outcomes: Leukaemias (143), NHL (34), secondary outcomes: ALL (113), AML (15)</td>
<td>Matching on hospital catchment area of birth, sex, year and month of birth</td>
<td></td>
</tr>
<tr>
<td>Ross et al. 2003</td>
<td>Amoxicillin (34), Ampicillin (49), Cefaclor (7), Cephalexin (13), Clotrimazole (24), Cotrimoxazole (11), Erythromycin (43), Metronidazole (8), Miconazole (61), Nitrofurantoin (12), Nystatin (11), Penicillin V potassium (11), Triple sulfa (10). Routinely collected health data</td>
<td>IL (243), ALL (157), AML (77)</td>
<td>Case-control (636). US, ages 0-18 months. Conditional logistic regression</td>
<td>Matching on birth year and telephone area, adjustment for maternal age, education and income</td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome Type of cancer (number of cases)</th>
<th>Materials and methods Design (sample size). Study characteristics. Statistical analysis</th>
<th>Results$^a$ Estimate of association (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medication (number of exposed). Ascertainment. Exposure window: during pregnancy, unless stated otherwise</td>
<td>Clotrimazole: IL: 0.9 (0.4;2.2), ALL: 1.0 (0.4;2.9), AML: 0.7 (0.1;3.9); Co-trimoxazole: IL: 1.8 (0.5;6.7), ALL: 11.2 (1.5;∞), AML: 0.3(0.03;2.3); Erythromycin: IL: 0.5 (0.3;1.1), ALL: 0.7 (0.3;1.6), AML: 0.4 (0.1;1.4); Metronidazole: IL: 1.4 (0.3;5.9), ALL: 2.0 (0.4;9.8), AML: -; Miconazole: IL: 0.9 (0.5;1.7), ALL: 0.7 (0.4;1.5), AML: 1.5 (0.6;4.1); Nitrofurantoin: IL: 0.8 (0.2;2.7), ALL: 1.1 (0.3;4.3), AML: -;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Results¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salonen et al. 1976</td>
<td>Sulphonamides (24), Penicillins (28), Tetracyclines (9).</td>
<td>Type of cancer (number of cases)</td>
<td>Method to account for confounding in the main analysis</td>
<td>Estimate of association (95% CI)</td>
</tr>
<tr>
<td></td>
<td><em>Routinely collected health data</em></td>
<td>Routine collection health data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Main outcomes:</em> Leukaemias (373), brain tumours (245)</td>
<td>Design (sample size). Study characteristics</td>
<td>Statistical analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nystatin: IL: 0.4 (0.7;1.7), ALL: 0.2 (0.03;1.9), AML: 1.3 (0.1;21.2); <em>Penicillin V potassium:</em> IL: 0.9 (0.2;3.1), ALL: 1.6 (0.3;9.2), AML: 0.3 (0.03;3.0); <em>Triple sulfa:</em> IL: 4.1 (0.8;20.9), ALL: 3.2 (0.6;18.1), AML: -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Exposure</td>
<td>Outcome</td>
<td>Materials and methods</td>
<td>Results^a</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>-------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Sanders &amp; Draper 1979 [290]</td>
<td>Streptomycin (9). Maternal retrospective report and medical records, findings from the two data sources are reported separately</td>
<td>Any cancer (27)</td>
<td>Case-controls (40). UK, Oxford Survey of Childhood Cancers, ages 0-15. All children were born to mothers with pulmonary tuberculosis. Descriptive statistics</td>
<td>35% of cases were exposed and 33% of controls were exposed</td>
</tr>
</tbody>
</table>

Brain tumours: Sulphonamides: 2 exposed of 245 cases and 14 exposed of 972 controls. Penicillins: 1 exposed of 245 cases and 12 exposed of 972 controls. Tetracyclines: 0 exposed of 245 cases and 6 exposed of 972 controls.
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Results$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shaw et al. 2004 [123]</td>
<td>Antibiotics (87). <em>Maternal retrospective report</em></td>
<td>ALL (789)</td>
<td>Case-control (1578). <em>Canada, age 0-14 years</em>. Conditional logistic regression</td>
<td>1.3 (0.8;2.0)</td>
</tr>
<tr>
<td>Shaw et al. 2006 [291]</td>
<td>Antibiotics (30). <em>Maternal retrospective report</em></td>
<td>CNS tumour (272)</td>
<td>Case-control (544). <em>Canada, ages 0-15 years</em>. Conditional logistic regression</td>
<td>1.7 (0.8;3.6)</td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Resultsa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medication (number of exposed). Ascertainment. Exposure window: during pregnancy, unless stated otherwise</td>
<td>Type of cancer (number of cases)</td>
<td>Method to account for confounding in the main analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Design (sample size). Study characteristics. Statistical analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Method to account for confounding in the main analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stålberg et al. 2010</td>
<td>Penicillins (55), Other antibiotics (60). Routinely collected health data</td>
<td>Brain tumour (512)</td>
<td>Matching on sex and birth year, adjustment for maternal age, parity, maternal country of birth, and level of hospital (primary, secondary, or tertiary)</td>
<td>Penicillins: 0.6 (0.3;1.2), Other antibiotics: 0.8 (0.5;1.5)</td>
</tr>
<tr>
<td>[88]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thapa et al. 1998</td>
<td>Metronidazole (79,716 person-years). Routinely collected health data</td>
<td>Main outcome: Any cancer (175), secondary outcomes: leukaemias (42), CNS tumours (30), neuroblastoma (28), other cancers (75)</td>
<td>Adjustment for maternal age, residence area, ethnicity, marriage status, maternal education, and parity</td>
<td>Leukaemias: 0 exposed cases, CNS tumours: 1.2 (0.3;5.2), Neuroblastoma: 2.6 (0.9;7.6), Other cancers: 0.6 (0.2;1.8)</td>
</tr>
</tbody>
</table>
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Resultsa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Duijn et al. 1994 [124]</td>
<td>Antibiotics in the year before or during pregnancy (8). Maternal retrospective report, validated by comparison to associations in another case-control study where cases of a different type of cancer did not report more medication use than controls</td>
<td>ANLL (80)</td>
<td>Case-control (320). The Netherlands, ages 0-14 years. The study reported results for both population- and cancer controls, here the population controls are reported. Conditional logistic regression</td>
<td>1.0 (0.2;5.7)</td>
</tr>
<tr>
<td>Van Steensel-Moll et al. 1985 [125]</td>
<td>Antibiotics (27). Maternal retrospective report</td>
<td>Leukaemias (519)</td>
<td>Case-control (1026). The Netherlands, ages 0-15 years. Logistic regression</td>
<td>1.1 (0.5;2.3)</td>
</tr>
<tr>
<td>Wen et al. 2002 [126]</td>
<td>Antibiotics in the year before pregnancy through pregnancy and breast-feeding (400). Maternal retrospective report</td>
<td>ALL (1842)</td>
<td>Case-control (3828). US, ages 0-14 years. Conditional logistic regression</td>
<td>Exposure both before and during pregnancy: 0.9 (0.4:1.8), only during pregnancy: 1.0 (0.8:1.3)</td>
</tr>
<tr>
<td>Reference</td>
<td>Exposure</td>
<td>Outcome</td>
<td>Materials and methods</td>
<td>Results&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Ye et al. 2019 [127]</td>
<td>Antibiotics (98,997). Routinely collected health data</td>
<td>Main outcome: Any cancer, excluding nonmelanoma skin cancer (361), secondary outcomes: Leukaemias (141), ALL (114)</td>
<td>Cohort (262,116). Canada, ages 0-19 years. Cox regression</td>
<td>Cancers diagnosed within age of 5 years: Leukaemias: Exposure anytime during pregnancy: 1.1 (0.7;1.7), First trimester exposure: 1.5 (0.9;2.5), Second trimester exposure: 0.9 (0.5;1.5), Third trimester exposure: 0.9 (0.5;1.6); ALL: Exposure anytime during pregnancy: 1.0 (0.6;1.6), First trimester exposure: 1.5 (0.9;2.6), Second trimester exposure: 0.8 (0.4;1.5), Third trimester exposure: 0.8 (0.4;1.6);</td>
</tr>
</tbody>
</table>

<sup>a</sup> Estimate of association (95% CI)
Table B.2: Studies on the association between prenatal exposure to antibiotics and childhood cancer, continued.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Materials and methods</th>
<th>Resultsa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medication (number of exposed), Ascertainment. Exposure window: during pregnancy, unless stated otherwise</td>
<td>Type of cancer (number of cases)</td>
<td>Design (sample size). Study characteristics. Statistical analysis</td>
<td>Estimate of association (95% CI)</td>
</tr>
</tbody>
</table>

Cancers diagnosed within age of 19 years: Leukaemias: Exposure anytime during pregnancy: 1.3 (0.9;1.8), First trimester exposure: 1.3 (0.9;2.0), Second trimester exposure: 1.1 (0.7;1.7), Third trimester exposure: 1.1 (0.7;1.7); ALL: Exposure anytime during pregnancy: 1.1 (0.7;1.6), First trimester exposure: 1.3 (0.8;2.1), Second trimester exposure: 0.9 (0.6;1.5), Third trimester exposure: 1.0 (0.6;1.7)

aWhen available, adjusted estimates from the most detailed exposure- and outcome classification. In studies that report results from more than one model, the model that does not adjust for intermediate factors is reported.

References


References


[34] Zhu, T. et al. ‘Meta-analysis of antenatal infection and risk of asthma and eczema’ in: Medicine vol. 95, no. 35 (2016), e4671.


[38] UK teratology information service. ANTIBIOTIC USE IN PREGNANCY. UKTIS. 2013. URL: https://www.medicinesinpregnancy.org/bumps/monographs/ANTIBIOTIC-USE-IN-PREGNANCY/ (visited on 12/02/2021).
References


[43] Manzoku-Kanja, K. *Pregnancy and flu - Colds and flu - Kaiser Permanente*. Kaiser Permanente. 2018. URL: https://healthy.kaiserpermanente.org/health/care/ld/p/a0/FchLDslgEADQs_QAwkh7CXXD15BYTehWEIgaBD1-uryPbR4Q8v0DjyVkJniz8Z5rr6cXY7bE4g3uMcXXtGiPQrtidBwBkfuf4f9HpQYXPZp16I/UWi4CuuywgpRylj0J9Cy1aqdR9YPClyX1mZrmC-Hs4SM/ (visited on 04/02/2021).


References


References


References


References


References


253
References


References


References


References


Boggess, T. and Risher, W. C. ‘Clinical and basic research investigations into the long-term effects of prenatal opioid exposure on brain development’. In: *Journal of Neuroscience Research* (2020).

Philippot, G. et al. ‘Short-term exposure and long-term consequences of neonatal exposure to delta(9)-tetrahydrocannabinol (THC) and ibuprofen in mice’. In: *Behavioural Brain Research* vol. 307 (2016), pp. 137–144.


[254] Wood, M. E. et al. ‘Longitudinal changes in neurodevelopmental outcomes between 18 and 36 months in children with prenatal triptan exposure: findings from the Norwegian Mother and Child Cohort Study.’ In: BMJ open vol. 6, no. 9 (2016), e011971.


References


