Risk of psychiatric disorders in children delivered by cesarean section or treated with antibiotics in infancy

– a national cohort study using sibling designs to evaluate causal effects attributed to potential alterations of the infant microbiota

PhD thesis

Paul Axelsson Bryde, MD

2019

Supervisors

MD, PhD, Professor Ellen Christine Leth Løkkegaard

MD, PhD, Associate professor Tine Dalsgaard Clausen

MD, PhD, Associate professor Thomas Bergholt

MD, Ida Hageman Pedersen

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen on the 27/9/2019

Defense date 24th of January 2020 at Nordsjællands Hospital Hillerød.

Assessment committee:

Lone Krebs, Professor, Consultant, MD, DMSc, Hvidovre Hospital

Søren Dalsgaard, Professor, Consultant, MD, PhD, Aarhus University

Olof Stephansson, Associate Professor, Consultant, MD, PhD, Karolinska Institutet, Sweden

Author: Paul Vignir Axelsson Bryde, MD

Department of Obstetrics and Gynaecology
Nordsjællands Hospital, Hillerød
Dyrehavevej 29, 3400 Hillerød

Title: Risk of psychiatric disorders in children delivered by cesarean section or treated with antibiotics in infancy – a national cohort study using sibling designs to evaluate causal effects attributed to potential alterations of the infant microbiota

Academic supervisors: Principal supervisor:

Ellen Christine Leth Løkkegaard, Professor, MD, PhD

Department of Obstetrics and Gynaecology
Nordsjællands Hospital, Hillerød
Dyrehavevej 29, 3400 Hillerød

Primary co-supervisor:

Tine Dalsgaard Clausen, Associate Professor, MD, PhD

Department of Obstetrics and Gynaecology
Nordsjællands Hospital, Hillerød
Dyrehavevej 29, 3400 Hillerød

Co-supervisors:

Ida Hageman Pedersen, MD, Deputy Director

Mental Health Services
Capital Region of Denmark
Kristineberg 3, 2100 Copenhagen

Thomas Bergholt, MD, PhD, MSc

Department of Obstetrics
Juliane Marie Centre, Rigshospitalet
Blegdamsvej 9, 2100 Copenhagen

Acknowledgements

There are many people that I am indebted to, but this project would not have been possible without my main supervisor Professor Ellen Christine Leth Løkkegaard. When I was still fresh out of medical school, she took me under her wings, guided my career and steered me towards this PhD project. With her vast knowledge of epidemiological research, always taking the time to answer my numerous questions, and her positive disposition, I could not have asked for a better supervisor.

This project and my education would not have been the same without my main co-supervisor Tine Dalsgaard Clausen, devoting so much of her time to guide me through the labyrinth that is register-based research. I have lost count of all the times she has diverted the project and myself from disastrous errors, with her meticulous and conscientious mind. The passion you and Ellen have for research has been an inspiration for my continued academic journey.

In my humble opinion, Anne Helby Petersen and Professor Niels Keiding each represent the brightest statistical minds of their respective generations. Were it not for them and their skills in communicating complex topics, the intricacies of interpreting causality in observational data would have been lost on me. Their work has been above and beyond what I could have hoped for; there was no cutting corners or settling for the next best thing. I have dutifully taken notes during our conversations and if you see a very eloquently and carefully worded phrase in my thesis, chances are that its origins are with these two good people. I consider myself very lucky to have had their guidance during this project.

To Professor Anja Pinborg and my supervisor Ida Hageman Pedersen, I would like to say thank you for coming up with the idea for the project, setting the mood with late evening Indian cuisine, and always be willing to expediently help with everything I needed. Professor Lars Vedel Kessing has also been instrumental in helping us understanding psychiatric register-based research, with all its pitfalls.

I would also like to thank my supervisor Thomas Bergholt, for all his help and support throughout the years.

No one knows the Danish Medical Birth Register like Steen Rasmussen and without him working his magic, this project would have been a lot more difficult. Thank you for the many hours spent together programming and binging on Icelandic candy.

The staff and colleagues at the Department of Clinical Research have also been most helpful in everything: from the daily practicalities (nothing gets done without a computer and a connection to the internet!), to help with securing funding, research approvals, arranging lecturers to come to give a talk, and keeping up morale. I am also extremely grateful to my wonderful colleagues at the Department of Obstetrics and

Gynaecology for ensuring my continued interest in becoming an OBGYN, and for the unwavering support of our Department Chief Peter Hornnes.

I am very grateful to the financial support of the Capital Region Denmark Research Fund, the Jascha Foundation, the Capital Region Denmark PhD-start Fund, the Nordsjælland Hospital Hillerød Research Fund, the Tvergaard Fund and the Gangsted Fund.

Last but not least, none of this would have been possible without the devoted support and encouragement of my loving, understanding and thoughtful wife, Linda Axelsson, and my sweet daughters Victoria and Erika - both delivered by cesarean. Thank you for bearing with me!

I have been extraordinarily privileged to have had your help and guidance. To all of you, a heartfelt thank you!

-Paul

Table of contents

Preface

This thesis represents the conclusion of my PhD fellowship at the department of Gynaecology and Obstetrics, Nordsjællands Hospital Hillerød, in the period 2016-2019.

The thesis is based on the following three studies:

Study I:

Axelsson PB, Clausen TD, Petersen AH, Hageman I, Pinborg A, Kessing LV, Bergholt T, Rasmussen SC, Keiding N, Løkkegaard ECL. **Investigating the effects of cesarean delivery and antibiotic use in early childhood on risk of later attention deficit hyperactivity disorder.** J Child Psychol Psychiatry 2019 Feb; 60(2):151-159.

Study II:

Axelsson PB, Clausen TD, Petersen AH, Hageman I, Pinborg A, Kessing LV, Bergholt T, Rasmussen SC, Keiding N, Løkkegaard ECL. **Relation Between Infant Microbiota and Autism?: Results from a National Cohort Sibling Design Study.** Epidemiology 2019 Jan; 30(1):52-60.

Study III:

Axelsson PB, Petersen AH, Hageman I, Pinborg A, Kessing LV, Bergholt T, Rasmussen SC, Keiding N, Clausen TD, Løkkegaard ECL. **Is cesarean section a cause of affective disorders? – A national cohort study using sibling designs.** Journal of Affective Disorders 2019 Jul [Submitted].

Abbreviations

AB:	Broader spectrum Antibiotics
ADHD:	Attention Deficit Hyperactivity Disorder
ADD:	Attention Deficit Disorder
ATC:	Anatomical Therapeutic Chemical
ASD:	Autism Spectrum Disorder
CS:	Cesarean Section
CSIP:	Intrapartum Cesarean Delivery
CSPL:	Pre-labor Cesarean Delivery
DMBR:	Danish Medical Birth Registry
DSM:	Diagnostic and Statistical Manual of mental disorders
DSOG:	Danish Society of Obstetricians and Gynecologists
GA:	Gestational Age
GDM:	Gestational Diabetes Mellitus
HR:	Hazard Ratio
ICD8:	International Classification of Diseases 8th revision
ICD10:	International Classification of Diseases 10th revision
MBR:	Medical Birth Registry
NCSP:	NOMESCO Classification of Surgical Procedures
NOMESCO:	Nordic Medico-Statistical Committee
Penc:	Penicillin
ROM:	Rupture Of Membranes
SGA:	Small for Gestational Age
Vag:	Vaginal delivery
WHO:	World Health Organization

Summary

There is no lack of hypotheses regarding the cause of the apparent increase in most psychiatric disorders. However, as the role of the gut microbiota in various health outcomes is becoming seemingly more apparent, there has been an increasing interest in investigating if the gut microbiota might in some part explain this increase in incidence. Associations have been found between specific compositions of the gut microbiota and common psychiatric disorders such as autism spectrum disorders (ASD) and affective disorders, with similar associations expected to be true for attention deficit hyperactivity disorder (ADHD).

Some believe that there is a critical developmental window in infancy where the gut microbiota has a greater influence on the neurodevelopment of the infant through biochemical signaling - the *gut-brain axis*. Studies have shown that children born by cesarean delivery have different gut microbiota than those born by vaginal delivery, with increased abundance of bacteria from typical hospital environment and decreased abundance of bacteria that are believed to be beneficial. What the effects of antibiotics in early life are on the gut microbiota is unclear at present. Yet, inferring from studies of the adult gut microbiota it stands to reason that antibiotics will cause some changes to the diversity and abundance of gut bacteria, with greater effect of broader spectrum antibiotics than penicillin - although these changes may only be temporary.

Observational studies have indeed found increased risk of ADHD and ASD in children born by cesarean delivery and increased risk of affective disorders and ADHD after antibiotic treatment. But as confounder control is imperative in observational studies, we are left uncertain and thinking that the effect might merely be due to residual confounding. Therefore, we want to investigate if these are causal effects.

The implication is that if cesarean delivery and antibiotic treatment can have detrimental effects on gut microbiota, it may also be affected with favorable treatments such as probiotics or microbiota transplantation. This has even prompted some parents of infants born by cesarean delivery to try to emulate vaginal birth by swabbing their infant in vaginal microbes from the mother. There is also ongoing research into the effects of fecal microbiota transplant to children with autism.

In three separate studies, we used information from the Danish national registries regarding parental characteristics, in particular socio-economic and psychiatric status at the time of the infant's birth as well as relevant obstetric factors, we followed children until an event of a diagnosis of ADHD, ASD or affective disorder, death, emigration, a more severe psychiatric disorder, or until the end of follow-up at 31 December 2014.

We divided the birth cohort into two overlapping periods. Those born in 1982-2001 were examined in relation to delivery mode and risk of affective disorders after the age of 13 years. Those born in 1997-2010

were examined both in relation to delivery mode as well as antibiotic exposure in the first two years of life and risk of either ADHD or ASD after the age of two years. Using three different statistical methods, a standard Cox regression model, a sibling stratified Cox model, and a between-within sibling model, we aimed to evaluate potential causal effects. The three different outcomes of either ADHD, ASD, or affective disorders, were addressed in separate papers.

In the first two papers, we studied the risk of ADHD and ASD respectively, depending on mode of delivery: defined as pre-labor cesarean delivery or intrapartum cesarean delivery, compared to vaginal delivery – as well as antibiotic exposure in the first two years of life: defined as exclusively penicillin treatment or broader spectrum antibiotic than penicillin, compared to no antibiotic treatment. We included almost 671,600 live-born singleton children born to Danish parents, that were still alive at their second birthday, not having emigrated from Denmark and not diagnosed with ADHD or ASD (as relevant to each study) at the start of follow-up. We adjusted for numerous parental characteristics and obstetric factors.

As previous studies have found, the rate of cesarean deliveries during the period increased by more than 60%, in particular due to prelabour cesarean deliveries. There did not appear to be a change in the pattern for antibiotic prescriptions for infants under the age of two years. However, the incidence of ADHD rose greatly until the year 2011, where it flattened out. Similarly, the incidence of ASD rose after the year 2008 with an even greater increase after 2012 without any signs of levelling out before end of follow-up.

For the first two papers regarding ADHD and ASD, the standard Cox model found significantly increased risk for all exposures, which were attenuated in the fully adjusted models. The sibling stratified Cox model found significant increased risk of ADHD for intrapartum cesarean delivery and increased risk of ASD for broader spectrum antibiotics. The between-within sibling model found no significant effects of any exposure on ADHD and ASD risk.

For the third paper regarding affective disorders, we only studied the effects of mode of delivery but in a larger population of 1,009,444 included children, of singleton birth, Danish descent, with complete information for all variables and alive at 30 days past their 13th birthday. There were considerable changes in the incidence of affective disorders between the years 1995 and 2015, with a hotspot between the years 2007-2012 and a generally a lower age at first diagnosis for the later years. The standard Cox model found significant but small effects of intrapartum and pre-labor cesarean delivery. In the sibling stratified Cox model, these effects disappeared for the pre-labor cesarean delivery, but the effect size was similar for intrapartum cesarean delivery with wider confidence intervals, thus being non-significant. In the between-within model, pre-labor cesarean delivery was not associated with increased risk, but intrapartum cesarean

delivery was associated with very small and only borderline-significant increased risk, fully adjusted hazard ratio 1.05 (95%CI 1.003-1.12).

The 95% confidence intervals of the between-within sibling model were considerably narrower than those of the sibling stratified Cox model, for both ADHD and ASD studies. However, this difference was less pronounced in the affective disorder study, as was expected due to the larger sample size. However, the conclusions of the papers would have been somewhat different had we only relied on the results of the sibling stratified Cox model, instead of our main model, the between-within sibling model. For the ADHD and ASD studies we would have concluded that there were significantly increased risks of intrapartum cesarean delivery and broader spectrum antibiotics, respectively. However, for the affective disorders study we would have concluded that the risks were not significantly increased for any exposure, which is contrary to what the between-within model found. Overall, the effect estimates from the between-within models were mostly small and it is unlikely that such findings are applicable in a clinical context.

We performed several sensitivity analyses, without their results altering our main conclusion. Notably, we found that there was no evidence of an interaction effect between antibiotic treatment and mode of delivery for the ADHD and ASD studies. For the study of affective disorder, we did not consider interaction effects as we were only working with the mode of delivery exposure. Additionally, although we found substantial birth and diagnostic year effects, these effects did not seem to confound the exposure-outcome relation in the sibling models. However, we saw some evidence of first-born children diagnosed with ASD compared to other first-born children, were less likely to have younger siblings which might be due to reproduction stoppage. In standard Cox sensitivity analysis, there were some differences in effect estimates between first-born children and those with older siblings and therefore the results of the ASD study may not be generalizable to comprise children without siblings.

In conclusion, we believe that despite having tested multiple avenues of the gut-brain hypothesis, we have found no reasonable evidence that supports causal relationship between cesarean delivery, antibiotic treatment and later increased risk of ADHD, ASD or affective disorders. This conclusion stands on the firm basis of a null-finding in almost all of our results, or findings that contradict the hypothesis. Therefore, we do not believe that treatments aimed at correcting the gut microbiota of children that have received antibiotics in infancy or born by cesarean delivery will have any preventative measure against these psychiatric disorders. The sibling models have proved that they can be a valuable tool in the methodology-arsenal of epidemiologists when evaluating causality, although there are some inherent limitations.

Dansk resumé (Danish summary)

Der mangler ikke hypoteser der forklarer den tilsyneladende stigning i mange psykiatriske lidelser. Men eftersom vi bedre forstår sammenhængen mellem tarmmikrobiota og adskillige sygdomsforhold, har der været øget interesse i at undersøge om tarmmikrobiota har en delvis forklarende rolle i denne forøgede forekomst af psykisk sygdom. Der er fundet sammenhænge mellem specifikke sammensætninger af tarmbakterier og psykiatriske sygdomme som fx autisme spektrum sygdom (ASD) samt depression og bipolar lidelse (affektive lidelser), og lignende sammenhænge forventes at gøre sig gældende for hyperkinetisk adfærds og opmærksomheds-forstyrrelse (ADHD).

Der menes at være et kritisk udviklingsvindue i den tidlige barndom hvor tarmmikrobiota har større indflydelse på den neurologiske udvikling hos barnet gennem biokemiske signaler, såkaldt *tarm-hjerne akse*. Desuden har studier vist at børn født ved kejsersnit har anden tarmmikrobiota end børn født vaginalt, med øget forekomst af bakterier der typisk kommer fra hospitalsmiljø og nedsat forekomst af bakterier man mener er sundhedsfremmende. Hvilken effekt antibiotika har på tarmmikrobiota i den tidlige barndom er ikke helt klart endnu. Men hvis man udleder fra studier vedrørende de voksnes tarmmikrobiota, fremstår det sandsynligt at antibiotika vil forårsage ændringer i diversitet og forekomst af tarmbakterier, med større effekt af mere bredspektret antibiotika end penicillin – selvom disse forandringer kan være midlertidige.

Observationsstudier har nemlig fundet øget risiko for ADHD og ASD hos børn født ved kejsersnit, samt øget risiko for affektive lidelser og ADHD efter antibiotika behandling. Men det er vigtigt at tage højde for confoundere (faktorer der påvirker både eksponering og udfald) i observationsstuderier, og derfor ved vi ikke om effekten de finder er blot forårsaget af resterende confounding. Vi vil derfor undersøge om der er tale om kausale effekter.

Tanken er at hvis forløsning ved kejsersnit og antibiotika behandling i tidlig barndom kan have negativ effekt på tarmmikrobiota, kan mikrobiota også påvirkes positivt med behandlinger i form af probiotika eller mikrobiota transplantation. Det har endda fået nogen forældre til børn født ved kejsersnit til at prøve at efterligne vaginal fødsel ved at smøre børnene ind i klud med vaginale bakterier fra moderen. Der foregår desuden aktiv forskning hvor man undersøger effekten af tarmmikrobiota transplantation til børn med autisme.

Ved at bruge information fra de danske landsdækkende registre vedrørende forældrenes karakteristika, især socioøkonomisk og psykiatrisk status ved barnets fødsel i tillæg til relevante fødselsfaktorer, fulgte vi børn indtil tilfælde af ADHD, ASD eller affektive lidelser, død, udvandring, mere alvorlig psykiatrisk lidelse, eller slutningen af vores opfølgning den 31. december 2014.

Vi delte fødselskohorten i to: dem født i årene 1982-2001 blev undersøgt i forhold til fødselsmåde og risiko for affektive lidelser efter 13-årsalderen. Dem født i årene 1997-2010 blev undersøgt både i forhold til fødselsmåde og antibiotika behandling i de to første leveår og risiko for enten ADHD eller ASD efter 2-årsalderen. Vi brugte tre forskellige statistiske metoder, en standard Cox regressionsmodel, søskende stratificeret Cox model, samt between-within søskende model, for at vurdere potentielle årsagssammenhænge. De tre psykiatriske lidelser blev undersøgt i tre adskilte artikler.

I de to første artikler undersøgte vi risikoen for ADHD og ASD afhængig af fødselsmåde: defineret som enten kejsersnit foretaget før fødsel eller kejsersnit foretaget under fødsel og sammenlignet med vaginal fødsel – samt antibiotika behandling i de to første leveår: defineret som udelukkende penicillin behandling eller bredere spektrum antibiotika end penicillin, sammenlignet med ingen antibiotika behandling. Vi inkluderede næsten 671.600 levendefødte, enkeltfødte børn med danske forældre, hvor børnene var stadigvæk i live ved deres anden fødselsdag og havende hverken udvandret fra Danmark eller fået diagnosen ADHD eller ASD (i forhold til hvert studie) ved begyndelsen af opfølgningen. Vi justerede for en række egenskaber vedrørende forældrene og forholdene omkring fødslen.

Som tidligere studier har vist, er frekvensen af fødsler ved kejsersnit i perioden steget med over 60%, især på grund af kejsersnit foretaget før fødsel. Der så ikke ud til at være ændringer i mønstret for antibiotika recepter til børn under to-årsalderen. Til gengæld har incidensen af ADHD øgedes kraftigt indtil året 2011, hvor den fladede ud. På samme måde har incidensen af ASD steget siden året 2008, med endnu kraftigere stigning efter 2012 uden nogen tegn på at flade ud inden vores opfølgning sluttede.

I de to artikler vedrørende henholdsvis ADHD og ASD, fandt standard Cox modellen signifikant øget risiko for alle eksponeringer, som blev dæmpet i den fuldt-justerede model. Søskende stratificerede Cox modellen fandt signifikant øget risiko for ADHD ved kejsersnit foretaget under fødsel og øget risiko for ASD ved behandling med bredere spektrum antibiotika. Between-within søskende modellen fandt ingen signifikante effekter af nogen eksponering på risikoen for ADHD og ASD.

I den tredje artikel vedrørende affektive lidelser undersøgte vi kun effekten af fødselsmåde, men til gengæld i en større population: 1.009.444 inkluderede enkeltfødte børn, af Dansk herkomst, med komplet information for alle variable, samt levende 30 dage efter deres 13-års fødselsdag. Der var betydelige ændringer i forekomst af affektive lidelser i årene 1995 til 2015, med særligt mange tilfælde i årene 2007-2012 og lavere alder ved første diagnose i de senere år. Standard Cox modellen fandt statistisk signifikante men smalle effekter af både kejsersnit under og før fødsel. I søskende stratificerede Cox modellen forsvandt effekterne for kejsersnit før fødsel, men effekt størrelsen var lignende den i between-within modellen for kejsersnit under fødsel dog med bredere konfidens intervaller, og derfor værende ikke-

signifikant. I between-within modellen var kejsersnit før fødsel heller ikke forbundet med øget risiko, men kejsersnit under fødsel var forbundet med meget lidt og kun knapt signifikant øget risiko, med fuldt justeret hazard ratio på 1.05 (95%CI 1.003-1.12).

Konfidensintervallerne for between-within søskende modellen var betydeligt smallere end dem i søskende stratificerede Cox modellen, for både ADHD og ASD studierne. Til gengæld var disse forskelle noget mindre tydelige i studiet af affektive lidelser, som forventet på grund af den større studie-population. Konklusionerne ville til gengæld have været noget anderledes havde vi udelukkende brugt resultaterne fra den stratificerede Cox model, istedet for vores hovedmodel, between-within modellen. I ADHD og ASD studierne ville vi have konkluderet at der var statitistisk signifikant øget risiko ved henholdsvis kejsersnit under fødsel og behandling med bredere spektrum antibiotika for de respektive sygdomme. I studiet af affektive lidelser havde vi dog konkluderet at der var ikke signifikant øget risiko ved kejsersnit, som er modsat af hvad vi fandt i between-within modellen. Samlet set, var effekt estimaterne fra between-within modellen oftest små og det er usandsynligt at så små effekter er klinisk relevante.

Vi testede robustheden af vores fund i flere følsomhedsanalyser, uden at de påvirkede vores hovedkonklusion. Især vil vi nævne at der var ingen interaktionseffekt mellem fødselsmåde og antibiotika behandling i ADHD og ASD studierne. I studiet vedrørende affektive lidelser, undersøgte vi kun en enkelt eksponering (fødselsmåde) og derfor var det ikke relevant at kigge på interaktionseffekter. Derudover, fandt vi betydelige effekter af fødsels og diagnose-år, men de så ikke ud til at confounde forholdet mellem eksponering og udkomme i søskende analyserne. Til gengæld så vi at der er noget der tyder på at første fødte børn der får ASD diagnosen i forhold til førstefødte børn der ikke bliver diagnosticerede, får sjældnere yngre søskende, muligvis på grund af intenderet reproduktionsstop. I en følsomhedsanalyse foretaget i standard Cox modellen, var der nogen forskelle i effektstørrelser mellem førstefødte børn og de børn der havde ældre søskende og derfor kan resultaterne i ASD studiet muligvis ikke være generaliserbare til børn uden søskende.

Trods at have undersøgt flere aspekter i tarm-hjerne hypotesen, har vi ikke fundet nogen rimelige beviser der understøtter årssagssamenhæng mellem fødselsmåde, antibiotika behandling i tidlig barndom og senere øget risiko for ADHD, ASD og affektive lidelser. Denne konklusion når vi frem til ved hjælp af præcise nul-fund i næsten alle vores resultater, eller fund der går imod vores oprindelige hypotese. Derfor tror vi ikke at der vil være nogen beskyttende effekt mod disse psykiatriske sygdomme af behandlinger der sigter mod at korrigere tarmmikrobiota hos børn der har fået antibiotika i tidlig barndom eller dem der er født ved kejsersnit. Søskende modellerne har vist sig at være et værdifuldt værktøj i metode-værktøjskassen for epidemiologer der vil underøge kausale sammenhænge, trods modellernes indbyggede begrænsninger.

Background

The hypothesis

There have been many speculations as to the cause of the increase in incidence of attention deficit hyperactivity disorder (ADHD), autism spectrum disorders (ASD) and affective disorders witnessed for most industrialized countries over the last couple of decades. It is often attributed to changes in diagnostic practices and treatment possibilities coupled with de-stigmatization and awareness (1,2).

However, a prevailing theory suggests that the gut microbiota can influence the development of the brain in infancy (3). The gut microbiota consists of microbes such as bacteria, fungi and viruses, and its composition is related to several factors, such as whether you are born by cesarean section or are taking antibiotics (4,5). The theory is that if you are born vaginally, you will be colonized with your mother's vaginal and perineal bacteria whereas when you are born into the sterile environment of the operating theater, your first bacteria will be those typical of the hospital environment. After vaginal delivery, women are discharged earlier than after a cesarean section, contributing to the colonization with hospital bacteria. Further, children born by cesarean section are routinely exposed to antibiotics right before their birth, as it is the recommendation of most obstetric societies to give antibiotics 30-60 minutes prior to incision as a prophylactic against maternal postpartum infections (6). The antibiotics usually used for prophylaxis could be considered broader spectrum antibiotics, meaning that compared to penicillin they are effective against multiple bacterial families or both gram positive and negative bacteria. Common examples are ampicillin and cephalosporins (7), and they have been shown to remain in infants' bloodstream for up to 24 hours (8).

For adults, antibiotics can alter the composition of the gut microbiota during and immediately after treatment but the microbiota mostly reverts to its original composition after a short while, although some bacterial strains may be permanently lost (9). If the same holds true for infants, and even though the changes may be short lived, they could be of great consequence for the infant. We would also expect that the maternal microbiota that normally colonizes the infant during vaginal delivery, should play a role in the infant's risk burden of psychiatric disorders. Therefore, antibiotic treatment during pregnancy should also have detrimental effects on the infant microbiota and increase risk of psychiatric disorders (10).

It is believed that it is the very early infant microbiota that is most important in for modulating the developing infant brain or the "critical developmental window"-theory (11). The duration of this window is not known exactly (12,13). However, we know that any differences between children's microbiota due to factors such as mode of delivery and breastfeeding start to disappear after 6 months of age and children gain an adult-like microbiota around the age of 2 years (14). Interestingly, it is also around or little after 2

years of age that the first symptoms of ASD and ADHD usually start to become evident (15,16). If we assume that the critical window theory is accurate, it is therefore probably closed by the age of 2 years.

The gut-brain axis

Information on how exactly the gut influences the brain is in-complete but it is speculated that several mechanisms are in play, such as the vagal nerve, hypothalamic-pituitary-adrenal axis, tryptophan metabolism, the immune system, or bacterial metabolites (17,18).

At first, the theory that our gut can influence our brain in any way may seem somewhat implausible. Nevertheless, researchers have found convincing evidence in animal studies that changes to the gut microbiota can modulate behavior (17). For example, when germ-free mice are transplanted with fecal microbiota from children with ASD compared to controls, the test subjects showed hallmark autistic behaviors (19). Such interpretations of the behavior in rodents and similarity to human psychiatric disorders can be difficult but affords the possibility to examine causal effects we cannot perform on humans. A study on mice that were exposed to chronic social defeat, which entails placing a larger more aggressive mouse of another species in the same cage as the smaller test subject, showed altered gut microbiota (20). They then treated half the mice with a probiotic called *Lactobacillus rhamnosus*, which resulted in decreased anxiety-like behavior and prevented deficits in social interaction with mice of the same species (21). In yet another study, the authors also tested the effects of chronic social defeat on mice treated with broad spectrum antibiotics for 14 days (22). In contrast to control mice, these mice did not show symptoms of anhedonia, a core symptom of major depressive disorder, indicating that symptoms of anhedonia depended on the presence of a gut microbiota. So, the researchers argued that stress can alter the gut microbiota but altering the gut microbiota can also affect symptoms of stress and depression, meaning that this may be a bidirectional relationship between the brain and the gut.

Not everyone may have the patience to wait for further evidence of the gut-brain axis playing a role in psychiatric disorders, before resorting to methods to alter the gut microbiota. It may range anywhere from scientific research into the effects of fecal microbiota transplantation to patients with psychiatric disorders, patients going to alternative medicine clinics to ask for such fecal capsules (23,24), to parents trying to emulate vaginal delivery by smearing a swab of vaginal bacteria on neonates born by cesarean delivery (25,26). As the harmful effects of such treatments are unknown, it would be prudent to first test the theory that the gut microbiota has a causal relation with psychiatric disorders. But is that possible?

The problem:

Are children at increased risk of psychiatric disorders such as ASD, ADHD and affective disorders, if they attain a certain “harmful” gut microbiota in infancy? How do we investigate this question? Even with a

prospective observational cohort study, where the infant gut microbiota was sampled and analyzed, you would need a follow-up up until their early adolescence. Adjusting for confounding factors in studies of the gut microbiota can be very difficult; social status and genetics are intertwined with other factors that affect the early gut microbiota, such as breastfeeding (27), diet (28), genetics (29,30) and the parental microbiota, which in turn is affected by even more factors such as socioeconomic status (31) and psychiatric disorders (17,32,33). Meaning the study would require enormous resources and after a decade or more of follow-up, we would perhaps only have biased results as the cohort was not randomized.

How can we solve it:

There is an indirect way of measuring the effect of adverse changes to the gut microbiota on infant risk of psychiatric disorders. Numerous studies have found that children born by cesarean delivery have different gut bacteria than children born vaginally (34,35), and the microbiota associated with cesarean delivery is described as being less favorable than for vaginal delivery, with greater relative abundance of skin and "hospital" bacteria. Antibiotic treatment, and especially with broader spectrum antibiotics, may alter the composition of the infant microbiota with decreased diversity, which is considered a sign of dysbiosis, and increased relative abundance of opportunistic pathogens (36). Thus, cesarean delivery and antibiotic treatment in the first two years of life, may serve as surrogate markers for adverse changes in the infant gut microbiota.

Register data

These surrogate markers give us an opportunity to study these effects through prospectively collected observational data, in the Danish national registries. The strength of the Danish registries is clear, with vast number of validated health and social variables available at an individual level (37), in particular the validity of the psychiatric variables we are interested in here: ADHD (38), Autism (39) and Depression (40,41).

Of the Scandinavian countries, the Danish national registries were the first to include information from out-patient contacts, in 1995, compared to Sweden where they first became partially available in 2001 with full coverage five years later, and information from the Norwegian patient registry was first available on an individual level since 2008-2009 (37,42). For information on medical prescriptions, these became available in 1995 in Denmark, with complete maternal and infant prescription separation since 1997 (43). However, in Norway and Sweden these first became available in 2004 (44) and 2005 (45) respectively, which currently limits follow-up of a large portion of those countries' populations that have been treated in secondary care with medication.

One study found almost no difference between individuals with depression based either on psychiatric diagnoses or prescriptions for antidepressants, when it came to behavior related to admissions to a hospital for preventable disease (46). This suggests that antidepressants are a good surrogate marker for depression diagnoses, as both types of patients have the same inappropriate delay in contact to the health care system, presumably due to factors related to moderate or severe depression such as loss of energy and interest in daily activities.

Causality in epidemiological studies

Today, hardly anyone questions the causal relation between smoking and risk of cancer. Nonetheless, no randomized trials on humans have ever been conducted. Before the late 19th century, lung cancer was so rare that most doctors were never expected to encounter the disease. In the early 20th century, Adler reported (47) that in the period from 1840 to 1905, the incidence of lung cancer had increased from 1 in 129 deaths to 1 in 17 deaths – more than a 10 fold increase. Even though tobacco was suspected to cause cancer, it was only in the middle of the last century that convincing evidence started to grow. Cigarette smoking and lung cancer saw a parallel rise – which on the surface might resemble the rise in psychiatric disorders, cesarean delivery and antibiotic use. Just as with psychiatric disorders, it was believed that changes and improvements to diagnoses of lung cancer was the main reason for this increase in cases. However, case control studies in the 1930's and 40's started to indicate that the rate of smokers was higher for people dying of lung cancer, than those dying of other causes (48).

The tobacco industry's scientists fired back, claiming correlation does not equate causation, which is perfectly true (48,49). Their argument was that there could have existed an unknown variable that effected both the desire for cigarettes and the risk of lung cancer: a confounder. However, several large prospective cohort studies in the 1950's showed that especially heavy smokers were at greater risk of lung cancer and that the risk was reduced with smoking cessation (50,51). Later, randomized trials on animals have shown that there is a generally increased risk of any type of cancer if exposed to cigarette smoke were published (52) (even though this had been known to the tobacco industry for some time). In addition, there is some evidence that there is an interaction effect of smoking with alcohol consumption on cancer risk (53).

In essence, this very brief history of studying the link between smoking and cancer teaches us three lessons about use of observational data to study possible causality:

A) Correlation supports causality (but is not sufficient on its own)
B) Dose response / interaction / cessation effect is further support of causality
C) There must be a plausible hypothesis (such as evidence from laboratory or animal studies)

These are derivative of *Hill's criteria for causality,* most of which also are applicable here but were intended to be used as a guide rather than a checklist (54). One such important, but maybe somewhat obvious, criteria is that of temporality, meaning the effect must occur after the cause. We get back to that later.

In addition, we expect that if (almost) all confounding can be removed, the effect that remains must be the true effect of the exposure on the outcome. However, this last step just happens to be the very holy grail of epidemiology and usually considered unattainable. There are too many variables to measure and many of them are simply unknown and therefore cannot be adjusted for. In real life we are necessarily left with attempting to capture the *most important* confounding.

Sibling models

This is where sibling models come in. It is nothing new to use siblings, especially twin siblings, to study risk of disease where it is either unethical or difficult to do a randomized trial. Siblings share so many factors, such as genetics and socio-economic environment in the upbringing, that affect almost all health outcomes in early life (55). But there are some issues with using sibling comparison models, which we will discuss further in later sections of this thesis. However, I will mention one relevant issue that is often criticized: that sibling models generally lack power to detect possibly real effects (56).

Contrary to the standard Cox model, where every observation usually contributes to the effect estimates, a sibling stratified Cox modes does not use all observations, by design. First, obviously children without siblings do not contribute to the estimates. Second, within any family there have to be a pair of siblings that are discordant in exposure. Third, the sibling with the longest follow-up may not be the only child who experiences the outcome. In Figure 1 we give an example of a sibling pair that would not contribute with information to the estimation procedure and a pair of siblings that do contribute with information in Figure 2, although there are many other possible scenarios.

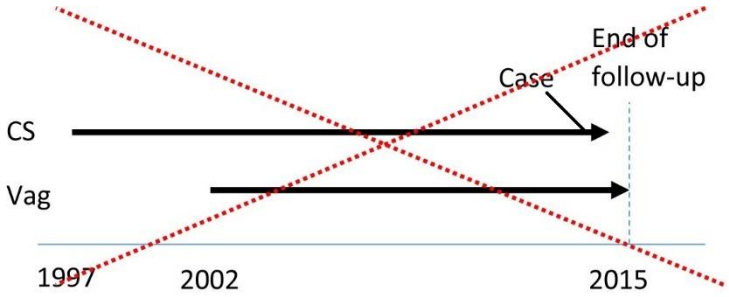


Figure 1: An example of a sibling pair not contributing with information to the effect estimates in the sibling models. One sibling was delivered by cesarean section (CS) and the other by vaginal birth (Vag) and thus discordant on exposure, with only one sibling experiencing the outcome (case), but the sibling not experiencing an event has a shorter follow-up.

Figure 2: An example of a sibling pair contributing with information to the effect estimates in the sibling models. Just like in the example given in figure 1, the siblings are both discordant on exposure (CS: Cesarean section, Vag: Vaginal birth) and outcome (Case), but here the sibling that did not experience an event has had a longer follow-up period.

It quickly becomes clear that as there is less information available for analysis in the stratified sibling model than in the standard Cox model, the effect estimates will become less precise. However, some children contribute with more information than others, meaning that the loss in precision may not be relative to the size of the total population used in the models. Furthermore, another sibling model suggested for survival analysis, the between-within model, had previously shown promise in simulations (57) and might provide greater precision than the stratified Cox model. As we are interested in evaluating causal effects and not population trends, the standard Cox model is less suited for the purpose as unobserved confounders are likely to affect the results (58). We discuss the statistical models further in the methods section.

Study objectives

We wished to examine the association between mode of delivery and psychiatric disorders, and if non-measurable factors related to familial confounding that are accounted for in a sibling model, would affect the results. We looked at the following psychiatric outcomes:

1. Attention deficit hyperactivity disorder (ADHD)
2. Autism spectrum disorders (ASD)
3. Depression and bipolar disorder (Affective disorders)

Methods

This thesis is based on three register-based cohort studies of live births in Denmark in the period 1982-2010 with follow-up until end of 2014. We used the between-within sibling model for our main analyses and compared these results to that of a standard Cox regression model and a sibling stratified Cox model.

We defined our cohorts based on each outcome.

- ADHD/ASD: the first cohort comprised children of singleton birth, born in the period 1997-2010, having Danish parents, complete information for all variables, still alive at their second birthday and not having experienced any events before that date.
- Affective disorders: the second cohort comprised children of singleton birth, born in the period 1982-2001, having Danish parents, complete information for all variables, and still alive at 30 days past their 13th birthday.

A total of 671,592 children were included in the analysis for the ADHD study, 671,606 children in the ASD study, and 1,009,444 children in the affective disorder study. Children were followed up until the first date of each respective outcome (ADHD, ASD, and affective disorders), or first emigration, death, a severe psychiatric disorder (ICD10 DF0, DF1, DF2) or the end of follow-up (31st December 2014). The process of defining the study population is described in Figure 1 of all three studies (58–60).

Data sources

We linked seven national registers to collect data for our study population: the Danish National Patient Registry (1977), the Medical Birth Registry (1973), Statistics Denmark (1978), The Fertility Database (1960), The Psychiatric Central Research Register (1969), The Register of Causes of Death (1970), and The Register of Medicinal Products Statistics (1995). Parentheses indicate first year of full availability of data. Further information is available in the supplementary material for papers I and II (58,59).

Definition of exposures

Mode of delivery

Information regarding cesarean delivery was attained from different variables. In the period 1982 to 1995, this information came from both the Medical Birth Registry (MBR) and the National Patient Registry (NPR). The information was coded manually by midwives for the MBR and by doctors for the NPR using a code classification developed by the Danish Health Ministry and first published in 1973 (61). For the period 1996 to 2010 we relied solely on the NPR, where codes were registered using the Nordic Medico-Statistical Committee (NOMESCO) classification of surgical procedures (NCSP)(37,62). Table 1 provides information on which diagnostic codes were used to define the two cesarean delivery variables and which register delivered the information.

Years	Register	Pre-labor cesarean delivery	Intrapartum cesarean delivery
1982-1995	MBR	Planned cesarean (b_i9), Cesarean before labor (b_sectiof)	Other operations (b_i10), Cesarean (b_i11), Cesarean during labor (b_sectiou)
1982-1995	NPR	Caesarea cervicalis ante partum (OPR65190)	CS Parva (OPR63620), CS Classica (OPR66020), CS Cervicalis (OPR66040), CS Vaginalis (OPR66060), CS classica (OPR78000)
1996-2010	NPR	CS in isthmus uteri performed as acute procedure before birth (KMCA10A), CS in isthmus uteri performed as planned procedure before birth (KMCA10B), CS before birth in isthmus uteri with exit technique (KMCA10C or KMCA10BX*), CS in isthmus uteri (KMCA10), CS in isthmus uteri ante partum (KMCA11)	CS in isthmus uteri during birth due to pregnancy complication/s (KMCA10D), CS in isthmus uteri during birth due to birth complication/s (KMCA10E), CS in isthmus uteri in partu (KMCA12), CS in isthmus uteri in partu complications graviditatis (KMCA12A), CS in isthmus uteri in partu complications partum (KMCA12B), CS corpus uteri (KMCA00), CS vaginalis (KMCA20), CS cervicalis (KMCA30), CS with total hysterectomy (KMCA33), CS unspecified (KMCA96)

Table 1: Definition of the exposure for mode of delivery. We used variables from the Medical Birth Register (MBR) and National Patient Register (NPR). Vaginal delivery was defined as any birth with no code for cesarean delivery. CS: Cesarean section. *Code "KMCA10BX" does not appear in the NPR.

Antibiotic treatment in the first two years

Our second exposure variable in papers I and II was antibiotic treatment in the first two years of life. As infants almost exclusively receive antibiotics as mixtures (save for intravenous antibiotics, but we do not have information regarding this type of administration) we could focus our attention to those 5 groups of antibiotics that are available in Denmark as mixtures (seen in sFigure C1 of the supplementary for paper I). In preliminary analysis we confirmed that 98.2% of all antibiotic prescriptions to infants under 2 years of age are within these 5 groups of antibiotic mixtures (ATC groups J01CE, J01CA, J01CR, J01FA, J01EA). Any topical antibiotics will appear in the group "other antibiotics".

As there are no agreed upon definitions of what constitutes a broad spectrum antibiotic, we wanted to group antibiotics to reflect the effect on the gut microbiota. At the time, no direct microbiota studies were available to us, so we used a surrogate marker for gut microbiota disturbance: diarrhea. Penicillin appeared to only rarely cause diarrhea (1.2%), whereas extended spectrum penicillin (8.1%), combination penicillin (19.8%), macrolides (>10%) and trimethoprim (approximately 6%) had greater effect (63–65).

We thus decided to divide antibiotic exposure into two categories: those children who had exclusively been exposed to *penicillin* (ATC group J01CE01), and those children who had been exposed to any other

antibiotic which we regarded to be a *broader spectrum antibiotic* as compared to penicillin (ATC groups J01CA, J01CR, J01FA, J01EA). In supplementary analyses we looked into the effect of each individual antibiotic, but we expected that each individual exposure to be rare and expected great loss of power for the sibling studies (visual representation in sFigure C1 of the supplementary for paper I). We also looked at antibiotic exposure in the first versus the second year of life.

Regarding conversion of ICD10 definitions to ICD8 classification

ICD10 diagnostic codes are not directly translatable to the older ICD8 codes, which presented a problem if we wanted to look at diagnoses before ICD10 was introduced in 1994. We discuss details of our considerations in this matter in the following paragraphs for those that are particularly interested, but provide a brief resume here: We concluded that we would only use ICD10 diagnostic codes to define our outcomes (ASD, ADHD, and affective disorders), whereas for confounder adjustment we used ICD8 and ICD10 codes. We relied on previous efforts for a conversion table, but sought to validate this table post-analysis. We also discovered that there were almost no parents with an ASD diagnosis and therefore chose to adjust for all parental psychiatric disorders instead, which we argue is just as valid as a confounder for the ASD and affective disorder studies. For further information, please read on.

First, we looked at how many children were diagnosed or treated for ASD or ADHD per year. Although ADHD is a term used by the Diagnostic and Statistical Manual of Mental Disorders (DSM) fourth and fifth edition used in the United States, our definition of ADHD diagnoses was based on the ICD10 criteria for hyperkinetic disorders, which are fairly similar (66). But as ICD10 was first introduced in 1994, that meant that if we wanted to look at the incidence before that point, we would have had to have been able to convert our ICD10 codes to ICD8, as ICD8 had been used since 1973. This proved no easy task, as the WHO had themselves experienced when trying to create official conversion tables back in the 1990's (67), which were later retracted. In 2014, Pedersen et al (68) suggested equivalent ICD8 codes for the most common ICD10 psychiatric diagnoses. When we reviewed the literature, we discovered that there was inconsistent use of conversion codes for any given psychiatric disorder (Table 3). Lacking official and uniform conversion tables, we decided to follow the suggestions of Pedersen et al, as one of the main authors of that research group, Aksel Bertelsen, had been heavily involved in the process of implementing ICD10 and an authority on the subject of translating ICD8 codes to ICD10.

As we were preparing our first paper, assessing offspring ADHD risk, we tried to validate our selection of ICD8 codes. As the offspring outcome of ADHD is based on ICD10 codes, this would only be relevant to confounder adjustment for parental ADHD-status, which could be based on both ICD8 and ICD10 codes. We were inspired by the study of Kessing from 1998 (69). There, he looked at patients admitted both in 1993

and 1994 (the latter year being when ICD10 was introduced in Denmark) with ICD10 code of affective disorder (DF3) and found that 70% had received the ICD8 code of 296.

First, we selected all persons with an ICD10 code of DF90 (ADHD) and any previous ICD8 code. We repeated the exercise for ICD10 code DF84 (ASD). We compared the resulting ICD8 codes to those suggested by Pedersen et al (68), WHO's unofficial conversion table (67) or the literature available at the time (references in Table 3). Presented below in Table 2 are those ICD8 codes, along with a few additional candidates of our own.

ICD8→	**299.00** Psychosis proto-infantilis	**299.01** Psychosis infantilis posterior	**299.02** Psychosis limitaris infantilis	**299.03** Psychosis infantilis non specificata	**299.05** Psychosis e causa dubia probabil. e conditio. non physicalis	**299.09** Psychosis	**310-315** Inferioritas, debilitas, imbecillitas, idiotia, et oligophrenia
ICD10↓							
DF90	0.1%	0.1%	1.3%	0.1%	0.4%	1.0%	2.4%
DF84	20.8%	2.1%	12.5%	2.7%	0.1%	0.9%	14.7%

ICD8→	**307.99** Reactiones maladaptivae transitoriae	**308.00** Neuroses infantiles	**308.01** Disordo personalitatis infantilis	**308.03** Reactio maladaptiva infantilis	**308.04** Reactio maladaptiva pubertatis	**308.05** Reactio maladaptiva alia	**308.06** Reactio maladaptiva
ICD10↓							
DF90	6.1%	2%	5.2%	6.2%	0.6%	0.3%	0.4%
DF84	1.8%	3.4%	4.5%	3.3%	0.2%	0.7%	0.1%

Table 2: ADHD and ASD conversion to ICD8 classification, based on persons with both an ICD10 diagnostic code of either DF90 (ADHD) or DF84 (ASD), and any ICD8 diagnostic code (here presented in Latin). ICD8 code 299 is usually associated with ASD diagnoses, where ICD8 code 308 is associated with ADHD. Presented are the percentages of the people with a specific ICD8 code as compared to number of people with any ICD8 code. Please note the table is split in two.

For ASD, 38.1% of all patients had an ICD8 code of 299.00-299.03 which seemed to fit nicely with all previously suggested conversions. It should be mentioned that many had received a diagnosis of a mental retardation (310-315), but this is a common unspecific comorbidity (70). Results were not as clear for ADHD, with no one previous suggestion covering particularly many ICD8 cases. However, when combining the suggestions to 308.01 and 308.03, and adding 307.99, 308.00, 308.04, 308.05 and 308.06, 21% of all

patients were covered by one of these ICD8 codes. Unfortunately, the codes were not specific to ADHD, as 14% of ASD cases also had one of these ICD8 codes. This relationship has been noted previously (71).

Put into context, this means that Pedersen et al's (68) suggestion of conversion from ADHD ICD10 diagnostic code of DF90 to ICD8 code 308.01 representing *infantile personality disorder*, could be expected to capture at least 5.2% of true cases before 1994. Meanwhile, the WHO conversion table suggestion of ICD8 code 308.3 would have captured 0% of cases. The use of the WHO conversion table is the only one we found to be used in the literature (besides that of papers by Pedersen et al).

ICD10	Description	ICD8 (WHO)	ICD8 (Pedersen)	ICD8 Literature
F30-F39	Affective disorders	296.09-296.99, 298.09, 298.19	296.*9 (excluding 296.89), 298.09, 298.19, 300.49, 301.19	296, 298, 300, 301.19 *Ref.:* (72–92)
F84	Pervasive developmental disorder (autism)	299.00-299.03	299.00-299.03	299.00-299.03, 299.10, 299.80 *Ref.:* (82,86,93–95)
F90	Attention Deficit Hyperactivity Disorder (ADHD)	308.3	308.01	308.3 *Ref.:* (72,74)

Table 3: Different conversions of our proposed main outcomes from ICD10 to ICD8, as according to the WHO's unofficial conversion list, the expert opinion of a group of researchers within the field of psychiatric register research and key persons involved in implementation of ICD10 (published in a paper by Pedersen et al in 2014), and how researchers using data from the Psychiatric Central Research Register convert ICD10 codes to ICD8 (Ref: References). Please note that the ICD10 codes listed here are not necessarily the specific codes we decided to use (refer to Table 7 for those).

However, there is no such ICD8 code listed in the ICD8 classification tables (96), although in ICD9 this code represented *acute reaction to stress,* which does not seem relevant to ADHD. This could however have been a "typo", as in ICD8 the code 308.03 represents *infantile maladaptive reaction*. Indeed, if this was the intended code, it would have captured 6.2% of cases. This realization, that the code could have been in error, did not dawn on us until after we had decided to go with Pedersen et al's suggestion of ICD8 code 308.01. In future endeavors, both 308.01 and 308.03 should perhaps be considered, along with 307.99 representing *transitory maladaptive reactions* and capturing additional 6.1% of ADHD cases, for a total of 12.3% (Table 2).

Meanwhile, when considering similar code conversions for ASD, we discovered that less than 0.06% of all parents of children born 1982 to 2010 had received an ASD diagnosis, compared to 1.1% of all children born in the same period (that would have had shorter time to be diagnosed than the parents). This might indicate that people with ASD were perhaps less likely to have offspring or at least that ASD had become more common after the parents had their offspring. This is especially likely when considering the rise in new cases per year over time, as seen in Figure 2 of paper II (58), and that out-patient diagnoses only first became available in 1995. In either case, it could have introduced more bias to adjust for parental ASD status than it would correct, as those parents with an ASD diagnosis and the oldest children born in the beginning of the period would presumably be a lot different from parents with an ASD code giving birth to the youngest children in the cohort. Instead, we chose to adjust for any parental psychiatric cases (ICD10: DF0-DF9 or ICD8: 290-315), when examining the outcome of offspring ASD risk. This was based on findings from previous studies (97–99) linking any parental psychiatric disorder to increased risk of autism in offspring, and that mothers with psychiatric disorders more often gave birth by cesarean delivery than mothers without a diagnosis (100). For the same reason and that parental mental health can increase risk of any psychiatric disorder in offspring (101), we also decided to adjust for any parental psychiatric disorder status in the affective disorder study.

Defining the population

Complete information regarding delivery mode became available in 1982. As the study cohorts were based on the same population for previous studies by my supervisors (102,103), the youngest children were born in 2010. We decided against updating the birth cohort with those born in 2011 and later, as we did not expect they would have added much information to the analyses due to too short follow-up for the outcomes of ADHD, ASD and affective disorders.

Using the definitions discussed above, the vast majority of cases for either ASD or ADHD happened after 1995. Between 1994 and 1995, the ICD10 classification was being introduced and the data considered to be less reliable as doctors were getting used to the new coding practice (37). In 1995 the records of the Register of Medicinal Product Statistics became fully digitally available. However, until 1997, prescriptions for an infant could sometimes be registered in the name of the parents. As we also discovered, 84% of ADHD cases were treated with ADHD medication and 21% of cases treated medically did not have a diagnostic code, making it very relevant that prescription information was available as we would otherwise have missed a considerable proportion of ADHD cases. As the validity of the information in the exposure variable “antibiotic treatment”, was also dependent on valid prescriptions, we decided for the ADHD and ASD studies to restrict our attention to those born in the years 1997-2010. However, in retrospect we could

see that there was very little difference in antibiotic prescriptions for children under 2 years of age in the years 1995 and 1996 compared to the following years (Figure 3). This might indicate that for antibiotic prescriptions, these were usually linked to the infant and not the mother even prior to the register becoming valid in 1997.

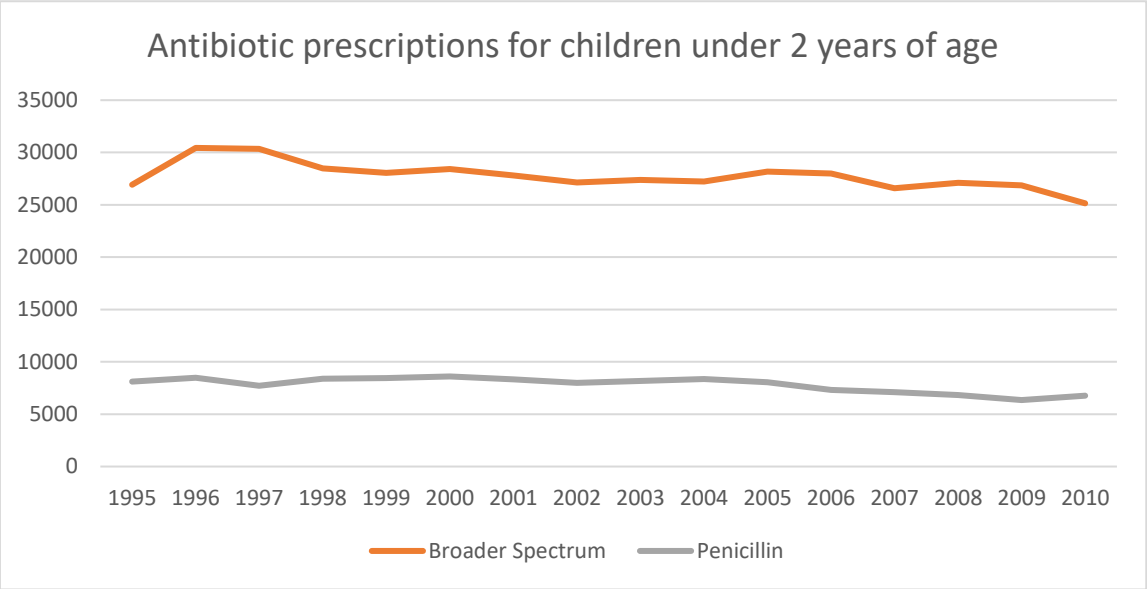


Figure 3: Number of antibiotic prescriptions per year, as either "Penicillin exclusively" or "Broader spectrum antibiotics" for all children born in Denmark from 1995-2010, during their first 2 years of life. Note that this population does not apply to any of the studies, because the study cohorts fulfill the inclusion criteria whereas here we apply no such limitations. There does not seem to be a drastic difference between the period 1995-1996 and 1997-2010, indicating perhaps that most antibiotic prescriptions were made out to the infants themselves and not the mothers, for both periods.

Affective disorders are not prevalent for ages under 15 years (study III, Figure 2)(60). This fact, coupled with ICD10 diagnostic codes becoming valid in 1995 and onwards and the Register of Medicinal Products Statistics becoming online the same year, as well as the oldest individual in our cohort being born in 1982, meant that we decided on starting our follow-up at 13 years of age to ensure a uniform outcome definition. However, to increase the validity of any cases based on prescriptions, we only included cases where medication was dispensed at 2 or more separate occasions for relevant drugs (Table 7). As prescriptions are usually made for one month at a time, it meant that we had to delay the start of follow-up to the 30th day past the 13th birthday, to allow for the possibility of an event of affective disorder.

Confounder adjustment

After careful consideration, the potential confounders selected for each outcome are listed in Tables 1 in each study (58–60) and Table 4 in this thesis. Not every previously identified potential confounder was available to us, such as paternal criminal history, severe marital discord, and placement in out-of-home care (104), although we did have information on marital status at each birth. Body mass index (or more specifically weight and height) at onset of pregnancy was only routinely recorded in the national registries since 2004 and was only adjusted for in sensitivity analyses.

Even though birthweight and gestational age have a tremendous predictive value for the future health of an infant, and often considered confounders in many studies, we opted not to adjust for these factors in our models. This is because birthweight and gestational age are affected by numerous other factors relating to the development of the fetus and medical decisions, such as mode of delivery. They might therefore act as colliders, and adjusting for them would potentially lead to increased bias in the models (103,105,106). However, we evaluated the effects of birthweight and gestational age in supplementary analyses for the outcome of ADHD but realized that the results of such sensitivity analyses would be inconclusive for the dramatically reduced sample sizes of children born preterm and small for gestational age (SGA). The sensitivity analyses were therefore not repeated in the ASD and affective disorder studies.

When considering the relationships between exposures, potential confounders and outcomes, our first impressions were that Apgar score at 5 minutes might not be a confounder, as it is measured after the exposure happens and therefore theoretically cannot affect the exposure. However, as Apgar score so soon after birth is actually an indicator of fetal stress before birth, it becomes a surrogate marker that can affect decisions for delivery mode. Similar arguments can be made for the variables asphyxia, CPAP and ventilator treatment, which were therefore all included as confounders.

The three studies did not use exactly the same set of confounder adjustment and we will explain the rationale behind this process in the following section.

Descriptive validity of confounder variables

When reviewing the variables we planned to use in our adjustment models, we discovered that there was a strange behavior for the Apgar score at 5 minutes (variable “v_apgar5” in the Danish Medical Birth Registry). For the years 1991 to 1996 there was a huge spike in incidence of Apgar scores under 7 (Figure 4). This coincided with an organizational restructuring of the DMBR in both 1991 and 1996. We therefore suspected that maybe the variable “v_apgar1” had been switched with “v_apgar5” for those years. But as we had not requested the “v_apgar1” variable to be included in our original dataset, we were unable to investigate this further until very recently, where this turned out not to be the explanation. Instead, we

suspect it may be due to a change in reporting to the MBR, where some missing values might have been reported as “0” instead of “99” as these years experienced a lot more 0’s and similarly fewer cases of 99’s, which would have artificially raised the number of low Apgar scores. If the latter scenario is indeed true, this will make it difficult to use the Apgar score as a confounder variable for any study using data from this period. For our three studies, this meant that we could only adjust for Apgar score as a confounder variable for the birth cohort born in 1997-2010 (ADHD and ASD studies) and not for the 1982-2010 cohort (affective disorders study). This problem with the Apgar variables will have to be looked into more closely by the custodians of the national registries.

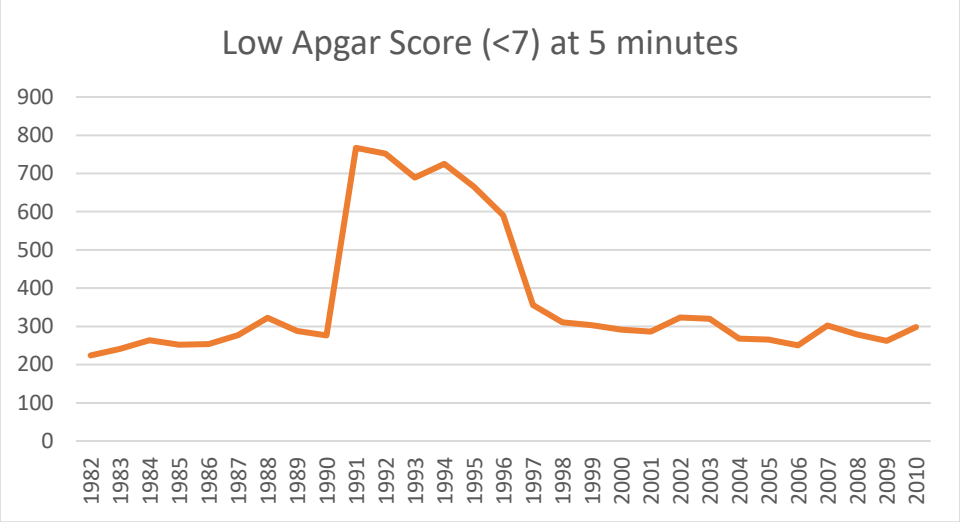


Figure 4: Number of infants born with a low Apgar score (<7) at 5 minutes postpartum, according to the Danish Medical Birth Register (DMBR). The sharp increase in cases in 1991 to 1996 coincides with restructuring of the DMBR and is probably in error.

Information on smoking during pregnancy only became available in 1991 and could therefore not be applied to confounder adjustment models in paper III. This variable sees a dramatic change in the period 1991 to 2010 (Figure 5), but it might be a natural phenomenon as there has been a growing public awareness regarding the risks of smoking during pregnancy, and similar trends can be seen in other countries (107). However as 10% of women reported smoking during pregnancy in 2010, but 20% of all Danish women were daily smokers (108), it raises the question whether pregnant women are reporting their smoking status. Nonetheless, we decided to include this variable in our adjustment models for the ADHD and ASD studies.

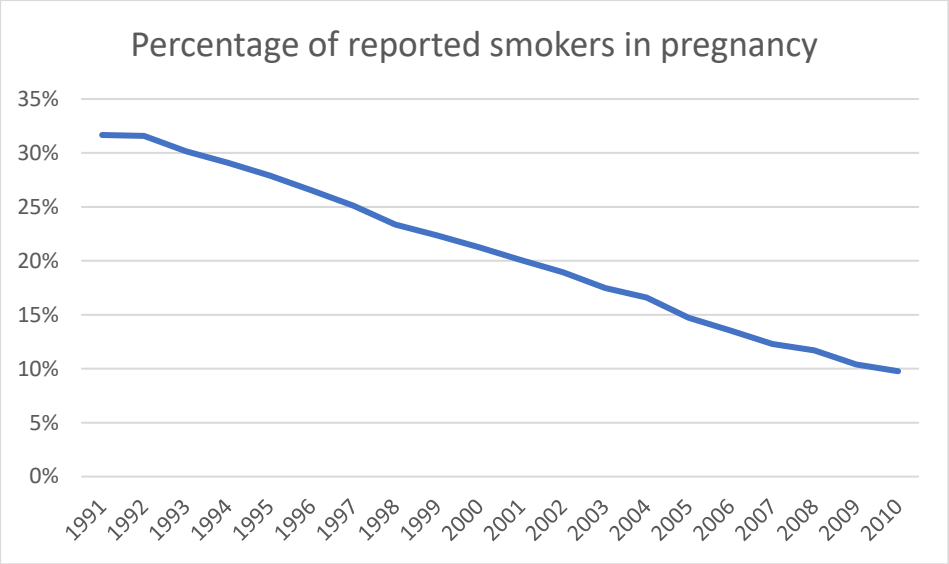


Figure 5: Percentage of all women with live-born singletons that reported smoking during pregnancy. Note that this period and population does not refer precisely to any of the study populations, as no inclusion criteria are applied.

We also discovered some issues with variables based on diagnostic codes. In 1994, Denmark went from using the 8th iteration of the ICD classification system to the 10th, and we saw some abrupt changes in prevalence of parental epilepsy, pre-eclampsia and gestational diabetes (Figure 6). In addition, the rate of these disorders had been changing drastically throughout the period, which we could not readily explain by natural forces. We suspected much of the change in prevalence was due to altered registration practices and diagnostic behavior. Therefore, we decided to leave these variables out of the adjustment models for the cohort in paper III, as it was in the period 1982-1997 that this change was often most noticeable. All the potential confounder variables as they pertain to each study can be seen in Table 4. The adjustment models themselves can be seen in papers I, II, III, Tables 1.

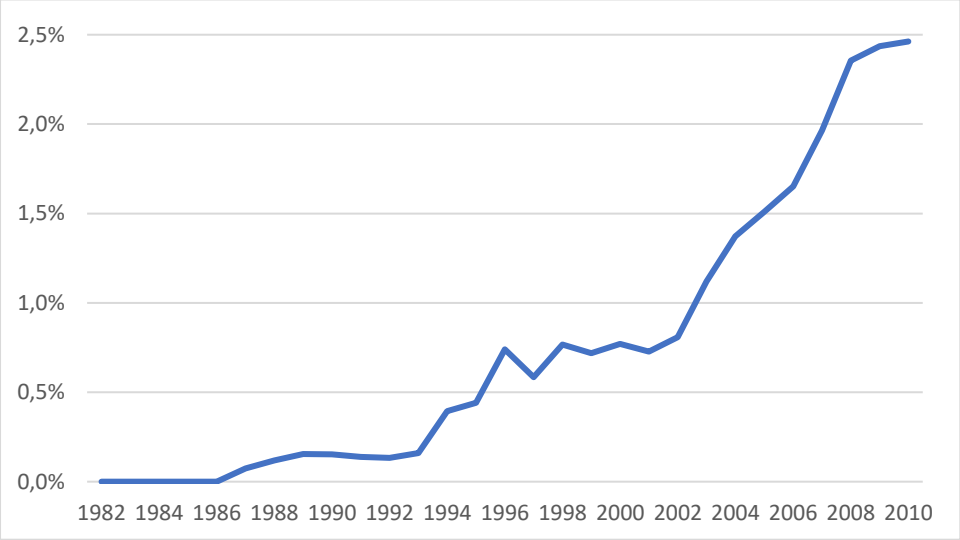


Figure 6: Percentage of women diagnosed with gestational diabetes (GDM). Before 1986 there were no registered cases of GDM, for which we have no explanation. In 2003 the DSOG published recommendations regarding screening protocols for GDM.

Finally, after we had already produced the results for our ADHD study, we realized that there might be issues with adjusting for the variables instrumental delivery, induction of labor and induction of contractions. First, we discovered that some births by pre-labor cesarean delivery were also delivered by instrumental delivery (165 children), had labors induced (2887 children), and had contractions induced (919 children), which does not make much clinical sense and is probably a coding/classification error. In on its own, this does not represent that big of a problem, as this applies to relatively few children out of the whole ADHD cohort.

Variables	ADHD	ASD	Affective Disorders
Childhood antibiotics use	*	*	
Mode of delivery	*	*	*
Maternal age at birth	*	*	*
Parental age difference	*	*	*
Parental education	*	*	*
Maternal marital status	*	*	*
Maternal smoking	*	*	
Infant sex	*	*	*
5-minute Apgar score	*	*	
Instrument use at delivery	*	(*)	
Use of CPAP or ventilator	*	*	
Asphyxia	*	*	
Parental epilepsy	*	*	
Preeclampsia or hypertension	*	*	
Gestational diabetes	*	*	
Parity	*	*	*
Induction of labour	*	(*)	
Induction of contractions	*	(*)	
Maternal antibiotics use during the pregnancy	*	*	
Maternal infections during the pregnancy	*	*	
Parental psychiatric history	*	*	*

Table 4: All confounder variables used in the adjustment models for our studies of ADHD, ASD and affective disorders. An asterisk marks if the variable was used for each given study outcome. Parentheses indicate that the variable was used in sensitivity analyses.

Secondly, as these variables strongly correlate with vaginal delivery, their effects would be difficult to separate from our effect of interest due to collinearity. Therefore, we decided to remove these variables from the main confounder adjustment model in the ASD study and include them in a sensitivity analysis instead. When reviewing the results of the ADHD study post hoc, we saw that including these variables in the adjustment models did not appear to affect the results as the effect-size of these variables were almost

1 and there were not that many cases. Therefore, we believe the conclusions of the ADHD study are robust regardless of having included these variables in the adjustment models.

Regarding the selection of ICD10 outcomes

Our primary outcomes were as follows: ADHD, ASD, and affective disorders. How we arrived at the precise ICD10 codes used to select cases representative of those outcomes, requires some explanation.

First, we discussed if we should only include primary diagnoses, or also include cases with a secondary diagnosis. Traditionally, it is assumed that primary diagnoses are more accurate and convey the certainty of the diagnosis, however if we assume that medical treatment for ADHD is a measure of severity (109), we can consider the children with a prescription for ADHD a golden standard and compare how many of those have either a primary or secondary diagnostic code or both. We also looked at the risk of parents having children treated with ADHD medication, for both primary and secondary parental ADHD diagnostic code. The results can be seen in Table 5.

	ADHD A	ADHD B	ADHD A AND B
CHILD ADHD MEDICATION	77%	66%	89%

Table 5: Proportion of offspring born 1982-2010 treated with ADHD medication, stratified by diagnostic code diagnostic code being either exclusively primary (A), exclusively secondary (B) or both (A and B) for the child. Note that the population does not directly refer to any of the study populations, as no inclusion criteria have been enforced.

	FATHER ADHD A	FATHER ADHD B	MOTHER ADHD A	MOTHER ADHD B
CHILD ADHD MEDICATION	7.6 (95% CI 7.0-8.3)	6.5 (95% CI 5.5-7.6)	11.7 (95% CI 10.8-12.6)	10.5 (95% CI 9.3-11.84)

Table 6: Odds ratios (OR) with 95% confidence intervals (CI) for parents having a child born in 1982-2010 treated with ADHD medication, stratified by parents' ADHD diagnoses being either exclusively primary (A) or exclusively secondary (B), with parents without an ADHD diagnosis as a reference. Parents with both a primary and secondary diagnoses are not included in any of the variables. Note that the population does not directly refer to any of the study populations, as no inclusion criteria have been enforced.

It could be argued that the relatively high number of children receiving ADHD medication, irrespective of their diagnostic code being either primary, secondary or both, is evidence for the secondary diagnoses also being indicative of more severe ADHD symptoms. As heritability is a strong risk factor for ADHD, the fact that there is only minimal difference between each parent's odds of having a child treated with ADHD medication, depending on primary or secondary ADHD diagnoses, supports the notion that these are

similar phenomena. However, it could also be argued that secondary diagnoses are usually less severe psychiatric disorders, as compared to whatever may be the primary diagnosis. So, it may simply serve as a proxy for any other severe psychiatric disorder, which may infer increased risk for offspring being treated with ADHD medication, as discussed above.

In either case, we opted for including both primary and secondary diagnoses for outcomes and confounders but looked more closely at these diagnoses in sensitivity analysis for each outcome. The results of these sensitivity analyses are discussed in supplementary chapters of our three papers (58–60), but there were no changes to our conclusions after excluding secondary diagnoses. This was somewhat expected as cases based on secondary diagnoses only accounted for 10% of all cases.

Similarly, we looked at the comparability of ADD cases to ADHD, assuming ADHD medication to reflect severity and similarity of the two related disorders. Of the 66,582 cases in a cohort of people born 1982-2010 (not directly comparable to any of the study populations as no inclusion criteria have been applied), 29,949 (45%) received both an ADHD (DF90) diagnosis and medication. There were 7017 persons with an ADD (DF988) diagnosis and 3870 (55%) received medication. However, 1762 of those with an ADD diagnosis also had an ADHD diagnosis. When taken together as a group, cases of ADHD/ADD received a prescription in 47% of cases. This indicated that ADD diagnoses were at least as “severe” as ADHD cases, in regard to treatment with ADHD medication.

We were also worried that if we did not include ADD cases, we would be introducing bias through the patient’s sex, as inattentiveness seems to dominate the clinical picture for females compared to hyperactivity in males (110). Therefore, we looked at differences between the male and female sexes. For ADHD (DF90), males were 2.3 times more likely to be diagnosed, whereas for ADD (DF988) this was somewhat less pronounced with males still being more likely to receive a diagnosis, at a risk ratio (RR) of 1.6. However, males were also a lot more likely to be treated with ADHD medication, RR 1.9. This indicated that there were disproportionately more females diagnosed with ADD than ADHD, assuming ADD is a different clinical manifestation of ADHD.

What we ended up on deciding as our main outcomes, was heavily influenced by the above discoveries. Besides the experts in our own group, we also consulted with pediatric psychiatrists with register research experience and asked for their advice. For example, we were told that during the clinical process of testing for ASD, the diagnostic code DF84.9 is mostly used by clinicians as a tentative diagnosis. We were therefore recommended to leave it out of our outcome definition, as we did, even though this meant a reduction in cases by 18%.

Outcome	ICD10 codes	ATC codes
Affective disorders	DF30-DF33 and DF38.00	N05AN1, N06A
ASD	DF84.0, DF84.1, DF84.5 and DF84.8	
ADHD/ADD	DF90 and DF988	N06BA02, N06BA04, B06BA09, N06BA12

Table 7: ICD10 and ATC codes used for outcomes reported in all three papers. We used both primary and secondary diagnostic codes.

Statistical analyses

As we have previously mentioned, we used two sibling comparison models in addition to a more traditional standard (non-stratified) Cox model that we have discussed in detail in papers I and II (58,59) and in the supplementary material. We chose survival analysis models as there is right censoring in the dataset. The thesis will only bring a brief description of the differences between the statistical models and leave out statistical formulas.

In a *non-stratified* Cox proportional hazards model (or *standard/descriptive* Cox model as we have referred to it in our papers) it is possible to estimate the association between a subject's hazard and any given exposure. Its pervasiveness in epidemiological research can hardly be overstated, as it explores categorical and quantitative variables alike and can estimate the effect of several risk factors on disease hazard simultaneously (111). Usually, researchers are interested in the direct (causal) effect of an exposure on risk of an outcome and try to adjust for any potential confounders that may affect both the exposure and the outcome. However, this situation is often not ideal, as many relevant confounders are either not available in whatever dataset you are working with or they are un-measured or un-measurable. An example of very relevant (partially) unmeasured confounding could be parental psychiatric status, as many parents with symptoms of any psychiatric disorders had probably not been diagnosed (Figure 5). Other relevant factors, such as upbringing are also difficult to measure and categorize.

Sibling models have been used in many research endeavors before, mostly case-control studies (112). That is, until more recently when national register data have become available. However, even with the large datasets available to us now that are generated from decades' worth of data from national cohorts, when studying associations of fairly rare disorders with uncommon exposures, you will need comparatively strong effects for the results to be significant if studied in traditional sibling models (56). Or phrased differently: the results of such studies might not be conclusive. This is why it was potentially rewarding to explore new avenues of statistical modelling.

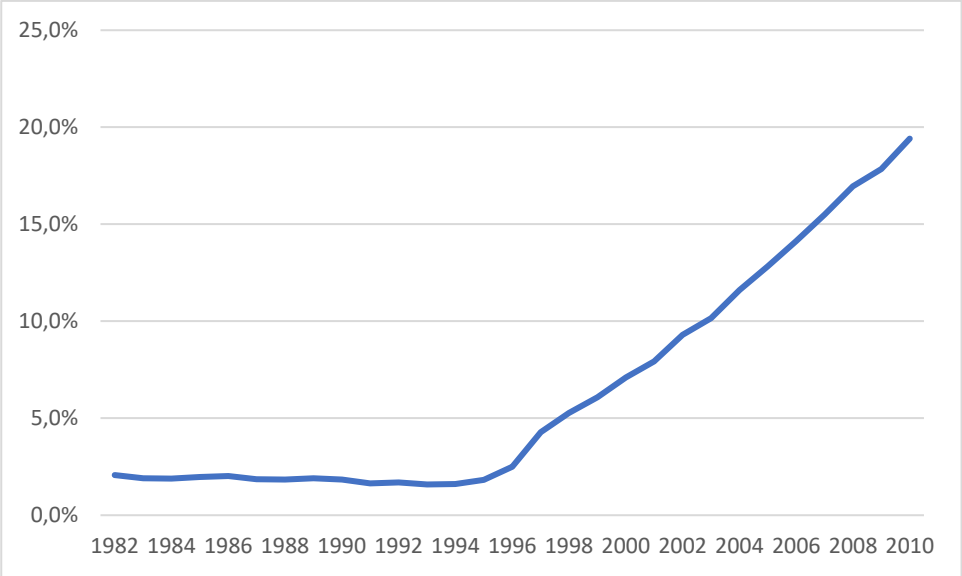


Figure 5: Percentage of mothers with any registered psychiatric diagnoses or medical treatment for psychiatric illness at the time of birth of singleton children in the years 1982-2010. There is a break in the data when out-patient contacts and prescription information become available in 1995, but that does not fully explain the drastic increase in prevalence after 1997. Note that the population does not directly apply to any of the study population, as no inclusion criteria have been enforced.

Every statistical model is dependent on different assumptions. One example could be “normality” - meaning that distributions are bell-shaped. Another assumption would be that observations are independent of each other. The more assumptions you make, the more restrictive the model becomes but it may also deliver certain advantages such as more precise estimates. This is the main difference between the stratified Cox model and the between-within model, the degree of assumptions. The former model assumes that the baseline-hazard can vary freely and is shared between the siblings – meaning that each family has their own baseline-hazard, but no other assumptions are made regarding this hazard. However, the latter model assumes that the variance in the family base-line hazard can be described with the gamma-distribution. Prior to our investigation, only a single simulation study had shown that the between-within model had potential for increased power compared to stratified Cox regression (57).

However, there might be intrinsic limitations to the sibling models, such as not every child having a sibling (Table 8). This means that the results may not be generalizable to single-child families (which we will discuss later) and there will be more individuals available for analysis in the standard Cox model. However, it is important to keep in mind that as the standard Cox model and the sibling models are estimating different effects, they should not be directly compared to one another. The sibling models are estimating effects within each family, whereas the standard Cox models estimate effects across individuals ignoring their relationship.

	No siblings	1 sibling	2 siblings	3 siblings	4+ siblings	Total
Not ASD	124,801	136,729	28,489	2317	205	292,541
ASD	1790	1986	440	37	6	4259
Total	126,591	138,715	28,929	2354	211	296,800

Table 8: Prevalence of ASD for first-born children, stratified by number of younger siblings. Notice that 1790 out of 8267 children (22%) with ASD in total have no siblings and therefore do not contribute to estimates in the sibling models.

Theoretically, the standard model and the between-within model use all observations for their estimation. But it does not mean that every observation is of equal value for the estimation or that these models are vastly more efficient than the stratified model. We refer to the reported confidence intervals as measures of the precision obtained in each of the three models.

Results

We saw the same increase in the rate of cesarean deliveries as has been reported previously (102,103). In the period between 1982 to 2010 in the affective disorder study cohort, there was indeed an 82% increase in cesarean sections, mostly due to a sharp increase in pre-labor cesarean deliveries after 2000 which levelled off in 2004 (103). If we consider the whole period, 87% were born by vaginal delivery, 7% by intrapartum cesarean delivery, and 5% by pre-labor cesarean delivery. However, this changes to 82%, 8%, and 9% respectively when only considering the birth-cohort 1997-2010 used in the ADHD and ASD studies.

For the three studies this thesis is based on, we used definitions for antibiotic groupings not reported previously for this period, albeit antibiotic consumption in the first two years of life has been published previously (102). We did not find any marked change in antibiotic prescriptions for the periods we studied (children born in 1997-2010), as seen in Figure 3. Of all the children in the cohort studied for antibiotic exposure, 28% received no antibiotics in their first two years of life, 16% received penicillin exclusively, and 56% received some form of broader spectrum antibiotics.

The number of new cases increased for all outcomes in the period studied (from 1995 for affective disorders and from 1990 for ADHD and ASD, until 2015 for all outcomes) as evident from Figures 2 in all papers. The pattern was not identical for the three outcomes, with the increase happening earlier for affective disorders and ADHD and then decreasing again, whereas the increase in ASD cases happened only in the last few years of follow-up and showed no sign of decreasing again.

Demographics can be seen in supplementary tables S3A and S3B for Paper I, eTable 3A and 3B for Paper II, and Table 2 for Paper III. The previously identified possible confounders all seemed to be unevenly distributed among children born by each delivery mode, as well as among children either not receiving any antibiotics in their first two years or receiving antibiotic treatment (58–60). However, when comparing the

non-adjusted and adjusted models, it seemed as if adjusting for offspring sex had the greatest influence on the exposure effect, with social variables being second most influential.

The standard Cox model

In all the standard (which we sometimes denoted *descriptive*) Cox models we found significantly increased risks of ADHD, ASD and affective disorders for all exposures. The effects were markedly smaller for antibiotic treatment and outcome of ADHD when applying more confounder adjustment. Conversely, the effects slightly increased when applying confounder adjustment for intrapartum cesarean delivery and affective disorders. All exposure effects remained significant regardless of the level of confounder adjustment. The exact effect estimates of all exposures from the fully adjusted model are shown in Table 9.

The sibling-stratified Cox model

For the outcome of ADHD, there was significantly increased risk for children born by intrapartum cesarean delivery when compared to vaginal delivery in the fully adjusted sibling stratified Cox model (Table 9). For the outcome of ASD there was also significantly increased risk for children treated by broader spectrum antibiotics compared with antibiotic treatment (Table 9). For all other exposures in the ADHD, ASD and affective disorders study, the effect estimates were non-significant and mostly small (Table 9).

The between-within sibling model

For the outcome of ADHD and ASD the between-within sibling model there were non-significant effects close to null when fully adjusted. For the outcome of affective disorders, there was slightly increased risk for children born by intrapartum cesarean delivery, but only when fully adjusted (Table 9). Generally, the estimates were closer to null than in the stratified sibling model, with smaller confidence intervals.

Robustness of the results

Below are the results of all sensitivity analyses we performed for the three outcomes. Some analyses were specific to an outcome, such as those for antibiotic exposures, reproduction stoppage and the dropout analysis.

Generally, we found nothing in the sensitivity analyses that changed our conclusions, with the exception of those for reproduction stoppage where the generalizability of the results to families without siblings or where the family size varied might not apply. We discuss this further in the section below.

Finally, for the ADHD outcome we performed a separate analysis where the infant's sex was added to the second adjustment model on its own. This resulted in estimates very close to that of the fully adjusted model and we therefore suspected that the rest of the variables in the third and fourth adjustment models (variables related to obstetric factors and maternal microbiota) did not have great confounding effects on the relationship between delivery mode, antibiotic exposure and ADHD.

Model	Exposure	ADHD		ASD		Affective Disorders	
		Estimate	95% CI	Estimate	95% CI	Estimate	95%CI
Standard Cox	Intrapartum CS	1.10	(1.04 – 1.16)	1.10	(1.02 – 1.19)	1.07	(1.05 - 1.10)
Standard Cox	Pre-labor CS	1.11	(1.05 – 1.17)	1.11	(1.03 – 1.20)	1.11	(1.08 – 1.15)
Standard Cox	Penicillin	1.13	(1.08 – 1.19)	1.11	(1.04 – 1.19)		
Standard Cox	Broader spectrum	1.23	(1.19 – 1.28)	1.10	(1.04 – 1.16)		
Stratified Cox	Intrapartum CS	1.21	(1.01 – 1.45)	1.11	(0.86 – 1.43)	1.06	(0.99 – 1.13)
Stratified Cox	Pre-labor CS	1.14	(0.96 – 1.35)	1.07	(0.84 – 1.34)	0.98	(0.91 – 1.06)
Stratified Cox	Penicillin	0.99	(0.88 – 1.12)	1.09	(0.91 – 1.29)		
Stratified Cox	Broader spectrum	1.02	(0.92 – 1.13)	1.16	(1.01 – 1.36)		
Between-within	Intrapartum CS	1.09	(0.97 – 1.24)	1.06	(0.89 – 1.26)	1.05	(1.003 – 1.12)
Between-within	Pre-labor CS	1.03	(0.91 – 1.16)	0.97	(0.83 – 1.15)	1.00	(0.94 – 1.07)
Between-within	Penicillin	0.98	(0.90 – 1.07)	1.05	(0.93 – 1.18)		
Between-within	Broader spectrum	0.99	(0.92 – 1.06)	1.05	(0.95 – 1.16)		

Table 9: Exposure effect hazard ratio estimates with pointwise 95% confidence intervals (CI) for ADHD, ASD and affective disorders, in the fully adjusted models for standard Cox, sibling stratified Cox, and the Between-within sibling model.

Interaction effect between delivery mode and antibiotic treatment

There were no significant interaction effects between cesarean delivery and antibiotics for either ADHD or ASD, as tested with a likelihood ratio test with p=0.58 and 0.94 respectively. Interaction effects in the ADHD study are shown in Figure 6 and for the ASD outcome in eFigure 2 of the published supplementary material (58). As we only looked at a single exposure in the affective disorder study, no interaction effects were tested.

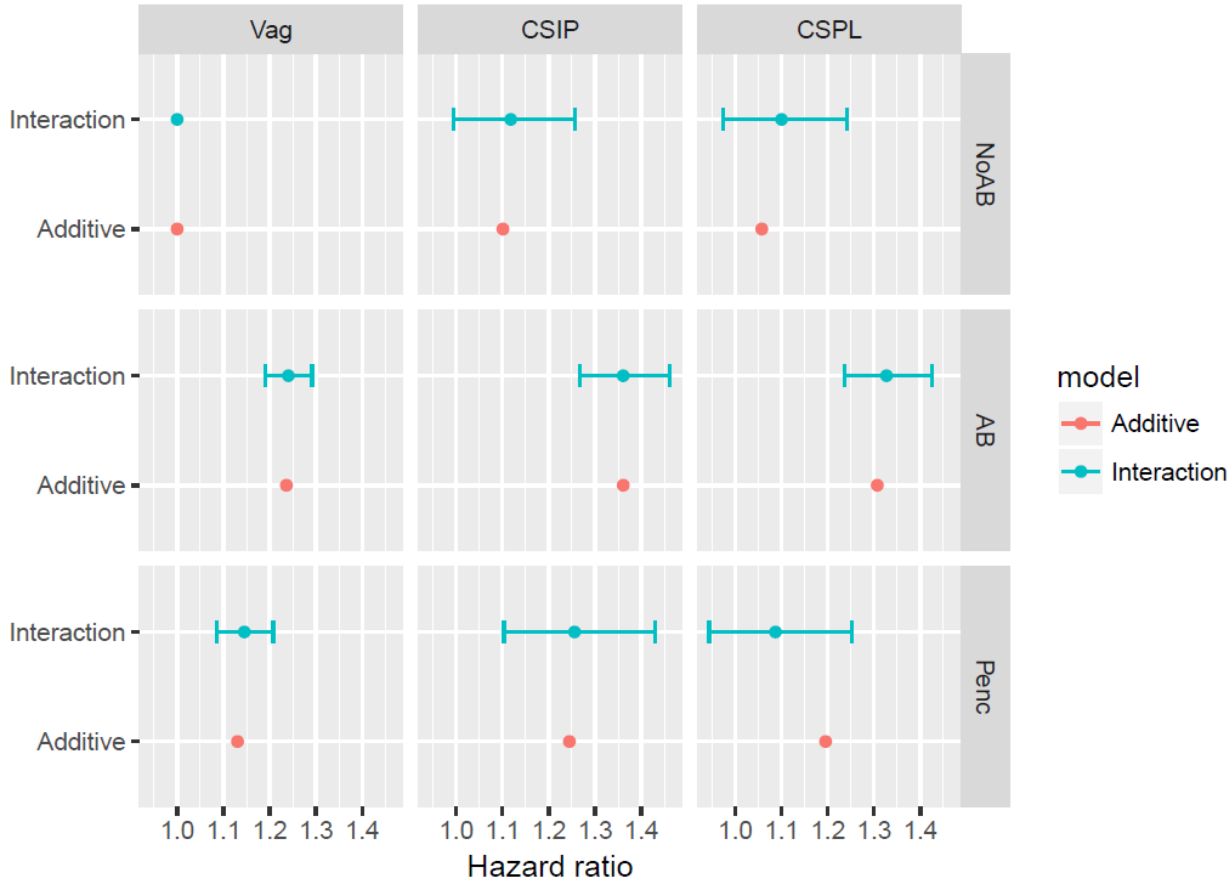


Figure 6: Interaction effects for the ADHD outcome, between mode of delivery and antibiotic treatment. Compared are the Hazard Ratios for additive effects and interaction effects, with only minor differences in point estimates. Please note that we have not calculated the confidence intervals for the additive effects as we did not deem it necessary for this visual representation. Vag: Vaginal delivery, CSIP: Intrapartum cesarean delivery, CSPL: Pre-labor cesarean delivery, NoAB: No antibiotic treatment, AB: Broader spectrum antibiotics, Penc: Exclusively penicillin treatment.

Antibiotic treatment divided into subgroups

For the ADHD outcome we sub-divided the antibiotic exposure into six groups: the five most commonly used antibiotics in infancy available as mixtures, and other antibiotics. These were Penicillin (ATC group J01CE), Combination penicillin (J01CR), Macrolides (J01FA), Trimethoprim (J01EA), Extended spectrum penicillin (J01CA), and other antibiotics. The results of the between-within model can be seen in sFigure C2 of the ADHD supplementary (59), but there were no significant effects of any of the subgroups of antibiotics.

Antibiotic treatment divided into 1st or 2nd year exposure

For both the ADHD and ASD outcome we sub-divided the antibiotic exposure period, depending on the first treatment happening in the infant's first or second year of life. To simplify these sensitivity analyses we examined the effects of any antibiotic on the outcomes, instead of both exclusive penicillin treatment and broader antibiotic treatment. When analyzed in the same model, we still saw insignificant effects of antibiotic treatment but any antibiotics in the first year now had slightly protective effects and antibiotic treatment in the second year had slightly increased risk in the ADHD study. However, when analyzed on its own, any antibiotic treatment in the first year had slight significantly protective effects in the ADHD study as seen in sFigure D3 of the ADHD online supplementary (59). In the ASD study, there were insignificant effects of any antibiotic treatment regardless of the exposure happening in the first or second year of life, as seen in eFigure 4 of the online supplementary (58).

Time effects

We had to investigate time effects carefully as there are clear trends of cohort and calendar time effects as can be seen in the Lexis-diagrams in all three Figures 2 of the published papers (58–60). Therefore, we examined whether adjusting for birth year or event year effects would change our conclusions from the primary analyses (supplementaries for papers I, II, and III (58–60)). Including event year in the sibling models was not possible due to current computational limitations of the servers at Statistics Denmark. However, comparing results from a resampling method and an adjustment model in the standard Cox model, we found that *only* adjusting for birth year effects would be an acceptable compromise when investigating time effects in the sibling models. Thus, we proceeded to include birth year effects in the sibling models which resulted in only minor differences. Additionally, we could conclude that it was necessary to adjust for at least birth year in the standard Cox model if trying to describe the association between mode of delivery and ASD, but no further adjustment for calendar time was needed for ADHD. For affective disorders, calendar time effects did not seem to confound the relationship with mode of delivery.

The sibling models seemed to control for a variable that was causally related to either exposure and calendar time or calendar time and outcome, thereby blocking the unknown confounder. This might be due

to the time between birthdates of a sibling pair is usually relatively short, thereby negating some of the effect of calendar time. It might also be that the parents most at risk of having a cesarean delivery, requesting antibiotics to treat their offspring, and having a disposition for psychiatric disorders might also be the parents that can explain some of the trends we see for these variables. Until we understand the interplay between calendar time and psychiatric disorder it will be necessary to consider these effects in future studies.

Reproduction stoppage in ASD-families

We wanted to test if there was any evidence of a reproduction stoppage in the Danish population if the first born child was later diagnosed with ASD, as had previously been found in the American population (113). This turned out to rather complex. First, we performed a descriptive analysis of the probability to have one or more younger siblings if you are the first-born child, stratified by ASD-diagnosis (Figure 7).

We speculated if the apparent tendency of a reproduction stoppage in Denmark was due to insufficient follow-up and differences in age distributions. It could also be that as symptoms of ASD may not be apparent until later in life and children are usually born within 3 years of each other, that the age at diagnosis played a role for the probability of siblings. We therefore plotted the age at diagnosis for first-born children diagnosed with ASD, stratified by birth year and sibling status (Figure 8), without finding any evidence to support this hypothesis.

Finally, we decided to look at the effect estimates for first born children compared to those who had older siblings. However, this could obviously not be tested in a sibling model, so we tried fitting the fully adjusted standard Cox model on two subsets of the dataset:

1. Only parity 1-children (n = 296,800)
2. Only children with parity greater than 1 (n = 374,806)

We found that there were some differences in exposure effects between the two groups of children, with less of an effect of antibiotic treatment and a larger effect of intrapartum cesarean delivery (eFigure 9 in the online supplementary Paper II).

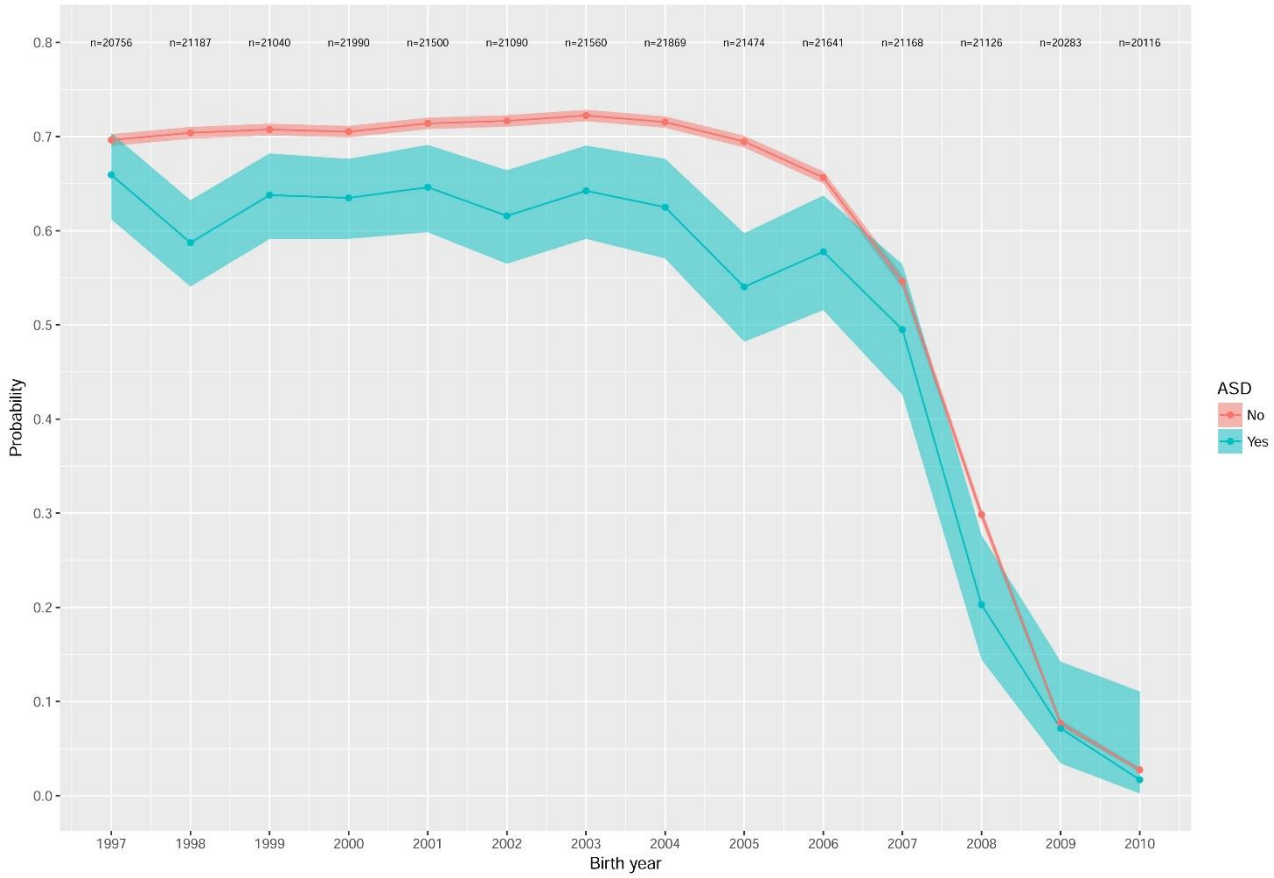


Figure 7: Probability of having a younger sibling, dependent on the first born child's ASD-status and birth year. We fitted separate logistic regression models for each cohort with ASD as the only covariate and calculated pointwise Wald confidence intervals on the log-odds-ratio-scale transformed back to a probability scale. There is a systematic tendency for first-born children with ASD to have less chance of having younger siblings than those first-born children never diagnosed with ASD.

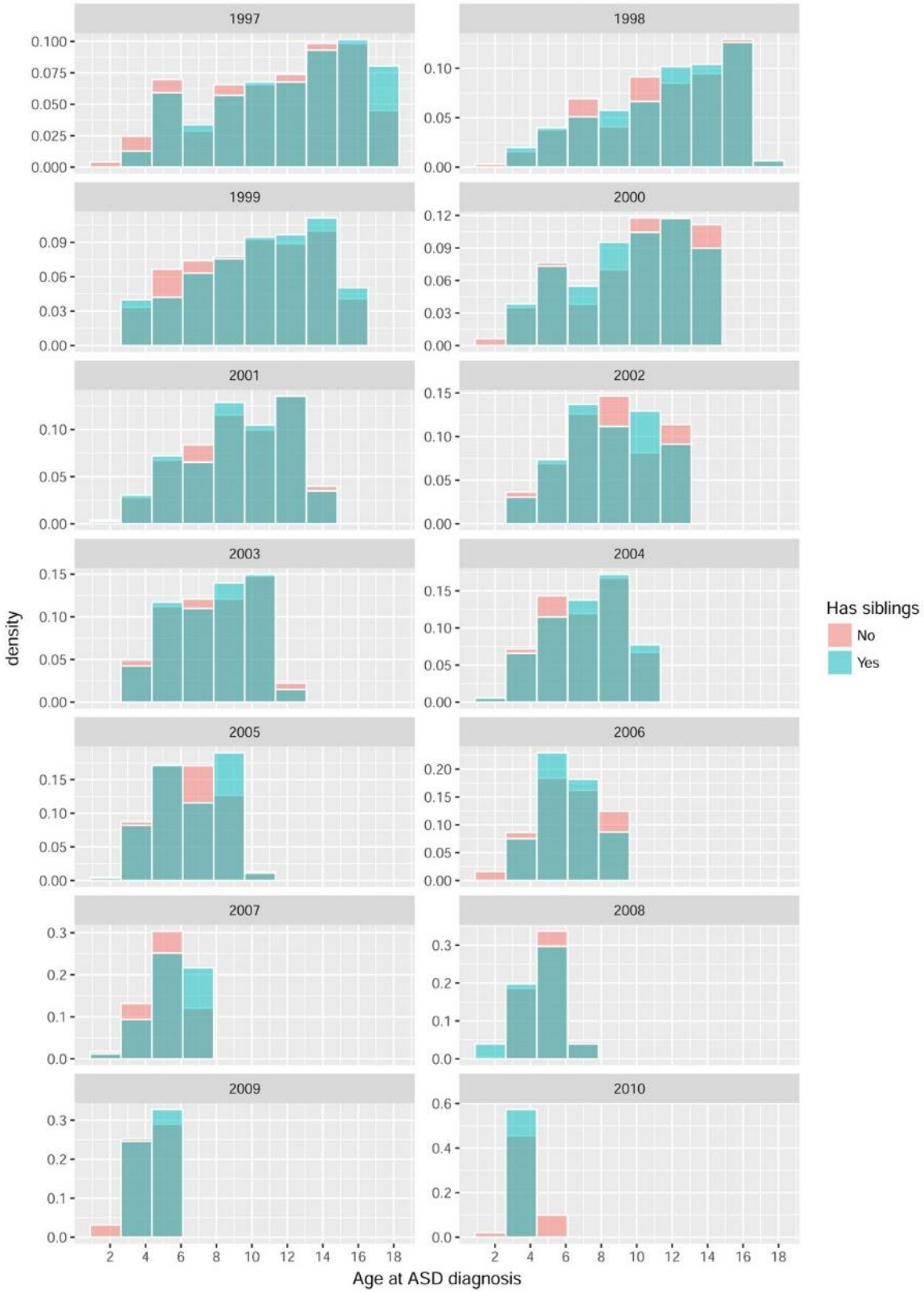


Figure 8: Age at diagnosis for first born children diagnosed with ASD, stratified by birth year and sibling-status. Each birth-year is represented by a column, and the two groups with or with-out siblings layered upon each other. There is no apparent tendency to only have siblings if the diagnosis of ASD occurs later than at 4 years of age.

Dropout analysis for ASD outcome

We used complete case analysis for our main outcomes, which means that we excluded any births where we had missing information for one or more variables. For the ASD study, this proportion was 5.1% of the original population. Therefore, we were asked by a reviewer of paper II to investigate whether there were systematic relationships between any of the completely observed covariates and the risk of having missing information. Almost all births with missing information were excluded due to five variables: Maternal smoking (3.3% of study population prior to exclusion due to missing information), Apgar score (1%), Paternal education (0.6%), Parity (0.6%), and Maternal education (0.3%). Some births had missing information for more than one variable, but none for more than five variables. For the affective disorders study the proportion of cases excluded due to missing information was considerably lower (1.7%) and therefore we did not investigate this cohort further. For the ASD study, we found the most increased odds of having missing information for children exposed to CPAP (OR 1.8 [95% CI 1.7-1.9]) or ventilator use (OR 3.8 [95% CI 3.2-4.5]) or maternal use of penicillin in the third trimester (OR 1.9 [95% CI 1.8-2.0]), but there were also moderately increased odds for mothers with gestational diabetes (OR 1.4 [95% CI 1.3-1.3]), cesarean delivery (OR 1.35 [95% CI 1.3-1.4]), or maternal age under 25 years as compared to maternal age 25-30 years (OR 1.4 [95% CI 1.3-1.4]). Therefore, children where these factors are relevant may be underrepresented in our study population and the results should be interpreted with this in mind. As the ASD and ADHD study cohorts are very similar, we assume these results apply to the ADHD study as well.

Birth weight and gestational age for ADHD outcome

Despite valid arguments against using birth weight and gestational age in our adjustment models for these outcomes, we wanted to increase the comparability of our results to previous studies that have included these variables in their analyses. Therefore, we performed sensitivity analyses where we looked at exposure effect estimates in the standard Cox model in subpopulations of children born either to term or preterm, small for gestational age or children with a normal to high birth weight. It is only for the highly imprecise estimate in the preterm subgroup (GA small) that we see qualitatively different estimation results. It is however not clear whether the results seen in Figure 9 are of any substantial nature as they could come about as a result of the dramatically reduced sample size or the potentially degenerated nature of the models due to collider-conditioning. Thus, we refrained from further explorations into effects of birth weight and gestational age on risk of ASD and affective disorders.

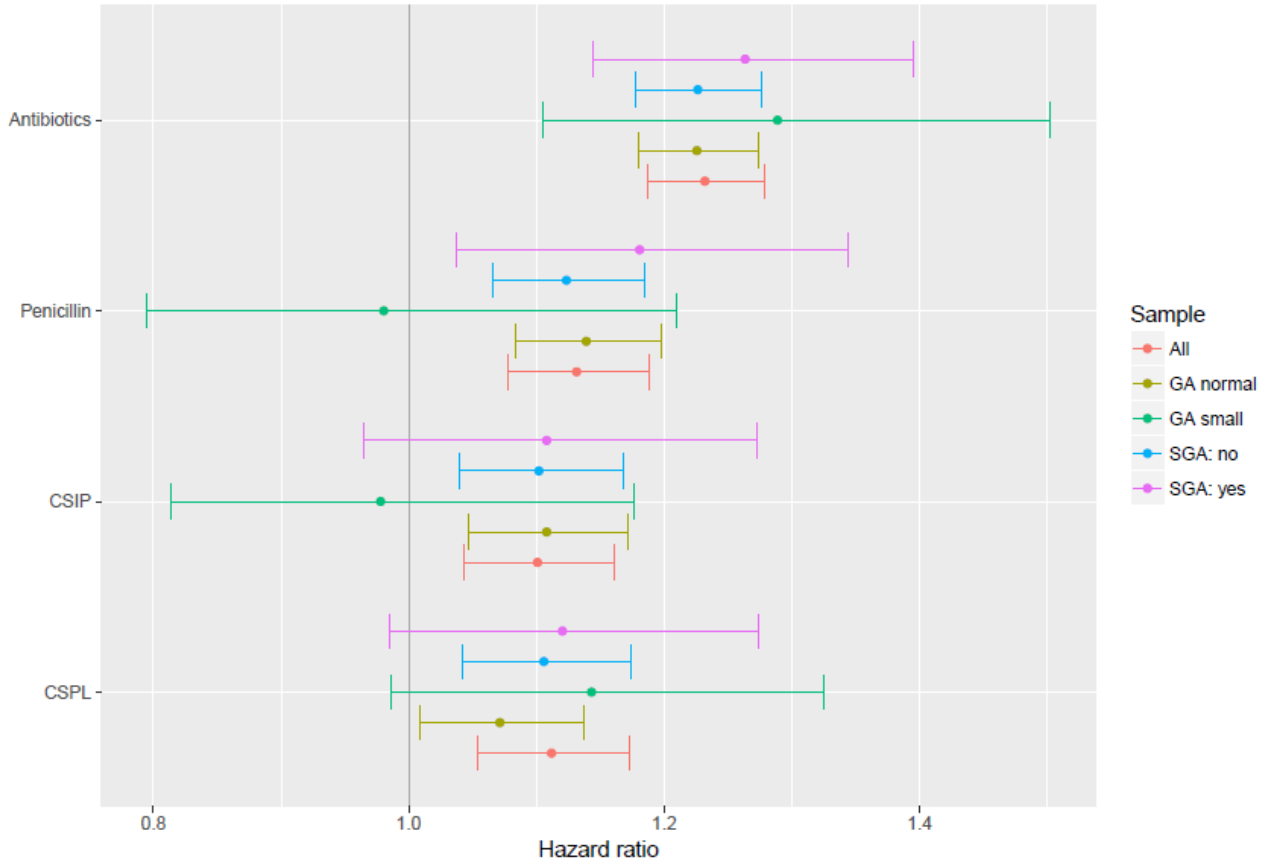


Figure 9: Exposure effect estimates from the standard Cox model for the ADHD study, where the population is divided into subpopulations of children born to term (GA normal), children born preterm (GA small), children with normal or high birthweight (SGA: no), and children that are small for gestational age (SGA: yes) as compared to the full population (All).

Definition of outcome

In the ADHD study we investigated the robustness of the between-within model results in regard to how the outcome variable was defined. In the main model we have used the first event of either a diagnostic code for ADHD (DF90) or ADD (DF988) or a prescription for an ADHD medication. Both primary (type A) and secondary (type B) diagnoses are included. In this way, we found 17,971 children with ADHD. For the sensitivity analyses we tried only looking at children that receive a primary diagnosis of ADHD, n=17,620. We then excluded the ADD diagnosis from the original outcome definition, n = 16,875. We also looked at children where both a diagnosis of either ADHD or ADD along with a prescription for an ADHD medication was required, n=10,481. We also looked at only cases that were diagnosed or treated before or at their 12th birthday, to account for DSM criteria for ADHD diagnosis, n= 14,802. Although we see some differences in point estimates, especially for the model requiring both a diagnosis and a prescription, the overall conclusions remain the same no matter which of the five ADHD outcome definitions is used (Figure 10).

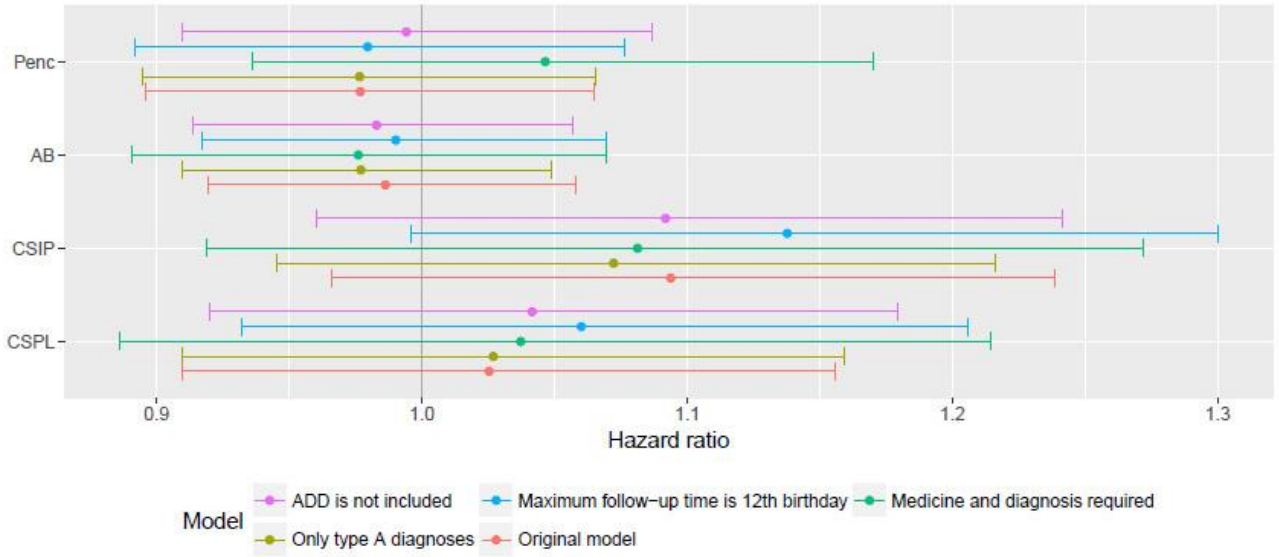


Figure 10: Exposure effect estimates of the fully adjusted between-within model, using five different variations of the ADHD outcome definition. Penc: Penicillin only, AB: broader spectrum antibiotics, CSIP: Intrapartum Cesarean Delivery, CSPL: Pre-labor Cesarean Delivery.

In the ASD study we also looked at the sensitivity of the results to changing the definition of an ASD case. This included two variations to the definition; first, we only included cases with a primary (type A) diagnosis, second we only included diagnoses of an infantile autism (DF840). In the original definition with both primary (A) and secondary (B) diagnoses and the entire autism spectrum diagnoses (DF84.0, DF84.1, DF84.5 and DF84.8), there were 8,267 included children. However, when being more restrictive with our definition of an outcome, there were 7,656 children with a primary diagnoses of ASD, and only 3,500 children with a diagnosis of infantile autism. This affected the confidence intervals somewhat (Figure 11). In short, the conclusion remained that there was no effect of delivery mode or antibiotics on the outcome.

We finally looked at the robustness of the results towards how the outcome variable for affective disorders was defined. In the main model we looked at both cases attained by information from prescriptions and diagnoses, and there were 92,371 such events in the dataset. When we only looked at people with affective disorders based on information from diagnoses, there were 22,067 such events. For people with affective disorders based on prescription information, there were 88,245 events. The results using only prescription information are very similar to those from the original outcome definition, as most events are medicine events. When the outcome is defined using only diagnosis events, the confidence intervals become rather broad as expected, but we also encounter a protective effect of being delivered by intrapartum cesarean delivery, although not a significant effect (Figure 12). The overall conclusions regarding no effect of pre-labor cesarean delivery and only minimal effect of intrapartum cesarean delivery on the risk of affective disorders thus still stands, though the point estimates vary a bit when comparing medicine events to diagnosis events.

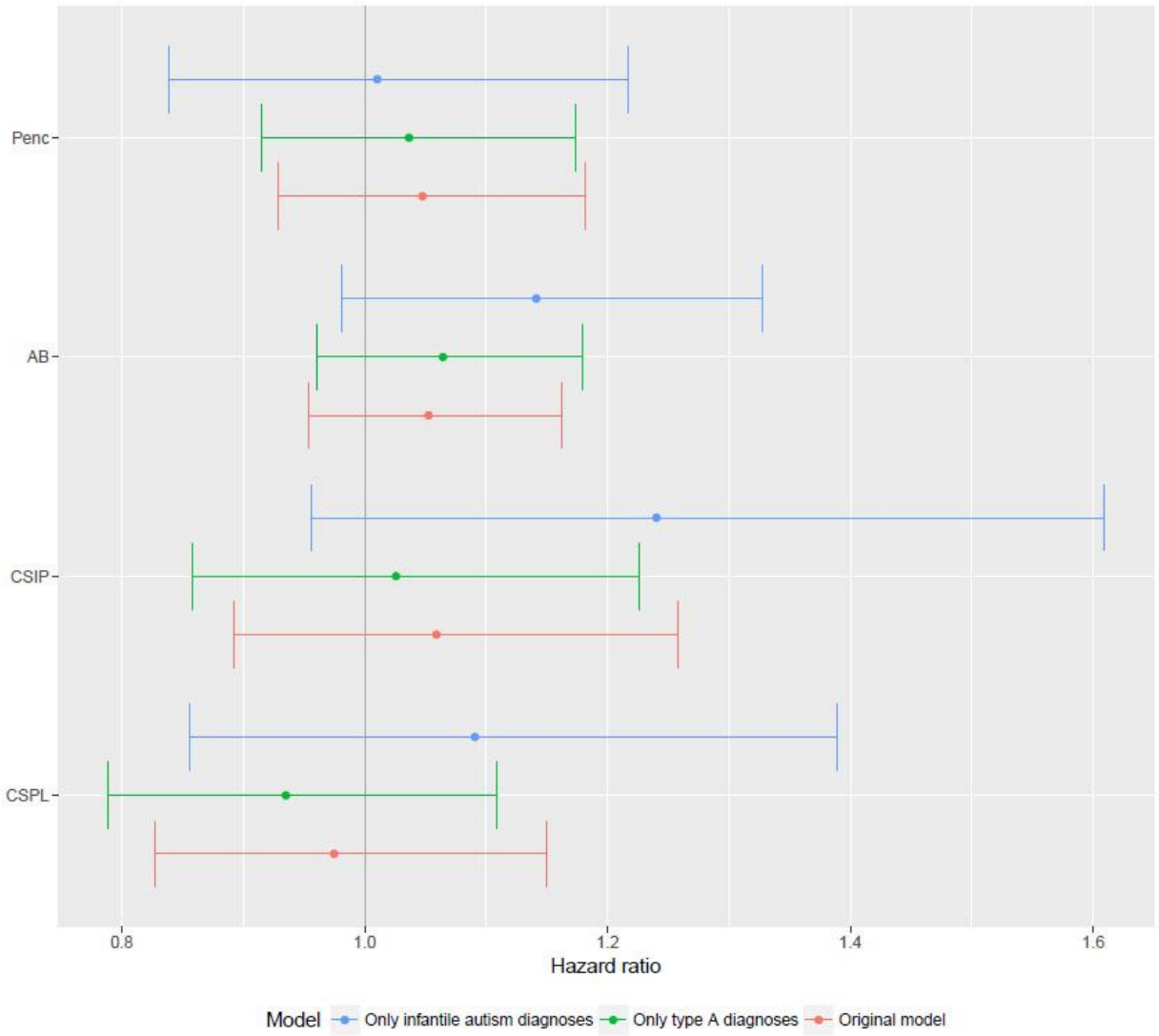


Figure 11: Exposure effect estimates of the fully adjusted between-within model in the ASD study, with varying definitions of the outcome variable.

Inclusion of additional obstetric variables for ASD outcome

As we have discussed in the methods-section, we became aware for the ASD-study that the variables induction of labor, induction of contractions and instrumental delivery had questionable validity and that they could not vary freely across the levels of the exposure variables. We therefore examined the effect of adding these variables to the fully adjusted between-within model. There was no observable change for the exposure effect estimates except for intrapartum cesarean delivery, which were slightly attenuated (Figure 13), but our conclusions remained the same. These variables were not available for the affective disorder study.

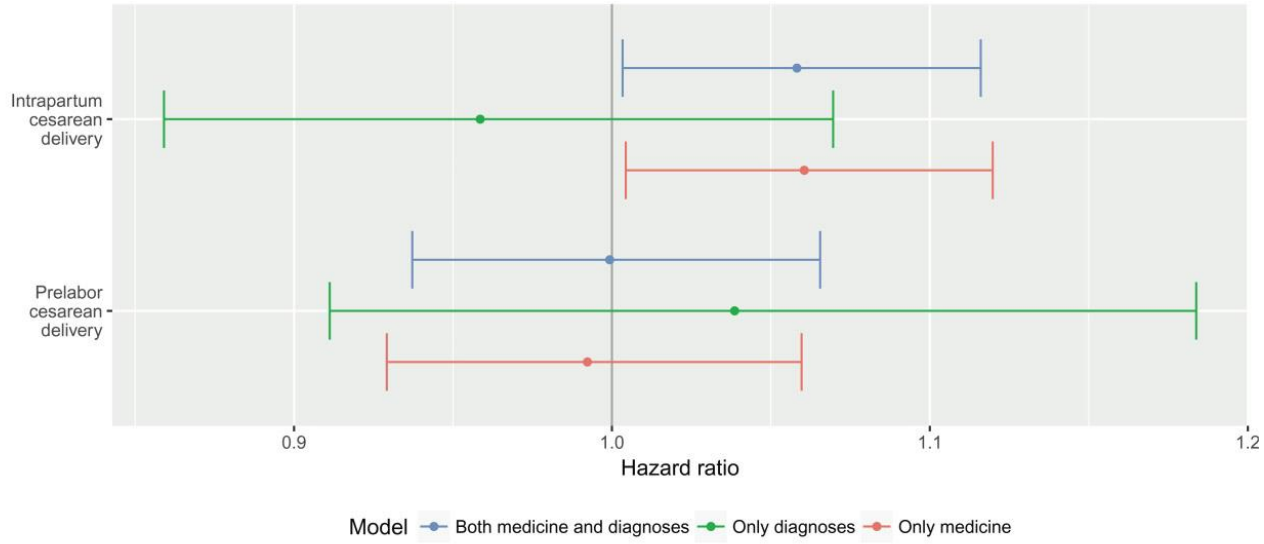


Figure 12: Exposure effect estimates for the affective disorder study from the between-within models, using varying definitions of the outcome variable.

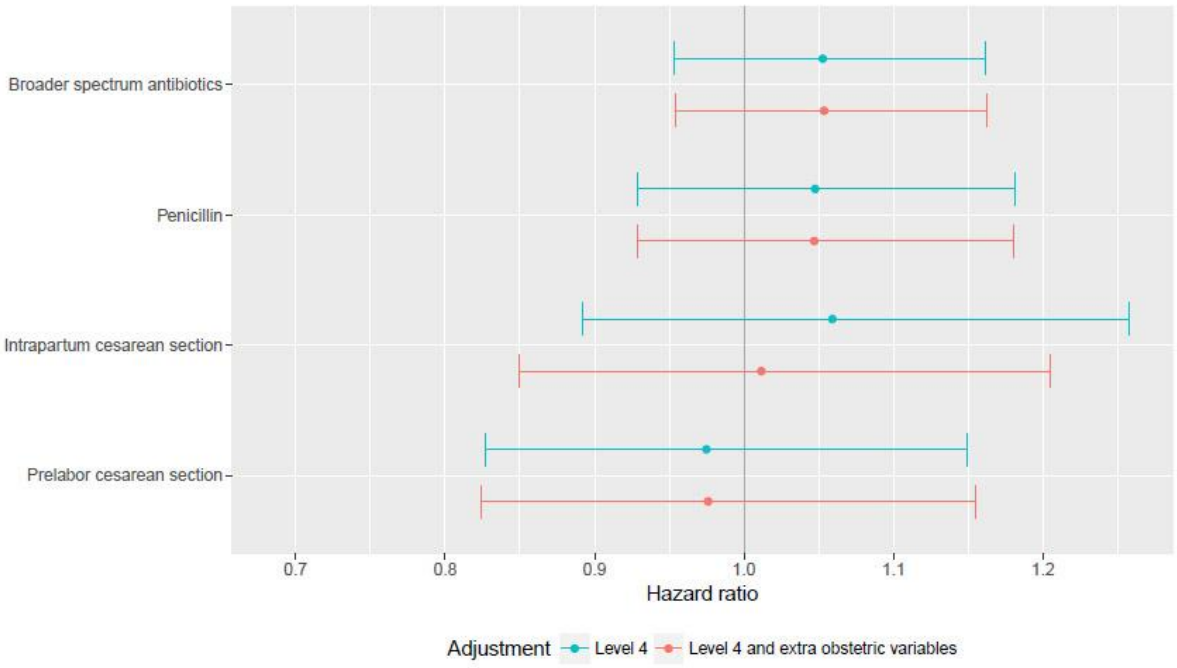


Figure 13: Exposure effect estimates of the between-within model in the ASD study with either full adjustment (Level 4) or the addition of the obstetric variables induction of labor, induction of contractions, and instrumental delivery.

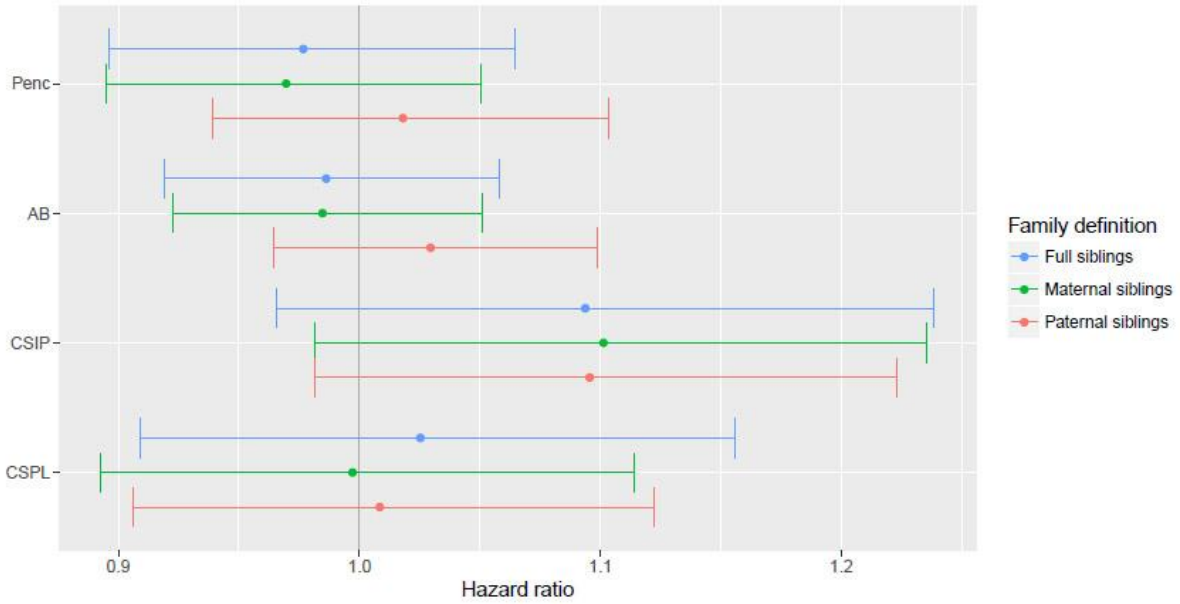


Figure 14: Fully adjusted exposure effect estimates from the between-within model using three different definitions of a family in the ADHD study: full siblings, maternal siblings sharing a mother, paternal siblings sharing a father. Penc: Penicillin only, AB: broader spectrum antibiotics, CSIP: Intrapartum Cesarean Delivery, CSPL: Pre-labor Cesarean Delivery.

Definition of siblingship

For the ADHD and ASD studies we looked at if changing the definition of the family from full-siblings (the main models) to maternal or paternal siblings. Maternal siblings are children that have the same mother but not necessarily the same father and vice versa for paternal siblings. The results of the fully adjusted between-within models are below in Figures 14 and 15. In short we found that all effects are still insignificant and quite stable for the three family definitions.

Fetal position for ADHD outcome

In the ADHD study we looked at the consequences of including a variable for fetal position in the fully adjusted between-within model, as previous studies have suggested breech position in delivery may influence risk of ADHD (114). We defined the variable as a dummy variable for birth in breech position (as opposed to vertex position). We saw a very high level of robustness to whether or not fetal position was included in the models.

Adjustment for parental ADHD or any psychiatric history

For the ADHD study we looked at the effects of adjusting for any psychiatric history of the parents, as compared with only adjusting for ADHD status as we do in the main standard Cox model. For the exposure effect estimates we found only very slightly different point estimates (Figure 16), which did not affect our conclusions.

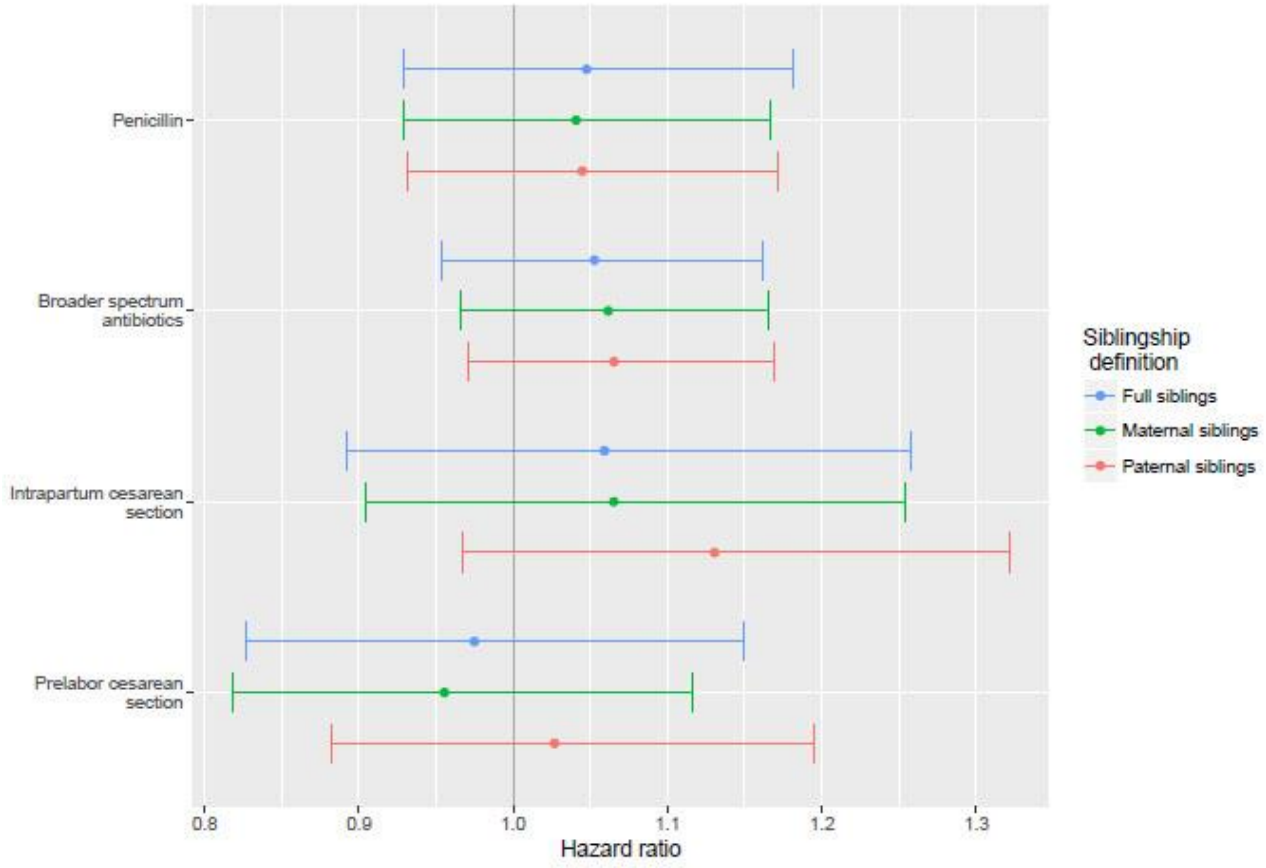


Figure 15: Fully adjusted exposure effect estimates from the between-within model using three different definitions of a family in the ASD study: full siblings, maternal siblings sharing a mother, paternal siblings sharing a father.

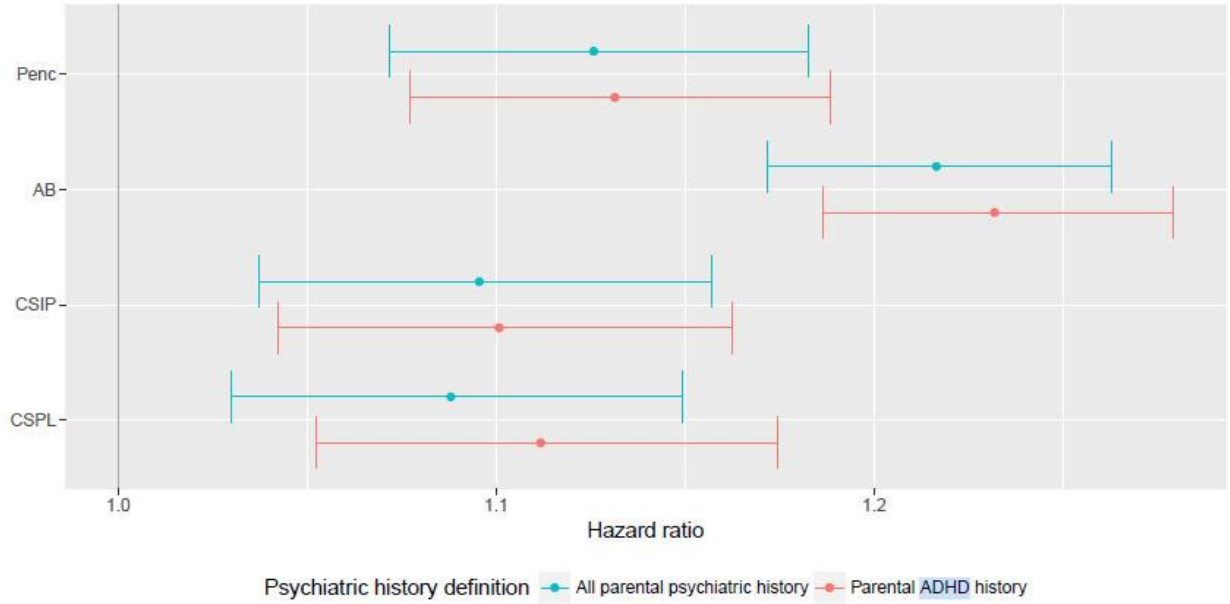


Figure 16: Comparison of fully adjusted standard Cox models with two different options for confounder adjustment for parental psychiatric history: either all psychiatric diagnosis and psychiatric medicine before childbirth are included, or only ADHD information. Penc: Penicillin only, AB: Broader spectrum antibiotics, CSIP: Intrapartum Cesarean Delivery, CSPL: Pre-labor Cesarean Delivery.

Discussion

In this national cohort study, we found little or no evidence that cesarean delivery was causally related to increased risk of ASD, ADHD, or affective disorders. We also found no evidence of causal relationship between antibiotic treatment in the first two years of life and ASD or ADHD. Our main statistical model, the between-within sibling model, found no effects of delivery mode and antibiotic treatment on ASD and ADHD risk, but found a slightly significant increase of affective disorders for children born of intrapartum cesarean and not of pre-labor cesarean. The sibling stratified Cox model was used as a benchmark model for the within-family effects, as we expected the between-within model to produce effect estimates similar to the stratified Cox model, albeit with more precision. We did find that the two sibling models were mostly in agreement regarding point estimates, with relatively small differences that were within the confidence intervals of the more precise between-within model. It is interesting to note that the estimates of the between-within model were always closer to the null-effect than the stratified Cox model, save for one estimate where the models produced estimates very close to one another. However, the between-within model is not intended to produce more accurate estimates, only increase precision.

The only statistically significant exposure effect found by the fully adjusted between-within models deserves a special mention. For the study of affective disorders, we were forced to abandon adjustment for several important obstetric confounders as the validity of the variables became questionable before 1997. As intrapartum cesarean delivery is often performed due to complications in vaginal delivery, we have not been able to adjust for confounding not shared between siblings. It should also be noted that the results of the between-within model can only be interpreted causally if non-shared confounders have been adjusted for (115). Therefore, we find it unlikely that this statistically significant finding is indicative of real causal effects.

The hypothesis that adverse changes to the gut microbiota affect the development of the infant brain, implies that pre-labor cesarean delivery should have greater detrimental effects than intrapartum cesarean delivery. This is because rupture of membranes (ROM) happens in about one fourth of all intrapartum cesarean deliveries, where the infant comes into contact with the (presumably) beneficial vaginal microbiota (116). However, we saw the opposite in our results: delivery by intrapartum cesarean produced greater risk of ADHD, ASD, and affective disorders than pre-labor cesarean delivery. Changing the exposure variable to either cesarean delivery with or without ROM only changed the estimates for the outcome of ADHD to slightly, but insignificant, protective effects. Similarly, under the hypothesis antibiotic treatment with broader spectrum antibiotics relative to penicillin should have had greater effects on risks of psychiatric disorders, as discussed previously. However, our findings in the between-within models showed

almost identical effect estimates for broader spectrum antibiotics and penicillin, and both were very close to null-effects. Under the gut-brain axis hypothesis we would also have believed that maternal antibiotic treatment in pregnancy, especially broader spectrum antibiotics in second or third trimester would have affected the infant's risk of psychiatric disorders, as it is the maternal microbiota that is passed on to the infant during delivery and early life (35). However, estimates for the parameters related to antibiotic treatment in pregnancy were close to null in both standard Cox models and sibling models, with estimates becoming even closer to null in the third trimester. These effects were not entered into the models as exposures and cannot be interpreted in such terms.

All the fully adjusted standard Cox models found significant effects of delivery mode on ADHD, ASD, and affective disorder risk. It is worth repeating that the effects estimated in the standard Cox models are not comparable with the within-family effects of the sibling models as they are not adjusted for family-shared confounding and therefore probably not much use when determining causality. The standard Cox model is therefore provided here as a sort of an epidemiological backdrop, representing what is currently accepted for similar efforts testing clinical associations in observational data. Importantly, it also illustrates that the effects seen in previous studies of these outcomes using standard regression models are possibly largely due to unobserved shared confounding (117).

Previous studies

The aim of this study was to test the hypothesis that adverse changes to the infant gut microbiota, through delivery mode and antibiotic treatment in early life increased risk of later psychiatric disorders. As our findings did not indicate any causal effect, it makes it less likely that the hypothesis is true. However, proving a negative based on observational data is hard if not impossible, which is why more than one study is needed, if we are to convincingly dismiss the hypothesis.

Previously, the theory that adverse changes to the early microbiota influences brain development has been supported by observational data. Children born by cesarean delivery have been found to be at increased risk of ADHD and ASD, most recently summarized in a meta-analyses of 38 cohort studies where our own results were included (117). Despite the authors of the meta-analysis being aware that studies with sibling comparison find attenuated or no association with cesarean delivery, their main results focused on traditional non-sibling stratified studies.

In a single study, children receiving antibiotics in their first two years of life were more likely to have symptoms that are typical of ADHD at the age of 11, as evaluated by their parents and teachers (119). Further, another population based cohort study found association with infections and antibiotic use at any age increased risk of mental disorders, in particular ASD and ADHD (120).

A very recent meta-analysis of early life exposure to antibiotics and ASD showed that there was limited evidence of an association (121), although our study (58) was included as one of the largest studies in the meta-analysis.

In sibling analyses, the risk of ASD, ADHD and affective psychoses (a subcategory of affective disorders) in relation to delivery mode has been evaluated previously (122–124). The author group used a sibling stratified Cox model, and found that there was slightly increased risk of ADHD (HR 1.13, 95%CI 1.01-1.26) for emergency cesarean section, but no other statistically significant increased risks for delivery by cesarean section. The precision of their results were similar to the results of our studies, with confidence intervals being either slightly broader or narrower, despite theirs being a much larger population.

We therefore conclude that our studies have contributed with increased knowledge regarding risk of ASD, ADHD, and affective disorders, especially regarding exposure to antibiotic treatment in infancy but also mode of delivery.

In addition, we have also set out to test if the between-within sibling model had merits when compared to more established methods in real-life application. There have been a few previous efforts using the between-within model in survival analyses (125–128), but none have made any effort to compare it to more established model designs, and it was only used in the main analysis for two studies (125,128) and as a sibling analysis in one (125). Therefore, we believe we are the first to use the between-within sibling model outside of simulation studies, to demonstrate that it has increased power when compared to the sibling stratified Cox model.

Efficiency of the between-within model

When we compared the results of two different sibling models, we found that the between-within model delivers more precise results than the stratified Cox model (57–59), as seen when comparing the width of the confidence intervals for each effect estimate. This fits with the findings of the simulation study of Sjölander at al (57) which we have mentioned previously, where greater power was found using the between-within model. The point estimates in both the sibling stratified Cox model and the between-within model resembled one another, which indicates it is the same subpopulation the models are describing and that the between-within model is correctly specified.

However, in the study regarding affective disorder risk the difference between the two models was not as pronounced. This is not surprising, since larger samples sizes generally imply more robustness towards the influence of specific model choices (129).

Dose response

We did not perform dose response analyses for antibiotic treatment, even though this might have been an interesting topic to explore. We could have performed analyses regarding the number of prescriptions, but we did not have information about the length of each course of treatment, the infant's weight and dosage, meaning we could not have analyzed dose-response effects accurately. Additionally, from a statistical standpoint, categorical groupings of antibiotic exposures might result in insufficient power to detect an effect. But, a null-finding regarding the overall effect of antibiotic treatment and ADHD/ASD risk entails one of three things regarding a dose-response relationship: 1) there is no dose-response relationship 2) there is a U-shaped dose-response curve 3) there is an adverse effect of increased antibiotic dosage but the vast majority of children have no more than 1 prescription for an antibiotic. We believe that the last two scenarios are unlikely.

Window of critical development

As was mentioned in the start of this thesis, it has been hypothesized that adverse changes to the gut microbiota have greatest effect in the very early life of the infant. However, sensitivity analysis involving whether children received their first prescription of antibiotics in the first or the second year of life, reflected no significant change in effect estimates for ADHD risk compared to when viewing the exposure period for the first two years combined. However, when only considering the first year of antibiotic exposure in the analysis, we found a small but significant lower risk of ADHD, which we found difficult to fully explain. With this effect being only marginally significant without correction for multiple testing, we assume it is merely sporadic and do not interpret it as a real protective effect of antibiotic treatment. Most importantly, we would have expected to see an increase in exposure effects if it had happened earlier in life and if the critical developmental window hypothesis was true.

Also, one might imagine that the effects of antibiotics was dependent on delivery mode – a double whammy within this critical window. However, we found no evidence of an interaction effect between antibiotic treatment and mode of delivery for the outcomes of ADHD and ASD. That implies that the effects of the antibiotic exposures do not depend on the value of the mode of delivery exposure variable, and vice versa.

Case severity

The difference in severity between cases based on prescriptions and diagnoses can be debated. For ADHD we would have expected the most severe cases were both seen by a psychiatrist working at either a hospital ward or at an out-patient service (meaning they had a diagnostic code ADHD), and that they would have been treated by ADHD medication. In sensitivity analyses, changing the ADHD outcome definition to

require both a diagnostic code and a prescription did produce slightly altered point estimates in particular for the exposure of Penicillin. However, none of the differences were significant and did not alter our conclusions of no effect of the exposures on ADHD risk.

In the ASD-study, we would have expected infantile autism to represent the more severe cases on the autism spectrum, as language skills and IQ are usually lesser than compared to Asperger's (130). Indeed, we did see somewhat increased risks of infantile autism for cesarean delivery, but they were statistically insignificant differences when compared to estimates using all autism spectrum disorder diagnoses. Moreover, the confidence intervals when only considering infantile autism diagnoses were considerably wider, especially for the cesarean delivery exposures.

For depression, we know from a previous Danish validation study (41), that only 16-18% of cases based on diagnoses are considered mild. In a large American validation study of prescription practices, those prescribed antidepressants were considered to have mild depression in only 12% of cases (131). As we have previously mentioned, people receiving antidepressants in Denmark seem to have the same inappropriate delay in seeking health services for preventable disease, as do those with a diagnosis of depression (46). These groups may therefore not be that dissimilar. Our hypothesis was however that those seen in a hospital setting, either admitted or in an out-patient clinic, had a more severe case of depression. However, we saw no evidence of a greater risk of affective disorders if they were solely based on diagnosis, on the contrary the point estimate for intrapartum cesarean delivery became protective, albeit statistically insignificant.

In conclusion, we can say that it is probably hard to define case severity from the information available in the national registries, but changing the outcome definitions did not reveal any significantly greater risk of these psychiatric disorders.

Generalizability

We have already covered some aspects that may limit the generalizability of our results to the whole population, such as underrepresentation of children treated by ventilator or CPAP, born to mothers treated with penicillin in the third trimester, in the ASD/ADHD cohort.

We have assumed that the differences we see in exposure estimates between the standard Cox model and the sibling models are due to shared confounding. However, there may be other factors that need considering. As the sibling studies are based on exposure discordance, the results can only be directly applied to such a population. The results may not be applicable to families with only one child, for example if parents with a higher social status would be more likely to have only one child.

We did see some differences in the risk of ASD for first born children compared to those who had older siblings when analyzed in the standard Cox model. However, we were unable to determine if these differences were substantive due to low precision and we have no reason to believe these effects would persist in a model where more complete confounding was accounted for. Regardless, we must conclude that results from the sibling models in this case may not be generalizable to children without siblings and advocate caution when generalized to the whole population, as there may be bias from reproduction stoppage (58,113).

The effect estimates of the sibling models are generated by mostly using information from families with exposure discordant siblings. Therefore, the results of the sibling studies may only apply to such subpopulations, but we have no reason to believe that they are particularly atypical constructions. However, if the firstborn is delivered by cesarean delivery most often the subsequent younger siblings are also born by cesarean delivery, but in such cases usually the firstborn is delivered by a intrapartum cesarean delivery whereas the younger siblings may be born by prelabour cesarean delivery (132). Mothers opting for a trial of labor after cesarean may perhaps be more psychologically robust than those who opt for repeat cesarean delivery. Similar trends might be true for antibiotic treatment of infants, where families of a lower socioeconomic position are more prone to plead for antibiotics and thus all infants in the family may have been exposed to antibiotics before they are two years of age (133).

Prevalence

We have seen trends of increased prevalence from the year 2009 as well as increasingly younger age at first treatment for all three psychiatric outcomes. We suggest this could be associated with the introduction of social media, with Facebook gaining popularity in 2008 in Denmark. Social media may affect people's self-esteem and thereby their mood, as it is inevitable to compare one's own life with the mostly positive moments from other people's lives (134,135). Other recent studies have found evidence in support of this hypothesis, such as a study of 4000 children in Canada, where increased screen time was related to increased signs of depression (136). Furthermore, a study of teenagers found that higher use of digital media increased risk of subsequent symptoms of ADHD (137). In addition, parents may have become more aware of ASD through posts on social media that become widely circulated (138), which is obviously is not the cause of ASD but rather perhaps explains part of the increase in diagnosed cases.

At first glance, it might appear to some as though the prevalence of ADHD was lower in our study than reported in previous studies (139). But the prevalence and incidence of ADHD cannot be read directly from the Lexis-diagram or the reported number of cases in the study follow-up period. As it is not the purpose of this study to report the prevalence of ADHD in Denmark, we have not done so. But the rate of ADHD cases

in our study population - 17,971 cases out of 671,592 children or 2.7% - is in line with a study of similar follow-up design based on Swedish data in almost the same period: 2.77% (122). In a recent meta-analysis, the world-wide ADHD prevalence was found to be 7.2%, but the vast majority of studies was conducted within school-populations (139). Had we similarly removed any children under the age of 6 from our study population, our case rate (and prevalence) would have been higher. In addition, the prevalence rates of ADHD in European studies are generally lower than when compared to North American studies (139). This could be due to the diagnostic criteria used in North America are somewhat broader than the ICD-10 criteria used in Europe (59,66,140). Indeed, the rate of both ADHD and ASD in our study population seems to be much in line with that reported in a recent reviews (141,142).

Temporality

As any outcome needs to come after the causal agent, we would intuitively expect that we should not be able to find any evidence of early manifestations of psychiatric disorders before children are exposed to delivery mode or antibiotics. However, the causality of psychiatric disorders is most probably multifactorial, and therefore it does not exclude the possibility that exposures at or after birth play a role for many ASD cases.

In a recent study the researchers were able to predict high risk of ASD in unborn children depending on blood metabolites of the mother during pregnancy (143). Currently, a study is underway using functional magnetic resonance imaging (MRI) of the fetal brain during pregnancy, and so far 24 unborn siblings of children with autism have been scanned. Preliminary results hints at atypical brain activity, with increased activity in regions involved in processing sensory stimuli (144).

There seem to be very specific structural differences in the brains of children with ADHD compared with control subjects (145). A study of 43 full-term, healthy infants' brains at two weeks postpartum by MRI showed "lower white matter development and lower functional connectivity in the brain default mode network" for children born by cesarean section when compared to those born vaginally. In children aged 3-60 months, there were also different trajectories of white matter myelination, with cesarean born infants having reduced myelin, however normalizing with age (146). In our opinion however, it does seem unlikely that changes in neuronal development due to delivery mode can be visualized already 2 weeks postpartum, and therefore do not believe these changes are caused by cesarean delivery. Yet, if cesarean delivery did cause the neuronal restructuring as seen in patients with ADHD (145), we would expect to see these changes in older children as well. However, in the same study there were no differences between another cohort of children at age 8 years, regardless of delivery mode (146). The explanation for these differences between children born of different delivery modes may be simple, as there were some demographic

differences between infants born by cesarean section compared to those delivered vaginally. Most notably greater maternal body mass index (BMI) and weight gain during pregnancy for children born by cesarean section.

So far, the evidence does not clearly indicate if there are manifestations of psychiatric disorders during pregnancy, although there are indications that it may be plausible. It will be interesting to follow future efforts within this line of research, as it may provide valuable puzzle pieces for understanding the neurodevelopment of the brain and the pathology of psychiatric disorders.

Strengths and limitations

There have been no previous efforts to study the potential causal relationship in a sibling model between antibiotic treatment in the first two years of life and later risk of ADHD and ASD. We were also the first to examine in a sibling model the risk of all affective disorders depending on delivery mode.

The study populations were large, with a relatively long follow-up for each outcome. Even though previous studies have reported the average age of first diagnosis of affective disorders to be little over 35 years (147) we can see from the Lexis-diagram in paper III (Figure 2)(60) that the rate of new cases declines before the end of follow-up for our study.

As a consequence of our large study populations and use of the between-within sibling model, our results were very precise. In addition, the numerous sensitivity analyses we performed indicated that the findings were robust. Much of the information available to us from the national registries has been validated, as we have mentioned previously. The Danish registries have also captured information regarding outpatient contacts and prescription medication for a longer time than most other national registries.

By only looking at children of Danish ancestry we limited our population size and reduced representativity, but as the amount of missing data for immigrants was considerably higher than for native Danes we avoided

There are inherent limitations to using any sibling comparison model, which we have covered in the "generalizability" section.

It has been cautioned not to interpret the results of sibling designs as causal effects such as those of a randomized controlled trials. The exception is when the results of sibling analyses adjusted for non-shared confounders are that of null-results, where the sibling analyses provide a well-defined interpretation of no causal effects (148).

We mostly found estimates close to null, so carry-over effects (such as the outcome of the older sibling affecting the exposure of the younger sibling) are unlikely to have affected our results. We have discussed previously that non-shared confounding makes inferring causal relations difficult – this is because non-shared confounding may still produce bias in sibling models. However, this is unlikely to be a problem for our study as there are only two possible scenarios: either the multiple contradicting effects cancels each other out, or there are no causal effects. The latter option is more likely in our opinion. Sibling models also control for shared hidden mediators – variables on the path from exposure to outcome, that are affected by the exposure and in turn affect the outcome. Controlling for such mediators will have us estimate the direct effect of the exposure on the outcome, which might also be of interest. However, since such mediators have to be shared by the siblings, we cannot think of any that are relevant to our studies.

There is also a very practical limitation to using the between-within sibling model. Currently, there are no statistical program packages available, meaning there is considerable programming work to be done each time a new hypothesis is tested. It also requires computing resources that challenge those currently available on the Statistics Denmark servers, as they are reset every week and some calculations can take that long to complete. This compared to the stratified Cox model, that could be finished in a few hours. These obstacles may however be overcome in the future, if the potential of the between-within model becomes more widely recognized.

Conclusions and future perspective

To be able to refute theories of causality, using observational studies, you need several different strong sources of evidence. However, we have failed to produce any such evidence for the theory of alterations to the infant gut microbiota, through mode of delivery and antibiotic treatment, influencing the neurodevelopment of the brain and resulting in major psychiatric disorders later in life. We argue that most limitations inherent to the sibling models in terms of causal understanding, are irrelevant as the null-findings for the outcomes of ADHD and ASD were precise and robust. Further, we suspect that the significant finding in the between-within model between intrapartum cesarean and affective disorders, is likely due to residual confounding. At least, it does not fit with the theory that prelabour cesarean delivery would affect the infant microbiota the most.

These precise negative findings mean that parents should not have to worry about increased risks of psychiatric disorders to the infant when considering delivery mode or antibiotic treatment. Derived from these results, we would not expect any benefit of probiotics or microbiota transfer treatments for the infant, when trying to reduce risk of these psychiatric disorders.

It should be noted that had we only investigated these topics using typical statistical methods commonly used at the time, be it standard or stratified Cox models, we would have concluded differently than we did with the results of the between-within model. When testing hypothesis of causality in observational data, we recommend considering the use of sibling models and preferably the between-within model, if there is any reason to believe there might be unobserved confounding by early familial environment and genetics. However, to gain more widespread use, the between-within model needs to become easier to apply on a popular statistical platform, such as for the open-source R program.

We hope that our findings will inspire others to consider sibling models for their epidemiological studies, although we recognize that not all exposures or outcomes can be examined in such a way. We are also optimistic that readers within the healthcare sector will become aware of the differences in strength of evidence when comparing statistical models that consider unmeasurable familial confounding and those models that do not.

The human microbiota is a fascinating subject and many researchers have pinned their hopes on making a breakthrough in finding the causal agent of common unexplained disorders by studying the gut microbiota. The reason for differences previously noted in gut microbiota between children born of different modes of delivery, seems like a good place to start. Currently we are investigating the very early human microbiota and how microbes are transferred from mother to offspring during birth in clinical trials. Hopefully we will be able to gain some insights into how and why the gut microbiota differs between infants this early in life.

References

1. Jensen CM, Steinhausen H-C. Time Trends in Incidence Rates of Diagnosed Attention-Deficit/Hyperactivity Disorder Across 16 Years in a Nationwide Danish Registry Study. J Clin Psychiatry. 2015 Mar 25;e334–41.

2. Hansen SN, Schendel DE, Parner ET. Explaining the increase in the prevalence of autism spectrum disorders: the proportion attributable to changes in reporting practices. JAMA Pediatr. 2015 Jan;169(1):56–62.

3. Mayer EA, Tillisch K, Gupta A. Gut/brain axis and the microbiota. J Clin Invest. 2015 Mar 2;125(3):926–38.

4. Stinson LF, Payne MS, Keelan JA. A Critical Review of the Bacterial Baptism Hypothesis and the Impact of Cesarean Delivery on the Infant Microbiome. Front Med. 2018;5:135.

5. Ianiro G, Tilg H, Gasbarrini A. Antibiotics as deep modulators of gut microbiota: between good and evil. Gut. 2016 Nov;65(11):1906–15.

6. Bollig C, Nothacker M, Lehane C, Motschall E, Lang B, Meerpohl JJ, et al. Prophylactic antibiotics before cord clamping in cesarean delivery: a systematic review. Acta Obstet Gynecol Scand. 2018 May;97(5):521–35.

7. Acar J. Broad- and narrow-spectrum antibiotics: an unhelpful categorization. Clin Microbiol Infect. 1997 Aug;3(4):395–6.

8. Zachariassen G, Hyldig N, Joergensen JS, Nielsen DS, Greisen G. The half-life and exposure of cefuroxime varied in newborn infants after a Caesarean section. Acta Paediatr Oslo Nor 1992. 2016 Sep;105(9):1074–8.

9. Palleja A, Mikkelsen KH, Forslund SK, Kashani A, Allin KH, Nielsen T, et al. Recovery of gut microbiota of healthy adults following antibiotic exposure. Nat Microbiol. 2018 Nov;3(11):1255–65.

10. Atladottir HO, Henriksen TB, Schendel DE, Parner ET. Autism After Infection, Febrile Episodes, and Antibiotic Use During Pregnancy: An Exploratory Study. PEDIATRICS. 2012 Dec 1;130(6):e1447–54.

11. Cox LM, Yamanishi S, Sohn J, Alekseyenko AV, Leung JM, Cho I, et al. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. Cell. 2014 Aug 14;158(4):705–21.

12. Bardin J. Neurodevelopment: Unlocking the brain. Nature. 2012 Jul;487(7405):24–6.

13. Meredith RM. Sensitive and critical periods during neurotypical and aberrant neurodevelopment: A framework for neurodevelopmental disorders. Neurosci Biobehav Rev. 2015 Mar;50:180–8.

14. Rodríguez JM, Murphy K, Stanton C, Ross RP, Kober OI, Juge N, et al. The composition of the gut microbiota throughout life, with an emphasis on early life. Microb Ecol Health Dis. 2015;26:26050.

15. Barbaro J, Dissanayake C. Autism spectrum disorders in infancy and toddlerhood: a review of the evidence on early signs, early identification tools, and early diagnosis. J Dev Behav Pediatr JDBP. 2009 Oct;30(5):447–59.

16. Brown HR, Harvey EA. Psychometric Properties of ADHD Symptoms in Toddlers. J Clin Child Adolesc Psychol Off J Soc Clin Child Adolesc Psychol Am Psychol Assoc Div 53. 2019 Jun;48(3):423–39.

17. Forsythe P, Kunze W, Bienenstock J. Moody microbes or fecal phrenology: what do we know about the microbiota-gut-brain axis? BMC Med. 2016 Apr 19;14:58.

18. Kelly JR, Clarke G, Cryan JF, Dinan TG. Brain-gut-microbiota axis: challenges for translation in psychiatry. Ann Epidemiol. 2016 May;26(5):366–72.

19. Sharon G, Cruz NJ, Kang D-W, Gandal MJ, Wang B, Kim Y-M, et al. Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. Cell. 2019 May 30;177(6):1600-1618.e17.

20. Bharwani A, Mian MF, Foster JA, Surette MG, Bienenstock J, Forsythe P. Structural & functional consequences of chronic psychosocial stress on the microbiome & host. Psychoneuroendocrinology. 2016 Jan;63:217–27.

21. Bharwani A, Mian MF, Surette MG, Bienenstock J, Forsythe P. Oral treatment with Lactobacillus rhamnosus attenuates behavioural deficits and immune changes in chronic social stress. BMC Med. 2017 Jan 11;15(1):7.

22. Wang S, Qu Y, Chang L, Pu Y, Zhang K, Hashimoto K. Antibiotic-induced microbiome depletion is associated with resilience in mice after chronic social defeat stress. J Affect Disord [Internet]. 2019 Sep [cited 2019 Sep 19]; Available from: https://linkinghub.elsevier.com/retrieve/pii/S0165032719321901

23. Hildebrandt S. Dagens Medicin - Ny teori; Tarmbakterier drejer hjernen ud af fatning [Internet]. 2017. Available from: https://dagensmedicin.dk/ny-teori-tarmbakterier-drejer-hjernen-fatning/

24. Kang D-W, Adams JB, Coleman DM, Pollard EL, Maldonado J, McDonough-Means S, et al. Long-term benefit of Microbiota Transfer Therapy on autism symptoms and gut microbiota. Sci Rep [Internet]. 2019 Dec [cited 2019 Sep 21];9(1). Available from: http://www.nature.com/articles/s41598-019-42183-0

25. Dominguez-Bello MG, De Jesus-Laboy KM, Shen N, Cox LM, Amir A, Gonzalez A, et al. Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. Nat Med. 2016 Feb 1;

26. Haahr T, Glavind J, Axelsson P, Bistrup Fischer M, Bjurström J, Andrésdóttir G, et al. Vaginal seeding or vaginal microbial transfer from the mother to the caesarean-born neonate: a commentary regarding clinical management. BJOG Int J Obstet Gynaecol [Internet]. 2017 Aug 22 [cited 2018 Mar 8]; Available from: http://doi.wiley.com/10.1111/1471-0528.14792

27. Ho NT, Li F, Lee-Sarwar KA, Tun HM, Brown BP, Pannaraj PS, et al. Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. Nat Commun. 2018 09;9(1):4169.

28. Wolters M, Ahrens J, Romaní-Pérez M, Watkins C, Sanz Y, Benítez-Páez A, et al. Dietary fat, the gut microbiota, and metabolic health – A systematic review conducted within the MyNewGut project. Clin Nutr [Internet]. 2018 Dec [cited 2019 Sep 10]; Available from: https://linkinghub.elsevier.com/retrieve/pii/S0261561418325925

29. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human Genetics Shape the Gut Microbiome. Cell. 2014 Nov;159(4):789–99.

30. Goodrich JK, Davenport ER, Clark AG, Ley RE. The Relationship Between the Human Genome and Microbiome Comes into View. Annu Rev Genet. 2017 Nov 27;51(1):413–33.

31. Bowyer RCE, Jackson MA, Le Roy CI, Ni Lochlainn M, Spector TD, Dowd JB, et al. Socioeconomic Status and the Gut Microbiome: A TwinsUK Cohort Study. Microorganisms. 2019 Jan 11;7(1).

32. Fond G, Boukouaci W, Chevalier G, Regnault A, Eberl G, Hamdani N, et al. The “psychomicrobiotic”: Targeting microbiota in major psychiatric disorders: A systematic review. Pathol Biol (Paris). 2015 Feb;63(1):35–42.

33. Mangiola F. Gut microbiota in autism and mood disorders. World J Gastroenterol. 2016;22(1):361.

34. Dunn AB, Jordan S, Baker BJ, Carlson NS. The Maternal Infant Microbiome: Considerations for Labor and Birth. MCN Am J Matern Child Nurs. 2017 Dec;42(6):318–25.

35. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe. 2015 Jun 10;17(6):852.

36. Dudek-Wicher RK, Junka A, Bartoszewicz M. The influence of antibiotics and dietary components on gut microbiota. Przeglad Gastroenterol. 2018;13(2):85–92.

37. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health. 2011 Jul;39(7 Suppl):30–3.

38. Mohr-Jensen C, Vinkel Koch S, Briciet Lauritsen M, Steinhausen H-C. The validity and reliability of the diagnosis of hyperkinetic disorders in the Danish Psychiatric Central Research Registry. Eur Psychiatry J Assoc Eur Psychiatr. 2016 May;35:16–24.

39. Lauritsen MB, Jørgensen M, Madsen KM, Lemcke S, Toft S, Grove J, et al. Validity of Childhood Autism in the Danish Psychiatric Central Register: Findings from a Cohort Sample Born 1990–1999. J Autism Dev Disord. 2010 Feb;40(2):139–48.

40. Kessing L. Validity of diagnoses and other clinical register data in patients with affective disorder. Eur Psychiatry J Assoc Eur Psychiatr. 1998 Dec;13(8):392–8.

41. Bock C, Bukh J, Vinberg M, Gether U, Kessing L. Validity of the diagnosis of a single depressive episode in a case register. Clin Pract Epidemiol Ment Health. 2009;5(1):4.

42. Troisi R, Bjørge T, Gissler M, Grotmol T, Kitahara CM, Myrtveit Sæther SM, et al. The role of pregnancy, perinatal factors and hormones in maternal cancer risk: a review of the evidence. J Intern Med. 2018 May;283(5):430–45.

43. Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. Int J Epidemiol. 2016 Oct 27;dyw213.

44. Norwegian Institute of Public Health. About the Norwegian Prescription Database [Internet]. Available from: http://www.norpd.no/

45. Wallerstedt SM, Wettermark B, Hoffmann M. The First Decade with the Swedish Prescribed Drug Register - A Systematic Review of the Output in the Scientific Literature. Basic Clin Pharmacol Toxicol. 2016 Nov;119(5):464–9.

46. Davydow DS, Fenger-Grøn M, Ribe AR, Pedersen HS, Prior A, Vedsted P, et al. Depression and risk of hospitalisations and rehospitalisations for ambulatory care-sensitive conditions in Denmark: a population-based cohort study. BMJ Open. 2015;5(12):e009878.

47. Adler I. Primary Malignant Growths of the Lungs and Bronchi. London: Longmans, 1912:22. London: Longmans. 1912(22).

48. Proctor RN. The history of the discovery of the cigarette–lung cancer link: evidentiary traditions, corporate denial, global toll: Table 1. Tob Control. 2012 Mar;21(2):87–91.

49. Messerli FH. Chocolate consumption, cognitive function, and Nobel laureates. N Engl J Med. 2012 Oct 18;367(16):1562–4.

50. Doll R, Hill AB. The Mortality of Doctors in Relation to Their Smoking Habits. BMJ. 1954 Jun 26;1(4877):1451–5.

51. Hammond EC, Horn D. The relationship between human smoking habits and death rates: a follow-up study of 187,766 men. J Am Med Assoc. 1954 Aug 7;155(15):1316–28.

52. Hecht SS. Carcinogenicity studies of inhaled cigarette smoke in laboratory animals: old and new. Carcinogenesis. 2005 Sep 1;26(9):1488–92.

53. Ferreira Antunes JL, Toporcov TN, Biazevic MGH, Boing AF, Scully C, Petti S. Joint and independent effects of alcohol drinking and tobacco smoking on oral cancer: a large case-control study. PloS One. 2013;8(7):e68132.

54. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerg Themes Epidemiol. 2015;12:14.

55. Frisell T, Öberg S, Kuja-Halkola R, Sjölander A. Sibling comparison designs: bias from non-shared confounders and measurement error. Epidemiol Camb Mass. 2012 Sep;23(5):713–20.

56. Gilman SE, Loucks EB. Another casualty of sibling fixed-effects analysis of education and health: An informative null, or null information? Soc Sci Med. 2014 Oct;118:191–3.

57. Sjölander A, Lichtenstein P, Larsson H, Pawitan Y. Between-within models for survival analysis. Stat Med. 2013 Aug 15;32(18):3067–76.

58. Axelsson PB, Clausen TD, Petersen AH, Hageman I, Pinborg A, Kessing LV, et al. Relation between infant microbiota and autism? - Results from a national cohort sibling-design study. Epidemiol Camb Mass. 2018 Sep 28;

59. Axelsson PB, Clausen TD, Petersen AH, Hageman I, Pinborg A, Kessing LV, et al. Investigating the effects of delivery by cesarean section and antibiotic use in early infancy on risk of ADHD later in life. Submitted for review;

60. Axelsson PB, Petersen AH, Hageman I, Pinborg A, Kessing LV, Bergholt T, et al. Is cesarean section a cause of affective disorders? - A national cohort study using sibling designs. Unpublished. 2019.

61. Danmark, Sundhedsstyrelsen. Operations- og behandlingsklassifikation. - - Tilgang og rettelser til operationsklassifikationen i 1989-1991. - [5 bl.] - - Tilgang og rettelser til operationsklassifikationen i 1989-1991. - [5 bl. Kbh.: Sundhedsstyrelsen; 1988.

62. Nordisk Medicinal-Statistisk Komité. NOMESCO classification of surgical procedures. Copenhagen: Nordic Medico-Statistical Committee; 2009.

63. Friis H, Hendel J, Dalhoff KimP, Bjerrum L. pro.medicin.dk [Internet]. Azithromycin. Available from: http://pro.medicin.dk/Medicin/Praeparater/5965

64. Goldman JL, Jackson MA, Herigon JC, Hersh AL, Shapiro DJ, Leeder JS. Trends in adverse reactions to trimethoprim-sulfamethoxazole. Pediatrics. 2013 Jan;131(1):e103-108.

65. Kuehn J, Ismael Z, Long PF, Barker CIS, Sharland M. Reported rates of diarrhea following oral penicillin therapy in pediatric clinical trials. J Pediatr Pharmacol Ther JPPT Off J PPAG. 2015 Apr;20(2):90–104.

66. Doernberg E, Hollander E. Neurodevelopmental Disorders (ASD and ADHD): DSM-5, ICD-10, and ICD-11. CNS Spectr. 2016 Aug;21(04):295–9.

67. WHO. Conversion table between section V (mental disorders) of the ICD-8 and chapter V (mental and behavioural disorders) of the ICD-10. 2003. Available from: http://www.who.int/msa/mnh/ems/icd10/convtabl/icd8-101.txt

68. Pedersen CB, Mors O, Bertelsen A, Waltoft BL, Agerbo E, McGrath JJ, et al. A Comprehensive Nationwide Study of the Incidence Rate and Lifetime Risk for Treated Mental Disorders. JAMA Psychiatry. 2014 May 1;71(5):573.

69. Kessing L. A comparison of ICD-8 and ICD-10 diagnoses of affective disorder -a case register study from Denmark. Eur Psychiatry J Assoc Eur Psychiatr. 1998 Nov;13(7):342–5.

70. Center for Disease Control. Prevalence of Autism Spectrum Disorders - Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States [Internet]. 2008. Available from: https://www.cdc.gov/mmwr/preview/mmwrhtml/ss6103a1.htm?s_cid=ss6103a1_w

71. Taurines R, Schwenck C, Westerwald E, Sachse M, Siniatchkin M, Freitag C. ADHD and autism: differential diagnosis or overlapping traits? A selective review. Atten Deficit Hyperact Disord. 2012 Sep;4(3):115–39.

72. Svahn MF, Hargreave M, Nielsen TSS, Plessen KJ, Jensen SM, Kjaer SK, et al. Mental disorders in childhood and young adulthood among children born to women with fertility problems. Hum Reprod Oxf Engl. 2015 Sep;30(9):2129–37.

73. Steinhausen H-C, Jakobsen H, Helenius D, Munk-Jørgensen P, Strober M. A nation-wide study of the family aggregation and risk factors in anorexia nervosa over three generations. Int J Eat Disord. 2015 Jan;48(1):1–8.

74. Andersen SM, Randers A, Jensen CM, Bisgaard C, Steinhausen H-C. Preceding diagnoses to young adult bipolar disorder and schizophrenia in a nationwide study. BMC Psychiatry. 2013;13:343.

75. Webb RT, Abel KM, Pickles AR, Appleby L, King-Hele SA, Mortensen PB. Mortality risk among offspring of psychiatric inpatients: a population-based follow-up to early adulthood. Am J Psychiatry. 2006 Dec;163(12):2170–7.

76. Mortensen PB, Pedersen CB, Hougaard DM, Nørgaard-Petersen B, Mors O, Børglum AD, et al. A Danish National Birth Cohort study of maternal HSV-2 antibodies as a risk factor for schizophrenia in their offspring. Schizophr Res. 2010 Sep;122(1–3):257–63.

77. Mortensen PB, Pedersen MG, Pedersen CB. Psychiatric family history and schizophrenia risk in Denmark: which mental disorders are relevant? Psychol Med. 2010 Feb;40(2):201–10.

78. Kläning U, Laursen TM, Licht RW, Kyvik KO, Skytthe A, Mortensen PB. Is the risk of bipolar disorder in twins equal to the risk in singletons? A nationwide register-based study. J Affect Disord. 2004 Aug;81(2):141–5.

79. Pedersen CB, Mortensen PB. Urbanicity during upbringing and bipolar affective disorders in Denmark. Bipolar Disord. 2006 Jun;8(3):242–7.

80. Mortensen PB, Pedersen CB, Melbye M, Mors O, Ewald H. Individual and familial risk factors for bipolar affective disorders in Denmark. Arch Gen Psychiatry. 2003 Dec;60(12):1209–15.

81. Lauritsen MB, Pedersen CB, Mortensen PB. Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. J Child Psychol Psychiatry. 2005 Sep;46(9):963–71.

82. Mouridsen SE, Rich B, Isager T, Nedergaard NJ. Psychiatric disorders in individuals diagnosed with infantile autism as children: a case control study. J Psychiatr Pract. 2008 Jan;14(1):5–12.

83. Mouridsen SE, Rich B, Isager T, Nedergaard NJ. Psychiatric disorders in the parents of individuals with infantile autism: a case-control study. Psychopathology. 2007;40(3):166–71.

84. Mouridsen SE, Hauschild K-M. A longitudinal study of schizophrenia- and affective spectrum disorders in individuals diagnosed with a developmental language disorder as children. J Neural Transm. 2008 Nov;115(11):1591–7.

85. Baldur-Felskov B, Kjaer SK, Albieri V, Steding-Jessen M, Kjaer T, Johansen C, et al. Psychiatric disorders in women with fertility problems: results from a large Danish register-based cohort study. Hum Reprod Oxf Engl. 2013 Mar;28(3):683–90.

86. Mouridsen SE, Rich B, Isager T. Psychiatric disorders in adults diagnosed as children with atypical autism. A case control study. J Neural Transm Vienna Austria 1996. 2008;115(1):135–8.

87. Mortensen EL, Sørensen HJ, Jensen HH, Reinisch JM, Mednick SA. IQ and mental disorder in young men. Br J Psychiatry J Ment Sci. 2005 Nov;187:407–15.

88. Thygesen LC, Dalton SO, Johansen C, Ross L, Kessing LV, Hvidt NC. Psychiatric disease incidence among Danish Seventh-day Adventists and Baptists. Soc Psychiatry Psychiatr Epidemiol. 2013 Oct;48(10):1583–90.

89. Sørensen HJ, Nielsen PR, Pedersen CB, Mortensen PB. Association between prepartum maternal iron deficiency and offspring risk of schizophrenia: population-based cohort study with linkage of Danish national registers. Schizophr Bull. 2011 Sep;37(5):982–7.

90. Khashan AS, McNamee R, Henriksen TB, Pedersen MG, Kenny LC, Abel KM, et al. Risk of affective disorders following prenatal exposure to severe life events: a Danish population-based cohort study. J Psychiatr Res. 2011 Jul;45(7):879–85.

91. Dalton SO, Laursen TM, Ross L, Mortensen PB, Johansen C. Risk for hospitalization with depression after a cancer diagnosis: a nationwide, population-based study of cancer patients in Denmark from 1973 to 2003. J Clin Oncol Off J Am Soc Clin Oncol. 2009 Mar 20;27(9):1440–5.

92. Andersen UA, Andersen M, Rosholm JU, Gram LF. Psychopharmacological treatment and psychiatric morbidity in 390 cases of suicide with special focus on affective disorders. Acta Psychiatr Scand. 2001 Dec;104(6):458–65.

93. Madsen KM, Hviid A, Vestergaard M, Schendel D, Wohlfahrt J, Thorsen P, et al. A population-based study of measles, mumps, and rubella vaccination and autism. N Engl J Med. 2002 Nov 7;347(19):1477–82.

94. Mouridsen SE, Hauschild K-M. Autism spectrum disorders in siblings of children with a developmental language disorder. Logoped Phoniatr Vocol. 2011 Dec;36(4):145–9.

95. Mouridsen SE, Hauschild K-M. A longitudinal study of autism spectrum disorders in individuals diagnosed with a developmental language disorder as children. Child Care Health Dev. 2009 Sep;35(5):691–7.

96. Danish Patient Safety Authority. Sundhedsdatastyrelsen - Klassifikationer (SKS) - ICD-8 klassifikationen [Internet]. 2019. Available from: https://sundhedsdatastyrelsen.dk/-/media/sds/filer/rammer-og-retningslinjer/klassisfikationer/sks-download/lukkede-klassifikationer/icd-8-klassifikation.txt

97. Daniels JL, Forssen U, Hultman CM, Cnattingius S, Savitz DA, Feychting M, et al. Parental psychiatric disorders associated with autism spectrum disorders in the offspring. Pediatrics. 2008 May;121(5):e1357-1362.

98. Jokiranta E, Brown AS, Heinimaa M, Cheslack-Postava K, Suominen A, Sourander A. Parental psychiatric disorders and autism spectrum disorders. Psychiatry Res. 2013 May 30;207(3):203–11.

99. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013 Sep;45(9):984–94.

100. Sydsjö G, Möller L, Lilliecreutz C, Bladh M, Andolf E, Josefsson A. Psychiatric illness in women requesting caesarean section. BJOG Int J Obstet Gynaecol. 2015 Feb;122(3):351–8.

101. McLaughlin KA, Gadermann AM, Hwang I, Sampson NA, Al-Hamzawi A, Andrade LH, et al. Parent psychopathology and offspring mental disorders: Results from the WHO World Mental Health Surveys. Br J Psychiatry. 2012 Apr;200(4):290–9.

102. Clausen TD, Bergholt T, Bouaziz O, Arpi M, Eriksson F, Rasmussen S, et al. Broad-Spectrum Antibiotic Treatment and Subsequent Childhood Type 1 Diabetes: A Nationwide Danish Cohort Study. PloS One. 2016;11(8):e0161654.

103. Clausen TD, Bergholt T, Eriksson F, Rasmussen S, Keiding N, L?kkegaard EC. Prelabor Cesarean Section and Risk of Childhood Type 1 Diabetes: A Nationwide Register-based Cohort Study. Epidemiology. 2016 Jul;27(4):547–55.

104. Østergaard SD, Larsen JT, Dalsgaard S, Wilens TE, Mortensen PB, Agerbo E, et al. Predicting ADHD by Assessment of Rutter's Indicators of Adversity in Infancy. Hay PJ, editor. PLOS ONE. 2016 Jun 29;11(6):e0157352.

105. Wilcox AJ. On the importance--and the unimportance--of birthweight. Int J Epidemiol. 2001 Dec;30(6):1233–41.

106. Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational age at birth. Am J Epidemiol. 2011 Nov 1;174(9):1062–8.

107. Reitan T, Callinan S. Changes in Smoking Rates Among Pregnant Women and the General Female Population in Australia, Finland, Norway, and Sweden. Nicotine Tob Res Off J Soc Res Nicotine Tob. 2017 Mar 1;19(3):282–9.

108. Ross CE. Sundhedsstyrelsen - Danskernes rygevaner 2018 [Internet]. 2019. Available from: https://www.sst.dk/da/Udgivelser/2019/Danskernes-rygevaner-2018

109. Visser SN, Danielson ML, Bitsko RH, Holbrook JR, Kogan MD, Ghandour RM, et al. Trends in the Parent-Report of Health Care Provider-Diagnosed and Medicated Attention-Deficit/Hyperactivity Disorder: United States, 2003–2011. J Am Acad Child Adolesc Psychiatry. 2014 Jan;53(1):34-46.e2.

110. Quinn PO, Madhoo M. A Review of Attention-Deficit/Hyperactivity Disorder in Women and Girls: Uncovering This Hidden Diagnosis. Prim Care Companion CNS Disord [Internet]. 2014 May 15 [cited 2019 Sep 10]; Available from: http://www.psychiatrist.com/pcc/article/pages/2014/v16n03/13r01596.aspx

111. Kassambara A. STHDA - Statistical tools for high-throughput data analysis - Cox Proportional-Hazards Model [Internet]. 2019. Available from: http://www.sthda.com/english/wiki/cox-proportional-hazards-model

112. Donovan SJ, Susser E. Commentary: Advent of sibling designs. Int J Epidemiol. 2011 Apr 1;40(2):345–9.

113. Hoffmann TJ, Windham GC, Anderson M, Croen LA, Grether JK, Risch N. Evidence of reproductive stoppage in families with autism spectrum disorder: a large, population-based cohort study. JAMA Psychiatry. 2014 Aug;71(8):943–51.

114. Getahun D, Rhoads GG, Demissie K, Lu S-E, Quinn VP, Fassett MJ, et al. In Utero Exposure to Ischemic-Hypoxic Conditions and Attention-Deficit/Hyperactivity Disorder. PEDIATRICS. 2013 Jan 1;131(1):e53–61.

115. Sjölander A, Zetterqvist J. Confounders, mediators or colliders – What types of shared covariates does a sibling comparison design control for?: Epidemiology. 2017 Mar;1.

116. Rehbinder EM, Lødrup Carlsen KC, Staff AC, Angell IL, Landrø L, Hilde K, et al. Is amniotic fluid of women with uncomplicated term pregnancies free of bacteria? Am J Obstet Gynecol. 2018 Sep;219(3):289.e1-289.e12.

117. Zhang T, Sidorchuk A, Sevilla-Cermeño L, Vilaplana-Pérez A, Chang Z, Larsson H, et al. Association of Cesarean Delivery With Risk of Neurodevelopmental and Psychiatric Disorders in the Offspring: A Systematic Review and Meta-analysis. JAMA Netw Open. 2019 Aug 28;2(8):e1910236.

118. Curran EA, O'Neill SM, Cryan JF, Kenny LC, Dinan TG, Khashan AS, et al. Research review: Birth by caesarean section and development of autism spectrum disorder and attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. J Child Psychol Psychiatry. 2015 May;56(5):500–8.

119. Slykerman RF, Thompson J, Waldie KE, Murphy R, Wall C, Mitchell EA. Antibiotics in the first year of life and subsequent neurocognitive outcomes. Acta Paediatr [Internet]. 2016 Oct [cited 2016 Nov 9]; Available from: http://doi.wiley.com/10.1111/apa.13613

120. Köhler-Forsberg O, Petersen L, Gasse C, Mortensen PB, Dalsgaard S, Yolken RH, et al. A Nationwide Study in Denmark of the Association Between Treated Infections and the Subsequent Risk of Treated Mental Disorders in Children and Adolescents. JAMA Psychiatry. 2019 Mar 1;76(3):271.

121. the SAWANTI Working Group, Łukasik J, Patro-Gołąb B, Horvath A, Baron R, Szajewska H. Early Life Exposure to Antibiotics and Autism Spectrum Disorders: A Systematic Review. J Autism Dev Disord. 2019 Sep;49(9):3866–76.

122. Curran EA, Khashan AS, Dalman C, Kenny LC, Cryan JF, Dinan TG, et al. Obstetric mode of delivery and attention-deficit/hyperactivity disorder: a sibling-matched study. Int J Epidemiol. 2016 Apr;45(2):532–42.

123. Curran EA, Dalman C, Kearney PM, Kenny LC, Cryan JF, Dinan TG, et al. Association Between Obstetric Mode of Delivery and Autism Spectrum Disorder: A Population-Based Sibling Design Study. JAMA Psychiatry. 2015 Sep 1;72(9):935.

124. O'Neill SM, Curran EA, Dalman C, Kenny LC, Kearney PM, Clarke G, et al. Birth by Caesarean Section and the Risk of Adult Psychosis: A Population-Based Cohort Study. Schizophr Bull. 2016 May;42(3):633–41.

125. Oberg S, Cnattingius S, Sandin S, Lichtenstein P, Iliadou A. Birth Weight-Breast Cancer Revisited: Is the Association Confounded by Familial Factors? Cancer Epidemiol Biomarkers Prev. 2009 Sep 1;18(9):2447–52.

126. Bini SA, Chan PH, Inacio MCS, Paxton EW, Khatod M. Antibiotic cement was associated with half the risk of re-revision in 1,154 aseptic revision total knee arthroplasties. Acta Orthop. 2016 Jan 2;87(1):55–9.

127. Sheth DS, Cafri G, Paxton EW, Namba RS. Bilateral Simultaneous vs Staged Total Knee Arthroplasty: A Comparison of Complications and Mortality. J Arthroplasty. 2016 Sep;31(9):212–6.

128. Cafri G, Paxton EW, Love R, Bini SA, Kurtz SM. Is There a Difference in Revision Risk Between Metal and Ceramic Heads on Highly Crosslinked Polyethylene Liners? Clin Orthop Relat Res. 2017 May;475(5):1349–55.

129. Goetgeluk S, Vansteelandt S. Conditional generalized estimating equations for the analysis of clustered and longitudinal data. Biometrics. 2008 Sep;64(3):772–80.

130. Pearce JMS. Kanner's infantile autism and Asperger's syndrome. J Neurol Neurosurg Psychiatry. 2005 Feb 1;76(2):205–205.

131. Simon GE, Rossom RC, Beck A, Waitzfelder BE, Coleman KJ, Stewart C, et al. Antidepressants Are Not Overprescribed for Mild Depression. J Clin Psychiatry. 2015 Dec 23;76(12):1627–32.

132. Quinlan JD, Murphy NJ. Cesarean delivery: counseling issues and complication management. Am Fam Physician. 2015 Feb 1;91(3):178–84.

133. Cantarero-Arévalo L, Hallas MP, Kaae S. Parental knowledge of antibiotic use in children with respiratory infections: a systematic review. Int J Pharm Pract. 2017 Feb;25(1):31–49.

134. Østergaard SD. Taking Facebook at face value: why the use of social media may cause mental disorder. Acta Psychiatr Scand. 2017 Nov;136(5):439–40.

135. Berry N, Emsley R, Lobban F, Bucci S. Social media and its relationship with mood, self-esteem and paranoia in psychosis. Acta Psychiatr Scand. 2018 Dec;138(6):558–70.

136. Boers E, Afzali MH, Newton N, Conrod P. Association of Screen Time and Depression in Adolescence. JAMA Pediatr. 2019 Sep 1;173(9):853.

137. Ra CK, Cho J, Stone MD, De La Cerda J, Goldenson NI, Moroney E, et al. Association of Digital Media Use With Subsequent Symptoms of Attention-Deficit/Hyperactivity Disorder Among Adolescents. JAMA. 2018 Jul 17;320(3):255.

138. Hoffman BL, Felter EM, Chu K-H, Shensa A, Hermann C, Wolynn T, et al. It's not all about autism: The emerging landscape of anti-vaccination sentiment on Facebook. Vaccine. 2019 Apr;37(16):2216–23.

139. Thomas R, Sanders S, Doust J, Beller E, Glasziou P. Prevalence of attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. Pediatrics. 2015 Apr;135(4):e994-1001.

140. Tripp G, Luk SL, Schaughency EA, Singh R. DSM-IV and ICD-10: a comparison of the correlates of ADHD and hyperkinetic disorder. J Am Acad Child Adolesc Psychiatry. 1999 Feb;38(2):156–64.

141. Thapar A, Cooper M. Attention deficit hyperactivity disorder. Lancet Lond Engl. 2016 Mar 19;387(10024):1240–50.

142. Lai M-C, Lombardo MV, Baron-Cohen S. Autism. Lancet Lond Engl. 2014 Mar 8;383(9920):896–910.

143. Hollowood K, Melnyk S, Pavliv O, Evans T, Sides A, Schmidt RJ, et al. Maternal metabolic profile predicts high or low risk of an autism pregnancy outcome. Res Autism Spectr Disord. 2018 Dec;56:72–82.

144. Zeliadt N. Spectrum - Brain Imaging Studies Seek Signs of Autism Before Birth [Internet]. 2017. Available from: https://www.spectrumnews.org/news/brain-imaging-studies-seek-signs-autism-birth/

145. Sun H, Chen Y, Huang Q, Lui S, Huang X, Shi Y, et al. Psychoradiologic Utility of MR Imaging for Diagnosis of Attention Deficit Hyperactivity Disorder: A Radiomics Analysis. Radiology. 2018 May;287(2):620–30.

146. Deoni SC, Adams SH, Li X, Badger TM, Pivik RT, Glasier CM, et al. Cesarean Delivery Impacts Infant Brain Development. AJNR Am J Neuroradiol. 2019 Jan;40(1):169–77.

147. Jensen CM, Steinhausen H-C. Time Trends in Lifetime Incidence Rates of First-Time Diagnosed Bipolar and Depressive Disorders Across 16 Years in Danish Psychiatric Hospitals: A Nationwide Study. J Clin Psychiatry. 2016 Dec;77(12):e1570–5.

148. Petersen AH, Lange T. What is the causal interpretation of sibling comparison designs? Epidemiol Accept.

Paper I

The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry **:* (2018), pp **–** doi:10.1111/jcpp.12961

Investigating the effects of cesarean delivery and antibiotic use in early childhood on risk of later attention deficit hyperactivity disorder

Paul Bryde Axelsson,[1] Tine Dalsgaard Clausen,[1] Anne Helby Petersen,[2] Ida Hageman,[3] Anja Pinborg,[4] Lars Vedel Kessing,[3] Thomas Bergholt,[5] Steen Christian Rasmussen,[6] Niels Keiding,[2] and Ellen Christine Leth Løkkegaard[1]

[1]Department of Gynaecology and Obstetrics, Nordsjaellands Hospital, University of Copenhagen, Copenhagen, Denmark; [2]Department of Public Health, Section of Biostatistics, University of Copenhagen, Copenhagen, Denmark; [3]Psychiatric Center Copenhagen, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [4]Department of Obstetrics and Gynaecology, Copenhagen University Hospital, Hvidovre Hospital, Hvidovre, Denmark; [5]Department of Obstetrics, University of Copenhagen, Rigshospitalet, Copenhagen, Denmark; [6]Department of Clinical Microbiology, Copenhagen University Hospital, Hvidovre Hospital, Hvidovre, Denmark

Background: Increasing attention deficit hyperactivity disorder (ADHD) incidence has been proposed to be caused by factors influencing microbiota in early life. We investigated the potential causality between ADHD and two surrogate markers for changes in children's microbiota: birth delivery mode and early childhood antibiotic use. **Method:** This population-based, prospective cohort study linked nationwide registers of data for native Danish singleton live births in Denmark from 1997 to 2010. Exposure variables were delivery mode and antibiotic use during the first 2 years of life. The main outcome measure was ADHD diagnosis or redeemed ADHD medication prescriptions. For statistical analysis, we used both advanced sibling models and a more traditional approach. **Results:** We included 671,592 children, followed from their second birthday in the period 1999–2014 for 7,300,522 person-years. ADHD was diagnosed in 17,971. In total, 17.5% were born by cesarean delivery, and 72% received antibiotic treatment within their first 2 years of life. In the adjusted between-within sibling survival model, mode of delivery or antibiotics had no effect on ADHD when compared with vaginal delivery or no antibiotic treatment as hazard ratios were 1.09 (95% confidence interval 0.97–1.24) for intrapartum cesarean, 1.03 (0.91–1.16) for prelabor cesarean, 0.98 (0.90–1.07) for penicillin, and 0.99 (0.92–1.06) for broader spectrum antibiotics. In a sibling-stratified Cox regression, intrapartum cesarean was associated with increased ADHD risk, but other exposures were not. In a descriptive, nonstratified Cox model, we found increased risk for ADHD for all exposures. **Conclusions:** Detailed family confounder control using the superior between-within model indicates that cesarean delivery or use of antibiotics during the first 2 years of life does not increase ADHD risk. Therefore, our study suggests that changes in children's microbiota related to cesarean delivery or antibiotic use, do not cause ADHD. **Keywords:** Attention deficit hyperactivity disorder; antibiotics; cesarean section; microbiota; sibling relations.

Introduction

Attention deficit hyperactivity disorder (ADHD) is a prevalent disorder in industrialized countries, and its incidence has increased in the past decade. This increase is likely at least partially due to changes in diagnostic practices, parental awareness, and improved treatment possibilities (Jensen & Steinhausen, 2015). Although ADHD is one of the most heritable psychiatric conditions with a heritability index of 0.76 (Faraone et al., 2005), many factors may influence risk, with familial environment probably being the second most important (Johnston, Mash, Miller, & Ninowski, 2012; Larsson, Sariaslan, Långström, D'Onofrio, & Lichtenstein, 2014). A prevailing hypothesis suggests that an unfavorable infant gut microbiota can affect brain development through epigenetic mechanisms (Mayer, Tillisch, & Gupta, 2015; Stilling, Bordenstein, Dinan, & Cryan, 2014). However, no direct evidence supports a relationship between gut microbiota composition and ADHD symptoms (Cenit, Nuevo, Codoñer-Franch, Dinan, & Sanz, 2017). In lieu of resource-demanding microbiota studies, information from a large national cohort can help shed light on this hypothesis.

Previous studies suggest that the gut microbiota of infants born by cesarean delivery is distinctly different from those born by vaginal delivery and that these changes often persist months after birth (van Best, Hornef, Savelkoul, & Penders, 2015). Data also suggest that antibiotic treatment adversely affects gut bacteria (van Best et al., 2015) in particular antibiotics with a broader spectrum (Korpela et al., 2016).

In this context, we assessed the possible causal association between cesarean delivery and early childhood use of antibiotics and the risk of developing ADHD later in life using three different survival models: a Cox regression model, a sibling-stratified Cox regression model, and a between-within survival model for siblings.

Method

Design

The investigation was designed as a population-based historically prospective cohort study. Data were obtained by register linkage via the unique personal registration numbers assigned to all Danes at birth. After linkage, the data were de-identified

Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA

to ensure data safety. We merged data from seven Danish nationwide registers (see Table S1).

Study population and follow-up

The cohort included 671,592 Danish singleton children, not diagnosed with ADHD and still alive and living in Denmark on their second birthday (Figure 1). These children were distributed among 442,002 full-sibling families, sharing both a Danish mother and a Danish father.

The children were censored at time of death, emigration, ICD-10 (International Classification of Diseases 10th revision) diagnosis of an Organic Mental Disorder, Mental Disorders due to Substance Abuse or Psychotic Disorders, or end of follow-up by December 31, 2014 (Figure 1), whichever occurred first. Parental psychiatric status was based on psychiatric diagnostic codes since 1969, since 1997 combined with information about psychiatric prescriptions. (see Table S2).

Exposure

Cesarean delivery was categorized as being either prelabor or intrapartum (see Table S2). Vaginal delivery was used as a reference. From 1997, information on outpatient-redeemed antibiotic prescriptions was available at the individual level. Antibiotic treatment during the first 2 years of life was hierarchically classified into either no antibiotics (reference), exclusively beta-lactamase–sensitive penicillin (hereafter called penicillin), or an antibiotic with a broader spectrum than penicillin (hereafter called broader spectrum antibiotics) (see Table S2). The reasoning behind grouping antibiotics together, is available in Appendix S4. Children exposed to broader

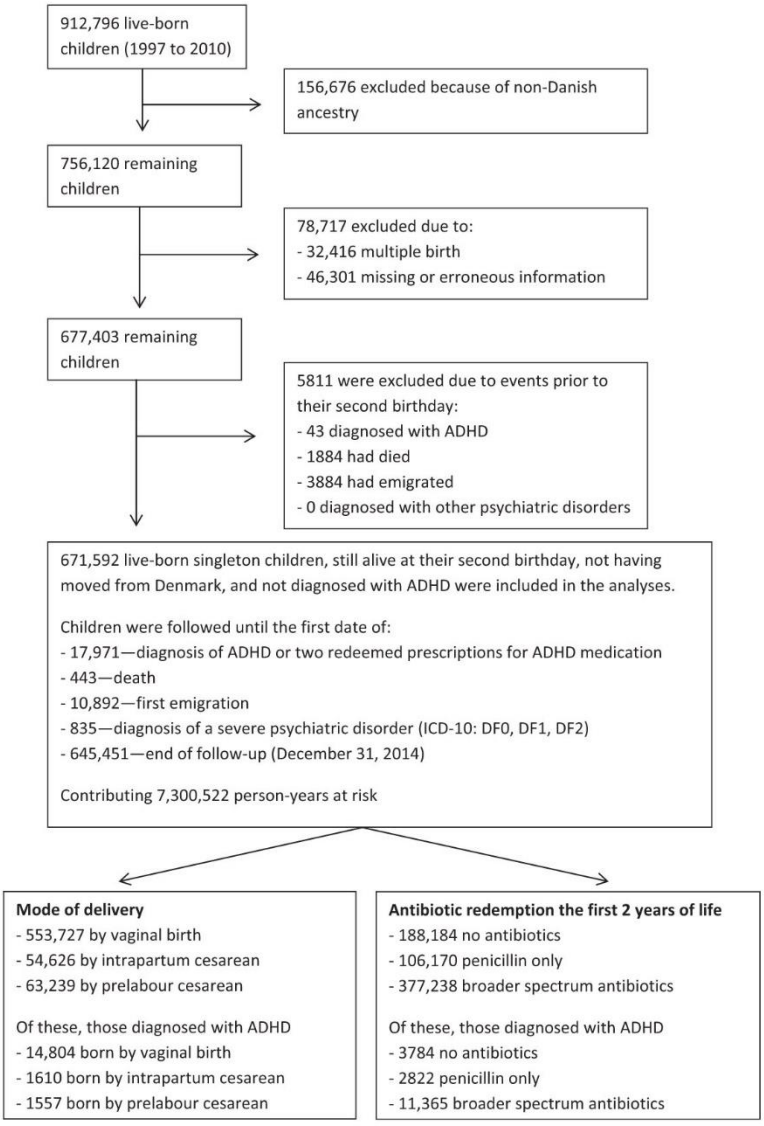


Figure 1 Inclusion, exclusion, and censoring. Shown are the number of children excluded, number of children in the final study population, number of events, and distribution of attention deficit hyperactivity disorder (ADHD) between exposure by either cesarean delivery or antibiotics

Table 1 Overview of the progression of confounder adjustment used in the models

Confounder adjustment models by covariate groups					
Variables	Unadjusted	Social	Obstetric	Maternal microbiota	Parental attention deficit hyperactivity disorder (ADHD)
Confounder adjustment level	1	2	3	4	5
Childhood antibiotics use	X	x	x	x	x
Mode of delivery	X	x	x	x	x
Maternal age at birth		x	x	x	x
Parental age difference		(x)	(x)	(x)	(x)
Parental education		x	x	x	x
Maternal marital status		x	x	x	x
Maternal smoking		x	x	x	x
Infant sex			x	x	x
5-minute Apgar score			x	x	x
Instrument use at delivery			x	x	x
Use of CPAP or ventilator			x	x	x
Asphyxia			x	x	x
Parental epilepsy			x	x	x
Preeclampsia or hypertension			x	x	x
Gestational diabetes			x	x	x
Parity			x	x	x
Induction of labor			x	x	x
Induction of contractions			x	x	x
Maternal antibiotics use during the pregnancy				x	x
Maternal infections during the pregnancy				x	x
Parental ADHD history					(x)

Confounder adjustment level 1 refers to the minimal model, including only the exposure variables (childhood antibiotic use and mode of delivery), but no confounders. Similarly, confounder adjustment level 4 corresponds to the fully adjusted model where all available and relevant confounders were included in the sibling analyses, and level 5 is the fully adjusted model for the descriptive Cox models. Parentheses indicate variables not used in the sibling models.

spectrum antibiotics were therefore not classified as exposed in the exclusively penicillin-treated group, even if they had received penicillin.

Outcome

The primary outcome was any ADHD case, defined as the assignment of an ADHD diagnosis, attention deficit disorder (ADD) diagnosis, or at least two redeemed prescriptions for ADHD/ADD medication on separate dates (see Table S2). Both patients admitted to the hospital and those treated in an outpatient hospital clinic were included, as were primary and secondary discharge diagnoses. Emergency room contacts and referral diagnoses were not included. We modeled time to the first event that constituted an ADHD case.

Confounder adjustment

Confounder adjustment was based on previously identified possible confounders for ADHD as well as theoretical reflections regarding factors hypothesized to confound the relationship between delivery mode, antibiotic use, and ADHD. Confounders and adjusted model levels are presented in Table 1. Codes for covariates and information on other variables are in Tables S2, S3A and B.

Statistical analyses

In the statistical analyses, we applied two sibling matched survival models: the stratified Cox model and the between-within gamma-Cox model. The first model targets within-family effects by allowing each family to have its own baseline hazard function. The between-within model keeps the singular baseline hazard assumption of the original Cox proportional hazards model but includes a family level shared random effect (frailty term), as well as a fixed between-family effect and a fixed within-family effect. This structure allows for consistent estimates of the effects of interest, namely the within-family effects (Sjölander, Lichtenstein, Larsson, & Pawitan, 2013). Moreover, simulation studies have shown that it has greater statistical power than the usual sibling-stratified Cox model (Sjölander et al., 2013).

To make the results reported here comparable with other studies, we also applied the usual, nonstratified Cox proportional hazards model. We refer to this model as the 'descriptive Cox model' to emphasize that it does not target the within-family effects of the sibling models but rather the descriptive parameters regarding population tendencies. Robust ('sandwich') standard error estimators were used to adjust for familial clustering in the variance structure for the descriptive Cox model. For further information about the statistical models, we refer to Appendix S1.

In all models, the underlying time was offspring age. Missing values were handled by complete case analysis. Interaction effects between the two exposure variables were assessed in the between-within model. Families were defined by full siblingship (both parents are the same for all siblings) for the clustering in the descriptive Cox models and the family structures in the stratified models and the between-within models.

A variety of sensitivity analyses were performed to assess the robustness of the overall conclusions with respect to central modeling choices. Because the fully adjusted between-within model serves as the primary option, the sensitivity analyses were conducted as extensions of this model whenever possible. The details of these sensitivity analyses are available in Appendices S2 and S3, available online. We considered the sensitivity toward alterations of the following six aspects of the primary model: definitions of exposure variables, ADHD definition, additional adjustment for confounders with a lot of

missing information (body mass index and fetal presentation), siblingship definition, adjustment level definitions, and time trends in ADHD.

All statistical analyses were performed in R, version 3.3.1 (The R Foundation, Vienna, Austria), open source, using the *survival* package with additional programming by the authors (available upon request).

Results

In the period of our cohort's births, from 1997 to 2010, cesarean deliveries increased by more than 60%, mostly because of a prelabor cesarean delivery increase of about 90% (Clausen et al., 2016). The pattern of antibiotic prescriptions for infants younger than age 2 years did not change discernibly in the same period (see Figure S1, available online). The incidence of ADHD did rise greatly for the same cohort but flattened out after 2011 (Figure 2). We have summarized outcome status by the end of follow-up, stratified by birth year in Appendix S5. In total, we have 17,971 cases of ADHD in the data.

In total, 17.5% of the cohort was born by cesarean delivery, and 72% received some type of antibiotic treatment within their first 2 years of life (see Tables S3A and B). Children born by cesarean delivery had a lower Apgar score, and their mothers were more likely to have gestational diabetes, preeclampsia, or hypertension. Children born by prelabor cesarean delivery were more likely to have received continuous positive airway pressure (CPAP) and/or have been respirator treated. They were also more often born to married women, women of advanced age, women with epilepsy, and women who had given birth before. Children born by intrapartum cesarean delivery were more often born by nonmarried women, with an infection during pregnancy, had labor or contractions induced, and were primiparous. Children who within the first 2 years of life received antibiotics were more often male and delivered by cesarean delivery. They had a higher Apgar score and younger parents with a shorter education. Their mothers were more often primiparous and more often had epilepsy, gestational diabetes, or ADHD. Their mothers also were more often smokers and had more often received antibiotics during pregnancy or been admitted to the hospital with infection during pregnancy. Children who specifically received broader spectrum antibiotics also received CPAP and/or were treated with a respirator more often.

Between-within models

The estimated effects of antibiotics use and cesarean delivery showed statistically nonsignificant effects with full confounder adjustment and slight significance for

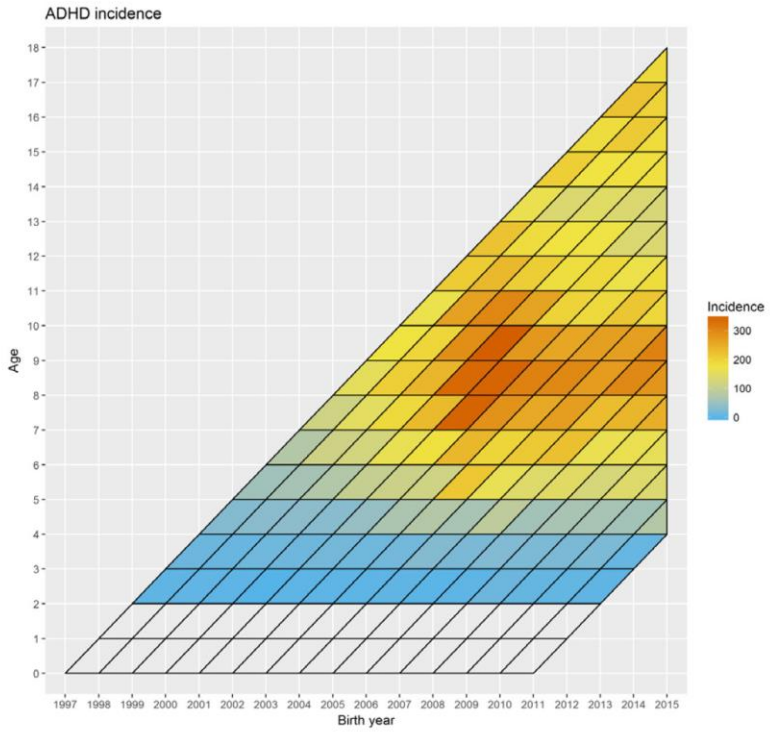


Figure 2 A Lexis heatmap of attention deficit hyperactivity disorder (ADHD) incidence, defined as either a diagnosis or a redeemed ADHD medication prescription, in one-year birth year and age intervals. There was an increase in incidence for birth-year cohorts 1999–2006, with higher age at diagnosis for earlier cohorts, a sharp increase in 2008 lasting until 2011, when it slowly leveled out. The incidence is therefore not stable over diagnostic-year, birth-year, or age

the models with no or little confounder adjustment in the between-within model (Figure 3). Generally, all the exposure effect estimates indicated null effects. For exact estimates for exposure effects, see Table S4.

The exposure effect estimates showed little sensitivity toward the adjustment for maternal antibiotic use in pregnancy, as evident from the similar effect estimates found in adjustment levels 3 and 4 (Figure 3). There were nonsignificant effects of most included covariates, except young maternal age, offspring sex, CPAP, and ventilator treatment (see Figure S2, available online).

Stratified Cox models

In the sibling-stratified Cox models, use of antibiotics in early childhood was not associated with statistically significant effects when full confounder adjustment was used and with only minor statistically significant effects when it was not. Compared to children born vaginally, the risk of ADHD was comparable among children born by prelabor cesarean delivery, whereas children born by intrapartum cesarean delivery had a significantly increased hazard ratio (HR) for ADHD (HR 1.21 (1.01–1.45) in the fully adjusted model (Figure 3).

The exposure effect estimates were again generally insensitive toward adjustment for maternal antibiotic use in pregnancy, just as in the between-within models. Because of the structure of the stratified Cox model, only some families provided information to its estimation (1.5%, see Table S5); which were those families with discordant explanatory variables and at least one ADHD event, experienced for a sibling

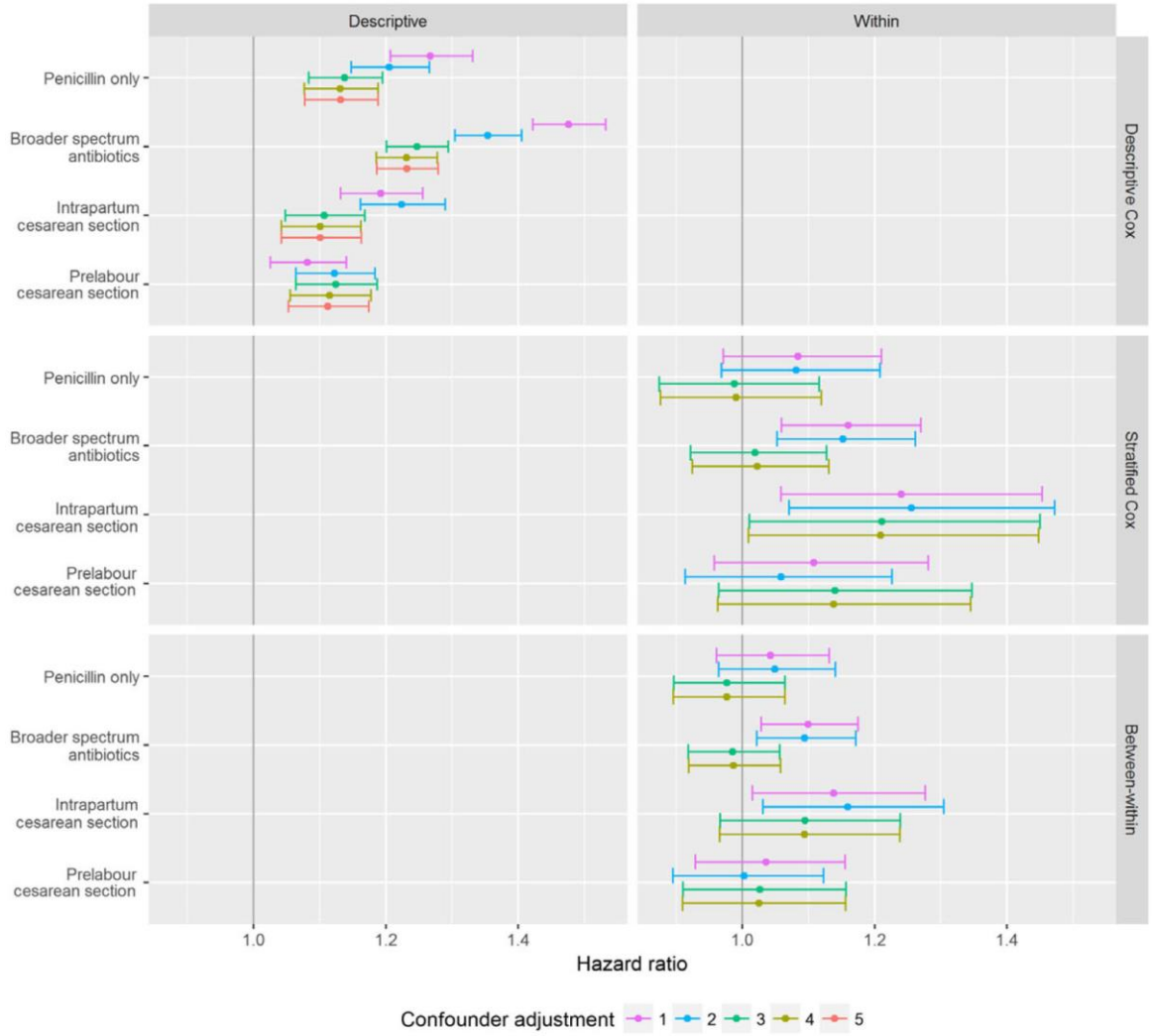


Figure 3 Main results from the three different statistical models. Level 4 adjustment was performed only on the descriptive Cox model (refer to Table 1 for confounder adjustment levels). Descriptive and Within effects cannot be compared and are thus depicted in separate columns. The parameter estimates are accompanied by 95% confidence intervals. The confidence intervals are point-wise and thus do not account for multiple testing

whose follow-up time was exceeded by a different sibling.

Descriptive cox models

There were statistically significant effects of all exposure variables in all of the descriptive Cox models, though the effects were smaller when more confounder adjustment was applied. In the fully adjusted model, ADHD risk was increased for children born by cesarean delivery compared to children born vaginally [HR 1.10 (95% confidence interval 1.04–1.16) and 1.11 (1.05–1.17)] for intrapartum and prelabor cesarean delivery, respectively. For antibiotics use, we found moderately larger effects when broader spectrum antibiotics were used [HR 1.23 (1.19–1.28)] relative to children not treated with antibiotics at all. For children treated exclusively with penicillin, we found a small but statistically significantly increased risk of ADHD [HR 1.13 (1.08–1.19)], compared with children who did not receive any antibiotics (Figure 3).

The parameter estimates were not sensitive toward adjustment for maternal antibiotic use in the pregnancy. Similarly, whether or not parental ADHD history information was included in the model did not change the conclusions regarding parameter estimates.

Robustness of the results

There were no significant interaction effects between the exposure variables 'cesarean delivery' and 'antibiotics' ($p = .94$). None of the sensitivity analysis models mentioned previously revealed any results that would have led to changes to our conclusions, except for one: When sex was added in the second adjustment on its own, the estimates changed so that they became close to those of the fully adjusted models. With further adjustment for the remaining covariates, however, the antibiotic exposure estimates remained relatively unchanged, and the delivery mode estimates changed only very little.

We also wish to draw attention to the sensitivity analyses regarding how the antibiotic exposure was defined. Even when using specific antibiotic groups there were no significant exposure effects (Appendix S3, available online).

The effects of time trends were evaluated, and there were clear, systematic effects of diagnosis year on ADHD occurrence and less clear effects of birth year. However, the estimates of the exposure effects were completely insensitive to whether year effects were included in the models or not. The full report on time trends in ADHD diagnoses is in Appendix S2.

Discussion

In a population of a little over 670,000 children, neither cesarean delivery nor antibiotic treatment in the first 2 years of life was associated with ADHD in the fully adjusted between-within survival model for siblings. In the fully adjusted sibling-stratified Cox model, we found that only exposure by intrapartum cesarean delivery produced a statistically significantly slightly increased hazard of ADHD, compared to vaginal delivery. Generally, the confidence intervals were considerably wider than they were in the between-within model. Point estimates for the between-within model were also consistently slightly closer to null values. The fully adjusted, descriptive Cox model gave different results, with all exposure effect estimates being statistically significant. However, the descriptive Cox model ignores unobserved confounding within families, so direct comparison of the estimated effects from sibling- and nonsibling models is not meaningful.

Please note that the efficiency of the three models can readily be compared by comparing the width of the confidence intervals they each produce. Thus, the mere 1.5% of children used for estimating the stratified Cox model is not far from the effective sample sizes utilized by the two other models: otherwise, we would have encountered larger differences in confidence interval widths. We found narrower confidence bands for all estimates in the between-within model compared to the stratified Cox model, indicating greater precision with the former. This finding corresponds well with the theoretical results regarding the between-within model as a more precise modeling tool for estimating the within-family effects that are also the target of the sibling-stratified Cox model.

Unobserved familial factors seem to drive the observed association between birth mode and antibiotics, and increased risk of ADHD. But even when adjusting for highly predictive factors for ADHD risk in the descriptive Cox model, the adjusted estimate will never be sufficiently adjusted as the information in the variables is incomplete. For example, it is very probable that information on parental psychiatric diagnoses is extensively lacking, as most parents with relevant symptoms would not have been diagnosed at the time when their children were born. Furthermore, some familial confounders are simply difficult to measure and record, and therefore do not figure in national registers. Other studies looking into the harmful effects of cesarean delivery and antibiotics on infant health, have made similar efforts to adjust for familial factors when using the descriptive Cox model. We generally advice caution in drawing causal associations from such studies.

To address the possibility of microbiota causally affecting ADHD risk, one would expect broader spectrum antibiotics to result in a larger HR compared to penicillin. The descriptive Cox model yielded this result, but in the stratified models and the between-within survival models, the association disappeared. In a sensitivity analysis, it made no difference to overall conclusions whether the

antibiotic exposure occurred in the first or the second year of life, as long as antibiotic use in both of the first 2 years of life was adjusted for. If the vaginal microbiota played a role in the etiology of ADHD, we would also expect smaller risk estimates for infants who have come in contact with the vaginal microbiota prior to cesarean delivery. We explored this possibility in another sensitivity analysis by changing the cesarean delivery exposure categories to either rupture of membranes or intact membranes prior to cesarean delivery. This change did not shift our conclusion of no causal association between cesarean delivery and ADHD. Furthermore, no interaction effect was found between cesarean delivery and antibiotics, making the potential microbiota alterations even less likely to be relevant.

Previous studies have highlighted the importance of social factors for ADHD risk, which we included in our second adjustment level (Østergaard et al., 2016). But when we further adjusted for obstetric factors and the child's sex in the second adjustment the estimates for intrapartum cesarean delivery and antibiotic exposure became statistically insignificant. To explore whether the difference between first and second confounder adjustment levels was mainly due to the inclusion of offspring sex in the latter, we performed a post hoc sensitivity analysis where the sex variable was added to first adjustment level. The amendment of the sex variable produced exposure effect estimates very similar to those of subsequent adjustment levels, indicating that obstetric factors may add no additional information in the sibling models (see SFigure C3).

Elaborating on the possible importance of non-shared confounders, adjusting for the mother's antibiotic treatment or infections in pregnancy did not seem to change the exposure effect estimates in any of the analyses. Although this result does not rule out an association with ADHD, it does not confound the relationship between cesarean delivery or antibiotics in early childhood and ADHD and thus further lends support to the conclusion that the mother's microbiota does not play a role in offspring ADHD.

Birth weight and gestational age may interact with potential causal pathways in several ways and are considered as *colliders* between certain exposures and outcomes (Wilcox, 2001; Wilcox Allen, Weinberg, & Basso, 2011) which we also propose to be the case for delivery mode, antibiotic treatment, and ADHD. The information represented by birth weight and gestational age is the summation of several unobserved pregnancy and obstetric factors. Therefore, they are not confounders to be adjusted for, which would produce biased estimates of the exposure effects (Clausen et al., 2016).

The present large, population-based cohort study is the first to use a sibling model analysis to explore the potential effects of antibiotic treatment in early life on the risk of a later diagnosis of ADHD. It is also the first study to look for an interaction effect between early life antibiotic use and mode of delivery on ADHD development. Furthermore, this study is the first to use the between-within survival model to evaluate whether the risk of ADHD depends on mode of delivery and to compare the results to those from more traditional models.

A strength of the study is that the validity of ADHD diagnoses in the Danish registers has previously been found to be high (Mohr-Jensen, Vinkel Koch, Briciet Lauritsen, & Steinhausen, 2016). Generally, the Danish registers encompass the whole population of Denmark and have a high degree of completeness (Lynge, Sandegaard, & Rebolj, 2011).

Another strength to the study is that we have explored the robustness of our findings in various sensitivity analyses. We found that even if we changed the detail of the antibiotic exposure to include information for the five individual antibiotics that were responsible for over 98% of all antibiotic prescriptions, it did not change our conclusions. Similarly we explored the effect of changing the definition of ADHD to reflect differences in case severity, but without changes to our conclusions.

A limitation to our study is that we could not look into the dose–response relationship between antibiotic treatment and the risk of ADHD, as we did not have the relevant data. However, in the light of our overall null-finding of any effect of antibiotic treatment on ADHD risk, we find it improbable that there is an increased risk with higher doses or multiple courses of antibiotics.

We restricted attention to native Danish children. Although slightly reducing representativity, we thereby obtained a homogeneous population free of the problem of information *missing not at random* among immigrants.

We caution against comparing the rate of ADHD cases in this study population to the reported prevalence in other studies, as it should be done using the same age groups. We have discussed this topic further in Appendix S5.

The largest published cohort study on the association between cesarean delivery and ADHD risk conducted analyses using both descriptive and sibling-stratified Cox models (Curran et al., 2016). Their results were much in line with our own findings for the descriptive and stratified Cox models: significantly higher risk for both planned (prelabor) and emergency (intrapartum) cesarean delivery compared to vaginal delivery in the descriptive model, but only exposure by emergency cesarean delivery remained statistically significant in the stratified model. Although the study was important for the discussion of causality between cesarean delivery and ADHD, concern over a decrease in association due to imprecision inherent in the stratified Cox model necessitated further attention. The same author group had previously included two matched case–control studies of ADHD risk by delivery mode in a systematic review, where the

pooled adjusted odds ratio estimate was insignificant (Curran et al., 2015). We believe that our results and the precision of the method used strongly support that cesarean delivery (or treatment by antibiotics in early childhood) is not associated with increased risk of ADHD.

The relationship between antibiotic use and ADHD risk has previously been tested in a case–control study of 493 children followed for 11 years. Infant antibiotic use was reported by the mother, and antibiotic use during the first year was associated with higher teacher- and parent-reported symptoms of ADHD at age 11 (Slykerman et al., 2017). These results might suffer from recall bias, not differentiating between different types of antibiotics, lack of confounder control, and 43% loss to follow-up.

In conclusion, we find that cesarean delivery and antibiotic treatment have no relation to risk of developing ADHD. We found no indication that potential changes in gut microbiota due to cesarean delivery or antibiotics use have any influence on ADHD risk. The between-within model emerged as superior, giving more precise estimates than the sibling-stratified Cox model. Unobserved familial confounding renders the descriptive Cox model inadequate for estimating potentially causal effects; its very precise estimates are thus not suited for this purpose because they are not precise around the parameter that we actually wish to estimate.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Statistical models.
Appendix S2. Time trends in ADHD diagnoses.
Appendix S3. Supplementary analysis.
Appendix S4. Antibiotic exposure.
Appendix S5. Age specific ADHD-counts.
Appendix S6. The RECORD statement – checklist of items extended from the STROBE statement.
Figure S1. Prevalence of redeemed antibiotics prescriptions.
Figure S2. Parameter estimates for all effects in the between-within model.
Table S1. Information on linked registers in our study.
Table S2. Specification of registers and codes used for defined variables.
Table S3A. Background characteristics, dependent on delivery mode.
Table S3B. Background characteristics, dependent on antibiotic treatment.
Table S4. Fully adjusted main effect estimates.
Table S5. Informative families and observations for the stratified Cox model.

Acknowledgements

The study was supported by grants from the Capital Region Denmark Research Fund, the Capital Region Denmark PhD-start Fund, the Jascha Fund, the Tvergaard Fund and the Gangsted Fund. The funders had no role in the analyses interpretation of results or the writing of this manuscript. The authors are fully independent from influence by the funders. We would also like to acknowledge the editing work done by San Francisco Edit for this manuscript. LVK has within the preceding 3 years been a consultant for Sunovion. All other authors have declared that they have no competing or potential conflicts of interest.

Correspondence

Paul Bryde Axelsson, Nordsjaellands Hospital, Hilleroed, c/o Gynaekologisk-Obstetrisk Afdeling, Dyrehavevej 29, 3400 Hilleroed, Denmark; Email: paulbryde@gmail.com

Key points

- A small observational cohort study found that antibiotic treatment during the first year of life increased the risk of ADHD symptoms. A large national cohort study found that emergency but not planned cesarean delivery was associated with an increased risk of ADHD, in a Cox regression analysis stratified by siblings.
- Using information from siblings, our large national cohort study shows there is no causal association between cesarean delivery nor treatment with either penicillin or broader spectrum antibiotics in the first two years of life, on the risk of developing ADHD later in life. The between-within sibling model may be advantageous compared to the Cox regression analysis stratified by siblings when studying potential causal associations based on observational data.
- Parents and clinicians faced with decisions for delivery mode and antibiotics use for young children, should not be concerned with increased ADHD risk. The apparent increased risk found when not considering information from siblings, can be explained by shared familial factors.

References

van Best, N., Hornef, M.W., Savelkoul, P.H.M., & Penders, J. (2015). On the origin of species: Factors shaping the establishment of infant's gut microbiota. *Birth Defects Research Part C, Embryo Today: Reviews*, *105*, 240–251.

Cenit, M. C., Nuevo, I. C., Codoñer-Franch, P., Dinan, T. G., & Sanz, Y. (2017). Gut microbiota and attention deficit hyperactivity disorder: New perspectives for a challenging condition. *European Child & Adolescent Psychiatry.*, *26*, 1081–1092.

Clausen, T.D., Bergholt, T., Eriksson, F., Rasmussen, S., Keiding, N., & Løkkegaard, E.C. (2016). Prelabor cesarean

section and risk of childhood type 1 diabetes: A nationwide register-based cohort study. *Epidemiology*, *27*, 547–555.
Curran, E.A., Khashan, A.S., Dalman, C., Kenny, L.C., Cryan, J.F., Dinan, T.G., & Kearney, P.M. (2016). Obstetric mode of delivery and attention-deficit/hyperactivity disorder: A sibling-matched study. *International Journal of Epidemiology*, *45*, 532–542.
Curran, E.A., O'Neill, S.M., Cryan, J.F., Kenny, L.C., Dinan, T.G., Khashan, A.S., & Kearney, P.M. (2015). Research Review: Birth by caesarean section and development of autism spectrum disorder and attention-deficit/hyperactivity disorder: A systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry*, *56*, 500–508.
Faraone, S.V., Perlis, R.H., Doyle, A.E., Smoller, J.W., Goralnick, J.J., Holmgren, M.A., & Sklar, P. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *57*, 1313–1323.
Jensen, C. M., & Steinhausen, H.-C. (2015). Time trends in incidence rates of diagnosed attention-deficit/hyperactivity disorder across 16 years in a nationwide Danish registry study. *The Journal of Clinical Psychiatry*, *76*, e334–e341.
Johnston, C., Mash, E.J., Miller, N., & Ninowski, J.E. (2012). Parenting in adults with attention-deficit/hyperactivity disorder (ADHD). *Clinical Psychology Review*, *32*, 215–228.
Korpela, K., Salonen, A., Virta, L.J., Kekkonen, R.A., Forslund, K., Bork, P., & de Vos, W.M. (2016). Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nature Communications*, *7*, 10410.
Larsson, H., Sariaslan, A., Långström, N., D'Onofrio, B., & Lichtenstein, P. (2014). Family income in early childhood and subsequent attention deficit/hyperactivity disorder: A quasi-experimental study. *Journal of Child Psychology and Psychiatry*, *55*, 428–435.
Lynge, E., Sandegaard, J.L., & Rebolj, M. (2011). The Danish national patient register. *Scandinavian Journal of Public Health*, *39*(7 Suppl), 30–33.
Mayer, E.A., Tillisch, K., & Gupta, A. (2015). Gut/brain axis and the microbiota. *The Journal of Clinical Investigation*, *125*, 926–938.
Mohr-Jensen, C., Vinkel Koch, S., Briciet Lauritsen, M., & Steinhausen, H.-C. (2016). The validity and reliability of the diagnosis of hyperkinetic disorders in the Danish Psychiatric Central Research Registry. *European Psychiatry: The Journal of the Association of European Psychiatrists*, *35*, 16–24.
Østergaard, S.D., Larsen, J.T., Dalsgaard, S., Wilens, T.E., Mortensen, P.B., Agerbo, E., … & Petersen, L. (2016). Predicting ADHD by assessment of Rutter's indicators of adversity in infancy. *PLoS ONE*, *11*, e0157352.
Sjölander, A., Lichtenstein, P., Larsson, H., & Pawitan, Y. (2013). Between-within models for survival analysis. *Statistics in Medicine*, *32*, 3067–3076.
Slykerman, R. F., Thompson, J., Waldie, K. E., Murphy, R., Wall, C., & Mitchell, E. A. (2017). Antibiotics in the first year of life and subsequent neurocognitive outcomes. *Acta Paediatrica (Oslo, Norway: 1992)*, *106*, 87–94.
Stilling, R.M., Bordenstein, S.R., Dinan, T.G., & Cryan, J.F. (2014). Friends with social benefits: Host-microbe interactions as a driver of brain evolution and development? *Frontiers in Cellular and Infection Microbiology*, *4*, 147.
Wilcox, A.J. (2001). On the importance–and the unimportance–of birthweight. *International Journal of Epidemiology*, *30*, 1233–1241.
Wilcox Allen, J., Weinberg, C.R., & Basso, O. (2011). On the pitfalls of adjusting for gestational age at birth. *American Journal of Epidemiology*, *174*, 1062–1068.

Accepted for publication: 18 July 2018

Paper I – supplementary information

Online Supplementary Material for: Investigating the effects of cesarean delivery and antibiotic use in early childhood on risk of later ADHD – Axelsson et al.

Includes

Appendix S1: Statistical models

Modelling siblings

There are two overall ways to construct sibling models, and other clustered models in general. From a philosophical point of view, we have to commit to one of two realities: Either believing that all individuals in our population are inherently independent, though siblings might have the same mean risk level for psychiatric disorders. Or we must believe that siblings are inherently more alike than non-siblings such that they cannot be uncorrelated. The former option is typically denoted a *fixed effect* model, while the latter is a random effect model. In the current setup, however, we do not have to commit to one of the two explanations a priori. In fact, as will be discussed further below, both of the two models actually allow us to estimate the same parameter of interest, namely the causal effect of cesarean delivery or antibiotics in early infancy on the risk of childhood ADHD. However, they do it in different ways, and perhaps a more formal description of the models will help shed light to the somewhat subtle differences.

Below, we present models that include only a single, binary exposure variable (cesarean delivery) and no other explanatory variables. However, the results can easily be generalized to include multiple exposures, as well as other explanatory variables.

A descriptive Cox model

Before turning to the sibling models, let us briefly dwell on what we would have done, had there been no information about siblings in the data. In this case, we might propose a Cox proportional hazards model, that is, a model on the form

$$\lambda_{ij}(t) = \lambda_0(t)\exp(\beta c_{ij})$$

where c_{ij} the indicator of cesarean delivery for the ijth observation (child). We organize the observations using double indices, *ij*, to underline the family structure that we have not yet included in the model. The first index, *i*, is the number of the family (ranging from 1 to some value *n*), while the second index, *j*, is the number of the child within the family, ranging from *1* to n_i, where n_i is the number of siblings in family *i*. In this model, we assume that all observations are independent.

The Cox model assumes that the hazard of each individual is given by a product of some shared hazard for everyone (the baseline hazard, $\lambda_0(t)$) and a term that depends on whether or not the child was born by cesarean delivery ($\exp(\beta c_{ij})$). Of course, more explanatory variables other than cesarean delivery could (and should) be included in the model as well, but in this document, we stick with the minimalist version for simplicity.

A fixed-effect sibling model: The stratified Cox model

The fixed effect model mentioned above extends the descriptive Cox model by allowing the baseline hazard to differ between families, while being constant within families. Thus, all the observations are still assumed to be statistically independent, though the baseline sibling hazard levels are assumed to be equal. Formally, the model becomes

$$\lambda_{ij}(t) = \lambda_i(t)\exp(\beta w c_{ij})$$

Note the explicit dependence of the baseline hazard on *i*. Note also that we have a new parameter in this model, namely βw. This parameter describes the difference in hazards *within* families (hence the W), whereas the β from the descriptive Cox model describes the difference in hazards between any two observations.

The fixed effect model is often denoted the *stratified* Cox model, as the data are split into strata (in our case, families) and the comparisons are now made within these strata. As this model has become the most widely used tool in studies similar to ours in the recent years, the stratified Cox model is a very natural starting point in our modelling endeavours. However, there is a caveat when using this model: Because we are comparing siblings to each other, families where siblings all have the same values in the explanatory variables (in this case, are all born by cesarean delivery or all born vaginally) cannot be used in the analysis. They simply have nothing to say about intra-family differences, as they know of no intra-family differences. This fact implies a potentially huge reduction in statistical power when comparing the descriptive Cox model to the stratified Cox model.

The siblings from a family will contribute with information to the estimation of the parameter of interest if and only if there exists a pair of siblings within that family such that the following two properties are both satisfied: First, the pair must be discordant in at least one of the explanatory variables included in the model. Secondly, at least one of the two children must experience the event (i.e. ADHD) and one of the children who experience the event (possibly the only one that does) must have a shorter survival time than the other child. The second property can occur in two different ways. Either, only one of the two children gets the psychiatric disorder of interest. In that case, the second property is fulfilled if we have a longer follow-up on the healthy child than the time it took for the ill child to get his/her diagnosis. Or, both children get the psychiatric disorder of interest, in which case, the second property is automatically fulfilled.

Especially if the event of interest is very uncommon or if it is highly correlated with familial confounders, the stratified Cox model will result in a very large reduction in the effective sample size used for modelling. This will often imply that p-values and confidence intervals cannot readily be compared with what is obtained using the descriptive Cox model. It should be stressed that the parameter estimates themselves should never be compared between the two models, unless we are certain there are no unobserved familial confounders.

A random effects sibling model: The between-within model

A different approach to modelling the sibling structure of the data is by introducing a random family effect. The between-within survival model is specified as

$$\lambda_{ij}(t) = \lambda_0(t) Z_i \exp(\beta_B \bar{c}_i + \beta_W \overline{(c_{ij} - \bar{c}_i)})$$

where several new terms have been introduced. First, $\bar{c}_i$ denotes the mean level of *c* (exposure) among the children in family *i*. Secondly, a new parameter, β_B has been introduced. This parameter is generally not of interest, but it describes the effect of the prevalence of cesarean deliveries for a mother on the risk of psychiatric disorders among her children. β_W is still in the model, and it should be noted that this really is the same parameter as in the stratified Cox model (57). The two effects, β_B and β_W, are referred to as between family- and within family effects, respectively, which also serves as the origin story of the name of the model.

Note that the baseline hazard is once again assumed to be shared among all the observations. However, these observations are no longer assumed to be independent. For each family, *i*, an observation from a random variable Z_i is drawn and multiplied onto the hazard. This means that all siblings within a family have the same observed value of Z_i, which renders them correlated. We still assume independence between families, though. The Z_is are furthermore assumed to all come from the same distribution, typically a gamma distribution. Thus, more parameters are needed to describe this model than the stratified model above, and therefore, we should regard the BW as a more restrictive alternative. However, as always, the more restrictive assumptions imply that the sample is used more efficiently. Whereas the stratified Cox model only really uses a very small fraction of the full dataset, the BW model uses more information from the data, so if it is correctly specified, it should be expected to yield a far more efficient estimate than the stratified model. Sjölander et al show that the BW model is generally rather stable under misspecifications of the distribution of the random effect, Z_i (also denoted the *frailty* term)(Sjölander et al., 2013). Thus, it seems like the extra parametric assumptions are not really that restrictive in practice and hence, the BW model is a very promising candidate for a model for describing the relationship between cesarean delivery and childhood psychiatric disorders.

Concluding remarks

The purpose of including all three models in this paper is to compare them and to get a better understanding of their differences and similarities. Note however, that we cannot compare all the β-parameters with each other. The β from the descriptive Cox model is neither β_W, β_B nor an average of the two. Quoting Neuhaus & Kalbfleisch (1998);

"When between [β_B] *and within-cluster* [β_W]*covariate effects are different, models that assume that these effects are the same* [e.g. the descriptive Cox model] *do not provide estimates of any*

substantive interest; the misspecified models measure neither the between- nor within-cluster covariate effects." (Neuhaus & Kalbfleisch, 1998)

The common practice in the current literature of comparing the estimates from descriptive Cox models with the estimates from stratified Cox models is thus generally unwarranted.

A final comment will be devoted to vocabulary: In survival modelling, a random effects model is typically referred to as a frailty model. However, we have chosen to use the more general term of a random effects model, as it is more clearly in contrast with the other option, namely the fixed effects model. The fixed effects, or stratified, model also has an alias that we have not yet mentioned. Due to rather technical properties of the model, it is sometimes referred to as the *conditional* model.

References for Appendix S1

Neuhaus, J. M., & Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. Biometrics, 54(2), 638–645.

Sjölander, A., Lichtenstein, P., Larsson, H., & Pawitan, Y. (2013). Between-within models for survival analysis. Statistics in Medicine, 32(18), 3067–3076.

Appendix S2: Time trends in ADHD diagnoses

Introduction

The ADHD diagnosis is a relatively new construct, and therefore, every analysis looking at ADHD as an outcome should investigate if the findings are sensitive to calendar time in one way or another. In the current study, we are using survival modelling methods and thus, we are already including one aspect of time in the models, namely the age of the children. However, if we suspect there to be more diagnoses for certain cohorts or in certain calendar years than others, and if these cohorts or calendar years also experience e.g. larger prevalence of cesarean deliveries, not adjusting for the time effects will induce bias in the model.

When looking at the raw incidences stratified by diagnosis year and age, we see definite tendencies towards more diagnoses in some years than others (sFigure 2). Especially the years 2009 and 2010 seem to have larger diagnosis occurrences. However, we also see a tendency towards most diagnoses around the ages of 7 to 10 for all but a few cohorts. If this is truly an effect of age, we have already adjusted for it through the underlying time variable in our survival models. But if the age effects differ by cohort or diagnosis year, there is no guarantee that we will not induce bias in our model, if we do not adjust for it.

As age/cohort/diagnosis year effects are very closely related by nature and quite complex to separate from each other, we will not spend more time on speculation as to what types of time effects are to blame for the patterns seen in the Lexis heat map (figure 2, main article). Instead, we will move on to do an empirical investigation by assessing whether the parameter estimates of interest, namely the effect of birth method and antibiotics usage in the first two years of life on subsequent ADHD occurrence, are sensitive to whether or not time adjustments are used in the models. Specifically, we will compare three different types of adjustment for time:

1. A model with a birth year effect (modelled categorically in one-year intervals)
2. A model with a diagnosis year effect (modelled categorically in one-year intervals)
3. A model with no extra time effects beyond age

Methods

The models are constructed as extensions of a descriptive Cox regression model, as described elsewhere. Aside from time effects, we include only the exposure effects in the models and they thus qualify as extensions of the confounder adjustment level 1 model. Note that the second method mentioned above involves using a time-dependent covariate, which increases the computational complexity of the estimation procedure dramatically, as the data are now arranged longitudinally. As we are already working with quite large datasets, this implies that we actually cannot perform the estimation procedure on the full dataset with the computer resources available at Statistics Denmark, where the data is located. However, this does not mean that we cannot assess whether we need to worry about diagnosis year effects. Instead we take advantage of the non-linearity of the computational complexity as a function of sample size and use repeated resampling instead. In layman's terms, though a longitudinal dataset with approximately 800,000 individuals is too large to work with using the available computer resources, estimating models on e.g. 80 longitudinal datasets of 10,000 individuals each is a heavy, but feasible, task, even though the same, total amount of individuals is covered. Specifically, we carry out 100 independent repetitions of the following procedure:

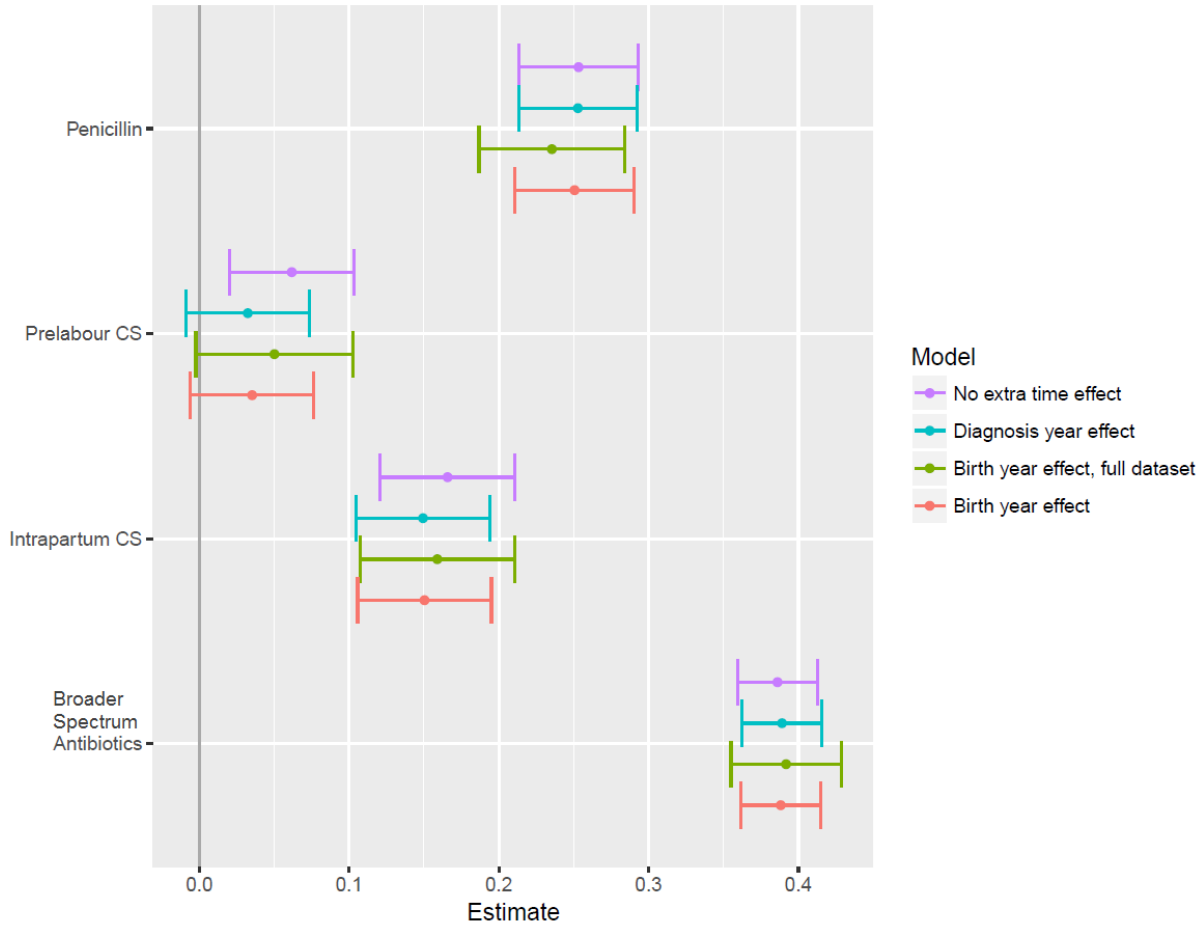
1. Draw 10,000 random observations independently from the full sample. This qualifies as the current dataset.

2. Fit each of the three models described above on the current dataset and the save the results.

By doing independent repetitions of the procedure, we create a situation where asymptotic results, such as the law of large numbers and the central limit theorem, come into play. Thereby, by looking at the distribution of the parameters obtained from these repeated resampling/re-estimation procedures, we will obtain distributions that are centered around the true parameter values while also getting an insight into the variability, or width, of the distributions. Thus, we will be able to conclude whether extra time effects are needed to obtain valid conclusions in the current study.

Non-time parameter estimates

In this section, we inspect the findings concerning the parameter estimates that are not directly related to time, i.e. the exposure variable effects. The exposure variables are birth method (vaginal birth, prelabour cesarean delivery or intrapartum cesarean delivery) and antibiotics use in the first two years of life (no antibiotics, only penicillin or broader spectrum antibiotics) modelled as categorical, additive effects. Vaginal birth and no antibiotics use serves as the reference category. sFigure B1 presents the parameter estimates from all the 100 resampling/re-estimation steps. More specifically, for each parameter estimate in each model type, we have calculated the mean estimate and added a 95 % Wald confidence interval where the variance is estimated as the empirical resampling variance. Due to the central limit theorem, we can assume asymptotic normality of the resampled parameter estimates and thus these confidence intervals should represent the sampling error adequately. For comparison, the results of the birth year model when run on the full dataset are also included in the figure. We see very little sensitivity towards whether or not extra time effects are included. Moreover, when slight time changes are found, such as those for the *prelabour CS* parameter, it does not seem to matter whether we use diagnosis year effects or birth year effects. This can be interpreted as an indication of no need for extra time adjustment, as these two concepts cannot agree unless what is needed is an age effect, and such an effect is already included in the Cox model by definition. The overall conclusion is therefore that the models seem to be insensitive to whether corrections for extra time effects are used or not.
These conclusions can also be reached by looking at the parameter estimates on a hazard ratio scale (i.e. by transformation using the exponential function). As the normality assumption no longer holds after the transformation, Wald confidence intervals cannot be produced here. Instead, for each of the 100 re-samplings, the estimate was represented by a transparent dot. Thus, the degree of opacity in different regions of the hazard ratio scale illustrated where most resample steps found each parameter value to be. We found quite similar distributions of the parameter estimates for the three model types, confirming the conclusion from the above on this new scale.



sFigure B1: Means of the parameter estimates from the 100 resampling/reestimation runs. The error bars mark 95% Wald confidence intervals based on the empirical resampling standard deviations. For comparison, the results of the birth year effect based on the full dataset are also included. Here, the error bar marks a 95% Wald confidence interval based on an estimated standard deviation. CS: Cesarean delivery.

Appendix S3: Supplementary analyses

For our primary model, the between-within survival model, we performed several sensitivity analyses, as mentioned in the main manuscript. Here follows the methodical details of these analyses. The results are reported either here, in the main manuscript or Appendix S2.

First, sensitivity towards the exact definitions of the exposure variables was examined. To this end, antibiotic use was divided into finer categories using the 5 antibiotic groups available in Denmark as mixtures (see supplementary sTable 2) and dichotomized as either none or some antibiotics (see sFigure C1 and sFigure C2). Moreover, a more detailed antibiotic exposure period effect was also investigated by including information about whether the first exposition to AB occurred in the first- or second year of life. For the mode of delivery, the prelabour/intrapartum cesarean delivery dichotomy was compared to differentiating between cesarean delivery before- or after rupture of membranes.

Secondly, we tested the robustness of the ADHD definition used in the outcome variable by use of four separate models: by limiting cases to primary diagnosis only, by restricting the maximum follow-up time to the 12^{th} birthday, by removing cases of ADD and by requiring cases to have both a diagnosis of either ADHD or ADD in addition to at least two redeemed ADHD-medication prescriptions.

Third, two potential confounders (maternal Body Mass Index (BMI) at onset of pregnancy (available from 2004) and fetal presentation) had a lot of missing information and were therefore excluded from the main analyses. The sensitivity towards this choice was also investigated, by adjusting for the confounders on a subset of the cohort born 2004 or later and a subset where the confounder was not missing, respectively.

Fourth, we investigated robustness towards alterations in the definition of siblingship by comparing full sibling models with either maternal- or paternal half sibling models.

Fifth, a post-hoc sensitivity analysis was conducted to assess robustness of the definitions of the adjustment levels presented in Table 1. More specifically, the variable *sex* was included among the variables in confounder adjustment level 1. The changes in effect estimates can be seen in sFigure C3.

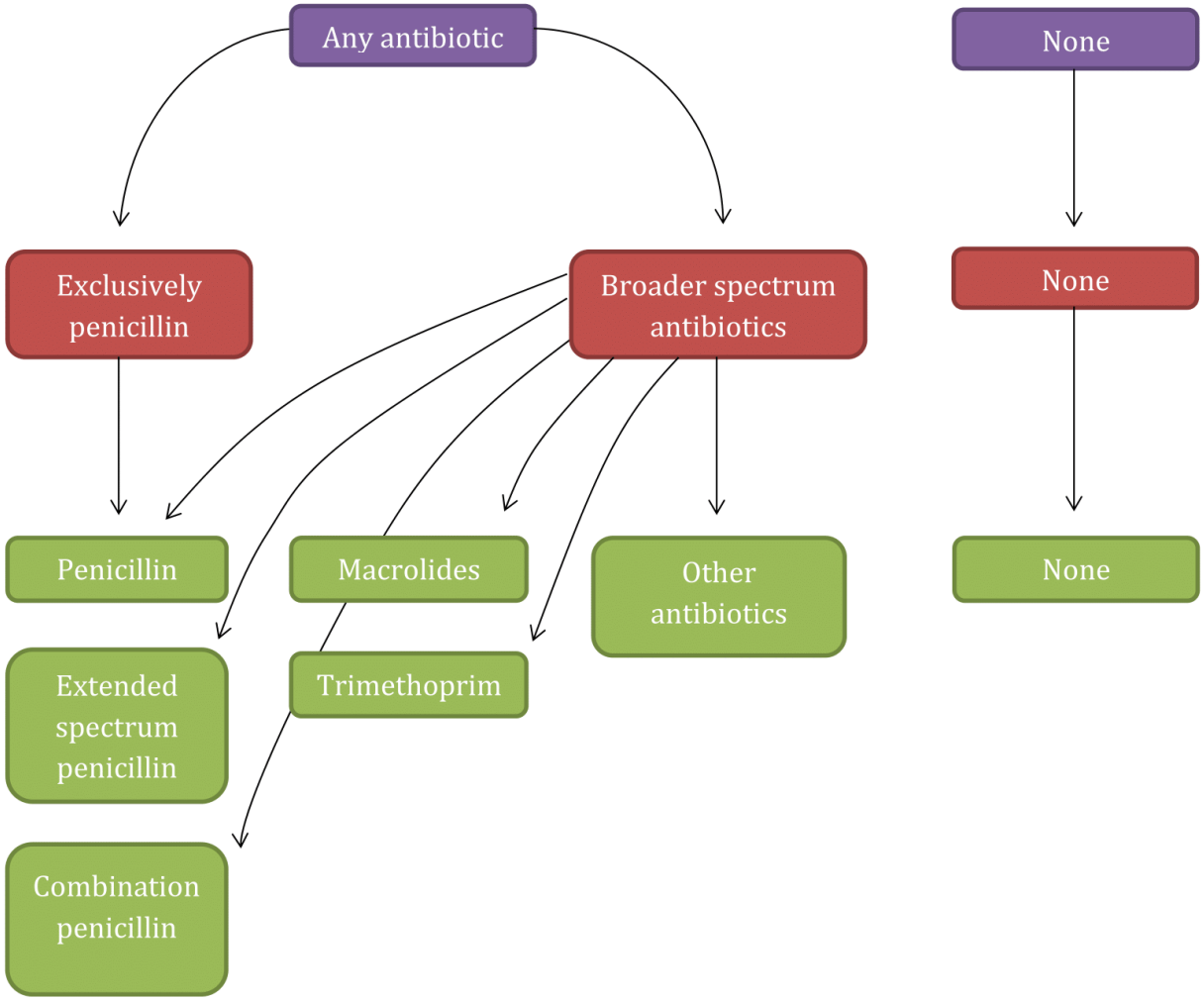
Lastly, the effects of time trends in ADHD diagnosis practices were evaluated in three different models, outlined in Appendix S2.

Descriptive Cox model

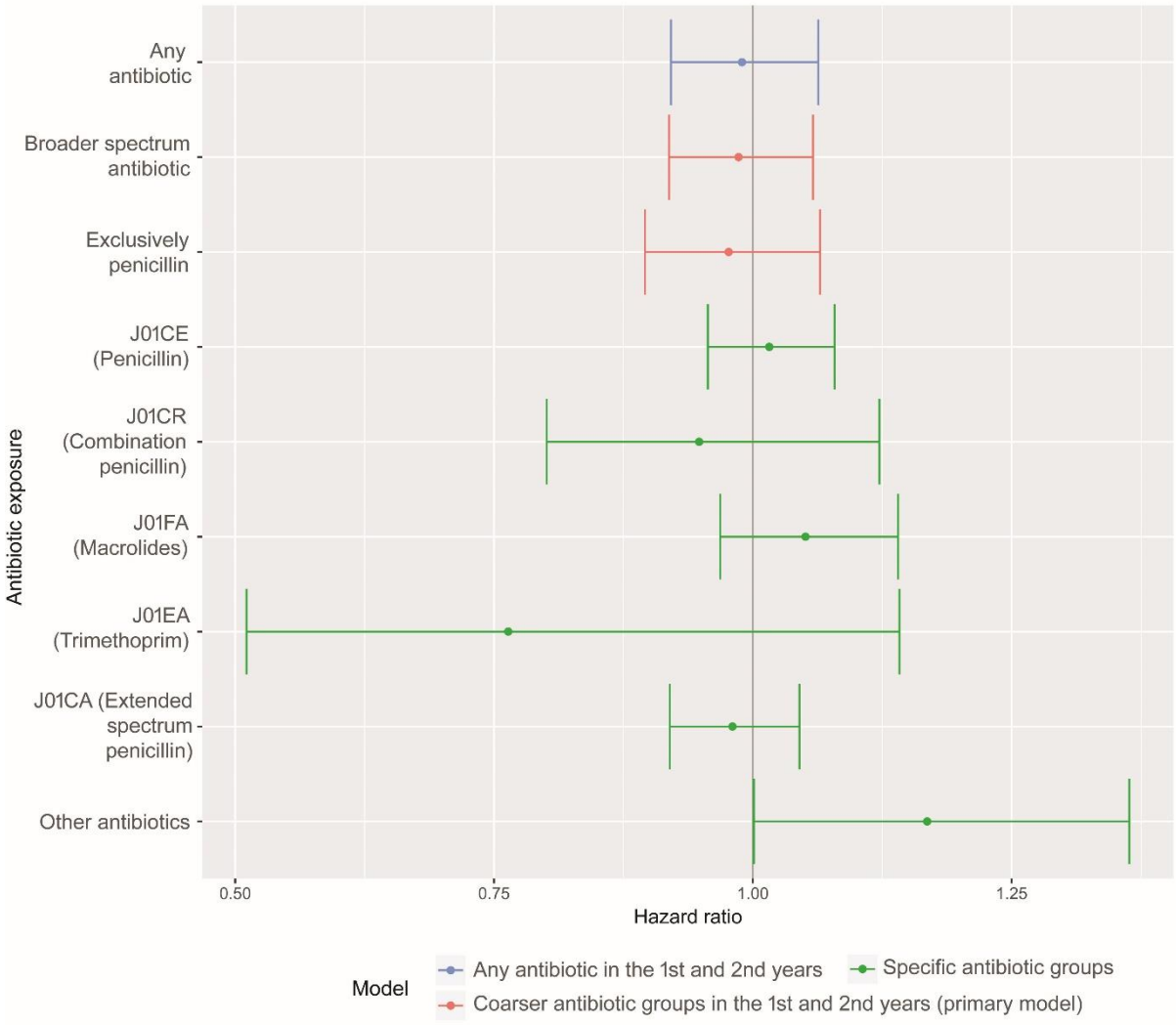
Before performing the statistical analyses, we planned on conducting sensitivity analyses that proved not to be applicable to the between-within model. However, the sensitivity analyses and if possible, the results, based on the descriptive Cox models are presented here.

1. For parental ADHD history in the fully adjusted model, we compared adjustment for ADHD diagnosis/medication with adjustment for *any* psychiatric disorder/medication. As parental ADHD history by definition is constant within a sibling-group, it cannot be evaluated in a sibling-model. The results, with slightly different exposure effect estimates, did not change our conclusions. The model is robust towards the choice of parental psychiatric history definition.

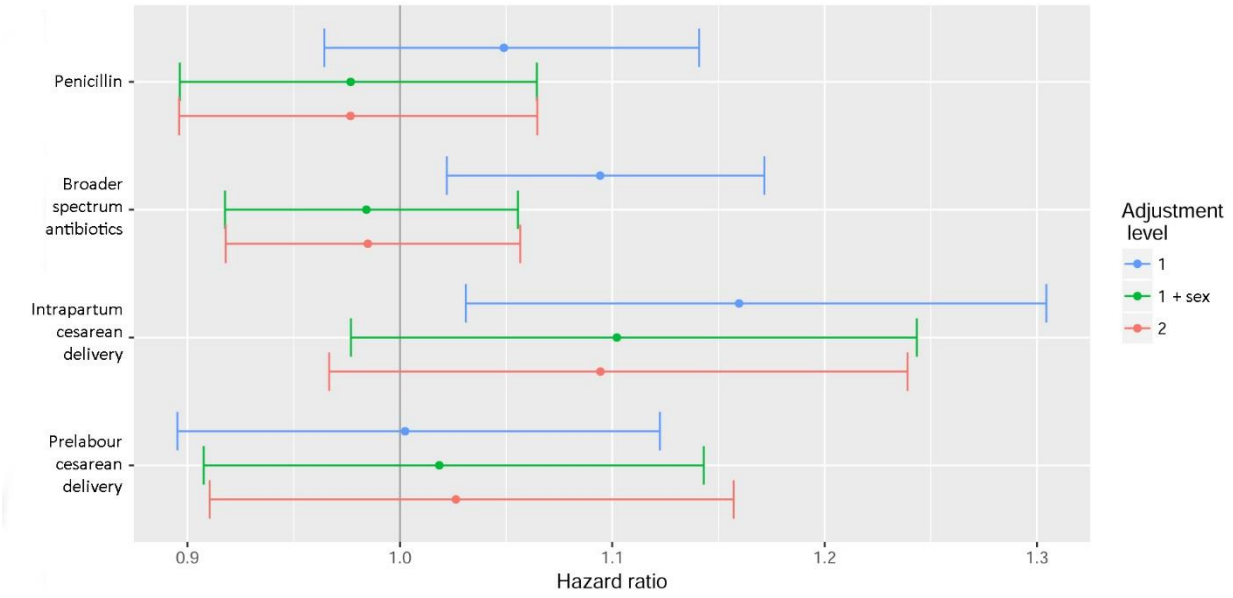
2. Sensitivity analysis was conducted using models fitted on subsamples of the data defined according the gestational age ("preterm" (<37 weeks) or 'term' (≥37 weeks) and birth weight relative to gestational age (Small for Gestational Age' (below the 10^{th} percentile) and "Average for Gestational Age" together with "Large for Gestational Age" (above the 10^{th} percentile)). Using subsampling in a sibling model is generally quite problematic, as one has to either split children across different subsamples (thereby destroying the family structure in the data) or only select siblings who are concordant in the variable used for subsampling (e.g. gestational age). Therefore, the descriptive Cox model was the basis of these sensitivity analyses. Generally, point estimates were similar to those found in the main analyses, except for the imprecise estimate found in the gestational age <37 weeks subsample. The results for this subsample lead to slightly different conclusions, with no effects of either penicillin treatment in the first two years of life or intrapartum cesarean delivery. It is however not clear whether these results are of any substantial nature, as they could come about as a result of the dramatically reduced sample size, the potentially degenerated nature of the models due to collider conditioning or perhaps a combination of both. Our conclusion remains the same, i.e. that gestational age and birthweight should not be adjusted for as confounders.
3. As mentioned above, the structure of sibling-models makes sensitivity-analyses based on subsampling strategies problematic. Therefore, we did not successfully conduct a sensitivity analysis regarding whether the children were born in the capital area or not.



sFigure C1: The three types of antibiotic exposure categorization. Children categorized as having received *broader spectrum antibiotics* in the red model, might be included in the *penicillin* category of the green model, if they receive both penicillin and a different type of an antibiotic. *None* antibiotics serve as a reference group for all categorizations.



sFigure C2: Within-effect parameter estimates and 95% confidence bands from three between-within models, varying by how the antibiotic exposure variable is defined. Only estimates corresponding to this exposure variable are shown.



sFigure C3: Exposure effects comparisons from the within-effects of the between-within model with confounder adjustment level 1, confounder adjustment level 1 with added effect of sex and confounder adjustment level 2.

Appendix S4: Antibiotic group definitions

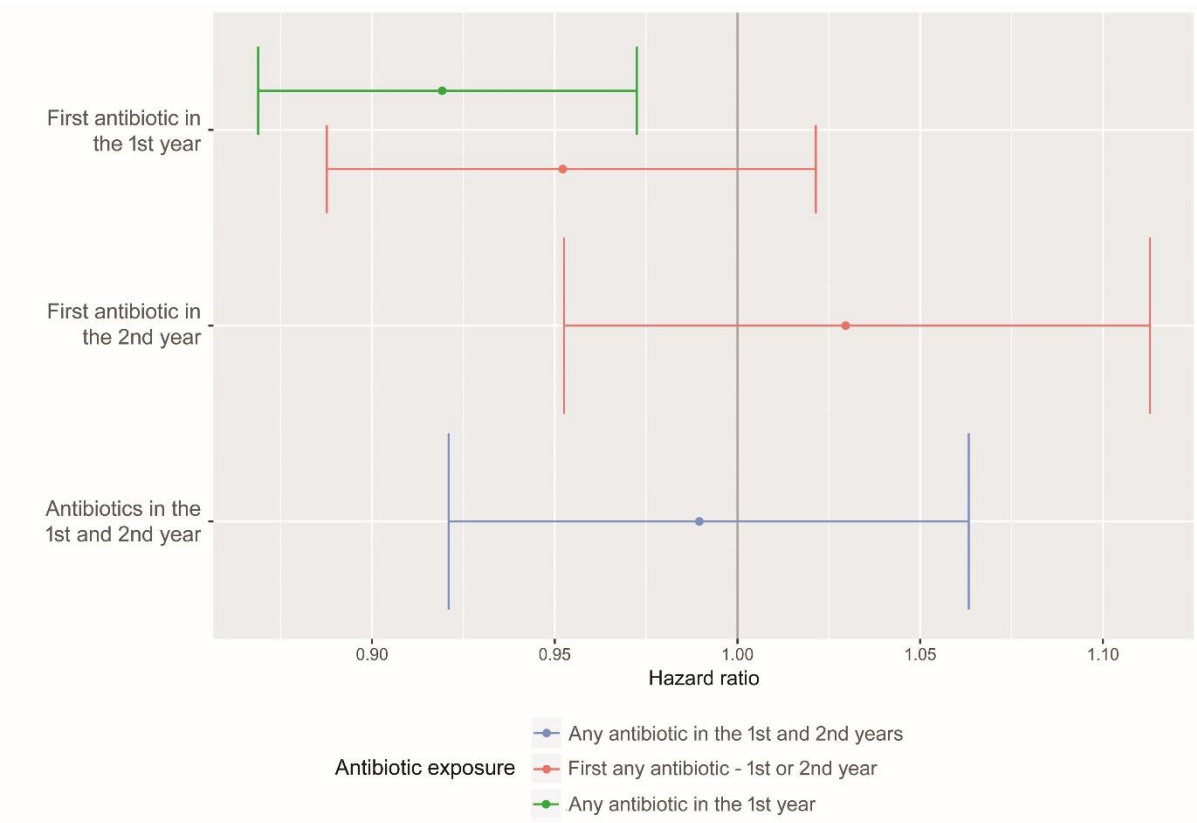
In this appendix, we wish to discuss our choice in defining the antibiotic exposure.

Infants most often receive antibiotics in the form of a mixture, so we presumed *a priori* that we could focus our attention to the 5 groups of antibiotics that are available in Denmark as mixtures (shown in sFigure D1 and ATC groups in Table S2) and group any remaining antibiotics into one group. Our presumptions were confirmed in preliminary descriptive analysis, as 98.2% of all antibiotic prescriptions for infants under 2 years of age belonged to the 5 groups of mixtures. However, the process of defining the exposure then becomes complicated as infants could theoretically have:

$$\sum_{K=0}^{6} \binom{6}{K} = 2^6 = 64$$

different possible combinations of antibiotic exposures. This presents challenges such as loss of statistical power, having to consider multiple doses of antibiotics and additionally eliminating the possibility to look into interaction effects with cesarean delivery. We therefore opted for simplifying the exposure to either exclusively penicillin, broader spectrum antibiotics than penicillin or no antibiotics, within the timeframe of the first two years of life. The grouping of the antibiotics was decided based on a surrogate marker for disturbance to the gut microbiota: reported side-effects in the form of diarrhea (Penicillin 1.2%, Extended spectrum penicillin 8.1%, Combination penicillin 19.8%, Macrolides >10%, Trimethoprim approximately 6%) (Friis, Hendel, Dalhoff, & Bjerrum, 2018; Goldman et al., 2013; Kuehn, Ismael, Long, Barker, & Sharland, 2015).

The Register of Medicinal Products Statistics, only contains information on prescriptions. We were therefore not able to evaluate the effects of intravenous antibiotics. Any topical antibiotic, figures in the group “other antibiotics” and not in the five main groups for mixtures (sFigure C1).



sFigure D3: Within-effect parameter estimates and 95% confidence bands from three between-within models with full confounder adjustment, depending on the exposure period for antibiotic use.

References for Appendix S4:

Friis, H., Hendel, J., Dalhoff, K., & Bjerrum, L. (2018, April 18). pro.medicin.dk. Azithromycin. Retrieved from http://pro.medicin.dk/Medicin/Praeparater/5965

Goldman, J. L., Jackson, M. A., Herigon, J. C., Hersh, A. L., Shapiro, D. J., & Leeder, J. S. (2013). Trends in adverse reactions to trimethoprim-sulfamethoxazole. Pediatrics, 131(1), e103-108.

Kuehn, J., Ismael, Z., Long, P. F., Barker, C. I. S., & Sharland, M. (2015). Reported rates of diarrhea following oral penicillin therapy in pediatric clinical trials. The journal of pediatric pharmacology and therapeutics: JPPT: the official journal of PPAG, 20(2), 90–104.

Appendix S5: Age specific ADHD-counts

In this appendix we wish to comment the rate of ADHD cases in our study population: A total of 17,971 cases out of 671,592 children or 2.7%. We note that this rate is very similar to that of Curran et al. 2016, based on Swedish data: 47,778 cases among 1,722,548 children, or a ratio of 2.77 %. Curran et al. used a similar follow-up design as the present study, explained in our Fig. 2.

The above rates might seem low when compared to world-wide ADHD prevalence, which has been reported to be 7.2% in a recent meta-analysis of 175 studies (Thomas, Sanders, Doust, Beller, & Glasziou, 2015). However, 74% of the studies were conducted within school populations, meaning that the youngest six age-groups do not count towards the denominator when calculating prevalence. The reported 'prevalence' in this survey is thus closer to the prevalence of the 7-16 year age group than to the prevalence among all children. They also noted that prevalence estimates were on average 2% lower in European studies when compared to North America. This could be due to the DSM (Diagnostic and Statistical Manual of Mental Disorders) criteria used in North America are somewhat broader than the ICD-10 criteria used in Europe (Doernberg & Hollander, 2016; Tripp, Luk, Schaughency, & Singh, 1999)

In order to obtain comparable measures of prevalence, we need to compare prevalence at certain ages, e.g. prevalence among 16 year olds or 14 year olds. We cannot compare such prevalence to prevalence from other studies, as they generally do not report age-specific prevalence, and comparing prevalence across study populations with different age distributions is not meaningful.

As the focus of this study is not to report the prevalence of ADHD in Denmark, which would be a study in itself, we instead provide age-specific ADHD counts by end of follow-up. As can be seen in sTable E1, less than 0.4% of children born in 2009 have been diagnosed with ADHD approximately 5 years later, at the end of follow-up. For children born in 1997 and at least 17 years of age, this number is more than 12 times higher.

Besides the age-specific ADHD counts, we also want to draw attention to Figure 2 in the main manuscript. It clearly illustrates that before the age of 5 there are very few cases of ADHD for all birth cohorts. Were we to further limit our attention to children aged 6-12 years, the prevalence of ADHD would be higher.

In conclusion, we have no reason to believe that the prevalence of ADHD in Denmark (or in our study) is abnormally low.

Birth year	Age at EoF	Number included at 2nd birthday	Number censored before EoF	Number of events by EoF	Proportion of events by EoF among included
2009	5 years	45,185	364	157	0.35%
2005	9 years	48,346	786	1162	0.24%
2001	13 years	48,558	991	1914	3.94%
1997	17 years	48,475	1576	2095	4.32%

sTable E1: Outcome distribution by end of follow-up on December 31, 2014, stratified by birth year. The last column counts the proportion of those included in the study from a given birth year that have developed ADHD by end of follow-up (EoF).

References for Appendix S5:

Curran, E. A., Khashan, A. S., Dalman, C., Kenny, L. C., Cryan, J. F., Dinan, T. G., & Kearney, P. M. (2016). Obstetric mode of delivery and attention-deficit/hyperactivity disorder: a sibling-matched study. International Journal of Epidemiology, 45(2), 532–542.

Doernberg, E., & Hollander, E. (2016). Neurodevelopmental Disorders (ASD and ADHD): DSM-5, ICD-10, and ICD-11. CNS Spectrums, 21(04), 295–299.

Thomas, R., Sanders, S., Doust, J., Beller, E., & Glasziou, P. (2015). Prevalence of Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-analysis. PEDIATRICS, 135(4), e994–e1001.

Tripp, G., Luk, S. L., Schaughency, E. A., & Singh, R. (1999). DSM-IV and ICD-10: a comparison of the correlates of ADHD and hyperkinetic disorder. Journal of the American Academy of Child and Adolescent Psychiatry, 38(2), 156–164.

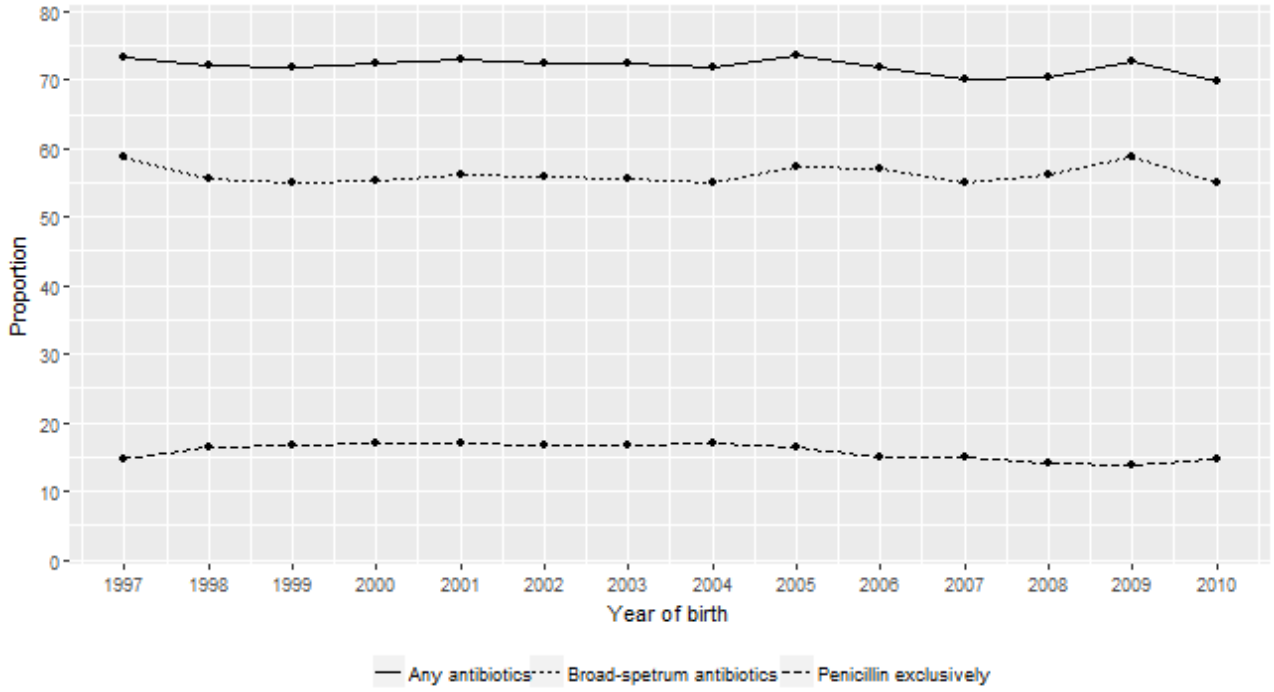


Figure S1: Prevalence of redeemed antibiotics prescriptions in the first two years of life for our cohort, consisting of 671,592 Danish children born in 1997-2010.

Nationwide Register	Data and function provided
The Medical Birth Registry (1973 to 2010)[1]	Date of birth, vital status at birth, mode of delivery, parity, multiple births, birth weight and gestational age at birth, Apgar score at 5 minutes postpartum, mother's weight in early pregnancy and smoking status
The Fertility Database (1960 to 2010) [2]	Linked children to their parents and provided data on offspring sex, and date of birth for the parents
The National Patient Registry (1977 to 2015)[3]	Hospital admission dates, codes related to complications during pregnancy, labor and delivery, and diagnoses and surgical procedures related to mode of delivery. Data on outpatient treatments were added in 1995.
The Psychiatric Central Research Register (1969 to 2015)[4]	Psychiatric discharge diagnostic codes for admissions to mental hospitals and psychiatric departments. Data on outpatient treatments were added in 1995.
The Register of Causes of Death (1970 to 2015)[5]	Vital status
The Register of Medicinal Products Statistics (1997 to 2015)	Redeemed prescriptions on psychiatric medications for parents and offspring, antibiotic medications for mothers during pregnancy and children during their first two years of life.
Statistics Denmark (1978 to 2015)	Information on emigration, parental country of origin and educational status.

Table S1: The specific information and function each register provided for our study.

References for Table S1:

1. Knudsen LB, Olsen J. The Danish Medical Birth Registry. Dan Med Bull. 1998;45(3):320-323.

2. Knudsen LB. The Danish Fertility Database. Dan Med Bull. 1998;45(2):221-225.

3. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health. 2011;39(7 Suppl):30-33. doi:10.1177/1403494811401482

4. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. Scand J Public Health. 2011;39(7 Suppl):54-57. doi:10.1177/1403494810395825

5. Helweg-Larsen K. The Danish Register of Causes of Death. Scand J Public Health. 2011;39(7 Suppl):26-29. doi:10.1177/1403494811399958

Variable	Code / Information	Register
Intrapartum cesarean delivery[a]	OP: KMCA00, KMCA10D, KMCA10E, KMCA12, KMCA12A, KMCA12B, KMCA20, KMCA30, KMCA33, KMCA96	MBR, NPR
Prelabour cesarean delivery	OP: KMCA10, KMCA10A, KMCA10B, KMCA10C, KMCA11	MBR, NPR
Obstetric factors	Birthweight, gestational age, year of birth.	MBR
Any psychiatric disorder	ICD8: 290 – 315 ICD10: DF0 - DF9	NPR
ADHD	ICD8: 308.01 ICD10: DF90	PCRR
ADD	ICD10: DF988	PCRR
Organic mental disorders	ICD8: 290.09, 290.10, 290.11, 290.18, 290.19, 292.*9, 293.*9, 294.*9, 309.*9 ICD10: DF00-DF09	PCRR
Mental disorders due to substance abuse	ICD8: 291.*9, 294.39, 303.*9, 303.20, 303.28, 303.90, 304.*9 ICD10: DF10-DF19	PCRR
Schizophrenia and related disorders	ICD8: 295.*9, 296.89, 297.*9, 298.29-298.99, 299.04, 299.05, 299.09, 301.83 ICD10: DF20-DF29	PCRR
Somatic Disorders		
Asphyxia	ICD10: DP2 for infant or DO68 for mother	NPR
Epilepsy	ICD8: 345 or ICD10: DG40, DG41	NPR
Preeclampsia, Eclampsia and Hypertension in pregnancy[b]	ICD10: DO13, DO14, DO15, DO16	NPR
Gestational Diabetes	ICD10: DO244	
Prenatal Infection	ICD10: DO23, DO411, DO753, DO98, DG0, DJ00-DJ06, DJ13-DJ18, DJ20-DJ22, DK35-DK37, DN10, DN12, DN30	NPR
Medication Use		
Any psychiatric medication	ATC-codes: N03AE, N05, N06	RMP
ADHD medication	ATC-codes: N06BA02, N06BA04, N06BA09, N06BA12	RMP
Penicillin[c]	ATC-code: J01CE	RMP
Extended Spectrum-Penicillins[c]	ATC-code: J01CA	RMP
Combination Penicillins[c]	ATC-code: J01CR	RMP
Macrolides[c]	ATC-code: J01FA	RMP
Trimethoprim[c]	ATC-code: J01EA	RMP
Broader Spectrum than Penicillin Antibiotics	ATC-codes: A07AA, J01A, J01CA, J01CF, J01CR, J01DB, J01DC, J01DD, J01DH, J01E, J01F, J01G, J01M, J01X, J04AB, P01AB	RMP

Table S2: Specification of registers and codes used for defined variables

Abbreviations:
ATC: Anatomical Therapeutic Chemicals according to WHO
ICD8: the International Classification of Diseases, 8th revision, from 1977 to 1993
ICD10: the International Classification of Diseases, 10th revision, since 1994
MBR: Medical Birth Registry since 1973
NPR: National Patient Registry since 1977
OP: "Nordic Classification of Surgical procedures" since 1996
PCRR: Psychiatric Central Research Register, since 1969
RCD: Register of Causes of Death, since 1970
RMP: The Register of Medicinal Product Statistics since 1995 and since 1997 on an individual level.
[a] Includes cesarean deliveries, where timing regarding onset of labor could not be distinguished
[b] 180 days before or 90 days after birth
[c] Types of antibiotics that are available as mixtures in Denmark.

L	Variable		Total	Vaginal birth		Intrapartum cesarean delivery		Prelabour cesarean delivery	
			n	n	(%)	n	(%)	n	(%)
	Number of children		671,592	553,727	(82.5%)	54,626	(8.1%)	63,239	(9.4%)
1	**No antibiotics**		188,184	158,629	(84.3%)	14,062	(7.5%)	15,493	(8.2%)
1	**Penicillin only**		106,170	88,016	(82.9%)	8,622	(8.1%)	9,532	(9.0%)
1	**Broader spectrum antibiotics**		377,238	307,082	(81.4%)	31,942	(8.5%)	38,214	(10.1%)
2	**Maternal age at birth (years)**								
	13-25<		79,019	68,550	(86.8%)	5,955	(7.5%)	4,514	(5.7%)
	25-30<		234,495	197,935	(84.4%)	19,401	(8.3%)	17,159	(7.3%)
	30-35<		247,387	202,278	(81.8%)	19,706	(8.0%)	25,403	(10.3%)
	35-61		110,691	84,964	(76.8%)	9,564	(8.6%)	16,163	(14.6%)
2	**Age difference (paternal – maternal, years)**								
	>-5 - +5		520,585	429,070	(82.4%)	42,527	(8.2%)	48,988	(9.4%)
	-5≤		16,138	12,257	(76.0%)	1,684	(10.4%)	2,197	(13.6%)
	>+5		134,869	112,400	(83.3%)	10,415	(7.7%)	12,054	(8.9%)
2	**Paternal education level**								
	Elementary/high school		113,155	93,498	(82.6%)	8,994	(7.9%)	10,663	(9.4%)
	Short education/skilled worker		34,003	27,847	(81.9%)	2,975	(8.7%)	3,181	(9.4%)
	Medium-length education		342,281	281,146	(82.1%)	28,143	(8.2%)	32,992	(9.6%)
	Long education		182,153	151,236	(83.0%)	14,514	(8.0%)	16,403	(9.0%)
2	**Maternal education level**								
	Elementary/high school		101,684	84,564	(83.2%)	7,682	(7.6%)	9,438	(9.3%)
	Short education/skilled worker		42,578	35,417	(83.2%)	3,488	(8.2%)	3,673	(8.6%)
	Medium-length education		264,247	216,088	(81.8%)	22,098	(8.4%)	26,061	(9.9%)
	Long education		263,083	217,658	(82.7%)	21,358	(8.1%)	24,067	(9.1%)
2	**Married/partnership**		340,984	280,853	(82.4%)	24,351	(7.1%)	35,780	(10.5%)
2	**Not married**		330,608	272,874	(82.5%)	30,275	(9.2%)	27,459	(8.3%)
2	**Smoking in pregnancy**	No	556,729	458,380	(82.3%)	45,559	(8.2%)	52,790	(9.5%)
		Yes	114,863	95,347	(83.0%)	9,067	(7.9%)	10,449	(9.1%)
3	**Sex**†	Male	344,375	282,007	(81.9%)	30,391	(8.8%)	31,977	(9.3%)
		Female	327,217	271,720	(83.0%)	24,235	(7.4%)	31,262	(9.6%)
3	**Apgar score at 5 minutes**	7-10	667,790	551,271	(82.6%)	53,785	(8.1%)	62,734	(9.4%)
		0-6	3,802	2,456	(64.6%)	841	(22.1%)	505	(13.3%)
3	**Instrumental delivery**	No	615,529	501,476	(81.5%)	50,979	(8.3%)	63,074	(10.2%)
		Yes	56,063	52,251	(93.2%)	3,647	(6.5%)	165	(0.3%)
3	**Treatment with CPAP or respirator**								
	No		658,526	547,318	(83.1%)	52,215	(7.9%)	58,993	(9.0%)
	CPAP only		12,144	6,034	(49.7%)	2,227	(18.3%)	3,883	(32.0%)
	Respirator		922	375	(40.7%)	184	(20.0%)	363	(39.4%)
3	**Signs of asphyxia* during or after birth**	No	575,348	480,808	(83.6%)	34,302	(6.0%)	60,238	(10.5%)
		Yes	96,244	72,919	(75.8%)	20,324	(21.1%)	3,001	(3.1%)
3	**Paternal epilepsy***	No	662,544	546,356	(82.5%)	53,829	(8.1%)	62,359	(9.4%)
		Yes	9,048	7,371	(81.5%)	797	(8.8%)	880	(9.7%)
3	**Preeclampsia or hypertension***	No	648,015	538,599	(83.1%)	51,141	(7.9%)	58,275	(9.0%)
		Yes	23,577	15,128	(64.2%)	3,485	(14.8%)	4,964	(21.1%)
3	**Maternal epilepsy***	No	661,141	545,415	(82.5%)	53,718	(8.1%)	62,008	(9.4%)
		Yes	10,451	8,312	(79.5%)	908	(8.7%)	1,231	(11.8%)
	…Continued								

L	Variable		Total	Vaginal birth		Intrapartum cesarean delivery		Prelabour cesarean delivery	
			n	n	(%)	n	(%)	n	(%)
3	**Gestational diabetes***	No	662,591	547,486	(82.6%)	53,531	(8.1%)	61,574	(9.3%)
		Yes	9,001	6,241	(69.3%)	1,095	(12.2%)	1,665	(18.5%)
3	**Primiparous**		296,785	238,792	(80.5%)	35,828	(12.1%)	22,165	(7.5%)
3	**1 prior child**		256,353	215,350	(84.0%)	14,357	(5.6%)	26,646	(10.4%)
3	**2 prior children**		92,957	78,140	(84.1%)	3,462	(3.7%)	11,355	(12.2%)
3	**≥3 prior children**		25,497	21,445	(84.1%)	979	(3.8%)	3,073	(12.1%)
3	**Induction of labor**	No	597,616	493,801	(82.6%)	43,463	(7.3%)	60,352	(10.1%)
		Yes	73,976	59,926	(81.0%)	11,163	(15.1%)	2,887	(3.9%)
3	**Induction of contractions**	No	538,597	442,030	(82.1%)	34,257	(6.4%)	62,310	(11.6%)
		Yes	132,995	111,697	(84.0%)	20,369	(15.3%)	929	(0.7%)
4	**Maternal antibiotic treatment during pregnancy**								
	No antibiotics, 1st trimester		594,922	491,706	(82.7%)	48,055	(8.1%)	55,161	(9.3%)
	Penicillin, 1st trimester		31,911	25,946	(81.3%)	2,427	(7.6%)	3,538	(11.1%)
	Broader spectrum AB, 1st trimester		44,759	36,075	(80.6%)	4,144	(9.3%)	4,540	(10.1%)
	No antibiotics, 2nd trimester		573,454	474,546	(82.8%)	46,236	(8.1%)	52,672	(9.2%)
	Penicillin, 2nd trimester		42,818	34,765	(81.2%)	3,213	(7.5%)	4,840	(11.3%)
	Broader spectrum AB, 2nd trimester		55,320	44,416	(80.3%)	5,177	(9.4%)	5,727	(10.4%)
	No antibiotics, 3rd trimester		587,996	485,454	(82.6%)	47,533	(8.1%)	55,009	(9.4%)
	Penicillin, 3rd trimester		29,915	24,469	(81.8%)	2,229	(7.5%)	3,217	(10.8%)
	Broader spectrum AB, 3rd trimester		53,681	43,804	(81.6%)	4,864	(9.1%)	5,013	(9.3%)
4	**Maternal infection in pregnancy***	No	646,579	534,927	(82.7%)	51,085	(7.9%)	60,567	(9.4%)
		Yes	25,013	18,800	(75.2%)	3,541	(14.2%)	2,672	(10.7%)
5	**Maternal ADHD history***	No	670,246	552,642	(82.5%)	54,508	(8.1%)	63,096	(9.4%)
		Yes	1,346	1,085	(80.6%)	118	(8.8%)	143	(10.6%)
5	**Paternal ADHD history***	No	669,447	551,982	(82.5%)	54,444	(8.1%)	63,021	(9.4%)
		Yes	2,145	1,745	(81.4%)	182	(8.5%)	218	(10.2%)

Table S3A: Background characteristics for singleton children born 1997 to 2010 and their parents, dependent on delivery mode.
The "L" column indicates adjustment level, where this variable was added as a confounder. *Diagnostic codes can be found in supplementary table S1. †Sex is also added separately, on its own, in a sensitivity model.

Children who within the first two years of life redeemed:

L	Variable		Total n	No antibiotics n	(%)	Penicillin only n	(%)	Broader spectrum antibiotic n	(%)
	Number of children		671,592	188,184	(28.0%)	106,170	(15.8%)	377,238	(56.2%)
1	**Vaginal birth**		553,727	158,629	(28.6%)	88,016	(15.9%)	307,082	(55.5%)
1	**Intrapartum cesarean delivery**		54,626	14,062	(25.7%)	8,622	(15.8%)	31,942	(58.5%)
1	**Prelabour cesarean delivery**		63,239	15,493	(24.5%)	9,532	(15.1%)	38,214	(60.4%)
2	**Maternal age at birth (years)**								
	13-25<		79,019	19,279	(24.4%)	12,420	(15.7%)	47,320	(59.9%)
	25-30<		234,495	63,695	(27.2%)	37,340	(15.9%)	133,460	(56.9%)
	30-35<		247,387	70,729	(28.6%)	39,105	(15.8%)	137,553	(55.6%)
	35-61		110,691	34,481	(31.2%)	17,305	(15.6%)	58,905	(53.2%)
2	**Age difference (paternal – maternal, years)**								
	>-5 - +5		520,585	145,981	(28.0%)	82,766	(15.9%)	291,838	(56.1%)
	-5≤		16,138	4,492	(27.8%)	2,424	(15.0%)	9,222	(57.1%)
	>+5		134,869	37,711	(28.0%)	20,980	(15.6%)	76,178	(56.5%)
2	**Paternal education level**								
	Elementary/high school		113,155	28,472	(25.2%)	17,515	(15.5%)	67,168	(59.4%)
	Short education/skilled worker		34,003	9,968	(29.3%)	5,378	(15.8%)	18,657	(54.9%)
	Medium-length education		342,281	91,136	(26.6%)	54,238	(15.8%)	196,907	(57.5%)
	Long education		182,153	58,608	(32.2%)	29,039	(15.9%)	94,506	(51.9%)
2	**Maternal education level**								
	Elementary/high school		101,684	25,271	(24.9%)	15,925	(15.7%)	60,488	(59.5%)
	Short education/skilled worker		42,578	12,346	(29.0%)	6,895	(16.2%)	23,337	(54.8%)
	Medium-length education		264,247	68,889	(26.1%)	41,960	(15.9%)	153,398	(58.1%)
	Long education		263,083	81,678	(31.0%)	41,390	(15.7%)	140,015	(53.2%)
2	**Married/partnership**		340,984	98,265	(28.8%)	54,071	(15.9%)	188,648	(55.3%)
2	**Not married**		330,608	89,919	(27.2%)	52,099	(15.8%)	188,590	(57.0%)
2	**Smoking in pregnancy**	No	556,729	161,800	(29.1%)	88,032	(15.8%)	306,897	(55.1%)
		Yes	114,863	26,384	(23.0%)	18,138	(15.8%)	70,341	(61.2%)
3	**Sex**†	Male	344,375	86,295	(25.1%)	54,853	(15.9%)	203,227	(59.0%)
		Female	327,217	101,889	(31.1%)	51,317	(15.7%)	174,011	(53.2%)
3	**Apgar score at**	0-6	3,802	971	(25.5%)	559	(14.7%)	2,272	(59.8%)
	5 minutes	7-10	667,790	187,213	(28.0%)	105,611	(15.8%)	374,966	(56.2%)
3	**Instrumental**	No	615,529	172,457	(28.0%)	97,280	(15.8%)	345,792	(56.2%)
	delivery	Yes	56,063	15,727	(28.1%)	8,890	(15.9%)	31,446	(56.1%)
3	**Treatment with CPAP or respirator**								
	No		658,526	184,984	(28.1%)	104,252	(15.8%)	369,290	(56.1%)
	CPAP only		12,144	2,996	(24.7%)	1,797	(14.8%)	7,351	(60.5%)
	Respirator		922	204	(22.1%)	121	(13.1%)	597	(64.8%)
3	**Signs of asphyxia***	No	575,348	161,589	(28.1%)	91,157	(15.8%)	322,602	(56.1%)
	during or after birth	Yes	96,244	26,595	(27.6%)	15,013	(15.6%)	54,636	(56.8%)
3	**Paternal epilepsy***	No	662,544	185,821	(28.0%)	104,738	(15.8%)	371,985	(56.1%)
		Yes	9,048	2,363	(26.1%)	1,432	(15.8%)	5,253	(58.1%)
3	**Preeclampsia or**	No	648,015	182,188	(28.1%)	102,512	(15.8%)	363,315	(56.1%)
	hypertension*	Yes	23,577	5,996	(25.4%)	3,658	(15.5%)	13,923	(59.1%)
3	**Maternal epilepsy***	No	661,141	185,638	(28.1%)	104,594	(15.8%)	370,909	(56.1%)
		Yes	10,451	2,546	(24.4%)	1,576	(15.1%)	6,329	(60.6%)
	…Continued								

L	Variable		Total	No antibiotics		Penicillin only		Broader spectrum antibiotics	
			n	n	(%)	n	(%)	n	(%)
3	**Gestational diabetes***	No	662,591	185,985	(28.1%)	104,781	(15.8%)	371,825	(56.1%)
		Yes	9,001	2,199	(24.4%)	1,389	(15.4%)	5,413	(60.1%)
3	**Primiparous**		296,785	83,985	(28.3%)	47,227	(15.9%)	165,573	(55.8%)
3	**1 prior child**		256,353	69,009	(26.9%)	40,257	(15.7%)	147,087	(57.4%)
3	**2 prior children**		92,957	27,101	(29.2%)	14,785	(15.9%)	51,071	(54.9%)
3	**≥3 prior children**		25,497	8,089	(31.7%)	3,901	(15.3%)	13,507	(53.0%)
3	**Induction of labors**	No	597,616	168,683	(28.2%)	94,675	(15.8%)	334,258	(55.9%)
		Yes	73,976	19,501	(26.4%)	11,495	(15.5%)	42,980	(58.1%)
3	**Induction of contractions**	No	538,597	151,671	(28.2%)	85,013	(15.8%)	301,913	(56.1%)
		Yes	132,995	36,513	(27.5%)	21,157	(15.9%)	75,325	(56.6%)
4	**Maternal antibiotic treatment during pregnancy**								
	No antibiotics, 1st trimester		594,922	170,685	(28.7%)	94,848	(15.9%)	329,389	(55.4%)
	Penicillin, 1st trimester		31,911	6,787	(21.3%)	4,767	(14.9%)	20,357	(63.8%)
	Broader spectrum AB, 1st trimester		44,759	10,712	(23.9%)	6,555	(14.6%)	27,492	(61.4%)
	No antibiotics, 2nd trimester		573,454	165,948	(28.9%)	91,674	(16.0%)	315,832	(55.1%)
	Penicillin, 2nd trimester		42,818	9,069	(21.2%)	6,310	(14.7%)	27,439	(64.1%)
	Broader spectrum AB, 2nd trimester		55,320	13,167	(23.8%)	8,186	(14.8%)	33,967	(61.4%)
	No antibiotics, 3rd trimester		587,996	168,903	(28.7%)	93,601	(15.9%)	325,492	(55.4%)
	Penicillin, 3rd trimester		29,915	6,384	(21.3%)	4,436	(14.8%)	19,095	(63.8%)
	Broader spectrum AB, 3rd trimester		53,681	12,897	(24.0%)	8,133	(15.2%)	32,651	(60.8%)
4	**Maternal infection in pregnancy***	No	646,579	182,161	(28.2%)	102,366	(15.8%)	362,052	(56.0%)
		Yes	25,013	6,023	(24.1%)	3,804	(15.2%)	15,186	(60.7%)
5	**Maternal ADHD history***	No	670,246	187,879	(28.0%)	105,958	(15.8%)	376,409	(56.2%)
		Yes	1,346	305	(22.7%)	212	(15.8%)	829	(61.6%)
5	**Paternal ADHD history***	No	669,447	187,623	(28.0%)	105,824	(15.8%)	376,000	(56.2%)
		Yes	2,145	561	(26.2%)	346	(16.1%)	1,238	(57.7%)

Table S3B: Background characteristics for singleton children born 1997 to 2010 and their parents, dependent on antibiotic treatment during the first two years of life.
The "L" column indicates adjustment level, where this variable was added as a confounder. *Diagnostic codes can be found in supplementary table S1. †Sex is also added separately, on its own, in a sensitivity model.

Model	Exposure	Effect Estimate	95% CI
Between Within			
	Penicillin	0.98	(0.90 to 1.07)
	Broader Spectrum antibiotics	0.99	(0.92 to 1.06)
	Intrapartum cesarean delivery	1.09	(0.97 to 1.24)
	Prelabour cesarean delivery	1.03	(0.91 to 1.16)
Stratified Cox			
	Penicillin	0.99	(0.88 to 1.12)
	Broader spectrum antibiotics	1.02	(0.92 to 1.13)
	Intrapartum cesarean delivery	1.21	(1.01 to 1.45)
	Prelabour cesarean delivery	1.14	(0.96 to 1.35)
Descriptive Cox			
	Penicillin	1.13	(1.08 to 1.19)
	Broader spectrum antibiotics	1.23	(1.19 to 1.28)
	Intrapartum cesarean delivery	1.10	(1.04 to 1.16)
	Prelabour cesarean delivery	1.11	(1.05 to 1.17)

Table S4: Fully adjusted main effect estimates with 95% Confidence Intervals (CI) for the three different statistical models: Between-Within survival model for siblings, stratified Cox by siblings, descriptive (or standard) Cox regression.

	Informative families		Informative observations	
Level 1	4062	(0.9%)	9481	(1.4%)
Level 2	5994	(1.4%)	13785	(2.0%)
Level 3	6817	(1.5%)	15458	(2.3%)
Level 4	6821	(1.5%)	15466	(2.3%)

Table S5: Counts of the number of families (and observations within those families) that contribute with information when estimating the parameters of each of the four adjustment levels in the stratified Cox model. The percentages in parentheses are calculated relative to the total number of families (n=442,002) or observations (n=677,360), respectively.

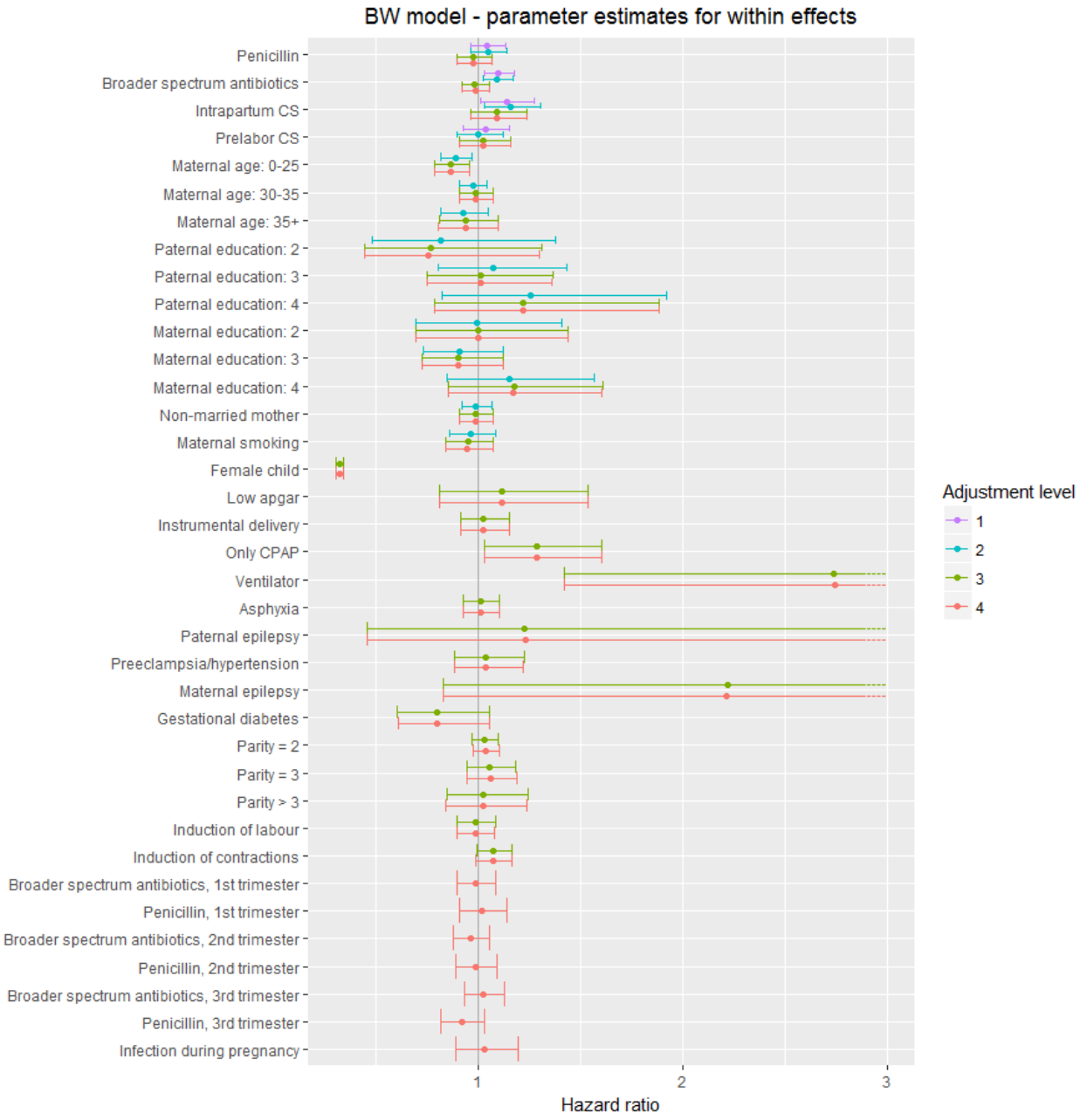


Figure S2: Parameter estimates for within-effects in the between-within model. To increase readability of the figure, 95% confidence intervals (CI) for the estimates are capped at hazard ratios above 3. For 'Ventilator' adjustment levels 2 and 3, the upper 95% CI limit is 5.3. For 'Paternal epilepsy' adjustment levels 2 and 3, the upper 95% CI limit is 3.3. For 'Maternal epilepsy' adjustment levels 2 and 3, the upper 95% CI limit is 5.9. CS: Cesarean delivery.

Paper II

Relation Between Infant Microbiota and Autism?

Results from a National Cohort Sibling Design Study

Paul Bryde Axelsson,[a] *Tine Dalsgaard Clausen,*[a] *Anne Helby Petersen,*[b] *Ida Hageman,*[c] *Anja Pinborg,*[d] *Lars Vedel Kessing,*[c] *Thomas Bergholt,*[e] *Steen Christian Rasmussen,*[f] *Niels Keiding,*[b] *and Ellen Christine Leth Løkkegaard*[a]

Background: Hypotheses concerning adverse effects of changes in microbiota have received much recent attention, but unobserved confounding makes them difficult to test. We investigated whether surrogate markers for potential adverse microbiota change in infancy affected autism risk, addressing unobserved confounding using a sibling study design.

Methods: This is a population-based, prospective cohort study including all singleton live births in Denmark from 1997 to 2010. The exposure variables were cesarean delivery and antibiotic use in the first 2 years of life. The outcome was a subsequent autism diagnosis. We used the between- and within-sibling model and compared it with sibling-stratified Cox models and simpler standard Cox models that ignored sibship.

Results: Of our study population including 671,606 children, who were followed for up to 15 years (7,341,133 person-years), 72% received antibiotics, 17.5% were delivered by cesarean, and 1.2% (8,267) developed autism. The standard Cox models predicted that both cesarean (compared with vaginal) delivery and antibiotics increased the risk of autism. In the sibling-stratified Cox model, only broader spectrum antibiotics were associated with increased risk of autism: hazard ratio (HR) = 1.16 (95% confidence interval = 1.01, 1.36). The between–within model estimated no exposure effects: intrapartum cesarean HR = 1.06 (0.89, 1.26); prelabor cesarean HR = 0.97 (0.83, 1.15); exclusively penicillin HR = 1.05 (0.93, 1.18); and broader spectrum antibiotics HR = 1.05 (0.95, 1.16).

Conclusions: The between–within model rendered more precise estimates than sibling-stratified Cox models, and we believe that it also provided more valid estimates. Results from these preferred models do not support a causal relation between antibiotic treatment during infancy, cesarean delivery, and autism. See video abstract at, http://links.lww.com/EDE/B432.

Keywords: Microbiota; Autistic Disorder; Cesarean Section; Anti-Bacterial Agents; Delivery; Proportional Hazards Models; Epidemiologic Research Design.

(*Epidemiology* 2019;30: 52–60)

Submitted March 14, 2018; accepted September 24, 2018.

From the [a]Department of Gynaecology and Obstetrics, Nordsjaellands Hospital, University of Copenhagen, Hillerød, Denmark; [b]Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark; [c]Psychiatric Center Copenhagen, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [d]Department of Obstetrics and Gynaecology, Copenhagen University Hospital, Hvidovre Hospital, Hvidovre, Denmark; [e]Department of Obstetrics, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark; and [f]Department of Clinical Microbiology, Hvidovre Hospital, Hvidovre, Denmark.

The study was supported by grants from the Capital Region Denmark Research Fund, the Capital Region Denmark PhD-start Fund, the Nordsjælland Hospital Hillerød Research Fund, the Jascha Fund, the Tvergaard Fund, and the Gangsted Fund. The funders had no role in the design and conduct of the study, analysis, interpretation of results, writing of the manuscript, approval of the manuscript, and decision to submit the manuscript for publication. The authors are fully independent from influence by the funders.

All authors have completed the International Comittee of Medical Journal Editors (ICMJE) uniform disclosure form and declare L.V.K. has within the preceding 3 years been a consultant for Sunovion. No other financial relations with any organizations that might have an interest in the submitted work in the previous 3 years. No other relations or activities that could appear to have influenced the submitted work. The other authors have no conflicts to report.

Data and computing code: By Danish law, the authors are not permitted to share person-level data. Anyone can request access to the data, by first acquiring permission from the Danish Data Protection agency and afterward the Danish Health Authority. However, these government instances have strict requirements to be allowed access to medicinal information that most foreign identities do not live up to, and therefore, cooperation with researchers from a Danish University or a Hospital is recommended if this particular dataset is of interest. The data are located at the Statistics Denmark servers and require an application defining the persons to be in the cohort and which variables to be extracted from the national registries. Data can only be accessed through the Statistics Denmark servers. The statistical program R is available at the Statistics Denmark servers. The code is available as an Appendix; links.lww.com/EDE/B418 to this publication.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Paul Bryde Axelsson, Nordsjællands Hospital, Hillerød c/o Gynækologisk-Obstetrisk Afdeling, Dr Paul Axelsson Dyrehavevej 29, 3400 Hillerød Denmark. E-mail: paulbryde@gmail.com.

ISSN: 1044-3983/19/3001-0052
DOI: 10.1097/EDE.0000000000000928

Autism spectrum disorder (hereafter referred to as autism) includes a range of developmental conditions, with genes and environment being strong risk factors.[1] Both the incidence and prevalence of autism have increased since the late 1990s.[2–4] This may be partly because of increased awareness and changes in diagnostic practices,[5] but the so-called "hygiene hypothesis" may also play a role. The hygiene hypothesis relates to the gut–brain axis, where the gut microbiota influences the development of the infant brain.[6] Indeed, differences in the composition and diversity of the gut microbiota have been detected between children with autism and healthy

controls.[6,7] The gut microbiota could be affected by cesarean delivery and antibiotic treatment,[8,9] but it is unclear whether such changes in microbiota are causally linked to autism risk.

Assessing causal relations from observational data involves making causal assumptions.[10] In particular, assumptions are made about observed or unobserved variables that may confound the effect of delivery mode or antibiotic use on autism. Sibling designs implicitly control for the main effects of all unobserved confounders shared within a sibling group without having to specify the nature or the functional form of these. Therefore, if confounding is mostly shared within a sibling group, a sibling study design can be used to estimate causal effects from the observational data. When considering exposure variables that occur early in life, associated confounding must occur at least as early. Therefore, the primary source of confounding may be linked to genetics or other parental features, which are shared among siblings.

For cohort studies with varying follow-up times and outcomes modeled as time to event, the most typical sibling design model is the sibling-stratified Cox regression model (hereafter the "Cox model"). However, a novel between–within gamma-Cox sibling model has recently been proposed as an alternative.[11] This model has shown promising results in terms of a boost in power in simulation studies.[11]

Our objectives were to investigate the causal relationship between both delivery mode and antibiotic use in infancy and the risk of developing autism later in life. We used three different statistical models: a sibling-stratified Cox model, a gamma-Cox sibling model, and a standard nonstratified Cox model for comparison with other studies.

METHODS

Design

This was a population-based prospective cohort study. Data from seven Danish nationwide registers (eTable 1; http://links.lww.com/EDE/B417) were obtained by registry linkage using unique personal registration numbers. We accessed the data in anonymized form through Statistics Denmark. Approval by the Danish Data Protection Agency (journal no.: 2012-58-0004) and the Danish Health Authority was attained. By Danish law, approval by an ethics committee is not required for research based on national registries.

Study Population and Follow-Up

We identified all live-born children born in Denmark to Danish parents between 1 January 1997 and 31 December 2010 (n = 912,796). In total, 671,606 Danish singletons who had not been diagnosed with autism and were living in Denmark at their second birthday were followed up (Figure 1). There were 439,718 full-sibling families, defined as groups of one or more children sharing both a father and a mother.

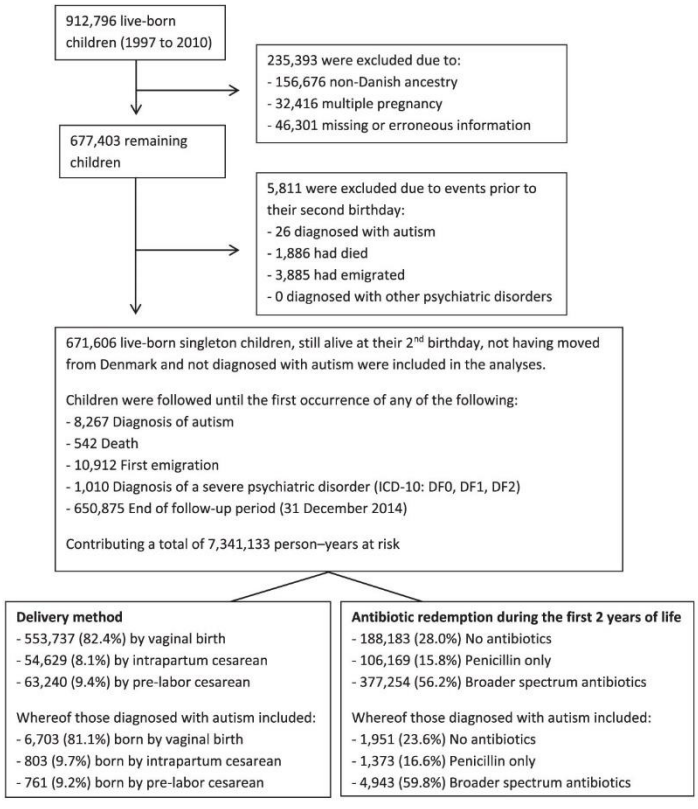


FIGURE 1. Study numbers. The number of children excluded, the number of children in the final study population, the number of events, and autism diagnoses according to delivery mode and exposure to antibiotics are shown. Children with autism were more likely to have been born by intrapartum cesarean delivery (9.7%) compared with the whole study population (8.1%). In addition, children with autism were also more likely to receive broader spectrum antibiotics (59.8%) compared with the whole study population (56.2%).

We censored children at time of death, emigration, International Classification of Diseases, 10th Revision, (ICD-10) diagnoses of organic mental disorders, mental disorders because of substance abuse, schizophrenia, schizotypal or delusional disorders, or the end of the follow-up period on 31 December 2014, whichever occurred first (Figure 1).

Exposure

We chose two surrogate markers for potential changes in a child's microbiota: delivery mode and antibiotic treatment during the first 2 years of life. Cesarean deliveries were performed either prelabor (elective) or intrapartum (after onset of labor). We used vaginal delivery as the reference category. Antibiotic use was classified into three mutually exclusive groups: treatment exclusively with β-lactamase-sensitive penicillin (penicillin), treatment with a broader spectrum antibiotic than penicillin (broader spectrum antibiotics), and no antibiotic treatment (reference category) (eTable 2; http://links.lww.com/EDE/B417).

Outcome

The outcome was time to first autism diagnosis, defined as the assignment of an autism spectrum disorder diagnosis (ICD-10 codes are listed in eTable 2; http://links.lww.com/EDE/B417). Both outpatient and inpatient diagnoses were included, as well as primary and secondary discharge diagnoses for both patient categories. Autism spectrum disorder diagnoses in the Danish Psychiatric Central Research Register have been validated previously, and 97% (95% confidence interval [CI] = 0.96, 0.99) of cases were confirmed by reviewing medical records.[12]

Confounder Adjustment

For the adjusted models, we selected covariates that might confound the effects of delivery mode and antibiotic use on the development of autism.[13–16] We applied a progressive confounder adjustment strategy, fitting several nested models with differing levels of confounder adjustment to assess the impact of different groups of potential confounders (Table 1).

We based parental psychiatric status on any psychiatric diagnostic code since 1969 and any prescription for psychiatric medication since 1995 (eTable 2; http://links.lww.com/EDE/B417). The diagnostic codes for covariates and information on other variables are shown in eTables 2; http://links.lww.com/EDE/B417, 3A, and 3B.

All potentially confounding covariates were measured at the time of birth.

Statistical Analyses

All statistical analyses were performed using R software (version 3.3.1; R Development Core Team, 2016) and the survival package with additional programming by the authors (eAppendix 1; links.lww.com/EDE/B418).

TABLE 1. Overview of the Progression of Confounder Adjustment Used in the Adjusted Models of This Study

	Confounder Adjustment Levels by Covariate Groups				
Variables	**Unadjusted**	**Social**	**Obstetric**	**Maternal Microbiota**	**Psychiatric Predisposition**
Confounder adjustment level	1	2	3	4	5
Childhood antibiotic use	x	x	x	x	x
Delivery mode	x	x	x	x	x
Maternal age at birth		x	x	x	x
Parental age difference		(x)	(x)	(x)	(x)
Parental education		x	x	x	x
Maternal marital status		x	x	x	x
Maternal smoking		x	x	x	x
Infant sex		x	x	x	x
5-min Apgar score			x	x	x
Use of CPAP or a ventilator			x	x	x
Asphyxia			x	x	x
Parental epilepsy			(x)	(x)	(x)
Preeclampsia or hypertension			x	x	x
Gestational diabetes			x	x	x
Parity			x	x	x
Maternal antibiotic use during the pregnancy				x	x
Maternal infections during the pregnancy				x	x
Parental psychiatric history					(x)

Parentheses indicate variables not used in the sibling models. Level 1: Unadjusted model including only antibiotic use during infancy and delivery mode. Level 2: Adjusted for social confounders. Level 3: Adjusted for confounders at the previous level as well as obstetric confounders. Level 4: Adjusted for confounders at the previous level and covariates affecting maternal microbiota. This is the fully-adjusted model for sibling analyses. Level 5: Adjusted for confounders at the previous level and parental psychiatric history. This is the fully-adjusted model for the "standard, nonstratified" analysis.

We operated within the Cox model framework and focused on two methods for estimating intrafamily effects, namely the stratified Cox model and the between–within gamma-Cox model. Let $\lambda_{ij}(t)$ denote the hazard of the jth subject of the ith family at time t. The stratified Cox model assumes a family-specific baseline hazard effect and thereby models the within-family effects in the linear predictor (denoted by subscript w), which can be represented by the following hazard function specification:

$$\lambda_{ij}^{\text{strat}}\left(t \mid \mathrm{X}_{ij}, \mathrm{U}_{ij}, i\right) = \lambda_i(t)\exp\left(\beta_w \mathrm{X}_{ij} + \eta_{ij}^{w}\right)$$

Here, we assume a single, numeric exposure, X_{ij}, for simplicity and therefore β_w is the within-family exposure effect. We let η_{ij}^{w} represent the contributions of other included covariates, U_{ij}, that are not constant within the family, e.g., sex. This model is typically estimated using maximum partial likelihood estimation, and it holds that a family (stratum) will provide information for the estimation procedure if and only if there is a pair of siblings within the family that are discordant in at least one covariate and one of the two siblings has an event at a point in time when the other sibling is still at risk. We refer to these informative families as the informative subpopulation. Immense reductions in computation time for fitting the stratified model can be achieved by a priori exclusion of the complement of the informative subpopulation before maximizing the partial likelihood. In our analyses, performed using the survival package in R, the reduction was approximately 200-fold, e.g., from 14 hours to four minutes when fitting the fully-adjusted stratified model. An R function that identifies the informative subpopulation is available in eAppendix 1; links.lww.com/EDE/B418.

An alternative method for estimating the within-family exposure effect is the between–within model. Here, we replace the family-specific baseline hazard with a population-constant baseline hazard and a family-shared frailty term, Z_i:

$$\lambda_{ij}^{bw}\left(t \mid \mathrm{X}_{ij}, \mathrm{U}_{ij}, \mathrm{Z}_i\right) = \lambda_0(t)\mathrm{Z}_i \exp\left(\beta_b \bar{\mathrm{X}}_i + \beta_w\left(\mathrm{X}_{ij} - \bar{\mathrm{X}}_i\right) + \eta_i^{b} + \eta_{ij}^{w}\right)$$

Now, $\bar{\mathrm{X}}_i$ denotes the intrafamily average in X for family i, and the new term, η_i^{b}, is the between-family effects of other included covariates, i.e., a linear function of their within-family averages $\bar{\mathrm{U}}_i$. The linear covariate effects have been parameterized such that the target of the within-family exposure effect parameter is the same as in the stratified Cox model by the introduction of between-family effects, β_b.[11] We assume that the Z_is are independent and identically Γ distributed. The purpose of introducing such parametric assumptions through the frailty term is to obtain increases in power. In simulations, this appears to be largely successful, and the model seems to be robust toward misspecifications of the frailty distribution or the function used to define the "between terms" (here: average).[11] Therefore, the stratified and between–within models will generally target the same parameter and will be asymptotically equivalent under only mild assumptions. However, the between–within model provides greater precision.

We also fitted a nonstratified Cox model, and we refer to this model as the standard model in the following in order to emphasize that it does not target the same parameters as the sibling models. This model estimates population-scale associations by comparing (assumed to be) unrelated individuals. For these models, we used robust (sandwich type) standard errors to account for the influence of familial clustering on the covariance structure.

Although we can identify the informative subpopulation for the stratified Cox model as a proper subset of the whole study population, this does not mean that we should compare the efficiency of the models by comparing the size of this informative subpopulation with the entire sample. Instead, efficiency comparisons should rely on estimated CI widths. Moreover, the between–within model must necessarily use the same informative subpopulation as the stratified model to estimate within-family effects, otherwise the two models could not be targeting the same parameter.

In all models, the underlying time was offspring age. Missing information was handled by complete case analysis. For the family structure in the sibling models and the clustering in the standard models, families were defined to consist of full siblings, with all siblings having the same mother and father.

Sensitivity analyses were performed to evaluate how robust our conclusions were with respect to key modeling choices. The sensitivity analyses were conducted as extensions of the fully-adjusted between–within model whenever possible. We investigated the sensitivity of the primary model results toward the following alterations (eAppendix 2; links.lww.com/EDE/B417):

- Including an interaction between the exposure variables (cesarean delivery and antibiotic treatment).
- Changing the definitions of the exposure variables (cesarean delivery dichotomized according to whether rupture of membranes occurred before the delivery or not, differential effects for antibiotics first used in the first and second years of life).
- Changing the definition of autism (restricting cases to either primary diagnoses only or infantile autism).
- Adding adjustments for instrument-assisted deliveries, induction of labor, and induction of contractions (because of low validity of the variables).
- Changing the definitions of families and siblings (using maternal and paternal half-siblings).
- Including first-born children only (conducted with the standard model).
- Including calendar time trends in autism diagnoses (diagnosis and birth-year effects).

RESULTS

During the birth period of our cohort, from 1997 to 2010, the number of cesarean deliveries increased by more

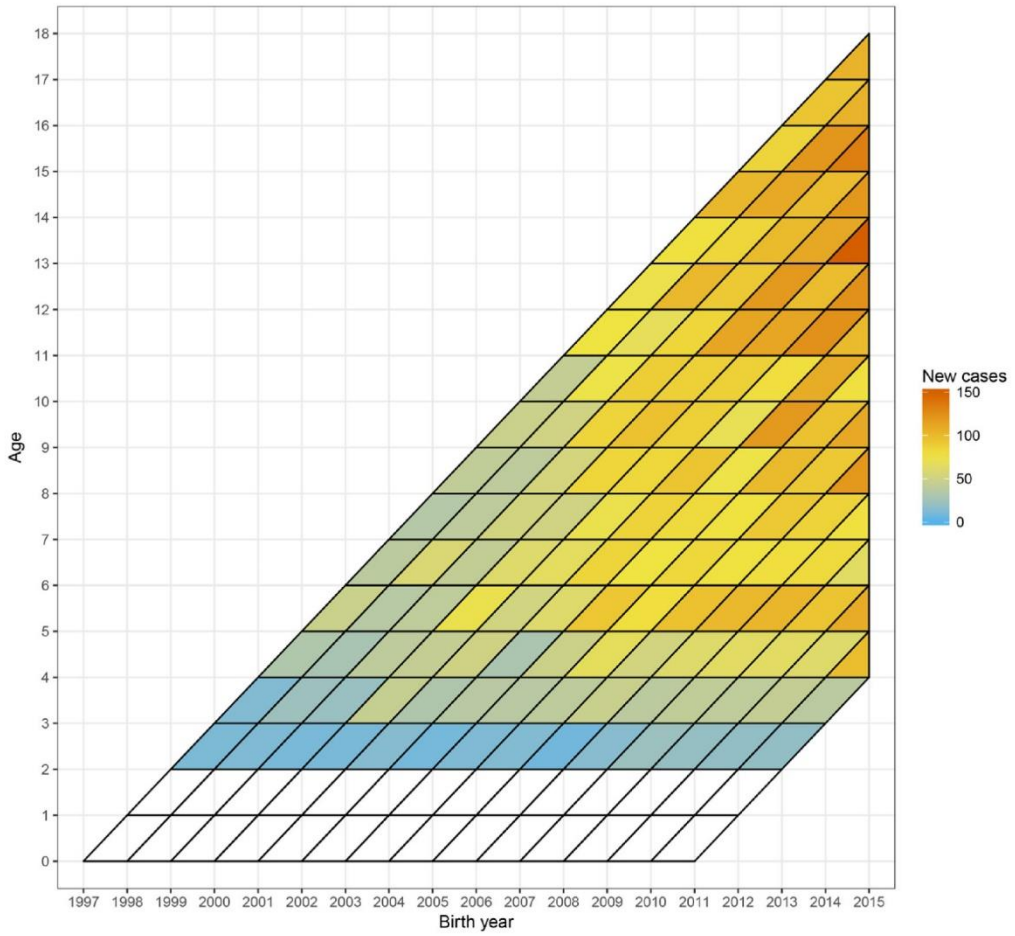


FIGURE 2. A Lexis heat map of new autism cases by 1-year birth year and age intervals. There is a general rise in incidence for all age groups after the diagnostic year 2008 and further rise in 2012 (color version available online). Age is presented in years.

than 60%, but there was no change in the pattern of antibiotic prescriptions for children younger than 2 years of age. Seventy-two percent of the children in our cohort received antibiotics in the first 2 years of life, and 17.5% were delivered by cesarean delivery. The incidence of autism in the cohort rose greatly, with marked increases in 2008 and later (Figure 2). In total, 1.2% (8,267) children developed autism.

Figure 1 shows that children born by intrapartum cesarean delivery were more likely to develop autism than those born by prelabor cesarean delivery or vaginal delivery. Children who received antibiotics, particularly broad-spectrum antibiotics, in the first 2 years of life were more likely to develop autism than those who did not. Further descriptive statistics are reported in eTables 3A and 3B; http://links.lww.com/EDE/B417.

Standard (Nonstratified) Models

In all the standard models, all exposures were associated with adverse effects, although these effects were attenuated when we applied more confounder adjustment. In the fully-adjusted model, we saw an increased risk of autism for cesarean compared with vaginal delivery (Table 2; Figure 3). We also found an increased risk of autism for children exposed to treatment with both categories of antibiotics (Table 2, Figure 3).

Stratified Models

In the fully-adjusted stratified model, the exposures generally had small, null effects (Figure 3). The largest effect was estimated for broader spectrum antibiotics, which was associated with a slightly increased risk of autism compared with no antibiotic use (HR = 1.16; 95% CI = 1.01, 1.36) (Table 2, Figure 3).

The exposure effect estimates were generally insensitive to adjustment for additional covariates, with the exception of offspring sex. There were 7632 informative observations in the fully-adjusted model (eTable 4; http://links.lww.com/EDE/B417).

TABLE 2. Exposure Effects for the Different Models

Model	Exposure	Hazard ratio	95% CI
Between-within			
	Penicillin	1.05	(0.93, 1.18)
	Broader-spectrum antibiotics	1.05	(0.95, 1.16)
	Intrapartum cesarean delivery	1.06	(0.89, 1.26)
	Prelabor cesarean delivery	0.97	(0.83, 1.15)
Stratified			
	Penicillin	1.09	(0.91, 1.29)
	Broader-spectrum antibiotics	1.16	(1.01, 1.36)
	Intrapartum cesarean delivery	1.11	(0.86, 1.43)
	Prelabor cesarean delivery	1.07	(0.84, 1.34)
Descriptive			
	Penicillin	1.11	(1.04, 1.19)
	Broader-spectrum antibiotics	1.10	(1.04, 1.16)
	Intrapartum cesarean delivery	1.10	(1.02, 1.19)
	Prelabor cesarean delivery	1.11	(1.03, 1.20)

Fully-adjusted exposure effect estimates with pointwise 95% CIs for the three different statistical models: the between-within, stratified, and descriptive models. The hazard ratios are computed for comparison using the reference levels vaginal delivery and no antibiotic use, respectively.

Between–Within Models

In general, we found no evidence that delivery mode or antibiotic treatment had an effect on the risk of autism in any of the confounder-adjusted models: intrapartum cesarean delivery HR = 1.06 (0.89, 1.26), prelabor cesarean delivery HR = 0.97 (0.83, 1.15), exclusively penicillin HR = 1.05 (0.93, 1.18); and broader spectrum antibiotics HR = 1.05 (0.95, 1.16) when compared with vaginal delivery and no antibiotic treatment (Table 2). In the unadjusted model, treatment with broader spectrum antibiotics produced a minor increased risk (HR = 1.16; 95% CI = 1.05, 1.27) (Figure 3).

We found that most of the covariates included produced effect estimates close to 1. The exceptions were female sex and higher maternal parity, which were both associated with a reduced risk of autism (eFigure 1; http://links.lww.com/EDE/B417).

Robustness of the Results

A number of sensitivity analyses showed that our overall conclusions were generally robust (eAppendix 2; http://links.lww.com/EDE/B418). In particular, we found no evidence of an interaction between the exposure variables. Two sensitivity analyses highlighted aspects of the modeling choices that somewhat altered the conclusions, namely the investigation into first-born children (eAppendix 2; http://links.lww.com/EDE/B418, section 7) and the influence of calendar time (eAppendix 2; http://links.lww.com/EDE/B418, section 8), which are discussed below.

Observations with missing or erroneous information were excluded and accounted for 5.1% of the original population. A dropout analysis found the most increased odds of having missing information for children exposed to continuous positive airway pressure (CPAP) or ventilator use, or maternal antibiotic use in the third trimester (eAppendix 3; http://links.lww.com/EDE/B419).

DISCUSSION

Our national cohort study found that no evidence that cesarean delivery, and little evidence that antibiotic treatment, was causally related to the risk of developing autism. This overall conclusion was based on estimates from two sibling models, the between–within model and the stratified model, which were mostly in agreement. The models differed on the estimated magnitude of the effect of broader spectrum antibiotic treatment (compared with no antibiotic treatment), with the stratified model finding a larger effect (HR = 1.16; 95% CI = 1.01, 1.36) than the between–within model (HR = 1.05; 95% CI = 0.95, 1.16). Sjölander et al.[11] found that the between–within model was more precise than the stratified model, which is consistent with this study. We did find somewhat different point estimates in the two models, but these differences were small and unlikely to indicate different underlying target parameters, which would be the case if the between–within model was misspecified. Therefore, the between–within model should be considered to be the most reliable model of the two and we will therefore only refer to this model in our conclusions hereafter.

For both exposure variables, the fully-adjusted standard model generally produces larger parameter estimates than the between–within model, e.g., an HR of 1.10 (95% CI = 1.04, 1.16) for broader spectrum antibiotic treatment, compared with no antibiotic treatment. However, the estimates obtained from the standard models should not be compared with those from the sibling models because they do not target the same underlying parameters. The parameters from the standard model have limited clinical relevance because generally, they do not relate to the results of interventions (e.g., changes in practice for cesarean deliveries or antibiotic use); instead, they describe current population tendencies which may well be affected by unobserved confounders. However, we need to consider whether this might also be the case for the results of the between–within model.

Frisell et al.[17] performed a careful and wide-ranging investigation into to what extent sibship analysis can control for confounding. Sibship analyses use information from families with specific discordancy properties (see the Statistical Analyses section) to form an informative subpopulation. Singleton families do not contribute intrasibship information and are not part of the informative subpopulation. Recent evidence[18–20] for reproductive stoppage, where families stop having children when a child shows symptoms of autism, is an example of a potential mechanism for inducing bias in the selection of the informative subpopulation. More specifically, singletons may be generated that have no information value for the sibship analysis.

We performed a crude sensitivity analysis to investigate this possibility. We compared the results from first-born

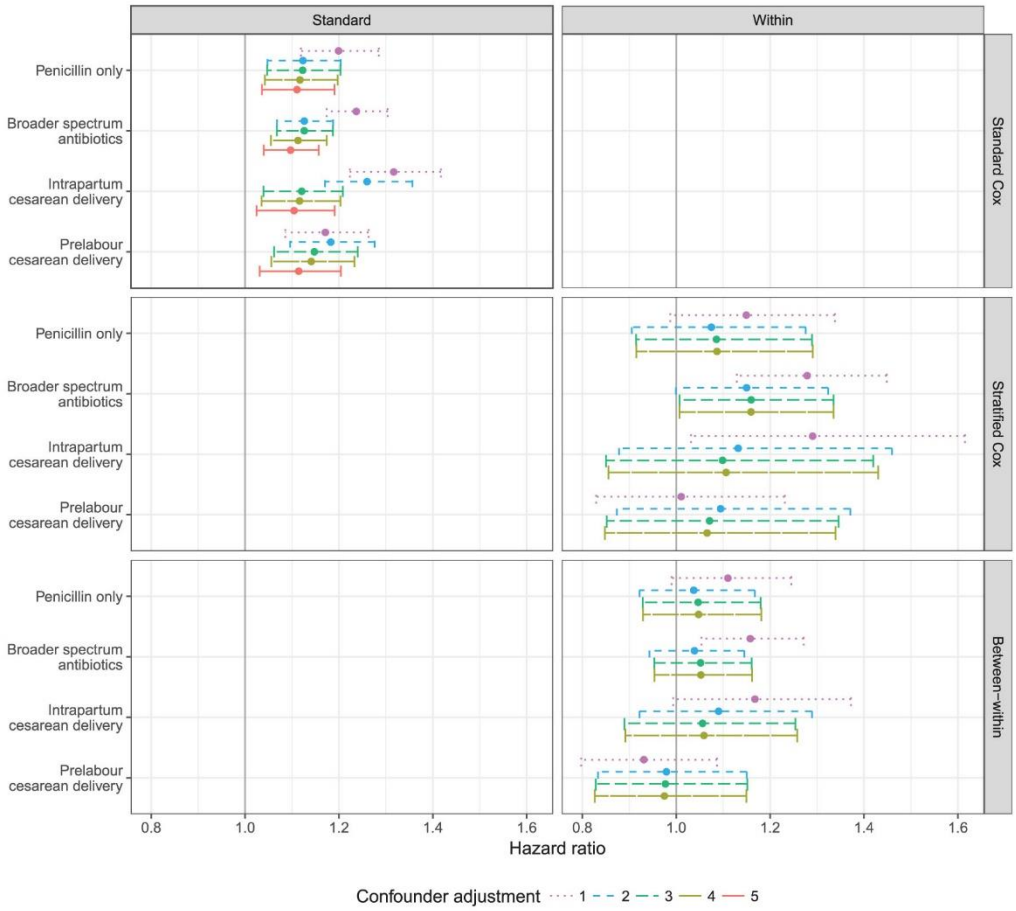


FIGURE 3. Results from the statistical models. The main results from the three different statistical models are shown. Level 5 adjustment was only performed on the standard Cox model (refer to Table 1 for confounder adjustment levels). "Standard" and "within" type effects cannot be compared and are therefore shown in separate columns. The estimated hazard ratios are accompanied by 95% CIs. The CIs are pointwise and therefore do not account for multiple testing (color version available online).

children only with those from other children, as estimated by standard models, because a sibling study design cannot be used to compare first-born children. We found some differences between these two groups of children, but low precision meant that we were unable to determine whether these differences were substantive (eAppendix 2; http://links.lww.com/EDE/B418, section 7). Overall, we found no evidence to support or refute the suggestion that reproductive stoppage imposes a generalizability issue in these analyses. Therefore, we advocate caution when generalizing the results of this study to the whole population where families differ in size and suggest further research. Moreover, we would like to emphasize that generalization to single-child families should also be conducted with caution.

Other caveats of the sibling design approach include the risk of carryover effects between siblings[21] and the impact of unobserved confounders that are not shared.[17] In particular, the causal interpretation of the within-family effect relies heavily on the untestable assumption of no unshared, unobserved confounders. If any of these issues apply, they will result in bias. However, because our study found estimates suggesting a null effect, problematic carryover or confounder bias could only be present if they canceled each other out almost perfectly, which is unlikely. Therefore, because we found evidence suggesting null effects, it is more likely that the sibling models were able to produce effects that can be interpreted causally.

A striking feature of the descriptive Lexis heat map (Figure 2) was the notable calendar variation in recorded autism diagnoses. In fact, the possible interplay among the effects of age, period, and cohort has received little previous attention. There has been a general increase over time in the incidence of autism for all cohorts and age groups. In our models, age effects were included by design as the underlying time variable, but birth year (cohort) and diagnosis year

(period) effects were not. If one or both of these time aspects function as a confounder of the effects of the exposures on the outcome, then not including them in the model would introduce bias. We investigated whether the modeling results were sensitive to including such additional time effects (eAppendix 2; http://links.lww.com/EDE/B418, section 8). We found that for the standard model, the observed effects of cesarean delivery, and especially prelabor cesarean delivery, were attenuated when calendar time effects were included in the model: HR = 1.11 (95% CI = 1.03, 1.20) in a fully-adjusted model with no calendar time effects; HR = 1.08 (95% CI = 0.996, 1.16) in the same model but with added birth-year effects. The degree of attenuation did not seem to depend on which type of calendar time effects we included. This was likely because of the rather short follow-up time, relative to the bulk of the age-at-diagnosis distribution, as shown in the Lexis heat map, which means that the diagnosis and birth years are highly correlated. In either case, these results suggest that calendar time confounds the effect of delivery mode on the risk of autism. However, when birth-year effects were introduced into the sibling models, the attenuation disappeared, suggesting that the sibling study design implicitly controls for something that is on the causal pathway from calendar time to delivery mode or from calendar time to autism, thereby blocking the confounding. Further studies investigating the nature of this mechanism will be an important contribution to understanding the etiology of autism. In particular, until we understand how calendar time and autism interact, all autism studies should include a consideration of the effects of calendar time.

Strengths and Limitations

This is the first sibling design study to investigate potential causal relations between antibiotic treatment in infancy and autism. Furthermore, it is the first study to investigate the effect of an interaction between antibiotic treatment in infancy and delivery mode on autism risk and to examine the implications of calendar time trends in autism diagnoses in epidemiologic studies.

The validity of autism diagnoses in the Danish Psychiatric Central Research Register[12] is strong, and information in the Danish registers[22] shows a high level of completeness. Additionally, in Denmark, almost all children with autism are diagnosed by public psychiatry centers.

It is well known that there are technical epidemiologic problems associated with including birth weight and gestational age in observational epidemiologic studies.[23,24] We have omitted these variables from our analyses.

We have only included native Danish children in this study. This was to reduce the amount of missing data and avoid introducing excess heterogeneity, because of the different relationships that native Danes and immigrants have with the health system.

A possible limitation of this study concerns the strategy for handling missing information, which was complete case analysis. This method will produce unbiased results if the data are missing completely at random. The dropout analysis found that missing completely at random assumption was unlikely to be satisfied and, in particular, frail children exposed to ventilator treatment at birth have highly increased odds of being missing, compared with other children. However, only 922 out of the 671,606 children in our study population were treated by ventilator and the proportion with missing information was quite low, so the use of complete case analysis probably did not greatly influence the results. Moreover, better methods for handling the missing data (e.g., multiple imputation or the Substantive Model Compatible-Fully Conditional Specification procedure[25]) were not feasible because of the computational complexity associated with fitting the between–within model.

A comment should be dedicated to limitations because of multiple testing. The CIs reported in this manuscript were pointwise, and thus, they did not include the influence of multiple testing. Therefore, their joint coverage level was below the nominal level of 95%. However, this does not alter the overall conclusion of a null effect. Obtaining CIs with the correct coverage a priori, by using bootstrapping for example, was not feasible in this study because of the computational complexity of fitting the between–within model.

Recently, Sjölander and Zetterqvist[26] suggested that results from sibling design models should be interpreted cautiously as these models control not only for shared confounders but also for shared mediators. This is, however, only problematic if there exist any shared mediators at the family level. We are currently not aware of any such variables for either antibiotic treatment or delivery mode.

Previous Studies

In a meta-analysis of birth by cesarean delivery and the risk of autism, Curran et al.[27] found a pooled adjusted odds ratio from 13 studies of 1.23 (95% CI = 1.07, 1.40), all using the standard Cox model.. The same authors conducted a survey study of 13,141 children, in whom autism was not associated with cesarean delivery.[28] However, the small number of cases meant that their estimates were quite imprecise. They subsequently used sibling-stratified Cox models to investigate any link between cesarean delivery and autism in a large Swedish population study.[29] Their results were consistent with our between–within model findings, although somewhat more precise, most likely because their population was almost four times larger. However, the Swedish population study included substantial variations in diagnostic information, with complete outpatient coverage from only 2006 onward, whereas our study included complete outpatient coverage from 1997. Furthermore, Curran et al.[27] did not investigate the effects of selection bias because of reproductive stoppage or the effects of calendar time.

Previously, only case studies have addressed the association between antibiotic treatment and the risk of autism. A study of five children reported an improvement in autism symptoms after 6 months of antibiotic treatment.[30] A

nonrandomized trial that included 18 children with autism, who had fecal microbiota transplants after 14 days of treatment with broad-spectrum antibiotics, antireflux medication, and laxatives, reported an improvement in symptoms.[31] However, these studies had several limitations, including small sample sizes and a lack of stringent methodologies and randomization.

CONCLUSIONS

We found no evidence in support of a causal relation between surrogate markers for alterations in the infant microbiota as a cause of subsequent development of autism. More specifically, when comparing siblings, the preferred models did not show that cesarean delivery and antibiotic treatment in the first 2 years of life increased the risk of autism later in life when compared with vaginal delivery and no antibiotic treatment. Because our estimates suggested a null effect, this conclusion is likely to hold even if our model suffers from unshared, unobserved confounding. However, there are potential pitfalls when interpreting results from sibling models, especially concerning generalizability. Therefore, these results may only be relevant for multiple-child families. But in such families, parents of children born by cesarean delivery or treated using antibiotics early in life should not be concerned about these factors causing autism. Although the standard (nonstratified) model found associations between autism and the two exposures, the clinical relevance of this result is limited because it suffers from familial confounding and therefore does not describe what would happen in response to interventions. We also found that the between–within model for siblings yields more precise effect estimates than the stratified Cox model.

ACKNOWLEDGMENTS

We wish to acknowledge medical writer Bobby Brown for proofreading the manuscript for language.

REFERENCES

1. Colvert E, Tick B, McEwen F, et al. Heritability of autism spectrum disorder in a UK population-based twin sample. *JAMA Psychiatry*. 2015;72:415–423.
2. Jensen CM, Steinhausen HC, Lauritsen MB. Time trends over 16 years in incidence-rates of autism spectrum disorders across the lifespan based on nationwide Danish register data. *J Autism Dev Disord*. 2014;44:1808–1818.
3. Diallo FB, Fombonne É, Kisely S, et al. Prevalence and correlates of autism spectrum disorders in Quebec. *Can J Psychiatry*. 2018;63:231–239.
4. Zablotsky B, Black LI, Maenner MJ, Schieve LA, Blumberg SJ. Estimated prevalence of autism and other developmental disabilities following questionnaire changes in the 2014 National Health Interview Survey. *Natl Health Stat Report*. 2015;87:1–20.
5. Zylstra RG, Prater CD, Walthour AE, Aponte AF. Autism: why the rise in rates? *J Fam Pract*. 2014;63:316–320.
6. Mangiola F. Gut microbiota in autism and mood disorders. *World J Gastroenterol*. 2016;22:361.
7. Forsythe P, Kunze W, Bienenstock J. Moody microbes or fecal phrenology: what do we know about the microbiota-gut-brain axis? *BMC Med*. 2016;14:58.
8. van Best N, Hornef MW, Savelkoul PH, Penders J. On the origin of species: Factors shaping the establishment of infant's gut microbiota. *Birth Defects Res C Embryo Today*. 2015;105:240–251.
9. Korpela K, Salonen A, Virta LJ, et al. Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nat Commun*. 2016;7:10410.
10. Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge, United Kingdom: Cambridge University Press; 2000.
11. Sjölander A, Lichtenstein P, Larsson H, Pawitan Y. Between–within models for survival analysis. *Stat Med*. 2013;32:3067–3076.
12. Lauritsen MB, Jørgensen M, Madsen KM, et al. Validity of childhood autism in the Danish Psychiatric Central Register: findings from a cohort sample born 1990–1999. *J Autism Dev Disord*. 2010;40:139–148.
13. Christensen J, Grønborg TK, Sørensen MJ, et al. Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA*. 2013;309:1696–1703.
14. Guinchat V, Thorsen P, Laurent C, Cans C, Bodeau N, Cohen D. Pre-, peri- and neonatal risk factors for autism. *Acta Obstet Gynecol Scand*. 2012;91:287–300.
15. Dodds L, Fell DB, Shea S, Armson BA, Allen AC, Bryson S. The role of prenatal, obstetric and neonatal factors in the development of autism. *J Autism Dev Disord*. 2011;41:891–902.
16. Modabbernia A, Velthorst E, Reichenberg A. Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. *Mol Autism*. 2017;8:13.
17. Frisell T, Öberg S, Kuja-Halkola R, Sjölander A. Sibling comparison designs: bias from non-shared confounders and measurement error. *Epidemiology*. 2012;23:713–720.
18. Grønborg TK, Hansen SN, Nielsen SV, Skytthe A, Parner ET. Stoppage in autism spectrum disorders. *J Autism Dev Disord*. 2015;45:3509–3519.
19. Wood CL, Warnell F, Johnson M, et al. Evidence for ASD recurrence rates and reproductive stoppage from large UK ASD research family databases. *Autism Res*. 2015;8:73–81.
20. Hoffmann TJ, Windham GC, Anderson M, Croen LA, Grether JK, Risch N. Evidence of reproductive stoppage in families with autism spectrum disorder: a large, population-based cohort study. *JAMA Psychiatry*. 2014;71:943–951.
21. Sjölander A, Frisell T, Kuja-Halkola R, Öberg S, Zetterqvist J. Carryover effects in sibling comparison designs. *Epidemiology*. 2016;27:852–858.
22. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health*. 2011;39(7 suppl):30–33.
23. Wilcox AJ. On the importance–and the unimportance–of birthweight. *Int J Epidemiol*. 2001;30:1233–1241.
24. Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational age at birth. *Am J Epidemiol*. 2011;174:1062–1068.
25. Bartlett JW, Seaman SR, White IR, Carpenter JR; Alzheimer's Disease Neuroimaging Initiative*. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. *Stat Methods Med Res*. 2015;24:462–487.
26. Sjölander A, Zetterqvist J. Confounders, mediators or colliders: what types of shared covariates does a sibling comparison design control for? *Epidemiology*. 2017;28:540–547.
27. Curran EA, O'Neill SM, Cryan JF, et al. Research review: birth by caesarean section and development of autism spectrum disorder and attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. *J Child Psychol Psychiatry*. 2015;56:500–508.
28. Curran EA, Cryan JF, Kenny LC, Dinan TG, Kearney PM, Khashan AS. Obstetrical mode of delivery and childhood behavior and psychological development in a British cohort. *J Autism Dev Disord*. 2016;46:603–614.
29. Curran EA, Dalman C, Kearney PM, et al. Association between obstetric mode of delivery and autism spectrum disorder: a population-based sibling design study. *JAMA Psychiatry*. 2015;72:935–942.
30. Kuhn M, Grave S, Bransfield R, Harris S. Long term antibiotic therapy may be an effective treatment for children co-morbid with Lyme disease and autism spectrum disorder. *Med Hypotheses*. 2012;78:606–615.
31. Kang DW, Adams JB, Gregory AC, et al. Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. *Microbiome*. 2017;5:10.

Erratum: Relation Between Infant Microbiota and Autism?
Results from a National Cohort Sibling Design Study

Regarding the article by Axelsson PB[1] et al. in the January 2019 issue of *Epidemiology*, an error was discovered in the following equation on page 55 after the issue published.

The equation was mistakenly published as:

$$\ll_{ij}^{strat}\left(t \mid X_{ij}, U_{ij}, i\right) = \lambda_i(t)\exp(\beta_w X_{ij} + \eta_{ij}^{w})$$

The correct equation is:

$$\lambda_{ij}^{strat}\left(t \mid X_{ij}, U_{ij}, i\right) = \lambda_i(t)\exp(\beta_w X_{ij} + \eta_{ij}^{w})$$

Reference

1. Axelsson PB, Clausen TD, Petersen AH, et al. Relation Between Infant Microbiota and Autism?: Results from a National Cohort Sibling Design Study, 1979-2015. *Epidemiology.* 2019;30(1):52-60.

Paper II – supplementary information

Online supplementary material

Nationwide Register	Data and function provided
The Medical Birth Registry (1973 to 2010) [1]	Date of birth, vital status at birth, mode of delivery, parity, multiple births, birth weight and gestational age at birth, Apgar score at 5 minutes postpartum, mother's weight in early pregnancy and smoking status
The Fertility Database (1960 to 2010) [2]	Linked children to their parents and provided data on offspring sex, and date of birth for the parents
The National Patient Registry (1977 to 2015) [3]	Hospital admission dates, codes related to complications during pregnancy, labor and delivery, and diagnoses and surgical procedures related to mode of delivery. Data on outpatient treatments were added in 1995.
The Psychiatric Central Research Register (1969 to 2015) [4]	Psychiatric discharge diagnostic codes for admissions to mental hospitals and psychiatric departments. Data on outpatient treatments were added in 1995.
The Register of Causes of Death (1970 to 2015[5]	Vital status
The Register of Medicinal Products Statistics (1997 to 2015)	Redeemed prescriptions on psychiatric medications for parents and offspring, antibiotic medications for mothers during pregnancy and children during their first two years of life.
Statistics Denmark (1978 to 2015)	Information on emigration, parental country of origin and educational status.

eTable 1: The specific information and function each register provided for our study.

References for eTable 1:

1. Knudsen LB, Olsen J. The Danish Medical Birth Registry. Dan Med Bull. 1998;45(3):320-323.

2. Knudsen LB. The Danish Fertility Database. Dan Med Bull. 1998;45(2):221-225.

3. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health. 2011;39(7 Suppl):30-33. doi:10.1177/1403494811401482

4. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. Scand J Public Health. 2011;39(7 Suppl):54-57. doi:10.1177/1403494810395825

5. Helweg-Larsen K. The Danish Register of Causes of Death. Scand J Public Health. 2011;39(7 Suppl):26-29. doi:10.1177/1403494811399958

eTable 2: Specification of registers and codes used for defined variables

Variable	Code / Information	Register
Intrapartum cesarean delivery[a]	OP: KMCA00, KMCA10D, KMCA10E, KMCA12, KMCA12A, KMCA12B, KMCA20, KMCA30, KMCA33, KMCA96	MBR, NPR
Prelabour cesarean delivery	OP: KMCA10, KMCA10A, KMCA10B, KMCA10C, KMCA11	MBR, NPR
Obstetric factors	Year of birth.	MBR
Any psychiatric disorder	ICD8: 290 - 315 ICD10: DF0 - DF9	NPR
Autism spectrum disorders	ICD10: DF840, DF841, DF845, DF848	PCRR
Infantile autism	ICD10: DF840	PCRR
Organic mental disorders	ICD10: DF00-DF09	PCRR
Mental disorders due to substance abuse	ICD10: DF10-DF19	PCRR
Schizophrenia and related disorders	ICD10: DF20-DF29	PCRR
Somatic Disorders		
Asphyxia	ICD10: DP2 for infant or DO68 for mother	NPR
Epilepsy	ICD8: 345 or ICD10: DG40, DG41	NPR
Preeclampsia, eclampsia and hypertension in pregnancy[b]	ICD10: DO13, DO14, DO15, DO16	NPR
Gestational diabetes	ICD10: DO244	
Prenatal infection	ICD10: DO23, DO411, DO753, DO98, DG0, DJ00-DJ06, DJ13-DJ18, DJ20-DJ22, DK35-DK37, DN10, DN12, DN30	NPR
Medication Use		
Any psychiatric medication	ATC-codes: N03AE, N05, N06	RMP
Penicillin	ATC-code: J01CE	RMP
Broader spectrum antibiotics than penicillin	ATC-codes: A07AA, J01A, J01CA, J01CF, J01CR, J01DB, J01DC, J01DD, J01DH, J01E, J01F, J01G, J01M, J01X, J04AB, P01AB	RMP

Abbreviations:
ATC: Anatomical Therapeutic Chemicals according to WHO
ICD8: the International Classification of Diseases, 8th revision, from 1977 to 1993
ICD10: the International Classification of Diseases, 10th revision, since 1994
MBR: Medical Birth Registry since 1973
NPR: National Patient Registry since 1977
OP: "Nordic Classification of Surgical procedures" since 1996
PCRR: Psychiatric Central Research Register, since 1969
RCD: Register of Causes of Death, since 1970
RMP: The Register of Medicinal Product Statistics since 1995 and since 1997 on an individual level.
[a] Includes cesarean section, where timing regarding onset of labor could not be distinguished
[b] 180 days before or 90 days after birth

Delivery mode

L	Variable		Total	Vaginal birth		Intrapartum cesarean delivery		Prelabour cesarean delivery	
			n	n	%	n	%	n	%
	Number of children		671,606	553,727	82.4%	54,629	8.1%	63,240	9.4%
1	**No antibiotics**		188,183	158,629	84.3%	14,062	7.5%	15,492	8.2%
1	**Penicillin only**		106,169	88,017	82.9%	8,621	8.1%	9,531	9.0%
1	**Broader spectrum antibiotics**		377,254	307,091	81.4%	31,946	8.5%	38,217	10.1%
2	**Maternal age at birth (years)**								
	13-25<		79,025	68,554	84.4%	5,957	8.3%	4,514	7.3%
	25-30<		234,497	197,939	86.7%	19,400	7.5%	17,158	5.7%
	30-35<		247,393	202,285	81.8%	19,706	8.0%	25,402	10.3%
	35-61		110,691	84,959	76.8%	9,566	8.6%	16,166	14.6%
2	**Age difference (paternal – maternal, years)**								
	>-5 - +5		520,596	429,079	82.4%	42,529	8.2%	48,988	9.4%
	-5≤		16,139	12,257	75.9%	1,684	10.4%	2,198	13.6%
	>+5		134,871	112,401	83.3%	10,416	7.7%	12,054	8.9%
2	**Maternal education level**								
	Elementary/high school		101,690	84,568	83.2%	7,683	7.6%	9,439	9.3%
	Short education/skilled worker		42,577	35,415	83.2%	3,489	8.2%	3,673	8.6%
	Medium-length education		264,255	216,093	81.8%	22,100	8.4%	26,062	9.9%
	Long education		263,084	217,661	82.7%	21,357	8.1%	24,066	9.1%
2	**Paternal education level**								
	Elementary/high school		113,165	93,506	82.6%	8,995	7.9%	10,664	9.4%
	Short education/skilled worker		34,003	27,848	81.9%	2,976	8.8%	3,179	9.3%
	Medium-length education		342,283	281,146	82.1%	28,144	8.2%	32,993	9.6%
	Long education		182,155	151,237	83.0%	14,514	8.0%	16,404	9.0%
2	**Married/partnership**		340,990	280,858	82.4%	24,351	7.1%	35,781	10.5%
2	**Not married**		330,616	272,879	82.5%	30,278	9.2%	27,459	8.3%
2	**Smoking in pregnancy**	No	556,739	458,386	82.3%	45,561	8.2%	52,792	9.5%
		Yes	114,867	95,351	83.0%	9,068	7.9%	10,448	9.1%
2	**Sex**	Male	344,381	282,008	81.9%	30,395	8.8%	31,978	9.3%
		Female	327,225	271,729	83.0%	24,234	7.4%	31,262	9.6%
3	**Apgar score at 5 minutes**	0-6	3,800	2,454	64.6%	841	22.1%	505	13.3%
		7-10	667,806	551,283	82.6%	53,788	8.1%	62,735	9.4%
3	**Treatment with CPAP or ventilator**								
	No		658,539	547,327	83.1%	52,217	7.9%	58,995	9.0%
	CPAP only		12,145	6,035	49.7%	2,228	18.3%	3,882	32.0%
	Ventilator		922	375	40.7%	184	20.0%	363	39.4%
3	**Signs of asphyxia* during or after birth**	No	575,362	480,818	83.6%	34,305	6.0%	60,239	10.5%
		Yes	96,244	72,919	75.8%	20,324	21.1%	3,001	3.1%
3	**Paternal epilepsy***	No	662,558	546,366	82.5%	53,832	8.1%	62,360	9.4%
		Yes	9,048	7,371	81.5%	797	8.8%	880	9.7%
3	**Preeclampsia or hypertension***	No	648,028	538,609	83.1%	51,144	7.9%	58,275	9.0%
		Yes	23,578	15,128	64.2%	3,485	14.8%	4,965	21.1%
3	**Maternal epilepsy***	No	661,154	545,425	82.5%	53,721	8.1%	62,008	9.4%
		Yes	10,452	8,312	79.5%	908	8.7%	1,232	11.8%
3	**Gestational diabetes***	No	662,605	547,496	82.6%	53,534	8.1%	61,575	9.3%
		Yes	9,001	6,241	69.3%	1,095	12.2%	1,665	18.5%
3	**Primiparous**		296,800	238,804	80.5%	35,830	12.1%	22,166	7.5%
3	**1 prior child**		256,352	215,349	84.0%	14,358	5.6%	26,645	10.4%
3	**2 prior children**		92,955	78,138	84.1%	3,462	3.7%	11,355	12.2%
3	**≥3 prior children**		25,499	21,446	84.1%	979	3.8%	3,074	12.1%

…Continued

L	Variable		Total n	Vaginal birth n	Vaginal birth %	Intrapartum cesarean delivery n	Intrapartum cesarean delivery %	Prelabour cesarean delivery n	Prelabour cesarean delivery %
4	**Maternal antibiotic treatment during pregnancy**								
	No antibiotics, 1st trimester		594,938	491,718	82.7%	48,057	8.1%	55,163	9.3%
	Penicillin, 1st trimester		31,911	25,945	80.6%	2,429	9.3%	3,537	10.1%
	Broader spectrum AB, 1st trimester		44,757	36,074	81.3%	4,143	7.6%	4,540	11.1%
	No antibiotics, 2nd trimester		573,458	474,551	82.8%	46,237	8.1%	52,670	9.2%
	Penicillin, 2nd trimester		42,825	34,769	80.3%	3,214	9.4%	4,842	10.4%
	Broader spectrum AB, 2nd trimester		55,323	44,417	81.2%	5,178	7.5%	5,728	11.3%
	No antibiotics, 3rd trimester		588,007	485,462	82.6%	47,535	8.1%	55,010	9.4%
	Penicillin, 3rd trimester		29,916	24,468	81.6%	2,230	9.1%	3,218	9.3%
	Broader spectrum AB, 3rd trimester		53,683	43,807	81.8%	4,864	7.5%	5,012	10.8%
4	**Maternal infection in pregnancy***	No	646,593	534,937	82.7%	51,088	7.9%	60,568	9.4%
		Yes	25,013	18,800	75.2%	3,541	14.2%	2,672	10.7%
5	**Maternal psychiatric history***	No	595,856	494,912	83.1%	47,544	8.0%	53,400	9.0%
		Yes	75,750	58,825	77.7%	7,085	9.4%	9,840	13.0%
5	**Paternal psychiatric history***	No	622,370	514,316	82.6%	50,249	8.1%	57,805	9.3%
		Yes	49,236	39,421	80.1%	4,380	8.9%	5,435	11.0%

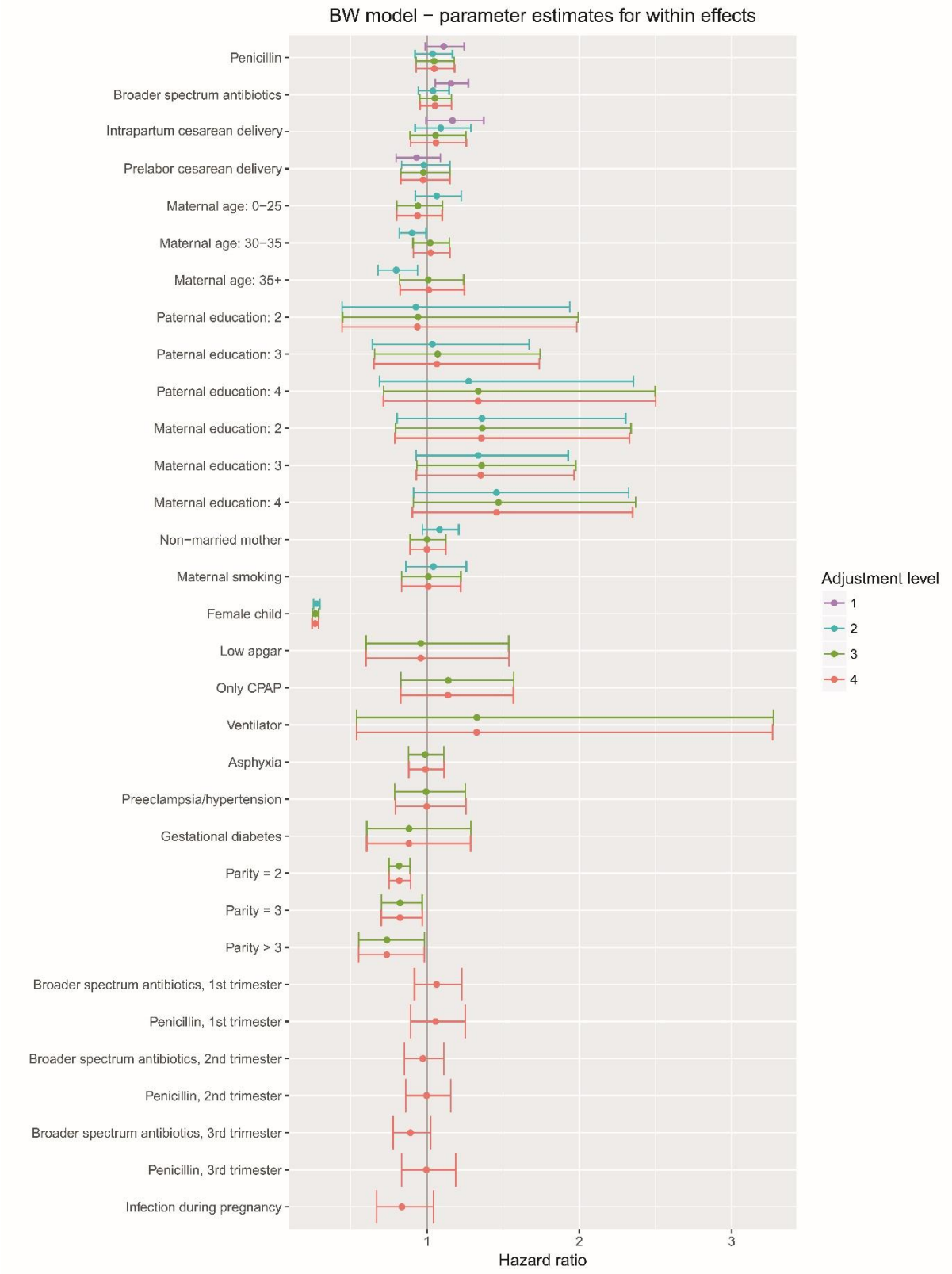
eTable 3A: Background characteristics for the study population and their parents, dependent on delivery mode. The "L" column indicates adjustment level, where this variable was added as a confounder. *Diagnostic codes can be found in online supplementary eTable 2.

Children who within the first two years of life redeemed:

L	Variable	Total n	No antibiotics n	No antibiotics %	Penicillin only n	Penicillin only %	Broader spectrum antibiotic n	Broader spectrum antibiotic %
	Number of children	671,606	188,183	28.0%	106,169	15.8%	377,254	56.2%
1	**Vaginal birth**	553,737	158,629	28.6%	88,017	15.9%	307,091	55.5%
1	**Intrapartum cesarean delivery**	54,629	14,062	25.7%	8,621	15.8%	31,946	58.5%
1	**Prelabour cesarean delivery**	63,240	15,492	24.5%	9,531	15.1%	38,217	60.4%
2	**Maternal age at birth (years)**							
	13-25<	79,025	19,282	24.4%	12,421	15.7%	47,322	59.9%
	25-30<	234,497	63,692	27.2%	37,339	15.9%	133,466	56.9%
	30-35<	247,393	70,730	28.6%	39,104	15.8%	137,559	55.6%
	35-61	110,691	34,479	31.1%	17,305	15.6%	58,907	53.2%
2	**Age difference (paternal – maternal, years)**							
	>-5 - +5	520,596	145,982	28.0%	82,765	15.9%	291,849	56.1%
	-5≤	16,139	4,492	27.8%	2,424	15.0%	9,223	57.1%
	>+5	134,871	37,709	28.0%	20,980	15.6%	76,182	56.5%
2	**Paternal education level**							
	Elementary/high school	113,165	28,473	24.9%	17,516	15.7%	67,176	59.5%
	Short education/skilled worker	34,003	9,969	29.0%	5,378	16.2%	18,656	54.8%
	Medium-length education	342,283	91,136	26.1%	54,235	15.9%	196,912	58.1%
	Long education	182,155	58,605	31.0%	29,040	15.7%	94,510	53.2%
2	**Maternal education level**							
	Elementary/high school	101,69	25,273	25.2%	15,926	15.5%	60,491	59.4%
	Short education/skilled worker	42,577	12,344	29.3%	6,895	15.8%	23,338	54.9%
	Medium-length education	264,255	68,890	26.6%	41,959	15.8%	153,406	57.5%
	Long education	263,084	81,676	32.2%	41,389	15.9%	140,019	51.9%
	…Continued							

L	Variable		Total	No antibiotics		Penicillin only		Broader spectrum antibiotics	
			n	n	%	n	%	n	%
2	**Married/partnership**		340,990	98,262	28.8%	54,069	15.9%	188,659	55.3%
2	**Not married**		330,616	89,921	27.2%	52,100	15.8%	188,595	57.0%
2	**Smoking in pregnancy**	No	556,739	161,797	29.1%	88,031	15.8%	306,911	55.1%
		Yes	114,867	26,386	23.0%	18,138	15.8%	70,343	61.2%
2	**Sex**	Male	344,381	86,293	25.1%	54,853	15.9%	203,235	59.0%
		Female	327,225	101,890	31.1%	51,316	15.7%	174,019	53.2%
3	**Apgar score at**	0-6	3,800	969	28.0%	559	15.8%	2,272	56.2%
	5 minutes	7-10	667,806	187,214	25.5%	105,610	14.7%	374,982	59.8%
3	**Treatment with CPAP or ventilator**								
	No		658,539	184,983	28.1%	104,252	15.8%	369,304	56.1%
	CPAP only		12,145	2,996	24.7%	1,796	14.8%	7,353	60.5%
	Ventilator		922	204	22.1%	121	13.1%	597	64.8%
3	**Signs of asphyxia***	No	575,362	161,589	28.1%	91,156	15.8%	322,617	56.1%
	during or after birth	Yes	96,244	26,594	27.6%	15,013	15.6%	54,637	56.8%
3	**Paternal epilepsy***	No	662,558	185,820	28.0%	104,737	15.8%	372,001	56.1%
		Yes	9,048	2,363	26.1%	1,432	15.8%	5,253	58.1%
3	**Preeclampsia or**	No	648,028	182,186	28.1%	102,511	15.8%	363,331	56.1%
	hypertension*	Yes	23,578	5,997	25.4%	3,658	15.5%	13,923	59.1%
3	**Maternal epilepsy***	No	661,154	185,637	28.1%	104,593	15.8%	370,924	56.1%
		Yes	10,452	2,546	24.4%	1,576	15.1%	6,330	60.6%
3	**Gestational diabetes***	No	662,605	185,984	28.1%	104,780	15.8%	371,841	56.1%
		Yes	9,001	2,199	24.4%	1,389	15.4%	5,413	60.1%
3	**Primiparous**		296,800	83,987	28.3%	47,227	15.9%	165,586	55.8%
3	**1 prior child**		256,352	69,005	26.9%	40,256	15.7%	147,091	57.4%
3	**2 prior children**		92,955	27,102	29.2%	14,785	15.9%	51,068	54.9%
3	**≥3 prior children**		25,499	8,089	31.7%	3,901	15.3%	13,509	53.0%
4	**Maternal antibiotic treatment during pregnancy**								
	No antibiotics, 1st trimester		594,938	170,685	28.7%	94,849	15.9%	329,404	55.4%
	Penicillin, 1st trimester		31,911	6,786	23.9%	4,766	14.6%	20,359	61.4%
	Broader spectrum AB, 1st trimester		44,757	10,712	21.3%	6,554	14.9%	27,491	63.8%
	No antibiotics, 2nd trimester		573,458	165,946	28.9%	91,673	16.0%	315,839	55.1%
	Penicillin, 2nd trimester		42,825	9,070	23.8%	6,310	14.8%	27,445	61.4%
	Broader spectrum AB, 2nd trimester		55,323	13,167	21.2%	8,186	14.7%	33,97	64.1%
	No antibiotics, 3rd trimester		588,007	168,905	28.7%	93,600	15.9%	325,502	55.4%
	Penicillin, 3rd trimester		29,916	6,382	24.0%	4,436	15.2%	19,098	60.8%
	Broader spectrum AB, 3rd trimester		53,683	12,896	21.3%	8,133	14.8%	32,654	63.8%
4	**Maternal infection in**	No	646,593	182,162	28.2%	102,365	15.8%	362,066	56.0%
	pregnancy*	Yes	25,013	6,021	24.1%	3,804	15.2%	15,188	60.7%
5	**Maternal Psychiatric**	No	595,856	170,321	28.6%	95,079	16.0%	330,456	55.5%
	history*	Yes	75,750	17,862	23.6%	11,090	14.6%	46,798	61.8%
5	**Paternal Psychiatric**	No	622,370	175,398	28.2%	98,683	15.9%	348,289	56.0%
	history*	Yes	49,236	12,785	26.0%	7,486	15.2%	28,965	58.8%

eTable 3B: Background characteristics for the study population and their parents, dependent on antibiotic treatment during the first two years of life. The "L" column indicates adjustment level, where this variable was added as a confounder. *Diagnostic codes can be found in online supplementary eTable 2.



eFigure 1: Parameter estimates with pointwise 95% CI from the between-within model, with the four confounder adjustment levels.

	Informative families		Informative observations	
Model 1	2,097	(0.5 %)	4,923	(0.7 %)
Model 2	3,163	(0.7 %)	7,266	(1.1 %)
Model 3	3,341	(0.8 %)	7,622	(1.1 %)
Model 4	3,346	(0.8 %)	7,632	(1.1 %)

eTable 4: Counts of the number of families (and observations within those families) that contribute with information when estimating the parameters of each of the four stratified models. The percentages in parentheses are calculated relative to the total number of families (n_{fam} = 439,718) or observations (n_{obs} = 671,606), respectively.

eAppendix 2

Robustness of the results

As stated in the main article we performed several sensitivity analyses, which are described in more detail here.

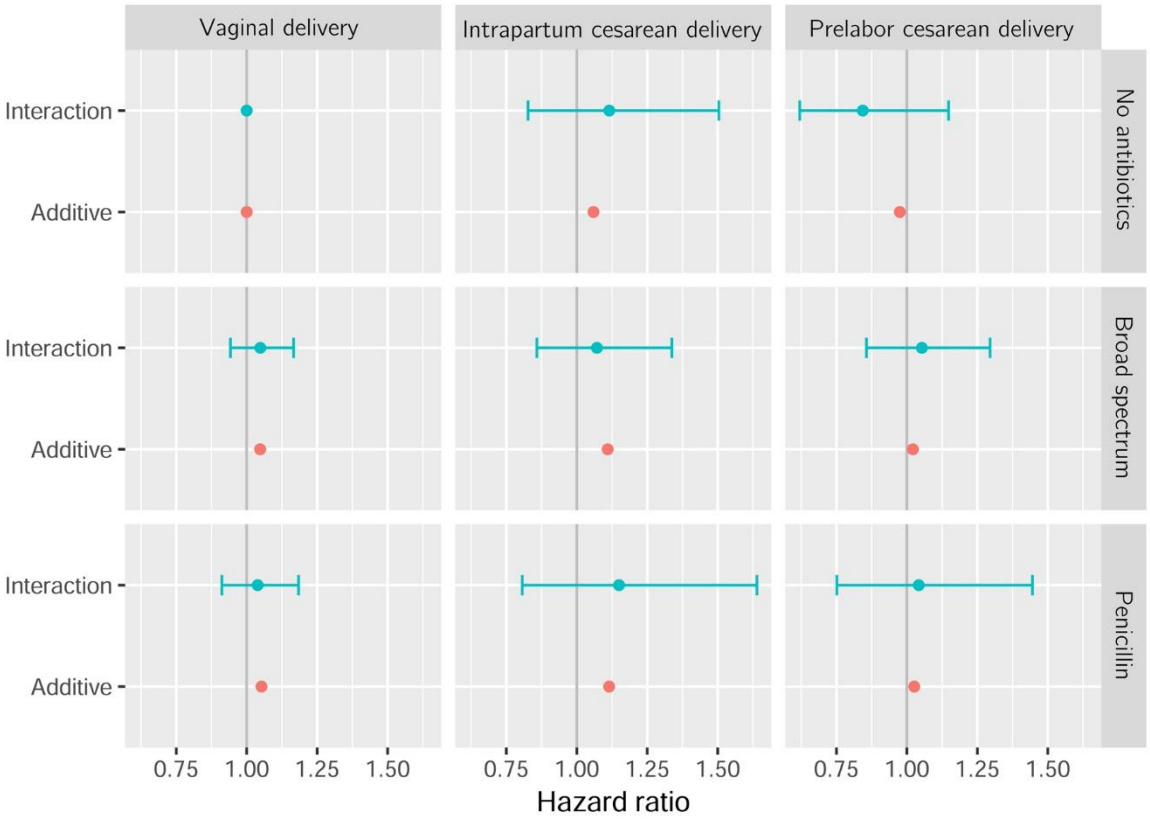
Generally, no notable degrees of sensitivity were found for the sensitivity analyses conducted on the primary model, i.e. the level 4 adjusted between-within model. Thus, the overall conclusions of the article were found to be rather robust: also with the alterations mentioned here, we still found neither an effect of mode of delivery nor early childhood antibiotics use on the risk of developing autism, when taking into account unobserved, familial confounding.

1. Sensitivity towards interaction effects in the exposure variables

We investigated whether the model results would have been any different, had we assumed an interaction effect between mode of delivery and antibiotics use instead of the additive model presented thus far. We took two approaches to this investigation: first, we made a qualitative assessment of whether or not we would have reached the same conclusions by inspecting parameter estimates and second, we conducted a formal test. The investigation was conducted based on the between-within model with full confounder adjustment.

For the parameter estimate comparison, we looked at hazard ratios for all nine distinct combinations of mode of delivery and antibiotics use. Here, we saw very similar estimates (eFigure 2), no matter if we fitted models with or without an interaction effect.

A Wald test of the interaction effect produced a test statistic of $W_{obs} = 11.8846$, which should be compared to a χ^2_8-distribution, as removal of the interaction effect frees 8 parameters (4 between parameters and 4 within parameters). We thus found $p = 0.16$, suggesting no significant interaction effect at a 5% level. We therefore concluded that the additive model is sufficiently nuanced to capture the relations between mode of delivery, antibiotics use and autism.



eFigure 2: Estimated hazard ratios relative to the reference category (delivery mode vaginal and no antibiotics use) from the two between-within models. The models differ by either containing only additive exposure effects or also an interaction effect between the exposure variables. The presented estimates here are for the within-effects. The interaction model estimates are presented with 95 % pointwise confidence limits.

2. Sensitivity towards how the antibiotics exposure variable is defined

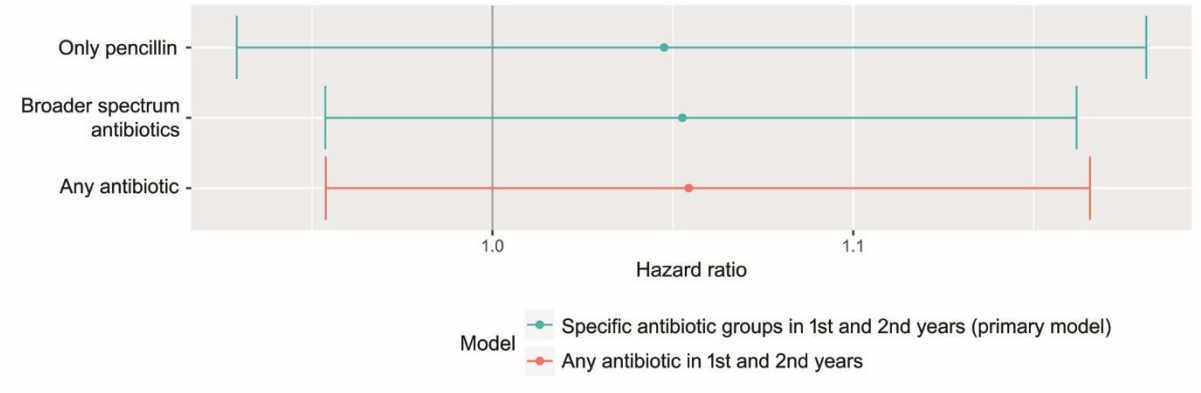
We investigated the sensitivity of how the exposure variable for early childhood antibiotics use is defined. More specifically, we looked into two aspects of the chosen definition:

a) The choice of differentiating between penicillin and broader spectrum antibiotics

b) The choice of *not* differentiating between antibiotics use in the first and second years of life.

We discuss each of these topics separately, starting with the antibiotic categorization. In the main analyses of our article, we modeled early childhood antibiotics use by using three mutually exclusive categories: No antibiotics, only penicillin, broader spectrum antibiotics (i.e. at least one type of antibiotics that is not penicillin). In this sensitivity analysis, we constructed coarser antibiotic categories by collapsing the two antibiotic groupings into one. By doing that, we only discriminated between children that had received *any* antibiotics and children that had received *no* antibiotics.

The study showed that the overall conclusions regarding the effect of antibiotics on autism was insensitive towards the coarseness of the antibiotic categories; no matter how we defined the antibiotic categories, we found a small, insignificantly increased autism risk for children receiving antibiotics (eFigure 3).



eFigure 3: Within-effect parameter estimates and 95% confidence bands for the antibiotic effects from the two models comparing categorizations into either two antibiotic groups (blue) or a single any antibiotic group (red).

We then looked at how sensitive the model results were to the definition of the exposure period for antibiotics use in early childhood. Since we concluded in the above that the results remained the same when we used coarser categories, we only looked at the coarsest antibiotic categories in this

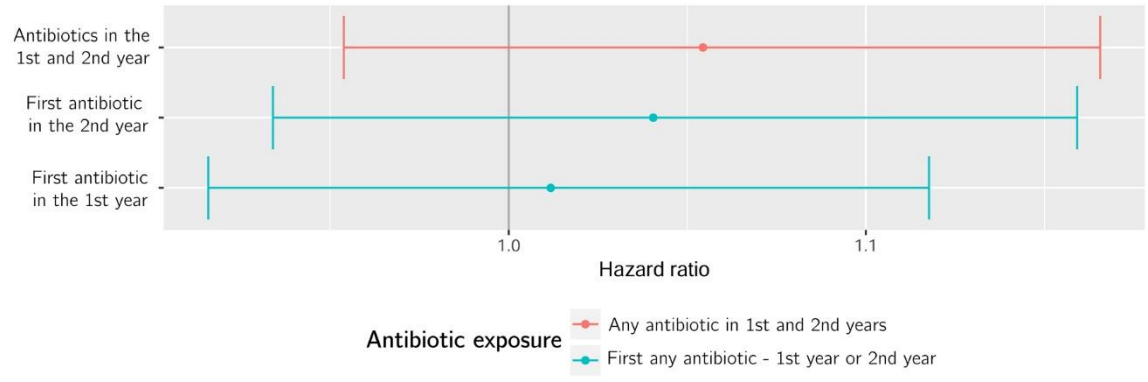
section, namely antibiotics vs. no antibiotics. This gave us a smaller number of parameter estimates to compare for examining the influence of the choice of exposure period.

We aimed to investigate whether the model results were affected by *when* the first antibiotic exposure occurred, in the first year of life or in the second year of life. Therefore, we compared models with the following variables for antibiotics:

- **First antibiotic in 1st year:** Only antibiotics use (any/no antibiotics) in the first year of life counted as an antibiotic exposure. Therefore, we were effectively modeling whether or not the first occurrence of antibiotic use was in the first year of life.
- **First antibiotic in 2nd year:** Only antibiotics use (any/no antibiotics) in the second year of life with no prior antibiotic history counted as an antibiotic exposure. Therefore, we were effectively modeling whether or not the first occurrence of antibiotic use was in the second year of life.
- **Antibiotics in 1st and 2nd year:** All occurrences of antibiotic use in the first and second years of life qualified as an antibiotic positive here. Therefore, we were modeling whether or not the child received any antibiotics at all in the first two years of its life.

Note that *first antibiotic in 1st year* and *first antibiotic in 2nd year* could meaningfully be included simultaneously in the same model, as the two categories are mutually exclusive. We thus had two models to compare: One that differentiated between 1st- and 2nd year use of any antibiotic, and one that only had a single (binary) variable for any antibiotic use.

The antibiotic exposure effects from these two models can be seen in eFigure 4. We saw that differentiating according to when the exposure occurs, leads to further attenuation of the already insignificant antibiotic exposure effect.



eFigure 4: Within-effect antibiotic hazard ratio estimates and pointwise 95 % confidence bands from the two fully adjusted between-within models that are identical except for how the antibiotics use exposure period is defined.

3. Sensitivity towards how mode of delivery is defined

We compared two definitions of mode of delivery, namely:

- Mode of delivery in three categories: Vaginal birth, prelabour cesarean delivery, intrapartum cesarean delivery (as used in the main analyses).
- Mode of delivery in three categories: Vaginal birth, cesarean delivery before rupture of membranes, cesarean delivery after rupture of membranes.

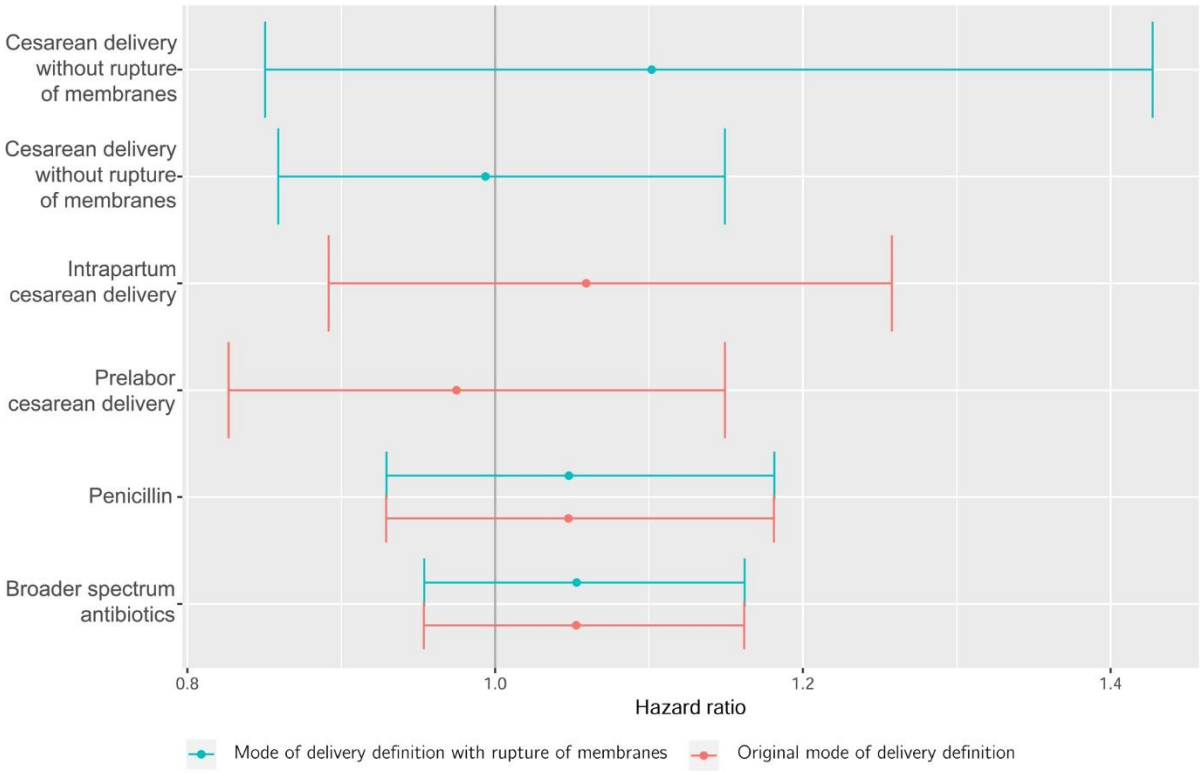
In eTable 5 we present a cross tabulation between the two variables. It should be noted that this table shows that the two variables were inconsistent: 19 children were categorized as vaginal births in one variable and cesarean delivery in the other. The variable on rupture of membranes was expected to be less reliable than the original mode of delivery variable.

New variable→ ***Original variable↓***	**Vaginal birth**	**CS with rupture of membranes**	**CS without rupture of membranes**	**Total**
Vaginal birth	553,718	0	19	553,737
Cesarean intrapartum	0	17,768	36,861	54,629
Cesarean prelabour	0	2,901	61,157	64,058
Total	553,718	20,591	97,297	671,606

eTable 5: A cross-tabulation of the two variables for birth method.

The exposure effect estimates obtained from two fully adjusted between-within models, that differ only by how their mode of delivery variables were defined, were almost identical for the antibiotics exposures (eFigure 5). For mode of delivery, we reached the same overall conclusion in both models (no effect of mode of delivery on autism), and there was also a large degree of pairwise similarities between the estimated effects of cesarean delivery without rupture of membranes and prelabour cesarean delivery, as well as the effects of cesarean delivery with rupture of membranes and intrapartum cesarean delivery. This corresponds well with the relation between the mode of delivery categorizations and the child's exposure to maternal gut microbiota for both choices of categorizations.

We concluded that there was no mentionable sensitivity towards the choice of defining mode of delivery relative to the timing of cesarean deliveries, rather than to the occurrence of rupture of membranes.



eFigure 5: Within-effect hazard ratio estimates and pointwise 95 % confidence bands from the two between-within models with confounder adjustment level 4, with the original mode of delivery definition (red) and the mode of delivery definition based on rupture on membranes (blue). Only exposure variable effects are shown. For both variables, the reference category is vaginal birth.

4. Sensitivity towards how the outcome variable is defined

We investigated the robustness of the model results towards how the outcome variable was defined. Specifically, we compared the following three variations of outcome variable definitions:

- **The original definition:** We registered the time to an autism diagnosis. Both type A (primary) and type B (secondary) discharge diagnoses were included. The number of events was then $n_e^{ori} = 8267$. We refer to this model as the *original* model below.
- **Only type A diagnoses:** Only children who at some point received a type A diagnosis were included as autism cases. The registered time of the event, however, was the time to the first diagnosis (A or B). If e.g. a child was diagnosed with autism at age 4 (type B diagnosis) and received a type A diagnosis of autism at age 7, the first time point counted at his/her event

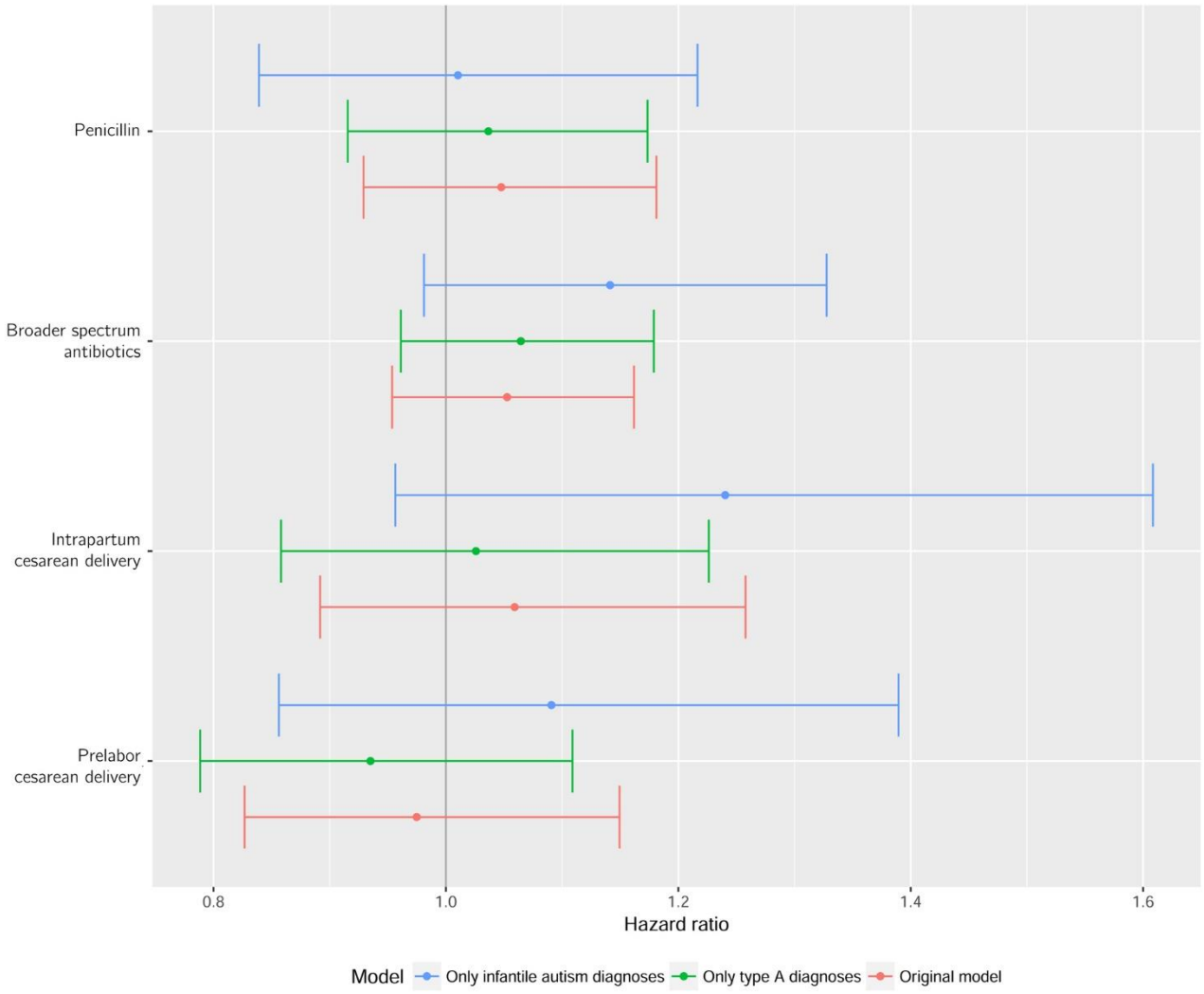
time. The number of events was $n_e^A = 7656$. We refer to this model as the *only A* model below.

- **Only infantile autism diagnoses:** We only registered the time to an infantile autism diagnosis, a subcategory of autism. Thus these diagnoses were also included in the original model. The number of events was $n_e^{aut} = 3500$. We refer to this model as the *only infantile* model below.

We fitted a fully adjusted between-within model for each of these outcome definitions. The within-effect estimates for the exposures are presented in eFigure 6. First of all, we saw broad confidence intervals for the *only Infantile* model, as expected. For all three models, the conclusion remained the same: No effects of neither mode of delivery nor antibiotics on the outcome.

For the antibiotics parameters, we see quite similar results across the three models and they agree on the direction of the relation between antibiotics use and the outcomes (antibiotics use implies an increased risk of the outcome event). For mode of delivery, we found that the *original* model and the *only A* model produced quite similar results, which was not very surprising, as the original model included only few outcome events that exclusively refer to a type B diagnosis (there are $n_e^{ori} - n_e^A = 8267 - 7656 = 611$ such cases). Their outcome variables were thus very similar. In the *only infantile* model, the direction of the relation with prelabour cesarean delivery was reversed: here, this mode of delivery implied an increased risk of the outcome. However, the effect was not significant and it was estimated with rather low precision, which was most likely due to the rarity of the infantile autism diagnosis.

In conclusion, we found differences in the estimated parameters, depending on how the outcome was defined, but these differences did not lead to differences in the overall conclusions of no effects of neither antibiotics nor mode of delivery on the outcomes.



eFigure 6: Within-effect hazard ratio estimates and pointwise 95 % confidence bands from the three fully adjusted between-within models with only the infantile autism diagnoses (blue), only type A diagnoses (green) and the original model (red). Only exposure effect estimates are displayed.

5. Sensitivity towards inclusion of additional obstetric variables

The dataset contained three obstetric variables that were suspected candidates for confounders, but whose validity was low. These variables were:

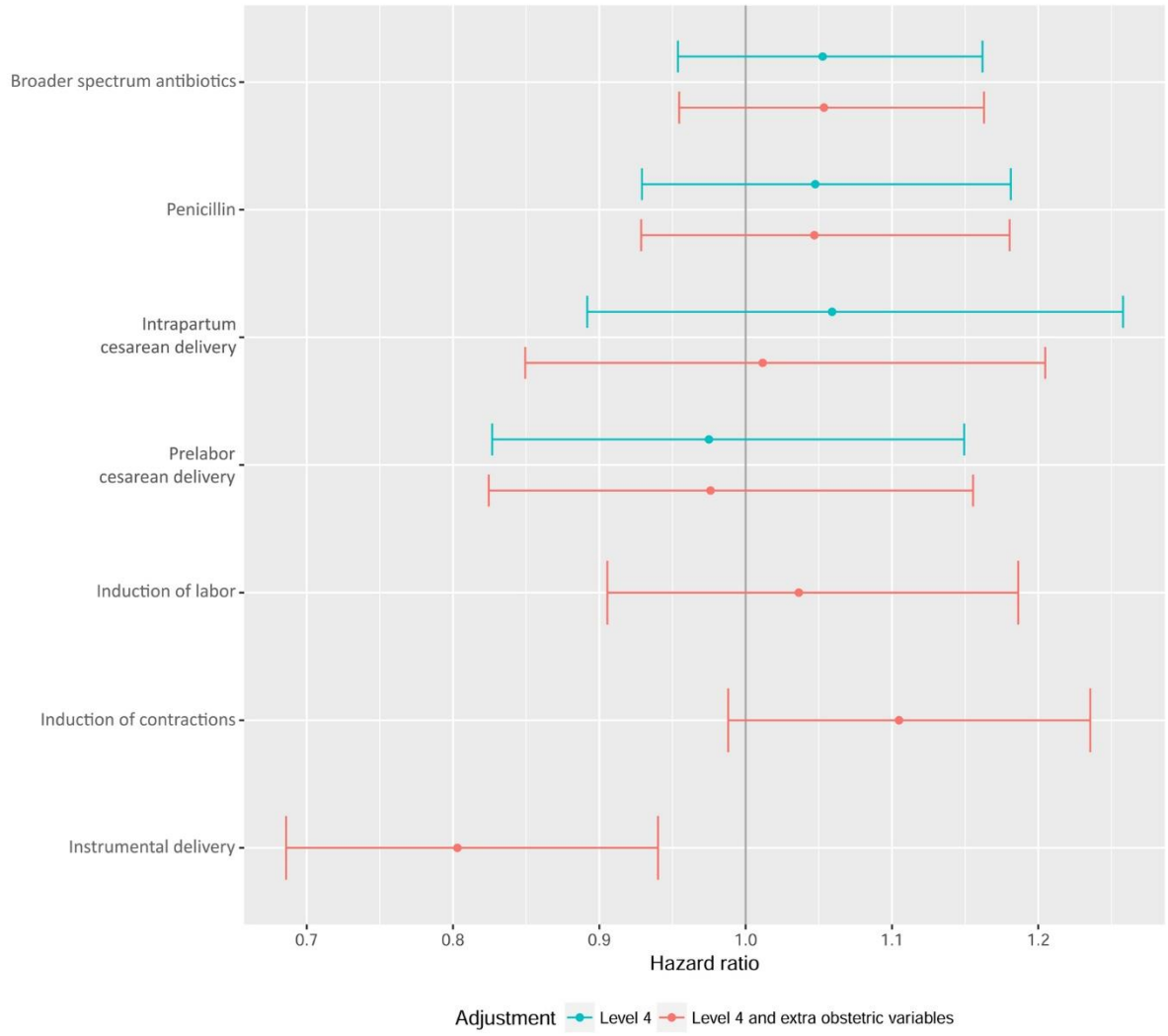
- Induction of labor
- Induction of contractions
- Instrumental delivery (use of vacuum extraction or forceps)

These variables could also not vary freely across the levels of the exposure variables. For instance, it was not quite clear what an instrumental, prelabour cesarean delivery would entail. Therefore, they were not included in the main analyses.

We investigated whether the results reported in the main article were sensitive towards the choice of *not* including these three additional obstetric variables in the primary models. We compared estimates from the usual fully adjusted between-within model with a similar model where the three extra variables have been added.

We found virtually no differences in the estimated exposure effects concerning antibiotic use and prelabour cesarean delivery (eFigure 7). For the effect of intrapartum cesarean delivery, we found only a slight attenuation of the effect estimate when the three extra variables were added to the model. In conclusion the modeling results were resistant to inclusion of the variables.

Regarding the estimates for the extra obstetric variables, we found no effects of either induction of labor or induction of contractions. However, we did find a significant, protective effect of instrumental delivery. However, it should be noted that our models were built to estimate unbiased effects of antibiotic use and mode of delivery and therefore, there is no guarantee that the effects of other included variables are estimated without bias, e.g. due to confounding. Therefore, one should hesitate to draw substantial conclusions regarding the effects of instrumental delivery on autism with reference to the results presented here.



eFigure 7: Within-effect hazard ratio estimates and pointwise 95 % confidence bands from two between-within models, one with confounder adjustment 4 (green) and one that also includes the three extra obstetric variables (induction of labor, induction of contractions and instrumental delivery) (red). Only the exposure effects and the new obstetric variable effects are shown.

6. Sensitivity towards how siblingship is defined

We investigated the sensitivity of our results in the main article towards how siblingship was defined. This was done by comparing three different versions of the fully adjusted (level 4) between-within model, differing only by how siblingship was defined. More specifically, we compared:

- Families defined by full siblings (the method used in the main article).
- Families defined by maternal siblings, i.e. children that have the same mother, but not necessarily the same father.
- Families defined by paternal siblings, i.e. children that have the same father, but not necessarily the same mother.

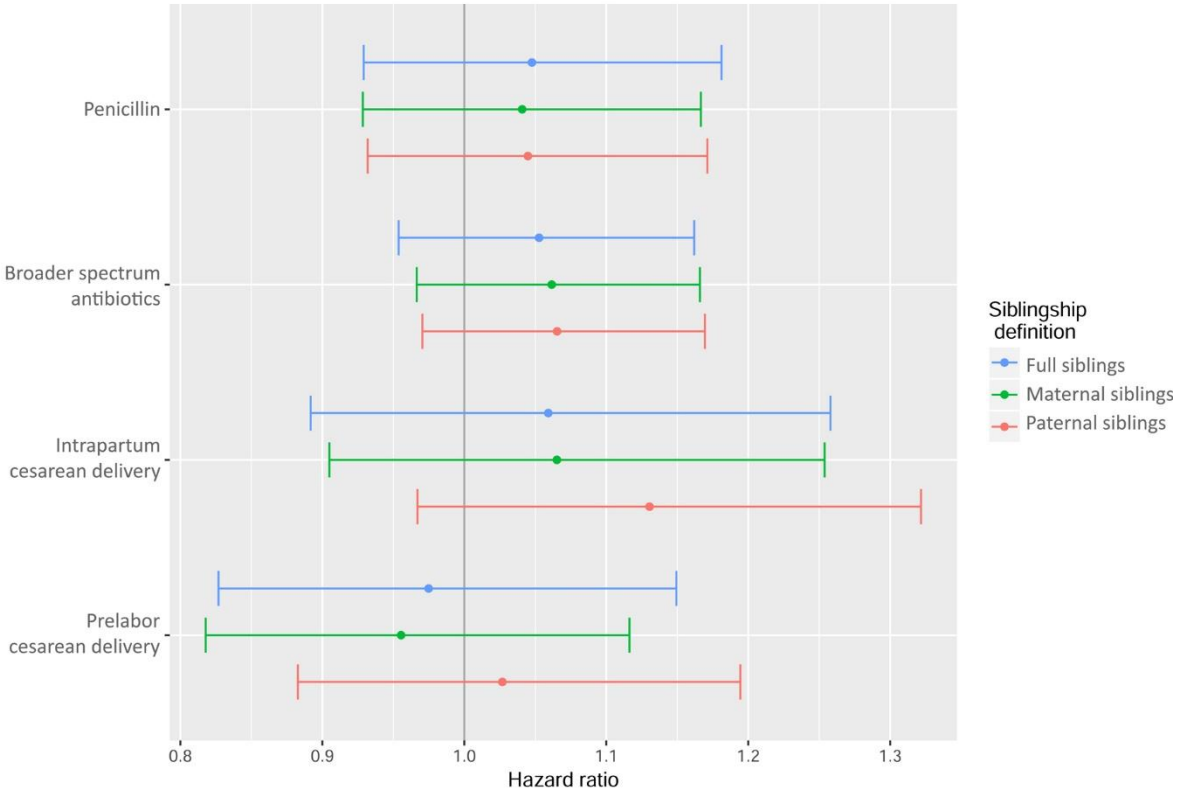
Note that as all full siblings were also both maternal- and paternal siblings, but not vice versa, there are more full sibling families than paternal- and maternal sibling families, respectively.

Information on how many families we had in the data of each type and the average sizes of these family types, can be found in eTable 6.

Estimates for the exposure effects obtained from models based on the three different family definitions can be seen in eFigure 8. For the antibiotics effects, we saw virtually no differences between the three models. With respect to the estimated effects of cesarean delivery, we found very similar results in the full- and maternal sibling models, but slightly different estimates in the paternal sibling model. This was unsurprising, considering that the cesarean deliveries are performed on different mothers. However, none of the models produced significant exposure effects and thus the conclusion from the main model was maintained: We found no significant effects of either antibiotics use or mode of delivery on the risk of developing autism, no matter how we defined siblingship.

	Full siblings	***Paternal siblings***	***Maternal siblings***
Number of families	439,718	422,248	420,137
Average family size (inc.s.)	1.53	1.59	1.60
Average family size (exc.s.)	2.19	2.22	2.22

eTable 6: The number of families of each type and the average number of children within one such family, including (inc.s.) or excluding (exc.s.) singleton families, respectively.



eFigure 8: Comparison of the within-effects of the fully adjusted between-within model, fitted three times with siblingship defined as full siblings (blue), maternal siblings (green) and paternal siblings (red). Only exposure effect estimates are shown.

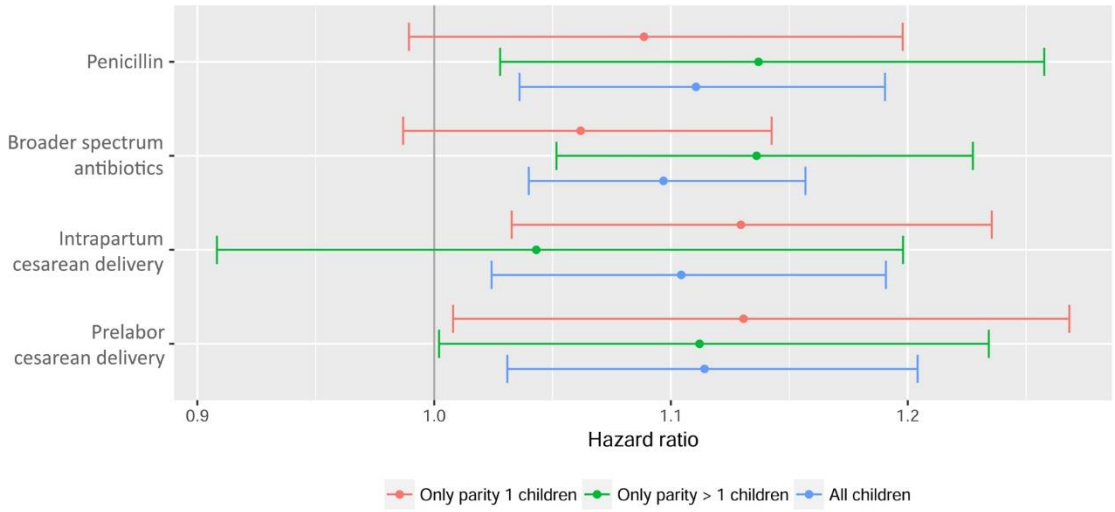
7. Investigating the influence of first born children

Other studies have found a connection between prevalence of autism in a child and the probability of that child getting younger siblings. In order to investigate whether this connection might impact our findings regarding the relation between mode of delivery, antibiotics use and autism, we have fitted the standard Cox model on two subsets of the dataset:

1. Only first born children (n = 296,800)
2. Only children with older siblings (n = 374,806)

We compared the results of these models with those of the "usual" fully adjusted (level 5) standard Cox model, fitted on the full dataset. These comparisons could not be conducted based on the sibling models, as it is not possible to fit sibling models on a subset consisting only of first born children; there can only be one such child per family and thus intra-family comparisons are not possible.

The exposure effect estimates can be seen in eFigure 9. We found differential exposure effects for the two subsets of the dataset, which indicated less of an effect of antibiotics and a larger effect of cesarean delivery for the first born children. However, there was no reason to believe that these effects would have persisted in a model that included information about e.g. family level unobserved confounders.



eFigure 9: Exposure effect estimates and pointwise 95 % confidence bands from three versions of the adjustment level 5 standard Cox model, varying by what subset of the data was used for estimation; only parity 1 children (first born children)(red), only children with parity > 1 (children with older siblings)(green) or all children (as done in the primary analyses)(blue).

8. Time trends in autism diagnoses

This last section is a first report on a detailed investigation of time trends which may later be elaborated into a separate study. It is included now to give background for the present manuscript.

Lastly, we investigated the robustness of the conclusions from the primary analyses towards the inclusion of additional calendar time effects. More specifically, we were interested in determining whether these conclusions would have been altered had we included birth year effects, diagnosis year effects or perhaps even both. This investigation was motivated by the patterns seen in the Lexis heatmap for autism incidence in the main article (Figure 2). The Lexis figure indicates some differences in autism incidence depending possibly on cohort as well as diagnosis year. Modeling such trends adequately might be important for the validity of the models if the time effects serve as confounders in the relation between autism, antibiotics and mode of delivery. If they do not,

however, they can safely be ignored. Note that as we were using time-to-event models, we were already including one time-aspect, namely offspring age.
Including diagnosis year effects in the sibling models was infeasible due to computational resources and little intra-family variation in diagnosis years. Therefore, we split this section into two parts: First, we used a resampling technique to investigate the interchangeability of birth year effects and diagnosis year effects in the standard Cox model.
Afterwards, we explored the necessity of including birth year effects in the full-data models with varying confounder adjustment levels. All in all, this gave us a dense, albeit pragmatic, description of the potential pitfalls associated with modeling the relation between mode of delivery, early childhood antibiotics use and autism without considering time effects beyond age.

Investigating the sensitivity towards inclusion of calendar time effects

We assessed whether the parameter estimates of interest, namely the effects of mode of delivery and antibiotics usage in the first two years of life on subsequent autism occurrence, were sensitive to whether or not calendar time adjustments were used in the models. Specifically, we compared four different types of adjustment for calendar time:

1. A model with a birth year effect (modeled categorically in one-year intervals)
2. A model with a diagnosis year effect (modeled categorically in one-year intervals)
3. A model with both a diagnosis year effect (modeled categorically in one-year intervals) and a birth year effect (modeled categorically in one-year intervals), entering additively
4. A model with no extra time effects beyond age

These models were constructed as extensions of the standard Cox regression model. Aside from time effects, we used two options for the inclusion of other explanatory variables:

- We included no other explanatory variables (confounder adjustment level 1)
- We included the variables used for confounder adjustment level 4, i.e. social variables, obstetric variables, variables related to maternal health and infections during the pregnancy

As stated above, the diagnosis year models could not be estimated as extensions of the sibling-models, as there was too little intra-family variation in the diagnosis years. Moreover, modeling diagnosis year involved using a time-dependent covariate, which increased the computational

complexity of the estimation procedure dramatically, as the data must then be arranged longitudinally. Because we were already working with quite large datasets, this implied that we actually could not perform the estimation procedure on the full dataset with the computer resources available at Statistics Denmark, where the data were located. However, there was another way of assessing if we needed to include the effects of diagnosis year. We took advantage of the non-linearity of the computational complexity as a function of sample size and used repeated resampling instead. Put more simply, although a longitudinal dataset with approximately 800,000 individuals was too large to work with using the available computer resources, estimating models on e.g. 80 longitudinal datasets of 10,000 individuals each was a heavy, but feasible, task, even though the same, total amount of individuals was covered.

Specifically, we carried out 100 independent repetitions of the following procedure:

1. Draw 10,000 random observations independently from the full sample. This qualifies as the current dataset.
2. Fit each of the four models described above on the current dataset and the save the results.

By doing independent repetitions of the procedure, we created a situation where asymptotic results, such as the law of large numbers and the central limit theorem, came into play. Therefore the distributions of the parameters obtained from these repeated resampling/reestimation procedures were centered around the true parameter values and the empirical variation in the parameters gave an insight into the variability, or width, of the population-level parameter distributions. Thus, we were able to conclude whether extra time effects were needed to obtain valid conclusions in the current study (see below).

Consequences for the exposure effect estimates

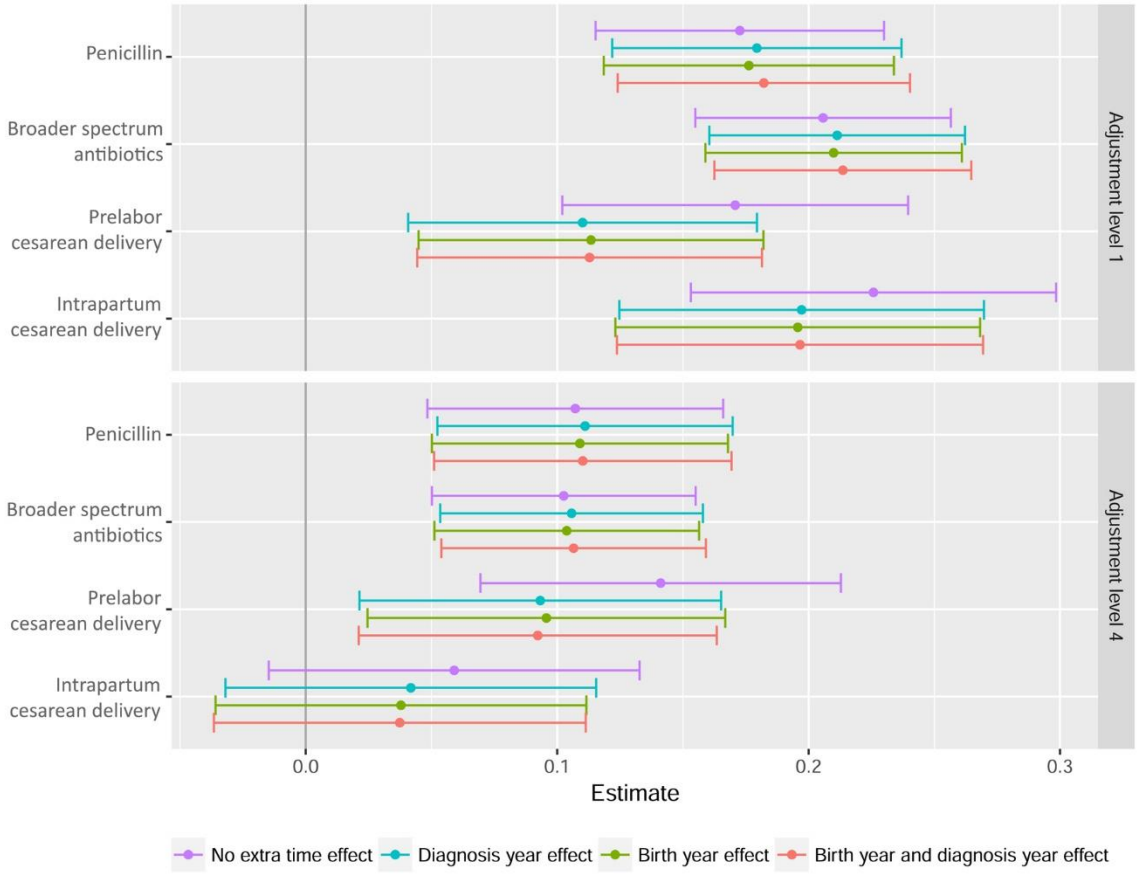
We inspected the findings concerning the exposure variable effects. The exposure variables were mode of delivery (vaginal birth, prelabour cesarean delivery or intrapartum cesarean delivery) and antibiotics use in the first two years of life (no antibiotics, exclusively penicillin or broader spectrum antibiotics) modeled as categorical, additive effects. Vaginal birth and no antibiotics use served as the reference categories.

The parameter estimates from all the 100 resampling/reestimation steps can be seen in eFigure 10. More specifically, for each parameter estimate in each model type, we have calculated the mean estimate and added a pointwise 95% Wald confidence interval where the variance was estimated as

the empirical resampling variance. Due to the central limit theorem, we could assume asymptotic normality of the resampled parameter estimates and thus these confidence intervals should represent the sampling error adequately.

The effect estimates related to mode of delivery were attenuated when we added either adjustment for diagnosis year or birth year. This suggests that the relation between this exposure and autism is likely to be confounded by such additional calendar time variables. This means that a model that does not take into account such variables is likely to produce biased exposure effect estimates. We found no systematic differences for the estimates related to antibiotic use. These tendencies were found in both the adjustment level 1- and adjustment level 4 models.

A natural next step, which we will address below, was to investigate whether it was necessary to include diagnosis year effects if the model was already adjusted for birth year. This was important, as it was not practically possible to estimate diagnosis year effects in the sibling models due to too little intra-family variation (which was likely to be related to how rare autism was in the full dataset).



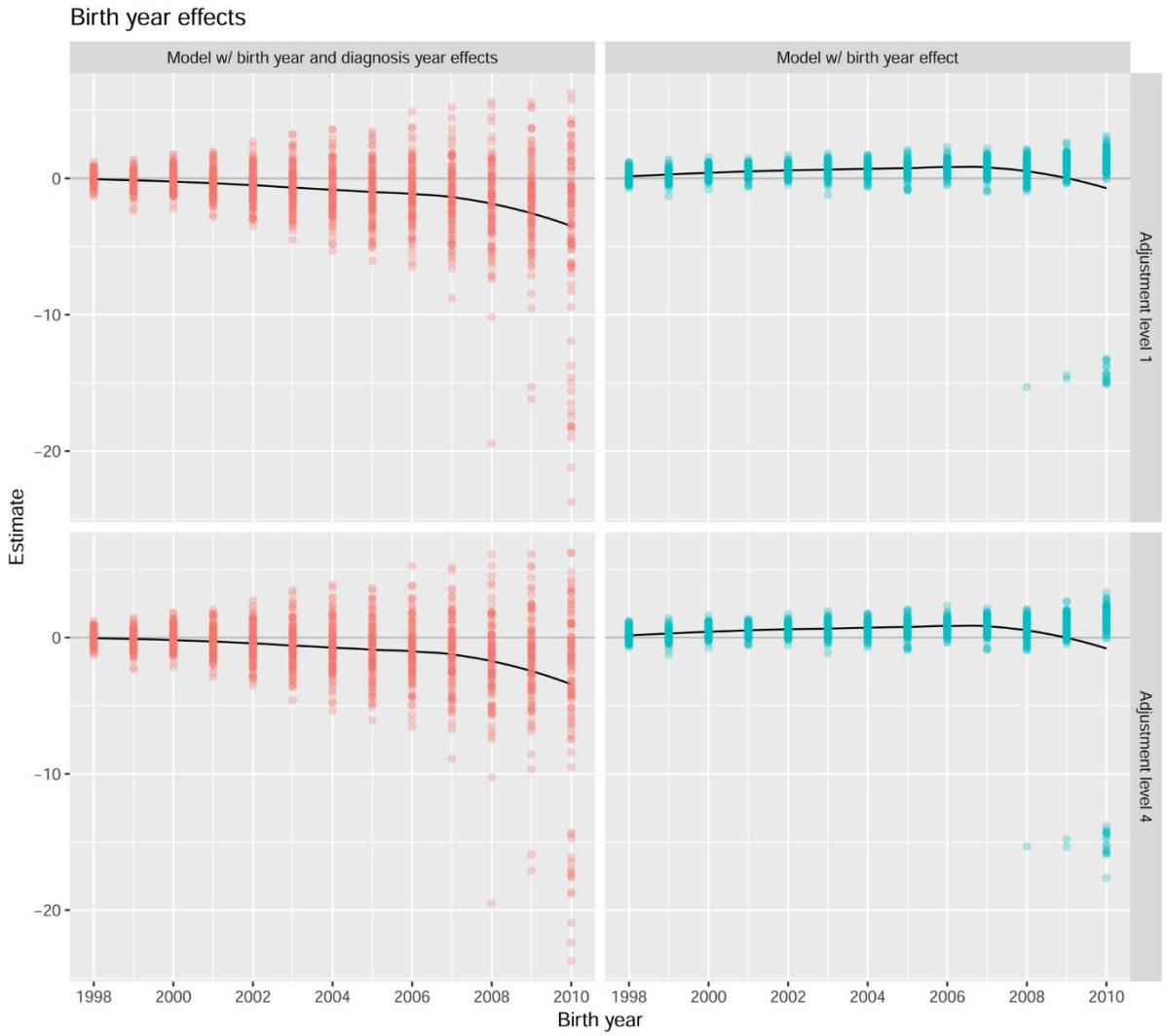
eFigure 10: Empirical means of the parameter estimates from the 100 resampling/reestimation runs using confounder adjustment level 1 and 4 and various strategies for handling extra time effects beyond offspring age: original model (purple), diagnosis year effect (blue), birth year effect (green), birth year and diagnosis year effect (red). The error bars mark pointwise 95% Wald confidence intervals based on the empirical resampling standard deviations.

Time effects

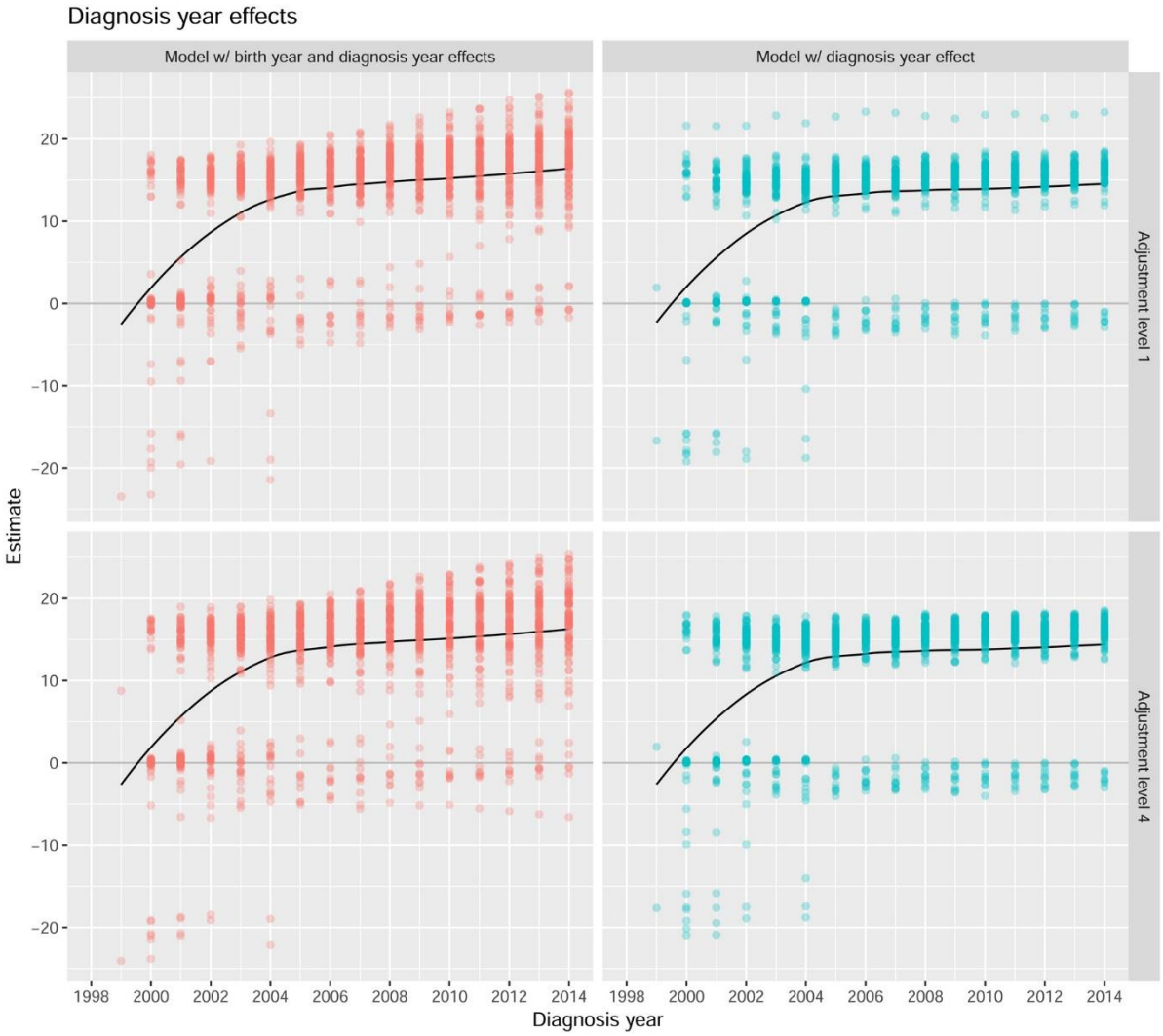
We then looked at the estimated time effects from the birth year- and diagnosis year models. They are illustrated in eFigure 11 (for birth year effects) and eFigure 12 (for diagnosis year effects). Each dot represents the parameter estimate from a single resampling/reestimation step.

We generally found little sensitivity towards the confounder adjustment level used. For both types of calendar time effects, we saw similar shapes across the two model types (either including only a single calendar time effect or both calendar time effects), but the dual time effect models generally produced numerically larger estimates, but with more variability.

Note that due to the resampling technique and the rareness of the outcome, some datasets did not have sufficient information to validly estimate all diagnosis year and birth year effects, especially not those near the boundaries. For the diagnosis year/birth year model with confounder adjustment level 4, 15 estimates were smaller than -50 and they have been excluded from the plot in order to enhance readability.



eFigure 11: Parameter estimates of the birth year effects on their original scale. Each dot represents one resampling/reestimation step. Loess smoothers have been added in black to emphasize the tendencies in the estimates.



eFigure 12: Parameter estimates of the diagnosis year effects on their original scale. Each dot represents one resampling/reestimation step. Loess smoothers have been added in black to emphasize the tendencies in the estimates. 15 estimates that were smaller than -50 have been excluded before plotting and smoothing. These all stem from the model with both birth year and diagnosis year effects.

Full sample models with birth year effects

We then inspected the consequences of including birth year effects in the full-sample models. We went through the three model classes (the standard Cox model, the stratified Cox model and the between-within model) in turn. Note that all parameter estimates in this section were on hazard ratio scale for comparability with the results reported in other chapters. This meant that the estimates here were exponential transformations of the estimates from the resampling models.

Standard Cox models with birth year effects

First, we discuss the interrelation between the usual confounders and birth year effects in the standard Cox model when using the full dataset. More specifically, for each level of confounder adjustment, we compare standard Cox models with and without a birth year effect. We focus on birth year effects because they were estimable in the sibling models that serve as the primary models. In eFigure 13, we present the findings comparing the ten standard Cox models (two for each confounder adjustment level). We saw results that were very similar to those of the resampling models in the above: Mode of delivery seemed to be confounded by birth year, but antibiotic use did not.

With increasing levels of adjustment for other confounders, the difference between the birth year and non-birth year models were somewhat attenuated, but they were still present in the fully adjusted model. We found more pronounced differences for prelabour cesarean delivery than for intrapartum cesarean delivery. The estimated birth year effects from the full-sample standard models with adjustment level 1 and 5 can be seen in eFigure 14. We saw a general rise in the risk of autism over time. This corresponded very well with the findings from the resampling models for the years up until 2007 (be aware of the difference in scales compared with the figure in the above). For births from 2007 onwards, the resampling models produced a trend towards a decrease in the risk of autism, but as is evident from eFigure 11, this decrease was mainly brought about from a few extreme resamplings that produced very small estimates (HRs of $\simeq \exp(-15) = 3 \cdot 10^{-7}$). This was very to be an artefact of the combination of the resampling strategy and the rarity of the outcome diagnosis, autism.

All in all, we found results very similar to those of the resampling method. Therefore, it is plausible that the results concerning diagnosis year effects, which were not possible to reproduce on the full sample due to computational resources, will also generalize to the full sample. This means that investigations into the birth year effects were acceptable compromises in the sibling models, where

diagnosis year effects cannot be identified. Moreover, it showed that (*at least*) birth year adjustments *must* be included if one is to describe the relation between mode of delivery and autism in a standard Cox model set-up.

Stratified Cox models with birth year effects

We investigated the impact of including birth year effects in the stratified Cox model. In eFigure 15, we compared eight stratified Cox models, two models for each confounder adjustment level: One with and one without birth year effects. We saw that the difference between the birth year adjusted and the non-birth year adjusted models decreased with increasing levels of confounder adjustment. This suggested that we might have adjusted for other confounders that are highly correlated with birth year, and thus we were possibly also adjusting for birth year indirectly already. Generally, we found only little differences in the birth year- and non-birth year models. This indicated that the causal pathway of birth year confounding has been broken by one or more unobserved intra-family confounders that were now taken into account by design. Therefore, not including time effects in the stratified Cox model seemed to be much less problematic compared to the standard Cox model, and - most importantly - the overall conclusions regarding the effects of mode of delivery and antibiotics on autism risk were unchanged.

The estimated birth year effects from the stratified Cox models are available in eFigure 16. Just like in the standard model, we saw little difference between the fully- and minimally adjusted models. Both models showed an overall increase in the risk of autism for children born up until 2003, but a decline hereafter. Note however that the uncertainty was also increasing a lot. So even though the survival models did take the short follow-up time of the later cohorts into account, there was not much data on which to base the estimates of the later birth year effects: Most children had not got old enough by end of follow-up to obtain an autism diagnosis. Most importantly, it should be noted that the confidence limits of the level 4 adjustment model did not exclude the possibility of null-effects for all birth year variables.

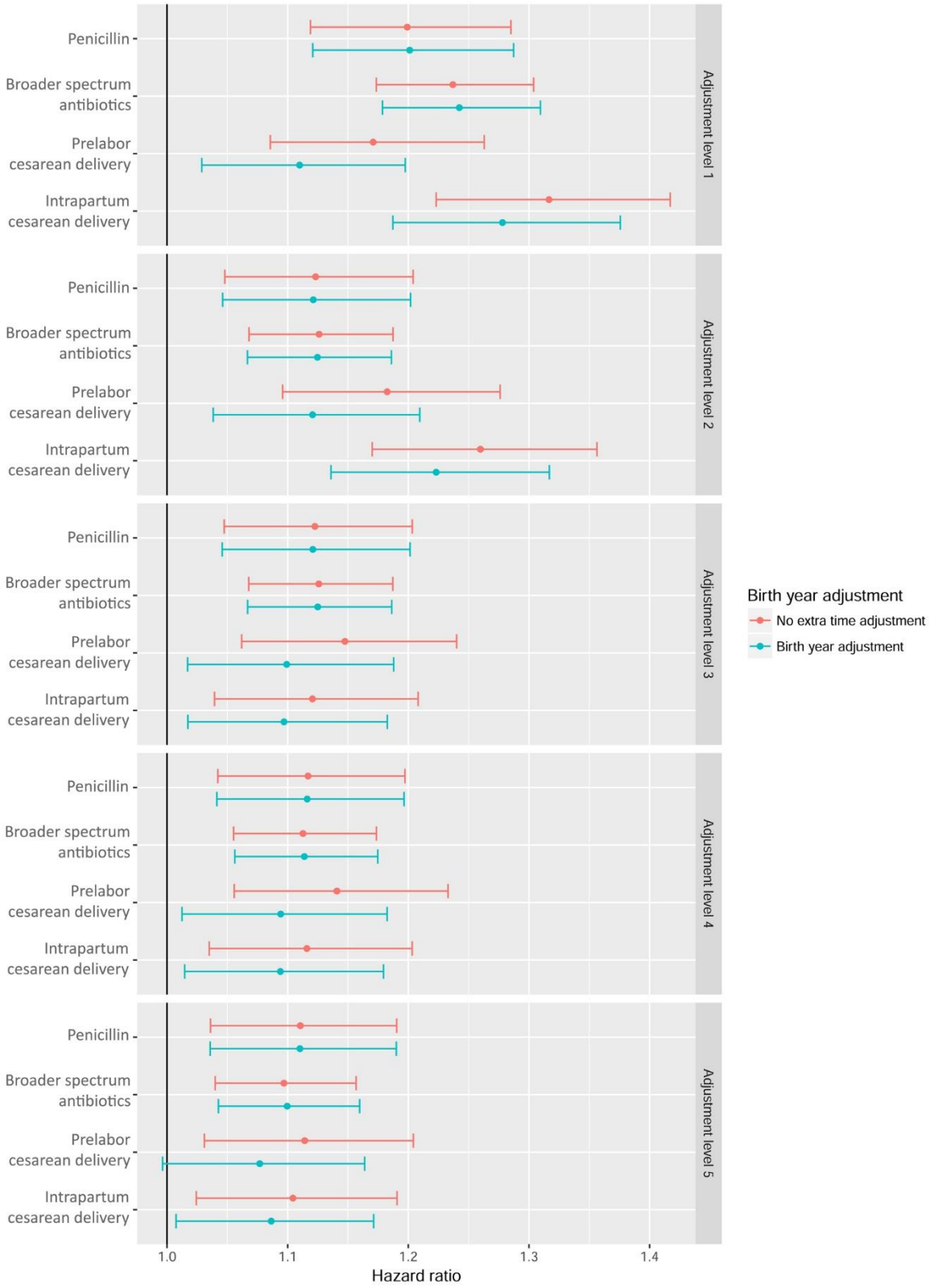
Between-within models with birth year effects

Lastly, we investigated the effects of including birth year effects in the between-within models. We present results from models with no other confounder adjustment (level 1) and full confounder adjustment (level 4). In eFigure 17 we present the within-type exposure effect estimates from these models with and without birth year effects. We found less of a tendency towards differences in the

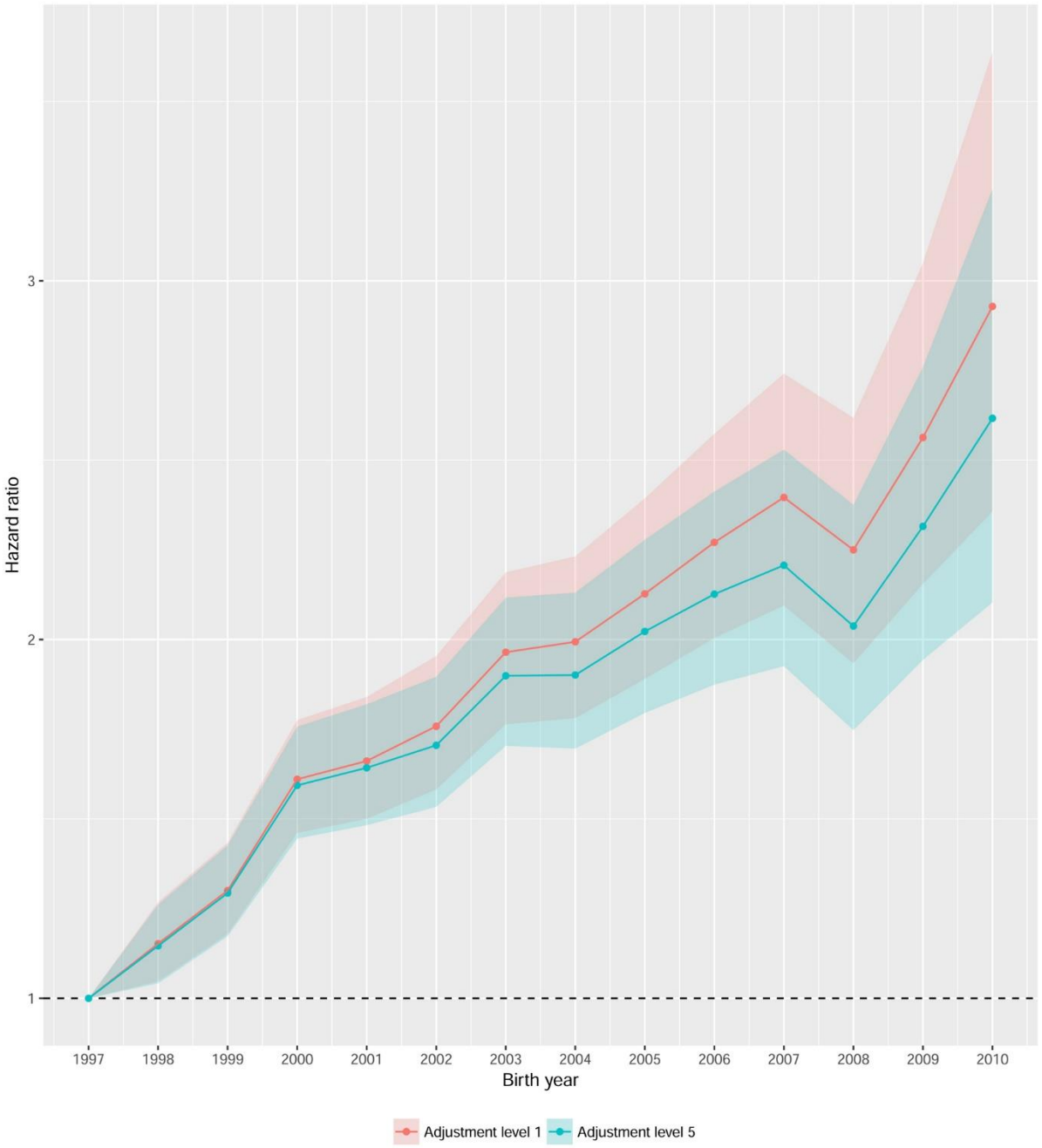
birth year/non-birth year models here, even for the level 1 adjustment model. This investigation thus also indicated that adjustments for birth year confounding was not necessary when intra-family unobserved confounders were taken into account in the sibling models. In eFigure 18 we present the estimated within-type birth year effects from the two between-within models including such effects. We saw a marked trend towards an increased risk of autism for later birth years, especially in the fully adjusted model.

Conclusion

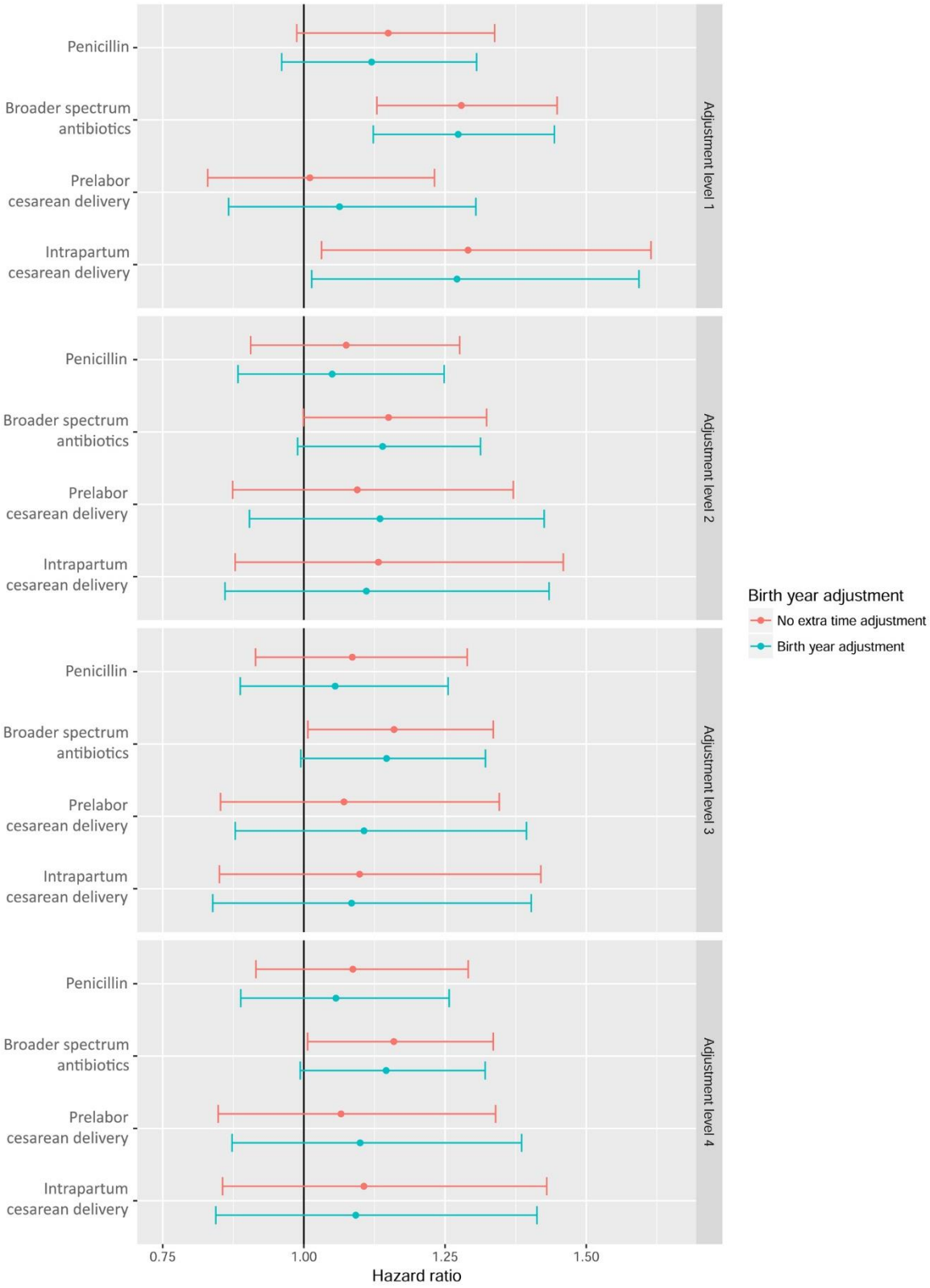
Using resampling methods for the standard model, the inclusion of either diagnosis year or birth year effects attenuated the observed effects of cesarean delivery on autism risk. The effects of including birth year effects in the sibling models however, showed little differences. This suggests that the sibling designs implicitly control for something that is either on the causal pathway from calendar time to mode of delivery or from calendar time to autism, thereby blocking the confounder. In particular, we urge that until a deep understanding of *how* calendar time and autism interplay is obtained, all studies of autism should include a consideration of calendar time effects.



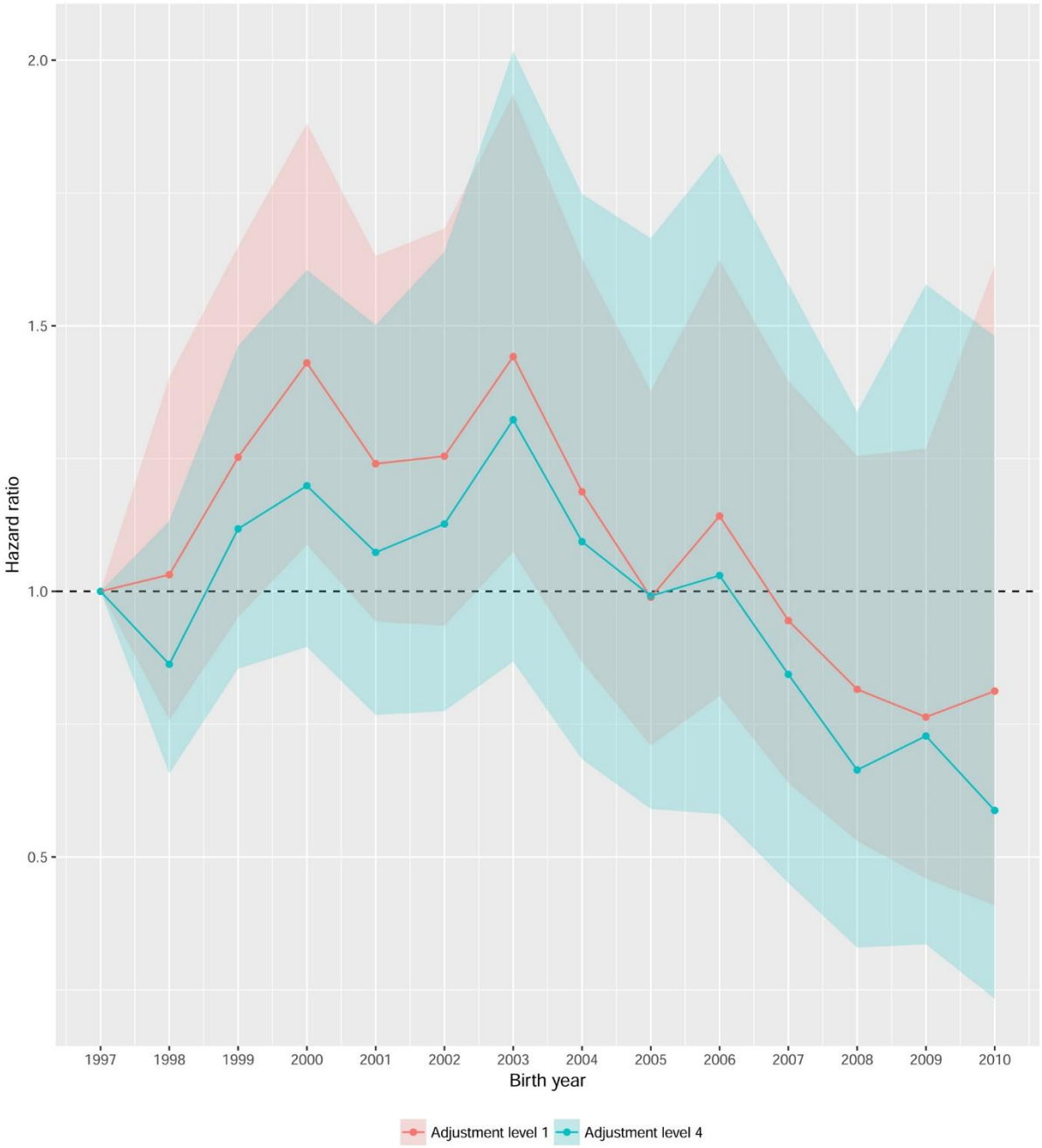
eFigure 13: Exposure effect estimates from ten standard Cox models, using all combinations of the five different options for confounder adjustment levels and the two options for including (blue) not including (red) extra adjustment for birth year.



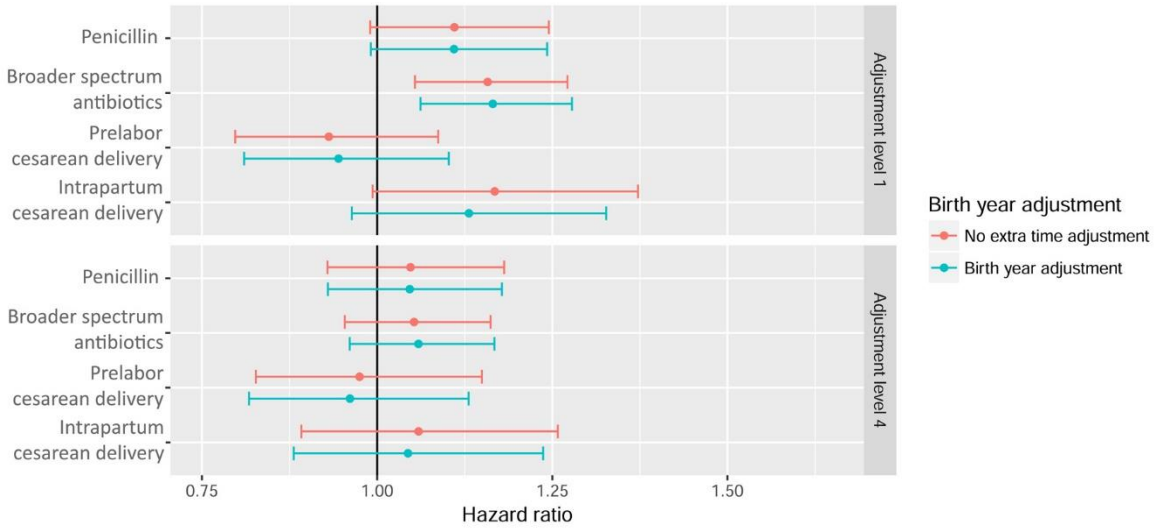
eFigure 14: Birth year effects, as estimated by the standard Cox model with varying degrees of adjustment for other confounders. The hazard ratio estimates are presented along with pointwise 95 % confidence limits in transparent ribbons. The birth years are modeled as categorical effects and the year 1997 serves as the reference category, which explains its (non-estimated) value of one.



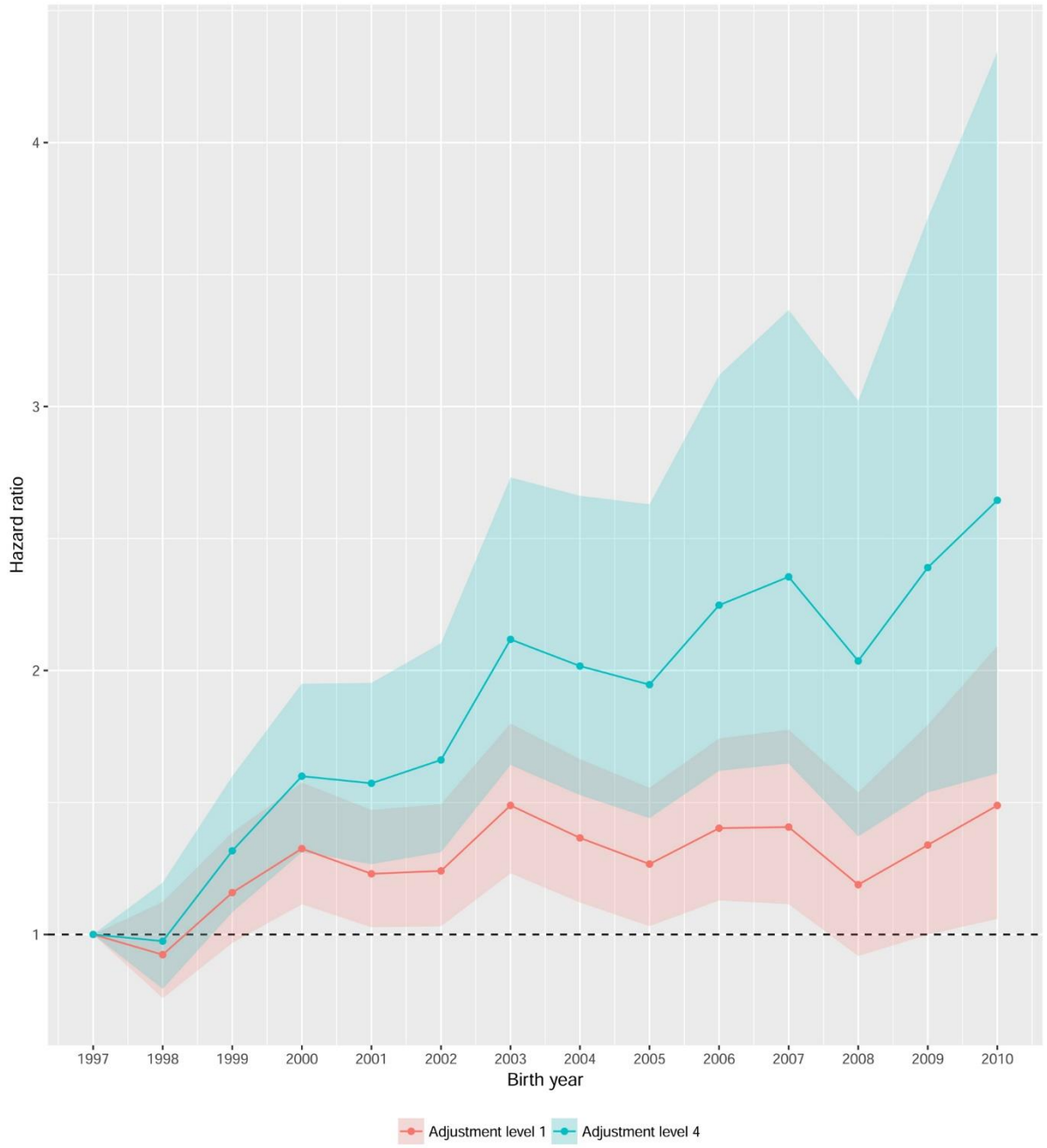
eFigure 15: Exposure effect estimates from eight stratified Cox models, using all combinations of the four different options for confounder adjustment levels and the two options for including (blue) or not including (red) extra adjustment for birth year.



eFigure 16: Birth year effects, as estimated by the stratified Cox model with varying degrees of adjustment for other confounders. The hazard ratio estimates are presented along with pointwise 95 % confidence limits in transparent ribbons. The birth years are modeled as categorical effects and the year 1997 serves as the reference category, which explains its (non-estimated) value of one.



eFigure 17: Within-type exposure effect estimates from four between-within models, using either no- or full confounder adjustment and the two options for including (blue) or not including (red) extra adjustment for birth year.



eFigure 18: Birth year effects, as estimated by the between-within model with varying degrees of adjustment for other confounders. The hazard ratio estimates are presented along with pointwise 95 % confidence limits in transparent ribbons. The birth years are modeled as categorical effects and the year 1997 serves as the reference category, which explains its (non-estimated) value of one.

Paper III

Title Page

Is cesarean section a cause of affective disorders?- A national cohort study using sibling designs

Authors and affiliations:

Paul Bryde Axelsson[1], Anne Helby Petersen[2], Ida Hageman[3], Anja Bisgaard Pinborg[4,6], Lars Kessing[3,6], Thomas Bergholt[5], Steen Christian Rasmussen[1], Niels Keiding[2], *Tine Dalsgaard Clausen[1,6], *Ellen Christine Leth Løkkegaard[1,6].

*Tine Dalsgaard Clausen and Ellen Christine Leth Løkkegaard should be considered joint senior author

1. Nordsjællands Hospital, University of Copenhagen, Department of Gynaecology and Obstetrics.
2. University of Copenhagen, Department of Public Health, Section of Biostatistics.
3. Rigshospitalet, Copenhagen University Hospital, Psychiatric Center Copenhagen.
4. Rigshospitalet, Copenhagen University Hospital, Fertility Department, The Juliane Marie Centre.
5. Rigshospitalet, Copenhagen University Hospital, Department of Obstetrics.
6. University of Copenhagen, Department of Clinical Medicine.

Corresponding author:

Name: Paul Bryde Axelsson
Address: Dyrehavevej 29, 3400 Hillerød, Denmark
Telephone: 00 45 2011 5532
E-mail: paulbryde@gmail.com

Is cesarean section a cause of affective disorders?- A national cohort study using sibling designs

Abstract

Background: The gut microbiota of children delivered by cesarean section differs from that of children delivered vaginally. In light of the gut-brain axis hypothesis, cesarean section may influence risk of affective disorders.

Methods: Population based prospective cohort study included Danish children born 1982 through 2001, with follow-up until 2015. The effect of delivery mode on the risk of affective disorders was assessed using a standard Cox model and two types of Cox sibling models. Diagnostic codes or prescriptions for antidepressants and lithium were used to define cases of affective disorders.

Results: 1,009,444 children were followed for 8,880,794 person-years from the age of 13 years, with relevant covariates available from birth. There are strong calendar time trends in the occurrence of affective disorders with an increasingly younger age at first diagnosis and with a hotspot between the years 2007-2012. Fully adjusted standard Cox models showed an increased risk of affective disorders for both pre-labor (hazard ratio [HR], 1.11; 95% confidence interval [CI], 1.08–1.15) and intrapartum (HR, 1.07; 95% CI, 1.05–1.10) cesarean section, compared to vaginal delivery. This effect disappeared in the between–within sibling model for pre-labor (HR, 1.00; 95% CI, 0.94–1.07) but not intrapartum (HR, 1.05; 95% CI, 1.00–1.12) cesarean section.

Limitations: Interpretation of results from sibling models may not be relevant to children without siblings.

Conclusions: These results do not support the hypothesis that a delivery-mode dependent change in gut microbiota is a cause of subsequent affective disorders, despite an apparent association with delivery mode.

Keywords

Microbiota; Depression; Mood Disorders; Cesarean Section; Epidemiologic Research Design; Proportional Hazards Models

Conflicts of interests:

All authors report no conflicts of interest.

Highlights

- The gut microbiota has been hypothesized to impact risk of affective disorders
- Changes to gut microbiota may manifest early in life and depend on delivery mode
- Using Danish national registers we compared siblings discordant on delivery mode
- We saw no evidence of a causal effect of pre-labor cesarean on affective disorders
- For intrapartum cesarean the risk was slightly increased, maybe due to confounding

1. Introduction

Recent studies have suggested that the gut microbiota may be associated with mood disorders (Coello et al., 2019; Liu and Zhu, 2018; Vinberg et al., 2019). In general, the etiology of mood disorders is regarded as multifactorial with genetic factors, traumatic life events, and alcohol and drug abuse among the most decisive factors (Bock et al., 2009; Boden and Fergusson, 2011; Di Florio et al., 2014; Etain et al., 2008; Lev-Ran et al., 2014, p.; Misiak et al., 2017; Wilde et al., 2014). However, over the past few decades, the incidence of affective disorders and the use of antidepressants may have increased (Bachmann et al., 2016; Jensen and Steinhausen, 2016; Medici et al., 2015; Pottegård et al., 2014; Weinberger et al., 2017). During the same period, the proportion of deliveries performed by cesarean section (CS) has also increased (Betrán et al., 2016), particularly elective CS deliveries (Clausen et al., 2016; Vogel et al., 2015).

These concurrent trends may be causally linked (Cenit et al., 2017; Dinan and Cryan, 2015; Neu and Rushing, 2011). One prominent theory, supported by animal studies (Forsythe et al., 2016; Leclercq et al., 2017), suggests that human gut microbiota may influence neurodevelopment and mood (Tognini, 2017; Winter et al., 2018; Yang et al., 2016). Indeed, a recent study showed that the gut microbiota of subjects with depression differed from those of subjects who did not have depression (Valles-Colomer et al., 2019). Another recent, but small, randomized trial found that subjects with an acute episode of mania were less likely to be readmitted to hospital when treated with oral probiotics (Dickerson et al., 2018).

These changes in the microbiota may occur early in life. The gut microbiota of children delivered by CS differ from those of children delivered vaginally. For those delivered by CS, the acquisition of gut microbiota is mainly mediated by the children's surroundings or by skin-to-skin contact with their mothers (Yang et al., 2016). In addition, some bacterial species, such as Clostridium difficile, more commonly colonize infants delivered by CS than those delivered vaginally (Rutayisire et al., 2016). Clostridium difficile occurs frequently in hospitals and can have detrimental effects on infant health (Pandey et al., 2012; Vael and Desager, 2009). Furthermore, the relative abundance of Bifidobacteria, a bacterial family thought to benefit the infant immune response (Underwood et al., 2015), is low in children delivered by CS. The long-term effects of these differences in microbiota have not been investigated, but accumulating evidence suggests they are associated with different health outcomes (O'Callaghan and van Sinderen, 2016; Yang et al., 2016).

Epidemiological studies on the effects of CS delivery on a child's long-term health have shown that this delivery mode is associated with an increased risk of asthma, allergies, inflammatory diseases, impaired cognitive function, and obesity (Aagaard et al., 2016; Bager et al., 2008; Huang et al., 2015; Polidano et al., 2017; Sevelsted et al., 2015). However, as with most observational studies, the question of causality remains. The objective of this study was to investigate the effect of a CS delivery on a child's risk of subsequent affective disorders.

We will do so by using a sibling comparison design. This modelling strategy aims to control for unmeasured confounding by utilizing the fact that siblings mostly share parental genetic makeup and the same early life environment. However, sibling models only use information from children who have siblings and only if there is sufficient intrafamily variation (Axelsson et al., 2018a). Therefore, these studies may lack statistical power. The large Danish national registers afford a valuable opportunity to follow an entire population for several decades and provide a strong basis for robust sibling analyses. In this study, we apply two different

models for sibling comparisons with time-to-event outcomes: the stratified Cox model and the between–within Cox model. The latter model is possibly more precise, but less commonly used (Sjölander et al., 2013). We compare the results from these two models with the results of a more traditional non-sibling stratified Cox model (hereafter the standard Cox model).

2. Method

2.1 Study population

Our cohort was established by linking seven Danish national registers (eTable 2) using personal registration numbers, which are unique for every Danish citizen. The data were pseudo-anonymized and could be accessed by the authors through Statistics Denmark (eTable 2).

The exclusion criteria and the procedure used to define the study population are shown in Figure 1. Initially, all children born alive in Denmark between 1 January 1982 and 31 December 2001 (N = 1,228,333) were identified. We then included the 1,009,444 Danish children who were singleton births, had Danish parents, had complete information for all variables of interest and had survived until the start of the follow-up period. These children were then followed from 30 days after their 13th birthday for a total of 8,880,794 person-years until an affective disorder was diagnosed or censoring at death, emigration, diagnosis of a higher ranking psychiatric disorder (ICD being hierarchical), or until the end of the follow-up period on 31 December 2014.

2.2 Study parameters

The mode of delivery exposure variable was separated into three categories: pre-labor CS, intrapartum CS, and the reference category of vaginal delivery (eTable 1).

The primary outcome was any episode of an affective disorder, based on either a diagnosis or a redeemed prescription. Diagnoses were selected according to the International Classification of Diseases (ICD) 10th revision. We used the classification codes DF30–DF33 and DF38.00. Admissions and outpatient hospital contacts were included, and both primary and secondary discharge diagnoses were used. Two redeemed prescriptions of either lithium or any antidepressant that were at least 30 days apart also constituted an episode of an affective disorder and were defined using the Anatomical Therapeutic Chemicals (ATC) codes N05AN1 and N06A, respectively.

Event times were defined as the earliest date of any case of an affective disorder. To ensure that all individuals could potentially experience both dimensions that define an event, we started the follow-up period at 30 days after the 13th birthday. To qualify as a case of an affective disorder, patients had to have two redeemed prescriptions on separate dates. Prescriptions are usually renewed at approximately 30 days interval. Additionally, information on prescriptions was only available from 1995 onwards and, at this time, children born in 1982 would be 13 years of age. Therefore, patients born on 1 January 1982 would not be at risk until 30 days after their 13th birthday.

2.3 Statistical analyses

All statistical analyses were performed using the survival package in R software (ver. 3.3.1; R Development Core Team, 2016) with additional programming by the authors.

We used three different statistical models: the standard Cox model, the sibling stratified Cox model (hereafter the stratified Cox model), and a between–within gamma-Cox model for siblings (hereafter the between–within model). We have previously discussed these models (Axelsson et al., 2018a, 2018b), and include only a short summary here. The sibling models target within-family effects (hereafter within-effects), which are generally not the same as the effects targeted by the standard Cox model because the former are adjusted for family-shared confounding, whereas the latter are not. This is because the standard Cox model bases its estimates on comparisons of individuals and ignores any intrafamily dependence. The stratified Cox and between–within models only compare individuals with their siblings, thereby eliminating bias due to unobserved confounding factors that are shared within families. On the other hand, this procedure reduces statistical precision, most prominently for the stratified Cox model. The between–within model makes more assumptions about the structure of the underlying hazard distribution, which potentially increase precision (Sjölander et al., 2013).

Observed potential confounding variables for the effects of mode of delivery on affective disorders were adjusted progressively at five nested adjustment levels (Table 1).

We conducted further analyses to test the robustness of our findings in the between–within models relative to the effect of calendar time on the outcome variables, the definition of an affective disorder outcome, and how the delivery modes were defined. These analyses are discussed in more detail in the online appendix.

2.4 Data Availability

By Danish law, the authors are not permitted to share person-level data. Anyone can request access to the data, by first acquiring permission from the Danish Data Protection agency and afterwards the Danish Health Authority. However, these government instances have strict requirements to be allowed access to medicinal information that most foreign identities do not live up to, and therefore cooperation with researchers from a Danish University or a Hospital is recommended if this particular dataset is of interest. The data is located at the Statistics Denmark servers and requires an application defining the persons to be in the cohort and which variables to be extracted from the national registries. Data can only be accessed through the Statistics Denmark servers.

3. Results

3.1 Demographic characteristics

In our study population, 7.4% of the live-born children were delivered by intrapartum CS, whereas 5.2% were delivered by pre-labor CS. Demographic data for the study population are presented in Table 2, stratified according to mode of delivery. Children delivered by intrapartum CS were more likely to be male, have older siblings, and be born to mothers with pre-eclampsia and a previous psychiatric history, compared to children born vaginally. Children delivered by pre-labor CS were more likely to have more than

one older sibling and be born to mothers who were older, married, have pre-eclampsia, a previous psychiatric history, and husbands who were more than 5 years older than they were, compared to children born vaginally.

Considerable changes in the incidence of affective disorders were observed over time for the cohort, with a peak in cases between the years 2007 and 2012 and a trend toward younger age at first diagnosis (Fig. 2). These patterns are mainly attributable to prescription practices. In contrast, diagnoses increase sharply for subjects aged 15–24 years from 2008 until the end of the follow-up period (eFigs. 3 and 4).

The unadjusted risks of affective disorders were similar for all modes of delivery: 10.8% for intrapartum CS, 9.5% for pre-labor CS, and 10.1% for vaginal delivery.

3.2 Sibling models

Delivery mode did not have a strong effect on the occurrence of affective disorders. For the fully adjusted between–within model, pre-labor CS showed no effect (hazard ratio [HR], 1.00; 95% confidence interval [CI], 0.94–1.07) and intrapartum CS showed a slightly significant increase (HR, 1.05; 95% CI, 1.00–1.12) for the risk of affective disorders, both compared to vaginal delivery (Fig. 3). eTable 3 shows exposure effect estimates for all the fully adjusted models.

The stratified Cox model showed a very similar pattern to that described for the between–within model. The estimates were slightly less precise, and all estimates remained nonsignificant across all levels of confounder adjustment: pre-labor CS (HR, 0.98; 95% CI, 0.91–1.06) and intrapartum CS (HR, 1.06; 95% CI, 0.99–1.13) for the fully adjusted models (Fig. 3).

Being married, older maternal age, greater parity, and female infants were all associated with an increased risk of affective disorders in the offspring. Other included covariates had no significant effect on outcomes (eFig. 1).

3.3 Standard Cox

In the standard model, the risks of affective disorders for both pre-labor (HR, 1.11; 95% CI, 1.07–1.15) and intrapartum CS (HR, 1.07, 95% CI, 1.05–1.10) were increased, compared to vaginal delivery. Adjustment for all available confounders had only a minor effect on the pre-labor CS estimate but an increased effect on the intrapartum CS estimate in the fully adjusted model (Fig. 3).

3.4 Robustness of the results

In general, the sensitivity analyses did not uncover anything to change the overall conclusions from the primary analyses. We did observe a very strong period (event year) effect on the incidence of affective disorders. However, this did not seem to confound the relationship between mode of delivery and the risk of affective disorders. When affective disorder outcomes were separated into cases based on prescriptions and diagnoses, we found that being delivered by intrapartum CS had a nonsignificant protective effect for cases based on diagnoses. Estimates for cases based on prescriptions remained similar to those calculated for the primary outcome definition. Changing the definition of the CS delivery exposure variable from pre-labor and intrapartum CS to now including information on either ruptured or intact membranes slightly

increased the risk of affective disorders, compared to vaginal delivery; however, this increase was not significant. Further information is available in the online appendix.

4. Discussion

This is the first study to explore the effects of mode of delivery on the subsequent risk of all affective disorders. When using standard Cox models, which do not adjust for unmeasured familial confounding variables, we found that both intrapartum and pre-labor CS produced a statistically significant increase in the risk of subsequent affective disorders in the children concerned, compared to those delivered vaginally. However, in sibling comparison models that do adjust for familial confounding variables, we found attenuated effect estimates and only the slight increase in risk that was associated with intrapartum CS remained significant.

Rupture of membranes occurs almost 20-fold more frequently during intrapartum CS than during pre-labor CS (Appendix, eTable 4). Because membrane rupture allows vaginal bacteria to enter the amniotic cavity (Rehbinder et al., 2018), the microbiota of children delivered by intrapartum CS should resemble that of children born vaginally more frequently than should the microbiota of children delivered by pre-labor CS (Stokholm et al., 2016) as we illustrate in Figure 4. We hypothesized that changes in the infant gut microbiota would increase the risk of affective disorders. Therefore, we presumed that delivery by pre-labor CS would be associated with a greater risk of subsequent affective disorders than delivery by intrapartum CS. However, when exposure was defined as CS with or without membrane rupture in our sensitivity analyses, we observed a small nonsignificant increase in risk for both types of CS, compared to vaginal delivery. Hence, neither the results from our main analyses or sensitivity analyses support the theory that contact with vaginal microbiota protects the infant from affective disorders, because the risk of these disorders increases in children born by intrapartum CS but not in children born by pre-labor CS.

Additionally, it is likely that we have been unable to adjust for all relevant unshared factors. Therefore, even the sibling model results may have bias due to unobserved confounding. Most of our study population were born before information for obstetric variables, such as Apgar score, signs of fetal asphyxia, need for airway support, maternal pre-eclampsia, or gestational diabetes was available or validated in the national registers. Therefore, we have been unable to include this information in the study. Such unshared factors are presumably particularly relevant for children born by intrapartum CS, as these children are more likely to be frail. Therefore, it is possible that the exposure variable includes such general obstetric frailty information and produces an increased risk that may not be due to the CS delivery itself. We cannot determine conclusively whether this is the case. However, we only found an increased risk of affective disorders in children who were simultaneously frail and had microbiota that should not have been significantly affected. These observations also fail to support the infant microbiota hypothesis for affective disorders.

If our findings are true effects of intrapartum CS, this delivery mode may increase a mother's risk of postpartum depression compared to other delivery modes, as previously reported (Xu et al., 2017). This, in turn, could affect an infant's risk of depression in later life. However, because the effect estimate is small, these interpretations are speculative.

We applied two different sibling models: the stratified Cox model and the between–within model. Previously, we have found that the between–within model delivers more precise results (Axelsson et al., 2018b, 2018a; Sjölander et al., 2013). However, in this study, the difference between the two models was smaller, although the between–within model remained superior.

Birthweight and gestational age were deliberately excluded from our analyses. They were considered potential colliders and their inclusion could have introduced bias (Wilcox, 2001; Wilcox et al., 2011).

The data from Figure 2, eFigures 3 and 4, and the analyses on the robustness of the results by calendar time in the online appendix show a pronounced increase in treatment with antidepressants in approximately 2007 that subsided in 2012. Similarly, psychiatric care contacts that resulted in affective diagnoses sharply increased in 2008 and did not decline thereafter. The possible reasons for this incidental finding, such as the global financial crisis (Van Hal, 2015) or the rising popularity of social media (Berry et al., 2018; Østergaard, 2017), may be investigated in future studies.

4.1 Previous studies

The risk of all affective disorders has not been assessed previously by delivery mode. However, a nested case–control study of 724 cases diagnosed with bipolar disorder compared to 1,419 controls, which examined 12 obstetric risk factors found that subjects who were delivered by elective CS had an adjusted odds ratio of 2.5 (95% CI, 1.32–4.78), indicating an increased risk of bipolar disorder (Chudal et al., 2014).

A review of records for 301 patients who were hospitalized in Scotland for affective psychosis, compared to matched cases, found no association with CS (Bain et al., 2000).

A large national cohort study based on the Swedish registers (O'Neill et al., 2016), investigated the most serious categories of affective disorders including bipolar disorder, severe depressive episodes with psychotic symptoms, and recurrent depressive disorder - current episodes with psychotic symptoms (O'Neill et al., 2016). The researchers found a significant association with elective CS (HR, 1.17; 95% CI, 1.05–1.31), which disappeared when sibling stratified Cox models were applied, similar to our study. However, we have investigated the entire spectrum of affective disorders, including patients prescribed antidepressants or lithium in primary care, as well as patients with an outpatient or inpatient diagnosis of affective disorders from psychiatric hospital care. We also have complete outpatient follow-up data from the start of our follow-up period in 1995, whereas outpatient data were only becoming available in Sweden from 2001, with complete coverage in 2006. Although the Swedish cohort was larger than ours (1.3 million versus 1.0 million), we had more than 10-fold as many cases of affective disorders as there were cases of affective psychoses in the Swedish cohort. Consequently, our results are considerably more precise.

5. Limitations

Sibling models are increasingly used to control for unmeasured confounding. As in our previous research, this study has shown that some apparent effects observed in simple analyses disappear when sibling models are applied (Axelsson et al., 2018b, 2018a). Nonetheless, sibling models do have limitations; for example, they rely on information with sufficient intra-family variation in the exposure, which implies that any conclusions drawn may not be relevant for children with no siblings. In addition, the effects of

exposures that are nearly constant within groups of siblings cannot be interpreted in terms of causality. However, we do not believe this is a direct limitation in our study.

On the other hand, a strength of this study was the large size of the cohort and the long follow-up period; with over 8.8 million person-years at risk and the oldest subject age 33 years at the end of the study period. In 2010, the average age for a first diagnosis of an affective disorder by contact with a Danish hospital ward was 35.3 years (Jensen and Steinhausen, 2016). If subjects from primary care were included on the basis of prescriptions, as in our study, this average age would presumably be younger. Nearly half of all patients diagnosed with bipolar disorder experience an episode of depression many years earlier (Medici et al., 2015). Figure 2 also shows that the number of new cases declined in patients aged over 25–30 years. Therefore, we have probably captured a substantial proportion of all the new cases of affective disorders that our cohort will experience.

Another strength of our study was that the sensitivity analyses that investigated changes to the definitions of outcomes and modes of delivery, as well as the event or birth time effects on the outcome variables did not change our overall conclusions.

6. Conclusions

Children delivered by CS had a slightly increased risk of affective disorders later in life; however, using a sibling model to adjust for genetic factors and familial environment suggested this was not due to changes in the microbiota. We found that pre-labor CS had no effect on the risk of subsequent affective disorders, whereas intrapartum CS was associated with a small but significant increase in risk for the fully adjusted between–within sibling model. However, there may be bias due to residual confounding because we lack information regarding relevant obstetric variables, such as fetal distress and maternal medical conditions. In conclusion, these results do not support the hypothesis that limited contact with a mother's vaginal microbiota during CS increases the risk of subsequent affective disorders, because infants are more likely to be exposed to the microbiota during intrapartum than during pre-labor CS. The incidental finding of marked changes in the rates of diagnoses and prescriptions for affective disorders over time suggest that further studies are needed.

References

Aagaard, K., Stewart, C.J., Chu, D., 2016. Una destinatio, viae diversae: Does exposure to the vaginal microbiota confer health benefits to the infant, and does lack of exposure confer disease risk? EMBO Rep. 17, 1679–1684. https://doi.org/10.15252/embr.201643483

Axelsson, P.B., Clausen, T.D., Petersen, A.H., Hageman, I., Pinborg, A., Kessing, L.V., Bergholt, T., Rasmussen, S.C., Keiding, N., Løkkegaard, E.C.L., 2018a. Relation between infant microbiota and autism? - Results from a national cohort sibling-design study. Epidemiol. Camb. Mass. https://doi.org/10.1097/EDE.0000000000000928

Axelsson, P.B., Clausen, T.D., Petersen, A.H., Hageman, I., Pinborg, A., Kessing, L.V., Bergholt, T., Rasmussen, S.C., Keiding, N., Løkkegaard, E.C.L., 2018b. Investigating the effects of cesarean delivery and antibiotic use

in early childhood on risk of later attention deficit hyperactivity disorder. J. Child Psychol. Psychiatry. https://doi.org/10.1111/jcpp.12961

Bachmann, C.J., Aagaard, L., Burcu, M., Glaeske, G., Kalverdijk, L.J., Petersen, I., Schuiling-Veninga, C.C.M., Wijlaars, L., Zito, J.M., Hoffmann, F., 2016. Trends and patterns of antidepressant use in children and adolescents from five western countries, 2005–2012. Eur. Neuropsychopharmacol. 26, 411–419. https://doi.org/10.1016/j.euroneuro.2016.02.001

Bager, P., Wohlfahrt, J., Westergaard, T., 2008. Caesarean delivery and risk of atopy and allergic disease: meta-analyses. Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol. 38, 634–642. https://doi.org/10.1111/j.1365-2222.2008.02939.x

Bain, M., Juszczak, E., McInneny, K., Kendell, R.E., 2000. Obstetric complications and affective psychoses. Two case-control studies based on structured obstetric records. Br. J. Psychiatry J. Ment. Sci. 176, 523–526.

Berry, N., Emsley, R., Lobban, F., Bucci, S., 2018. Social media and its relationship with mood, self-esteem and paranoia in psychosis. Acta Psychiatr. Scand. 138, 558–570. https://doi.org/10.1111/acps.12953

Betrán, A.P., Ye, J., Moller, A.-B., Zhang, J., Gülmezoglu, A.M., Torloni, M.R., 2016. The Increasing Trend in Caesarean Section Rates: Global, Regional and National Estimates: 1990-2014. PloS One 11, e0148343. https://doi.org/10.1371/journal.pone.0148343

Bock, C., Bukh, J.D., Vinberg, M., Gether, U., Kessing, L.V., 2009. Do stressful life events predict medical treatment outcome in first episode of depression? Soc. Psychiatry Psychiatr. Epidemiol. 44, 752–760. https://doi.org/10.1007/s00127-008-0491-1

Boden, J.M., Fergusson, D.M., 2011. Alcohol and depression. Addict. Abingdon Engl. 106, 906–914. https://doi.org/10.1111/j.1360-0443.2010.03351.x

Cenit, M.C., Nuevo, I.C., Codoñer-Franch, P., Dinan, T.G., Sanz, Y., 2017. Gut microbiota and attention deficit hyperactivity disorder: new perspectives for a challenging condition. Eur. Child Adolesc. Psychiatry. https://doi.org/10.1007/s00787-017-0969-z

Chudal, R., Sourander, A., Polo-Kantola, P., Hinkka-Yli-Salomäki, S., Lehti, V., Sucksdorff, D., Gissler, M., Brown, A.S., 2014. Perinatal factors and the risk of bipolar disorder in Finland. J. Affect. Disord. 155, 75–80. https://doi.org/10.1016/j.jad.2013.10.026

Clausen, T.D., Bergholt, T., Eriksson, F., Rasmussen, S., Keiding, N., L?kkegaard, E.C., 2016. Prelabor Cesarean Section and Risk of Childhood Type 1 Diabetes: A Nationwide Register-based Cohort Study. Epidemiology 27, 547–555. https://doi.org/10.1097/EDE.0000000000000488

Coello, K., Hansen, T.H., Sørensen, N., Munkholm, K., Kessing, L.V., Pedersen, O., Vinberg, M., 2019. Gut microbiota composition in patients with newly diagnosed bipolar disorder and their unaffected first-degree relatives. Brain. Behav. Immun. 75, 112–118. https://doi.org/10.1016/j.bbi.2018.09.026

Di Florio, A., Craddock, N., van den Bree, M., 2014. Alcohol misuse in bipolar disorder. A systematic review and meta-analysis of comorbidity rates. Eur. Psychiatry J. Assoc. Eur. Psychiatr. 29, 117–124. https://doi.org/10.1016/j.eurpsy.2013.07.004

Dickerson, F., Adamos, M., Katsafanas, E., Khushalani, S., Origoni, A., Savage, C., Schweinfurth, L., Stallings, C., Sweeney, K., Goga, J., Yolken, R.H., 2018. Adjunctive probiotic microorganisms to prevent rehospitalization in patients with acute mania: A randomized controlled trial. Bipolar Disord. 20, 614–621. https://doi.org/10.1111/bdi.12652

Dinan, T.G., Cryan, J.F., 2015. The impact of gut microbiota on brain and behaviour: implications for psychiatry. Curr. Opin. Clin. Nutr. Metab. Care 18, 552–558. https://doi.org/10.1097/MCO.0000000000000221

Etain, B., Henry, C., Bellivier, F., Mathieu, F., Leboyer, M., 2008. Beyond genetics: childhood affective trauma in bipolar disorder. Bipolar Disord. 10, 867–876. https://doi.org/10.1111/j.1399-5618.2008.00635.x

Forsythe, P., Kunze, W., Bienenstock, J., 2016. Moody microbes or fecal phrenology: what do we know about the microbiota-gut-brain axis? BMC Med. 14, 58. https://doi.org/10.1186/s12916-016-0604-8

Huang, L., Chen, Q., Zhao, Y., Wang, W., Fang, F., Bao, Y., 2015. Is elective cesarean section associated with a higher risk of asthma? A meta-analysis. J. Asthma Off. J. Assoc. Care Asthma 52, 16–25. https://doi.org/10.3109/02770903.2014.952435

Jensen, C.M., Steinhausen, H.-C., 2016. Time Trends in Lifetime Incidence Rates of First-Time Diagnosed Bipolar and Depressive Disorders Across 16 Years in Danish Psychiatric Hospitals: A Nationwide Study. J. Clin. Psychiatry 77, e1570–e1575. https://doi.org/10.4088/JCP.15m10276

Leclercq, S., Mian, F.M., Stanisz, A.M., Bindels, L.B., Cambier, E., Ben-Amram, H., Koren, O., Forsythe, P., Bienenstock, J., 2017. Low-dose penicillin in early life induces long-term changes in murine gut microbiota, brain cytokines and behavior. Nat. Commun. 8, 15062. https://doi.org/10.1038/ncomms15062

Lev-Ran, S., Roerecke, M., Le Foll, B., George, T.P., McKenzie, K., Rehm, J., 2014. The association between cannabis use and depression: a systematic review and meta-analysis of longitudinal studies. Psychol. Med. 44, 797–810. https://doi.org/10.1017/S0033291713001438

Liu, L., Zhu, G., 2018. Gut-Brain Axis and Mood Disorder. Front. Psychiatry 9, 223. https://doi.org/10.3389/fpsyt.2018.00223

Medici, C.R., Videbech, P., Gustafsson, L.N., Munk-Jørgensen, P., 2015. Mortality and secular trend in the incidence of bipolar disorder. J. Affect. Disord. 183, 39–44. https://doi.org/10.1016/j.jad.2015.04.032

Misiak, B., Stramecki, F., Gawęda, Ł., Prochwicz, K., Sąsiadek, M.M., Moustafa, A.A., Frydecka, D., 2017. Interactions Between Variation in Candidate Genes and Environmental Factors in the Etiology of Schizophrenia and Bipolar Disorder: a Systematic Review. Mol. Neurobiol. https://doi.org/10.1007/s12035-017-0708-y

Neu, J., Rushing, J., 2011. Cesarean versus vaginal delivery: long-term infant outcomes and the hygiene hypothesis. Clin. Perinatol. 38, 321–331. https://doi.org/10.1016/j.clp.2011.03.008

O'Callaghan, A., van Sinderen, D., 2016. Bifidobacteria and Their Role as Members of the Human Gut Microbiota. Front. Microbiol. 7, 925. https://doi.org/10.3389/fmicb.2016.00925

O'Neill, S.M., Curran, E.A., Dalman, C., Kenny, L.C., Kearney, P.M., Clarke, G., Cryan, J.F., Dinan, T.G., Khashan, A.S., 2016. Birth by Caesarean Section and the Risk of Adult Psychosis: A Population-Based Cohort Study. Schizophr. Bull. 42, 633–641. https://doi.org/10.1093/schbul/sbv152

Østergaard, S.D., 2017. Taking Facebook at face value: why the use of social media may cause mental disorder. Acta Psychiatr. Scand. 136, 439–440. https://doi.org/10.1111/acps.12819

Pandey, P.K., Verma, P., Kumar, H., Bavdekar, A., Patole, M.S., Shouche, Y.S., 2012. Comparative analysis of fecal microflora of healthy full-term Indian infants born with different methods of delivery (vaginal vs cesarean): Acinetobacter sp. prevalence in vaginally born infants. J. Biosci. 37, 989–998.

Polidano, C., Zhu, A., Bornstein, J.C., 2017. The relation between cesarean birth and child cognitive development. Sci. Rep. 7. https://doi.org/10.1038/s41598-017-10831-y

Pottegård, A., Zoëga, H., Hallas, J., Damkier, P., 2014. Use of SSRIs among Danish children: a nationwide study. Eur. Child Adolesc. Psychiatry 23, 1211–1218. https://doi.org/10.1007/s00787-014-0523-1

Rehbinder, E.M., Lødrup Carlsen, K.C., Staff, A.C., Angell, I.L., Landrø, L., Hilde, K., Gaustad, P., Rudi, K., 2018. Is amniotic fluid of women with uncomplicated term pregnancies free of bacteria? Am. J. Obstet. Gynecol. 219, 289.e1-289.e12. https://doi.org/10.1016/j.ajog.2018.05.028

Rutayisire, E., Huang, K., Liu, Y., Tao, F., 2016. The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. BMC Gastroenterol. 16, 86. https://doi.org/10.1186/s12876-016-0498-0

Sevelsted, A., Stokholm, J., Bonnelykke, K., Bisgaard, H., 2015. Cesarean Section and Chronic Immune Disorders. PEDIATRICS 135, e92–e98. https://doi.org/10.1542/peds.2014-0596

Sjölander, A., Lichtenstein, P., Larsson, H., Pawitan, Y., 2013. Between-within models for survival analysis. Stat. Med. 32, 3067–3076. https://doi.org/10.1002/sim.5767

Stokholm, J., Thorsen, J., Chawes, B.L., Schjørring, S., Krogfelt, K.A., Bønnelykke, K., Bisgaard, H., 2016. Cesarean section changes neonatal gut colonization. J. Allergy Clin. Immunol. 138, 881-889.e2. https://doi.org/10.1016/j.jaci.2016.01.028

Tognini, P., 2017. Gut Microbiota: A Potential Regulator of Neurodevelopment. Front. Cell. Neurosci. 11, 25. https://doi.org/10.3389/fncel.2017.00025

Underwood, M.A., German, J.B., Lebrilla, C.B., Mills, D.A., 2015. Bifidobacterium longum subspecies infantis: champion colonizer of the infant gut. Pediatr. Res. 77, 229–235. https://doi.org/10.1038/pr.2014.156

Vael, C., Desager, K., 2009. The importance of the development of the intestinal microbiota in infancy. Curr. Opin. Pediatr. 21, 794–800. https://doi.org/10.1097/MOP.0b013e328332351b

Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E.F., Wang, J., Tito, R.Y., Schiweck, C., Kurilshikov, A., Joossens, M., Wijmenga, C., Claes, S., Van Oudenhove, L., Zhernakova, A., Vieira-Silva, S., Raes, J., 2019. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat. Microbiol. https://doi.org/10.1038/s41564-018-0337-x

Van Hal, G., 2015. The true cost of the economic crisis on psychological well-being: a review. Psychol. Res. Behav. Manag. 8, 17–25. https://doi.org/10.2147/PRBM.S44732

Vinberg, M., Ottesen, N.M., Meluken, I., Sørensen, N., Pedersen, O., Kessing, L.V., Miskowiak, K.W., 2019. Remitted affective disorders and high familial risk of affective disorders associate with aberrant intestinal microbiota. Acta Psychiatr. Scand. 139, 174–184. https://doi.org/10.1111/acps.12976

Vogel, J.P., Betrán, A.P., Vindevoghel, N., Souza, J.P., Torloni, M.R., Zhang, J., Tunçalp, Ö., Mori, R., Morisaki, N., Ortiz-Panozo, E., Hernandez, B., Pérez-Cuevas, R., Qureshi, Z., Gülmezoglu, A.M., Temmerman, M., 2015. Use of the Robson classification to assess caesarean section trends in 21 countries: a secondary analysis of two WHO multicountry surveys. Lancet Glob. Health 3, e260–e270. https://doi.org/10.1016/S2214-109X(15)70094-X

Weinberger, A.H., Gbedemah, M., Martinez, A.M., Nash, D., Galea, S., Goodwin, R.D., 2017. Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. Psychol. Med. 1–10. https://doi.org/10.1017/S0033291717002781

Wilcox, A.J., 2001. On the importance--and the unimportance--of birthweight. Int. J. Epidemiol. 30, 1233–1241.

Wilcox, A.J., Weinberg, C.R., Basso, O., 2011. On the pitfalls of adjusting for gestational age at birth. Am. J. Epidemiol. 174, 1062–1068. https://doi.org/10.1093/aje/kwr230

Wilde, A., Chan, H.-N., Rahman, B., Meiser, B., Mitchell, P.B., Schofield, P.R., Green, M.J., 2014. A meta-analysis of the risk of major affective disorder in relatives of individuals affected by major depressive disorder or bipolar disorder. J. Affect. Disord. 158, 37–47. https://doi.org/10.1016/j.jad.2014.01.014

Winter, G., Hart, R.A., Charlesworth, R.P.G., Sharpley, C.F., 2018. Gut microbiome and depression: what we know and what we need to know. Rev. Neurosci. 29, 629–643. https://doi.org/10.1515/revneuro-2017-0072

Xu, H., Ding, Y., Ma, Y., Xin, X., Zhang, D., 2017. Cesarean section and risk of postpartum depression: A meta-analysis. J. Psychosom. Res. 97, 118–126. https://doi.org/10.1016/j.jpsychores.2017.04.016

Yang, I., Corwin, E.J., Brennan, P.A., Jordan, S., Murphy, J.R., Dunlop, A., 2016. The Infant Microbiome: Implications for Infant Health and Neurocognitive Development. Nurs. Res. 65, 76–88. https://doi.org/10.1097/NNR.0000000000000133

	CONFOUNDER ADJUSTMENT LEVEL				
	1	2	3	4	5
MODE OF DELIVERY	x	x	x	x	x
OFFSPRING SEX		x	x	x	x
MATERNAL AGE AT BIRTH			x	x	x
PATERNAL AGE DIFFERENCE			(x)	(x)	(x)
PATERNAL EDUCATION			x	x	x
MATERNAL EDUCATION			x	x	x
MATERNAL MARITAL STATUS			x	x	x
PARITY				x	x
MATERNAL PSYCHIATRIC HISTORY					(x)
PATERNAL PSYCHIATRIC HISTORY					(x)

Table 1: Variables included at each level of confounder adjustment, with level 1 referred to as *unadjusted* and levels 4 and 5 as *fully adjusted* for the sibling models and standard Cox models, respectively. Variables marked with parentheses are not used in the sibling models.

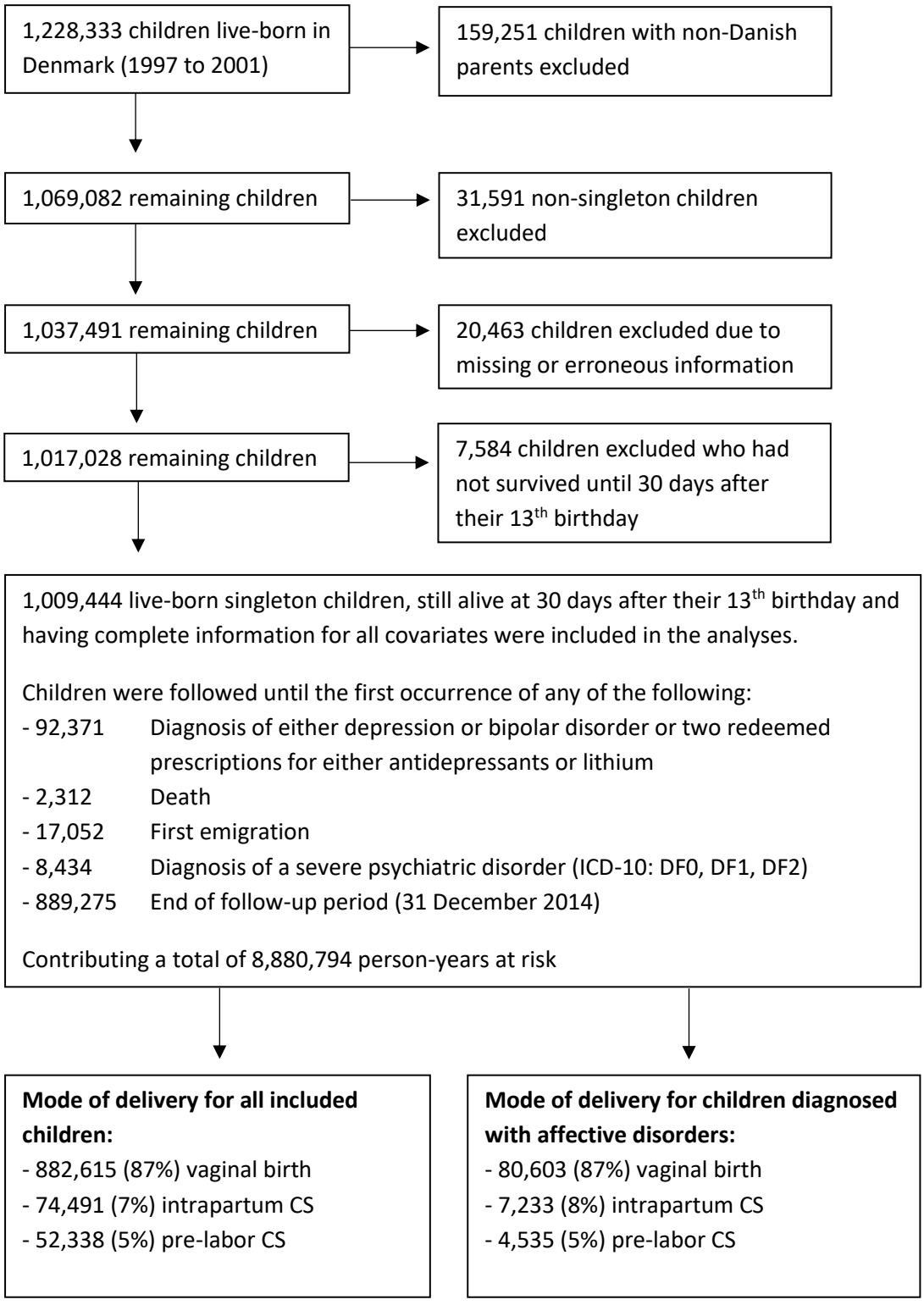


Figure 1. Inclusion, exclusion, and censoring
Flowchart showing the number of children excluded, the number of children in the final study population, the number of events, and the distribution of affective disorders according to mode of delivery.
Abbreviations: ICD-10, International Classification of Diseases - 10th Revision; CS, cesarean section.

		Vaginal delivery		Intrapartum CS		Pre-labor CS		Total
L	**Variable**	***n***	**%**	***n***	**%**	***n***	**%**	***n***
	Number of children	882,615	87.4%	74,491	7.4%	52,338	5.2%	1,009,444
2	**Sex**							
	Male	450,614	87.0%	41,349	8.0%	26,181	5.1%	518,144
	Female	432,001	87.9%	33,142	6.7%	26,157	5.3%	491,300
3	**Maternal age**							
	10–24 years	195,112	89.2%	16,328	7.5%	7,302	3.3%	218,742
	25–29 years	361,739	88.2%	29,717	7.2%	18,611	4.5%	410,067
	30–34 years	242,332	86.7%	19,983	7.1%	17,322	6.2%	279,637
	35 years or older	83,432	82.6%	8,463	8.4%	9,103	9.0%	100,998
3	**Maternal education**							
	Elementary/high school	258,221	87.4%	22,236	7.5%	15,105	5.1%	295,562
	Short education/skilled worker	57,988	88.1%	4,698	7.1%	3,102	4.7%	65,788
	Medium-length education	337,996	87.2%	29,063	7.5%	20,690	5.3%	387,749
	Long education	228,410	87.7%	18,494	7.1%	13,441	5.2%	260,345
3	**Paternal education**							
	Elementary/high school	214,723	87.3%	18,508	7.5%	12,655	5.1%	245,886
	Short education/skilled worker	38,754	87.4%	3,255	7.3%	2,307	5.2%	44,316
	Medium-length education	447,159	87.3%	38,018	7.4%	26,822	5.2%	511,999
	Long education	181,979	87.8%	14,710	7.1%	10,554	5.1%	207,243
3	**Maternal marital status**							
	Married/Partnership	461,279	87.6%	35,664	6.8%	29,915	5.7%	526,858
	Not married	421,336	87.3%	38,827	8.0%	22,423	4.6%	482,586
3	**Parental age difference (paternal – maternal)**							
	–5 years or lower	17,190	82.0%	2,122	10.1%	1,654	7.9%	20,966
	–4 to +4 years	674,709	87.5%	56,752	7.4%	39,818	5.2%	771,279
	+5 years or greater	190,716	87.8%	15,617	7.2%	10,866	5.0%	217,199
4	**Parity**							
	Primiparous	395,674	85.4%	47,126	10.2%	20,309	4.4%	463,109
	1 prior child	335,875	89.4%	19,729	5.3%	20,039	5.3%	375,643
	2 prior children	117,031	88.5%	5,804	4.4%	9,344	7.1%	132,179
	≥3 prior children	34,035	88.4%	1,832	4.8%	2,646	6.9%	38,513
5	**Maternal psychiatric history**							
	No	864,749	87.5%	72,653	7.4%	50,804	5.1%	988,206
	Yes	17,866	84.1%	1,838	8.7%	1,534	7.2%	21,238
5	**Paternal psychiatric history**							
	No	866,004	87.5%	72,970	7.4%	51,255	5.2%	990,229
	Yes	16,611	86.4%	1,521	7.9%	1,083	5.6%	19,215

Table 2: Demographic data for the study population and their parents. The "L" column indicates the level at which covariates were added to the adjustment models.
Abbreviation: CS, cesarean section.

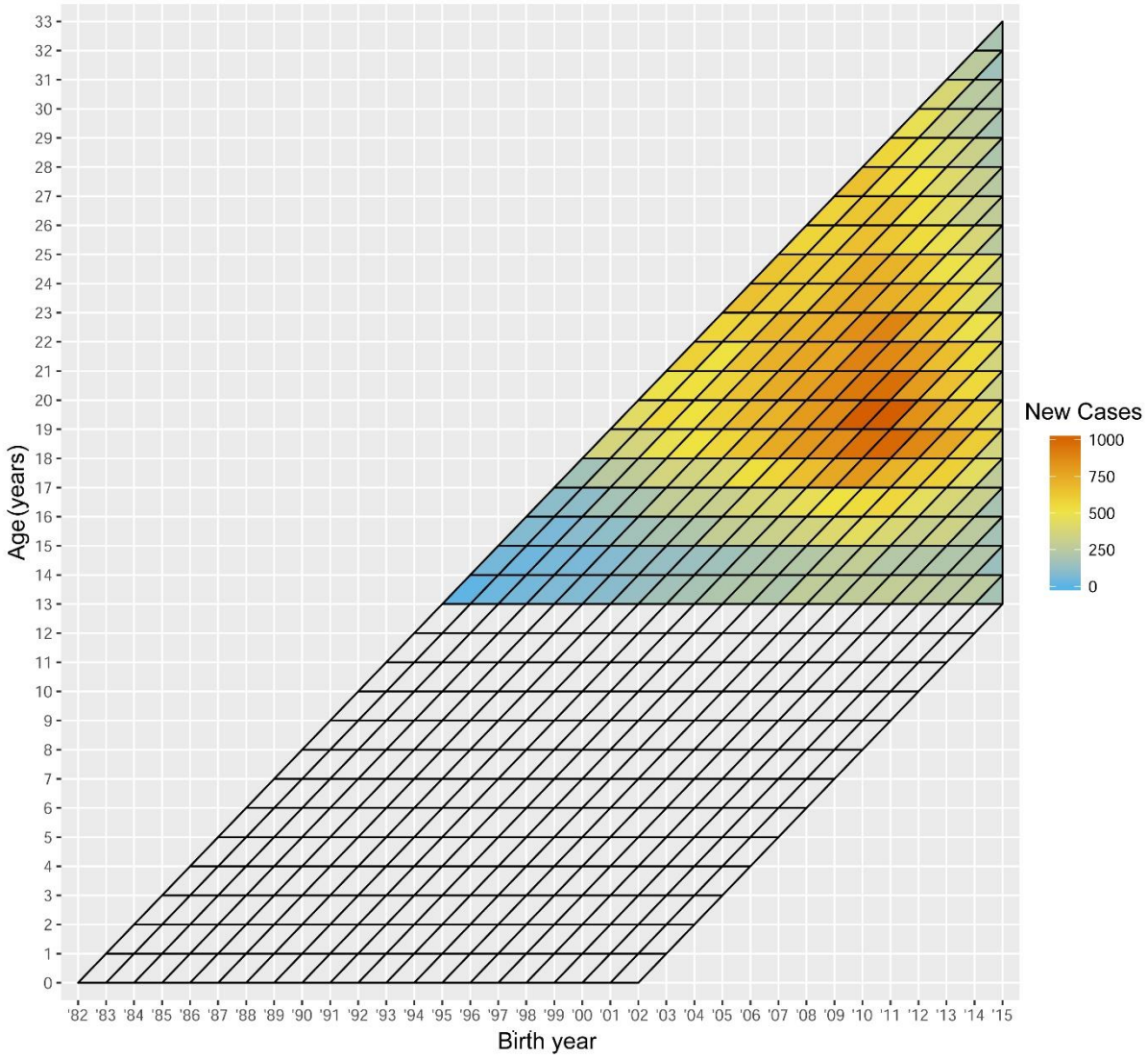


Figure 2. Lexis heat map showing the number of new affective disorder cases
New affective disorder cases are defined as either a diagnosis or a redeemed medication prescription and are shown at 1-year intervals for birth year and age. The rate of new cases was not stable over diagnostic-year, birth-year, or age. There was an increase in the number of new cases between the years 2007 and 2012, with a clear tendency toward younger age at first diagnosis.

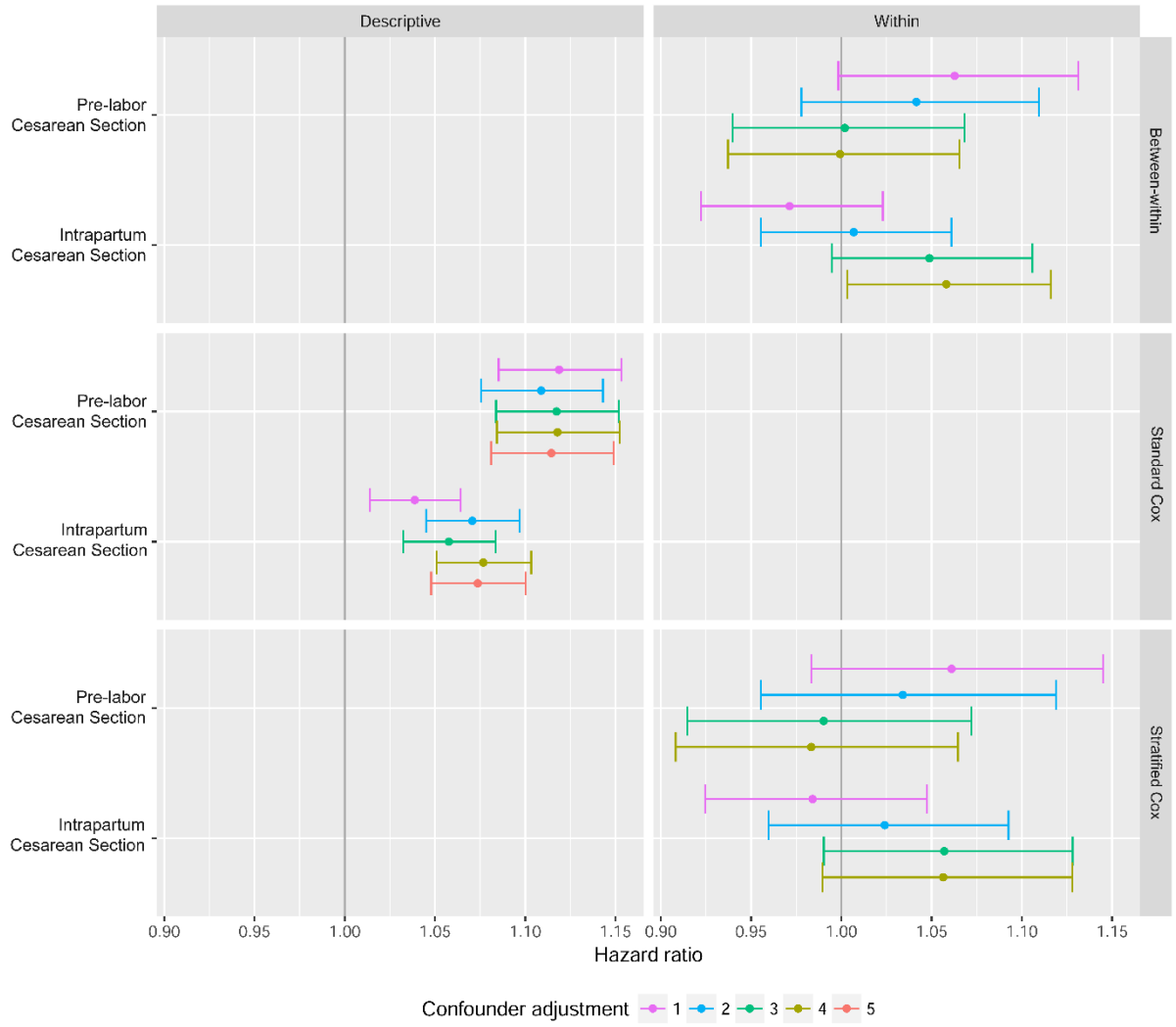


Figure 3. Results from the statistical models

The main results from the three different statistical models are shown. Level 5 adjustment was performed only on the standard Cox model (refer to Table 1 for confounder adjustment levels). Within-family effects are adjusted for unobserved confounding that is shared within the family. Therefore, descriptive and within effects cannot be compared and are depicted in separate columns. The parameter estimates with 95% confidence intervals (CIs) are shown. Because the CIs are pointwise, they do not account for multiple testing.

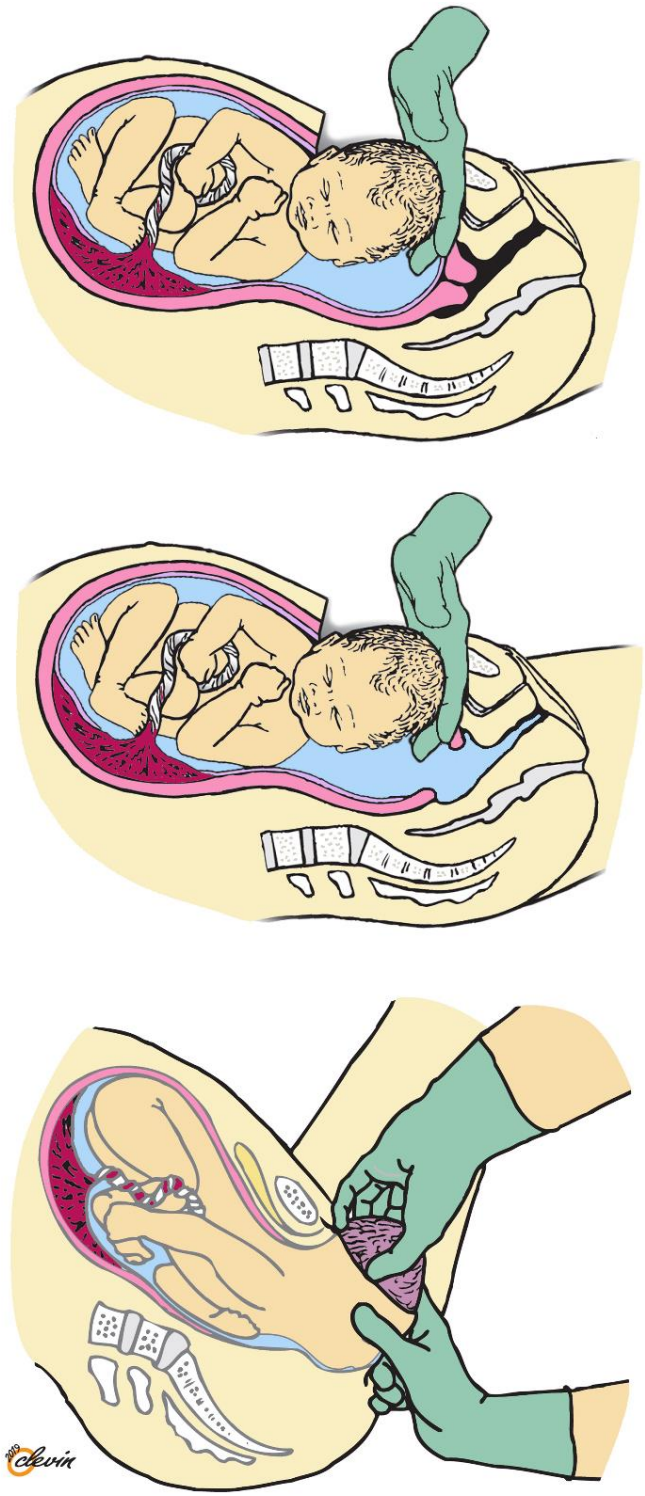
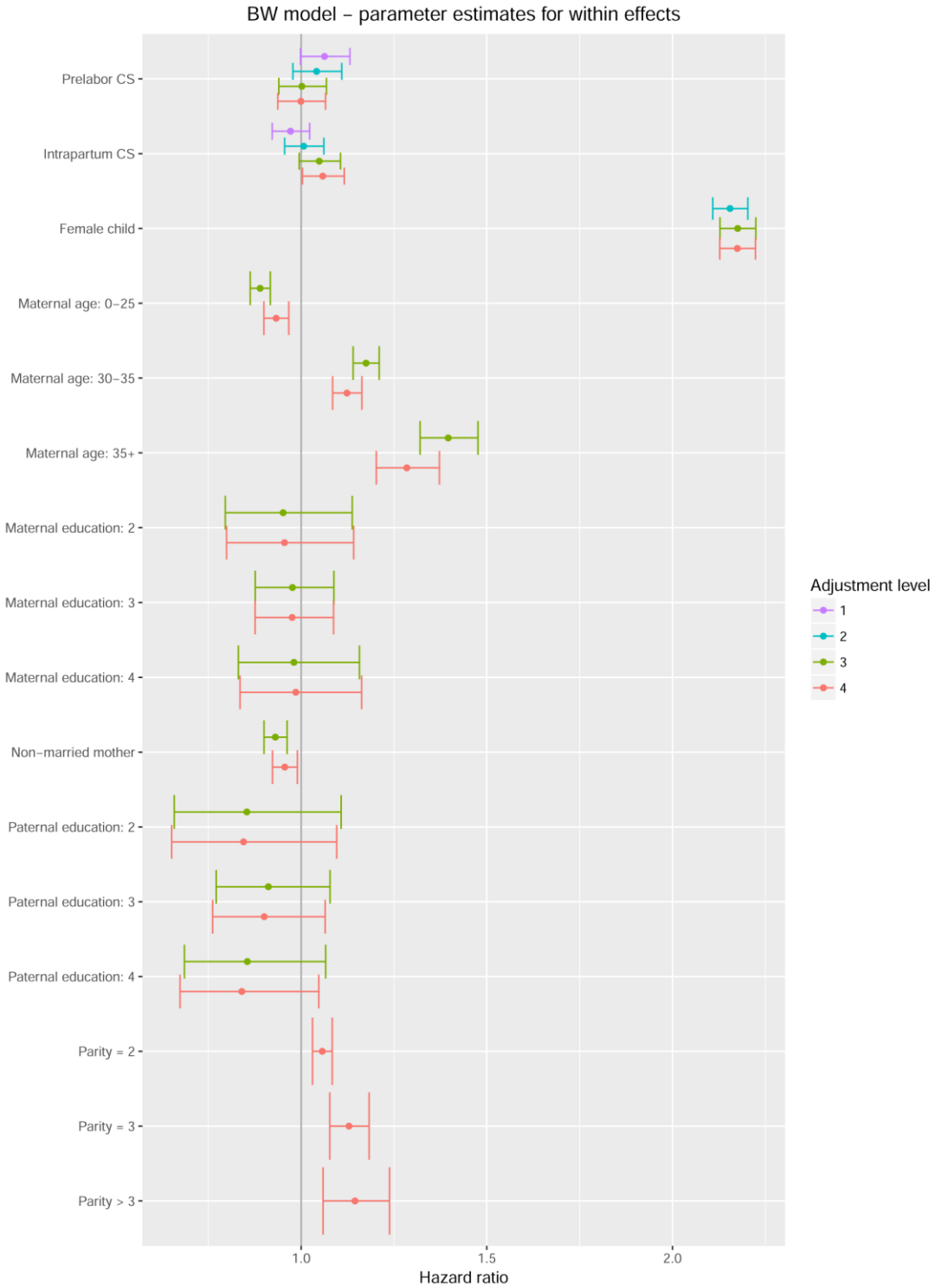


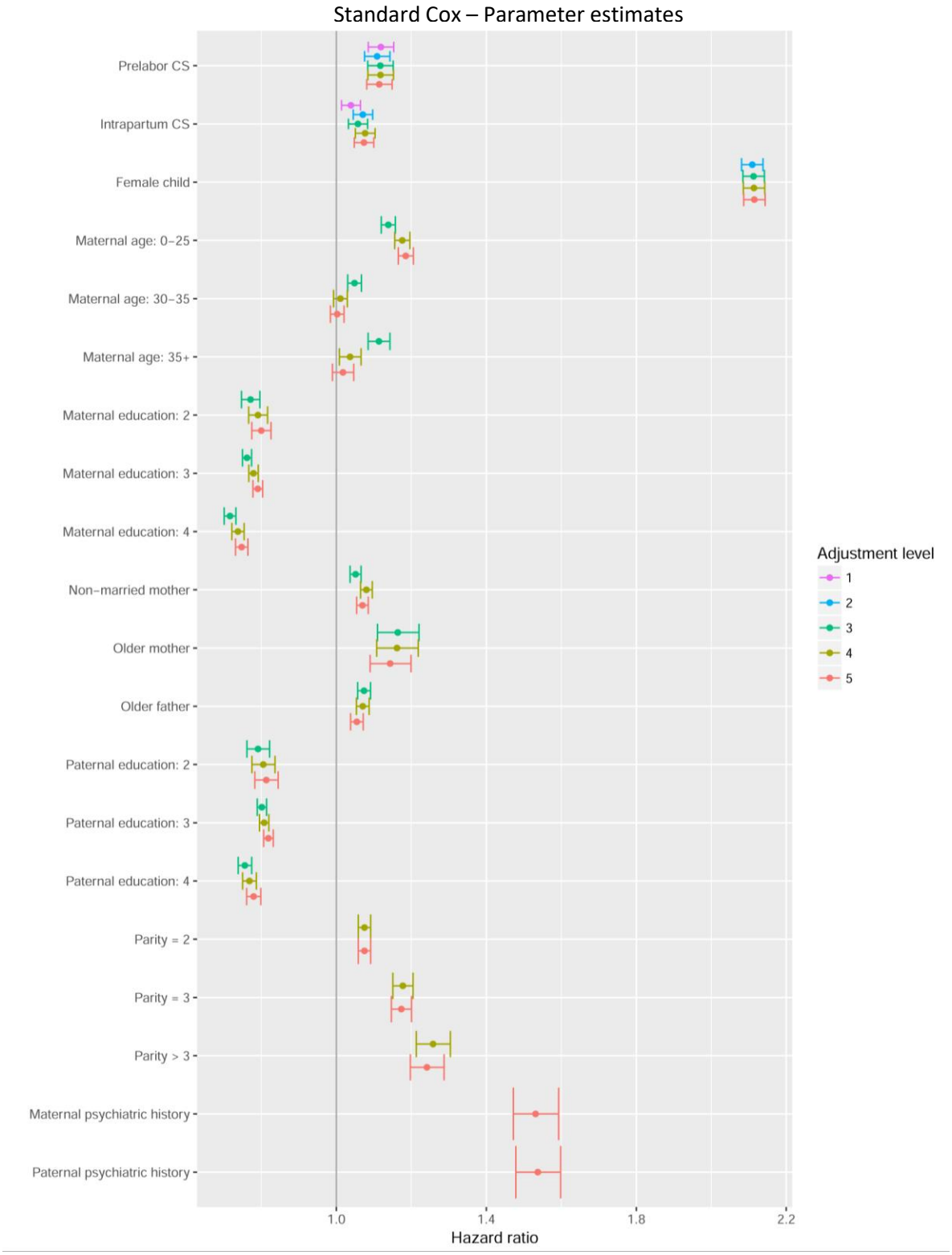
Figure 4: Mode of delivery and early infant microbiota.
The three different modes of delivery discussed in our paper: pre-labor cesarean section, intrapartum cesarean section and vaginal birth. For each delivery mode the first infant microbiota composition is expected to be on average more “beneficial” in vaginal delivery and more “pathogenic” in planned cesarean section, with intrapartum cesarean section resulting in microbiota that is a mixture of the other two delivery modes, as the infant has probably had contact with vaginal microbes.

Paper III – supplementary information

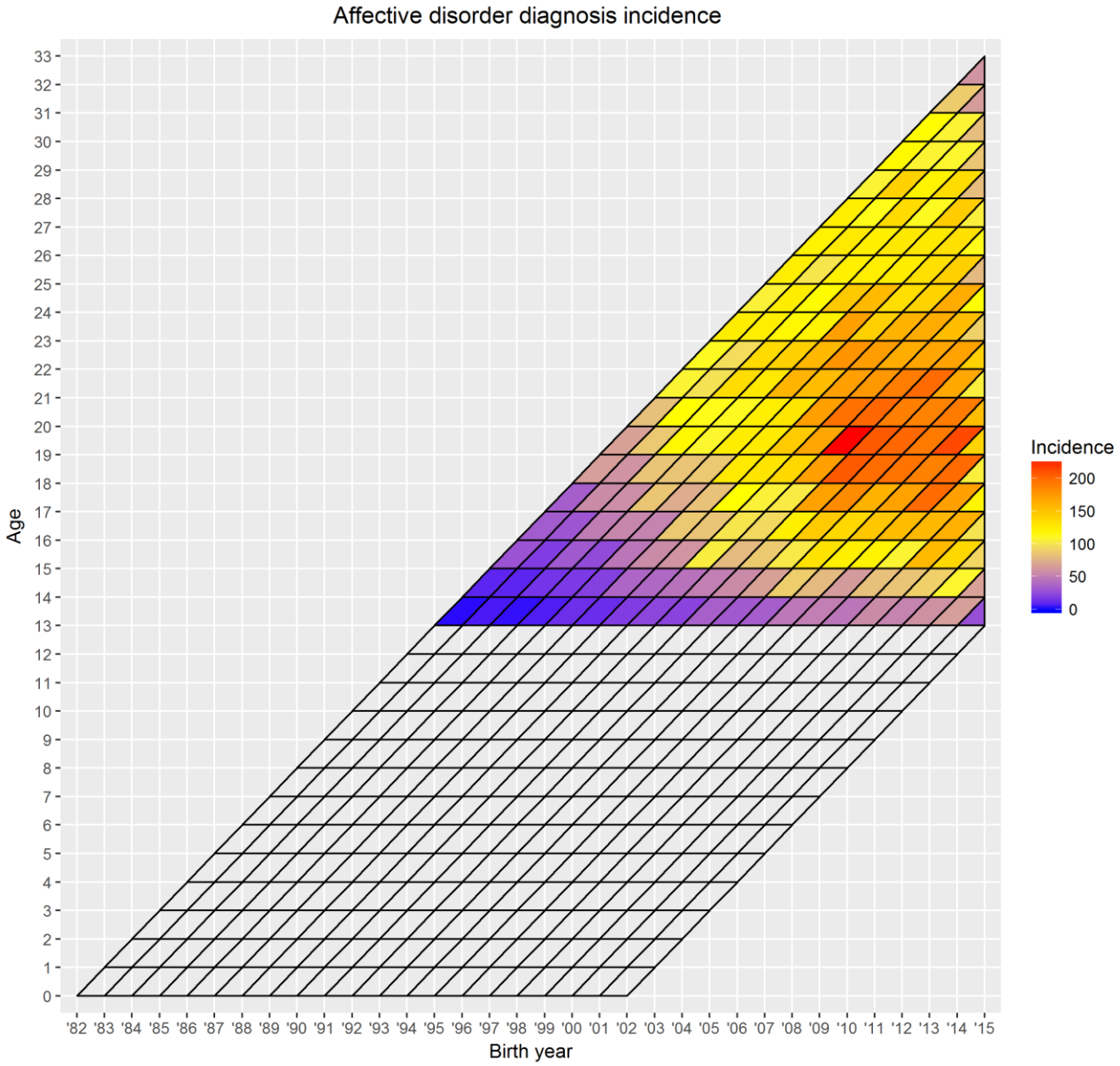
eAppendix



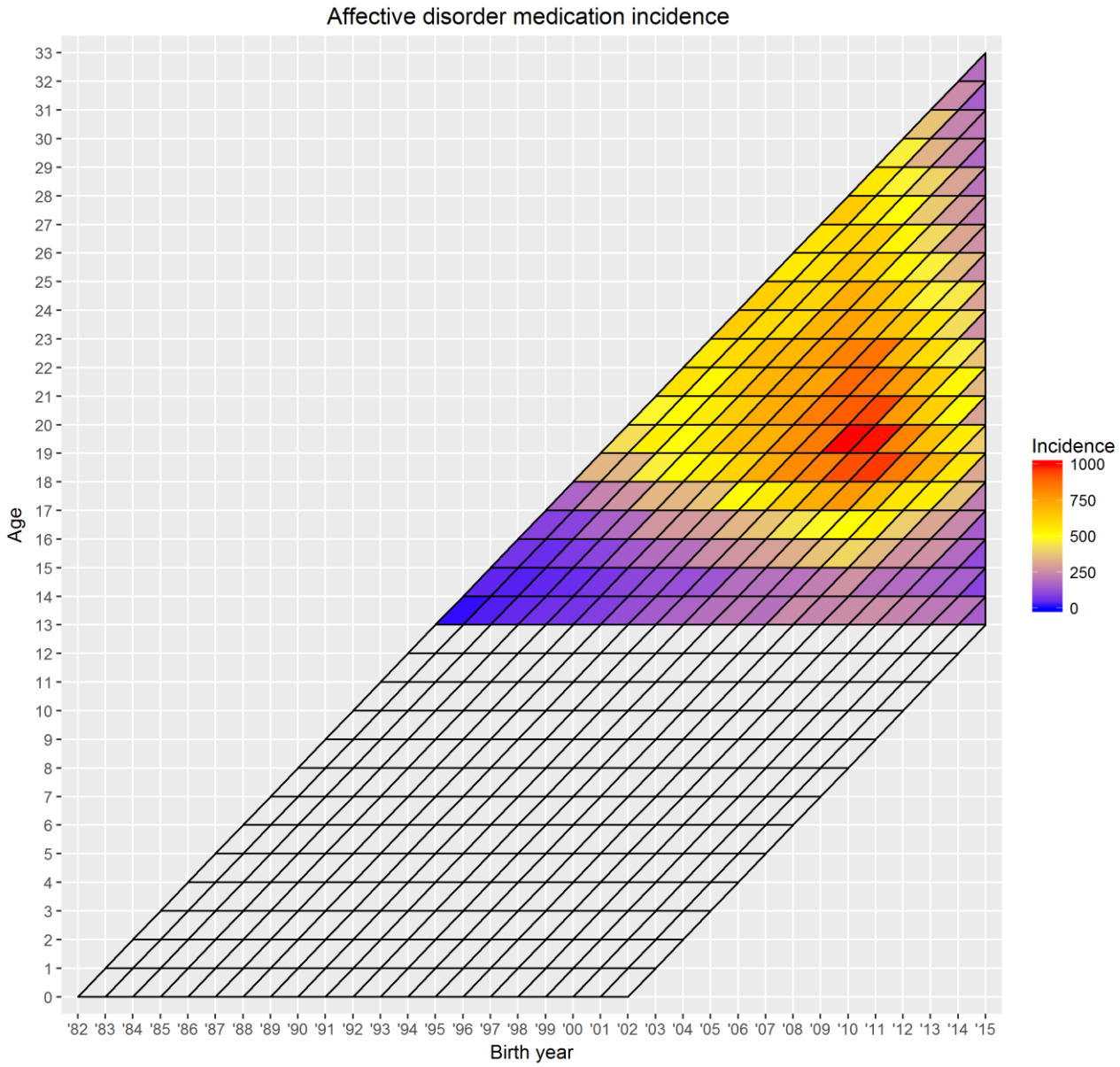
eFigure 1: Estimated hazard ratios of the within-effect of the between-within (BW) model with lines marking 95% confidence bands. Abbreviations: CS = Cesarean section, Education 2, 3 and 4 = short, medium-length or long education. Refer to table 1 in the main article for confounder adjustment levels.



eFigure 2: Estimated hazard ratios using the standard Cox model with lines marking 95% confidence bands. Abbreviations and clarifications: CS = Cesarean section, Education 2, 3 and 4 = short, medium-length or long education, Older father/mother = an age difference of 5 years. Refer to table 1 in the main article for confounder adjustment levels.



eFigure 3: A Lexis heatmap for new cases based on diagnoses only.



eFigure 4: A Lexis heatmap for new cases based on prescriptions for antidepressants and lithium only.

Variable	Code / Information	Register
Intrapartum cesarean delivery[a]	OP: KMCA00, KMCA10D, KMCA10E, KMCA12, KMCA12A,	MBR,
	KMCA12B, KMCA20, KMCA30, KMCA33, KMCA96	NPR
Prelabour cesarean delivery	OP: KMCA10, KMCA10A, KMCA10B, KMCA10C, KMCA11	MBR, NPR
Obstetric factors	Year of birth.	MBR
Any psychiatric disorder	ICD8: 290 - 315	NPR
	ICD10: DF0 - DF9	
Bipolar Disorder	ICD10: DF30-DF31, DF38.00	PCRR
Depression	ICD10: DF32-DF33	PCRR
Organic mental disorders	ICD10: DF00-DF09	PCRR
Mental disorders due to substance abuse	ICD10: DF10-DF19	PCRR
Schizophrenia and related disorders	ICD10: DF20-DF29	PCRR
Medication Use		
Any psychiatric medication	ATC-codes: N03AE, N05, N06	RMP
Antidepressants	ATC-code: N06A	RMP
Lithium	ATC-code: N05AN1	RMP

eTable 1: Specification of registers and codes used for defined variables, hosted by Statistics Denmark, a government statistics bureau.

Abbreviations:
ATC: Anatomical Therapeutic Chemicals according to WHO
ICD8: the International Classification of Diseases, 8th revision, from 1977 to 1993
ICD10: the International Classification of Diseases, 10th revision, since 1994
MBR: Medical Birth Registry since 1973
NPR: National Patient Registry since 1977
OP: "Nordic Classification of Surgical procedures" since 1996
PCRR: Psychiatric Central Research Register, since 1969
RCD: Register of Causes of Death, since 1970
RMP: The Register of Medicinal Product Statistics since 1995 and since 1997 on an individual level.
[a] Includes cesarean section, where timing regarding onset of labor could not be distinguished
[b] 180 days before or 90 days after birth

Nationwide Register	Data and function provided
The Medical Birth Registry (1973 to 2010)[1]	Date of birth, vital status at birth, mode of delivery, parity, multiple births, birth weight and gestational age at birth, Apgar score at 5 minutes postpartum, mother's weight in early pregnancy and smoking status
The Fertility Database (1960 to 2010)[2]	Linked children to their parents and provided data on offspring sex, and date of birth for the parents
The National Patient Registry (1977 to 2015)[3]	Hospital admission dates, codes related to complications during pregnancy, labor and delivery, and diagnoses and surgical procedures related to mode of delivery. Data on outpatient treatments were added in 1995.
The Psychiatric Central Research Register (1969 to 2015)[4]	Psychiatric discharge diagnostic codes for admissions to mental hospitals and psychiatric departments. Data on outpatient treatments were added in 1995.
The Register of Causes of Death (1970 to 2015)[5]	Vital status
The Register of Medicinal Products Statistics (1997 to 2015)	Redeemed prescriptions on psychiatric medications for parents and offspring, antibiotic medications for mothers during pregnancy and children during their first two years of life.
Statistics Denmark (1978 to 2015)	Information on emigration, parental country of origin and educational status.

eTable 2: The specific information and function each register provided for our study.

References for eTable 2:

1. Knudsen LB, Olsen J. The Danish Medical Birth Registry. Dan Med Bull. 1998;45(3):320-323.

2. Knudsen LB. The Danish Fertility Database. Dan Med Bull. 1998;45(2):221-225.

3. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health. 2011;39(7 Suppl):30-33. doi:10.1177/1403494811401482

4. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. Scand J Public Health. 2011;39(7 Suppl):54-57. doi:10.1177/1403494810395825

5. Helweg-Larsen K. The Danish Register of Causes of Death. Scand J Public Health. 2011;39(7 Suppl):26-29. doi:10.1177/1403494811399958

Model	Exposure	Effect Estimate	95% CI
Between–within			
	Intrapartum CS	1.05	(1.003 to 1.12)
	Pre-labor CS	1.00	(0.94 to 1.07)
Stratified Cox			
	Intrapartum CS	1.06	(0.99 to 1.13)
	Pre-labor CS	0.98	(0.91 to 1.06)
Standard Cox			
	Intrapartum CS	1.07	(1.05 to 1.10)
	Pre-labor CS	1.11	(1.08 to 1.15)

eTable 3: Fully adjusted main effect hazard ratio estimates with pointwise 95% confidence intervals (CI) for the three different statistical models: Between–within survival model for siblings, stratified Cox by siblings, and standard Cox regression.
Abbreviation: CS, cesarean section.

Robustness of the results

In this appendix, we investigated whether or not the results of the main analyses were robust with regards to three key modeling decisions. More specifically, we investigated the sensitivity of the model results towards:

- Calendar time effects in the outcome variable
- What types of affective disorder cases are included in the outcome
- How mode of delivery is defined

Each topic is presented in turn below. Generally, we do not find any results that change the overall conclusions in the primary analyses. However, we do see differences especially concerning the direction of the insignificant relationship between mode of delivery and affective disorders, with estimates close to null-value.

Calendar time trends in affective disorder occurrence

In this section, we investigate the robustness of the conclusions from the primary analyses towards the inclusion of additional calendar time effects. More specifically, we are interested in determining whether these conclusions would be altered had we included birth year effects or event year effects. We use the term *event year* to refer to the year in which the affective disorder event occurs (I.e. either a diagnosis or two consecutive redemptions of medication prescriptions). This investigation is motivated by the trends seen in the Lexis diagram (Figure 2) for affective disorder incidence in the *Data description*. In the Lexis diagram, we see a marked increase in incidences in the years 2009-2012, especially for people aged 18-23. If the mechanism causing this increased incidence is also a cause of delivery by cesarean section (which, of course, occurs at a different calendar time, namely around 18-23 years earlier), it is a confounder of the effects of interest of the current study and therefore, it should be adjusted for.
It is not feasible to include event year effects (i.e. time-varying covariates) in the sibling models due to lack of computational resources. Therefore, we will only investigate the possible impact of calendar time effects in the descriptive Cox model. Moreover, we focus on the crude model with no confounder adjustment in order to reduce the computational burden. However, even in this model, our sample size makes inclusion of event year effects impossible. Therefore, we use a resampling scheme and refit models multiple times on smaller subsets of the dataset. More specifically, we carry out 100 independent repetitions of the following procedure:

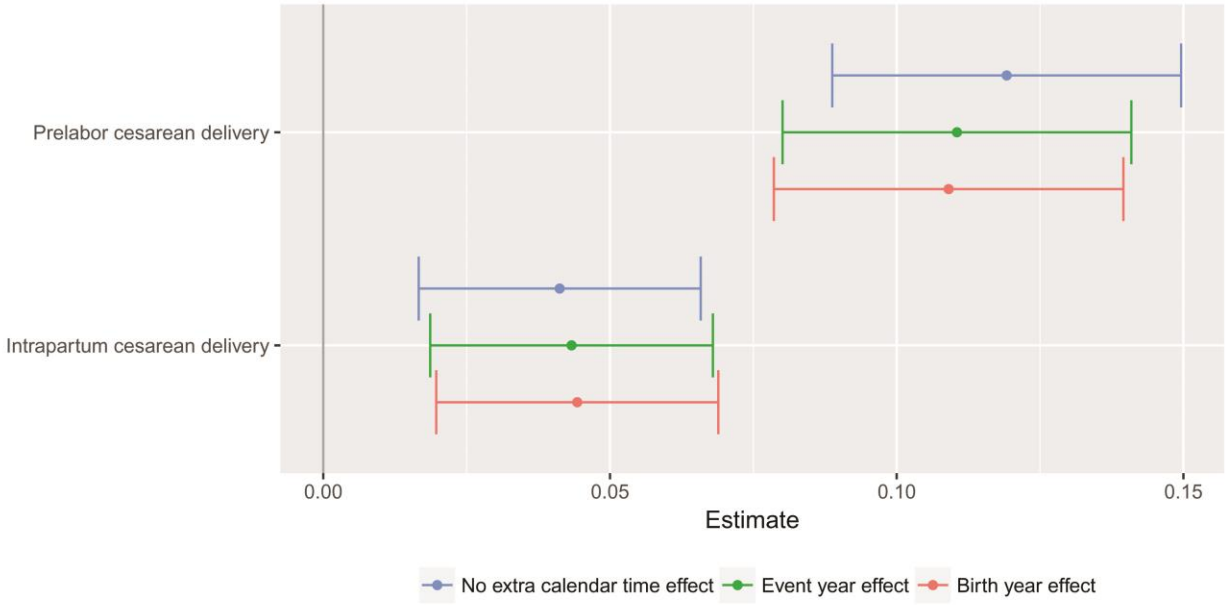
1. Draw 10000 random observations independently from the full sample. This qualifies as the current dataset.
1. Fit each of the following three models on the current dataset and the save the results:
    (a) A model with a birth year effect (modelled categorically in one-year intervals).
    (b) A model with an event year effect (modelled categorically in one-year intervals).
    (c) A model with no extra time effects beyond age.

By performing independent repetitions of the procedure, the law of large numbers and the central limit theorem will ensure that the resampling strategy will result in asymptotically consistent estimates that follow a normal distribution and that their variances can be estimated by use of the sampling variance.

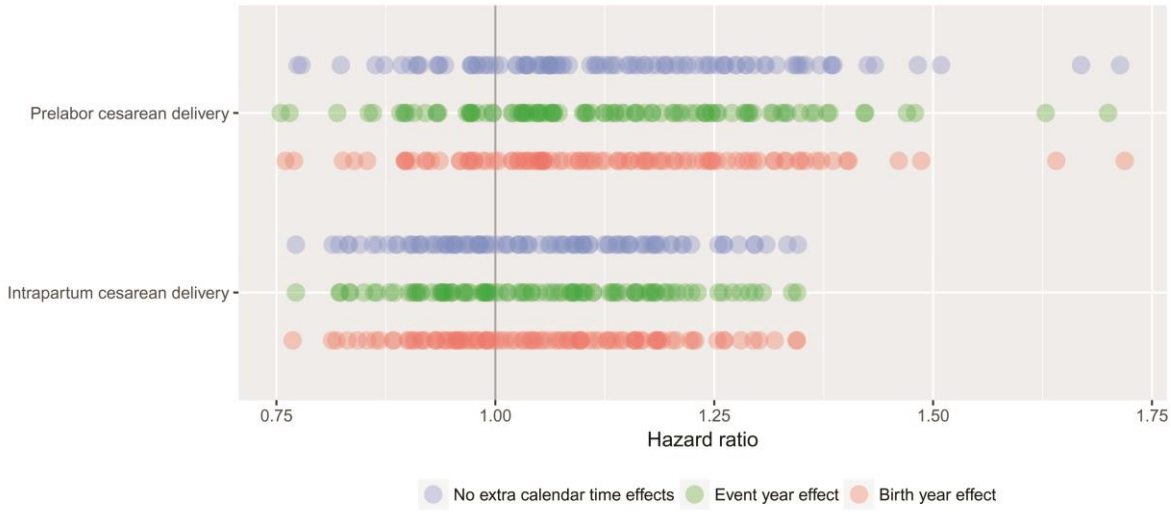
Results for the exposure effect estimates
In eFigure 5 we present the exposure effect estimates from all the 100 resampling/reestimation steps. More specifically, for each parameter estimate in each model type, we have calculated the mean estimate

and added a 95 % Wald confidence interval where the variance is estimated as the empirical resampling variance. Due to the central limit theorem, we can assume asymptotic normality of the resampled parameter estimates and thus these confidence intervals should represent the sampling error adequately. We find a very small attenuation in the estimated effect of pre-labor cesarean delivery when birth year- or event year effects are added to the model. For intrapartum cesarean delivery, the point estimates are almost identical, no matter if birth year-, event year- or no extra calendar time effects are included. All in all we conclude that calendar time does *not* seem to confound the relationship between cesarean delivery and risk of affective disorders, and if it does, the effect is very modest. The primary models are therefore likely to be robust towards (lack of) inclusion of calendar time effects.

In eFigure 6 we provide an overview of the results of all the resampling steps, represented by one transparent dot for each resampling. Please note that the estimates on this figure are on hazard ratio scale, in contrast to those of eFigure 5.



eFigure 5: Means of the parameter estimates from the 100 resampling/reestimation runs using the standard Cox model with no confounder adjustment and various strategies for handling extra time effects beyond offspring age. The error bars mark 95 % Wald confidence intervals based on the empirical resampling standard deviations.

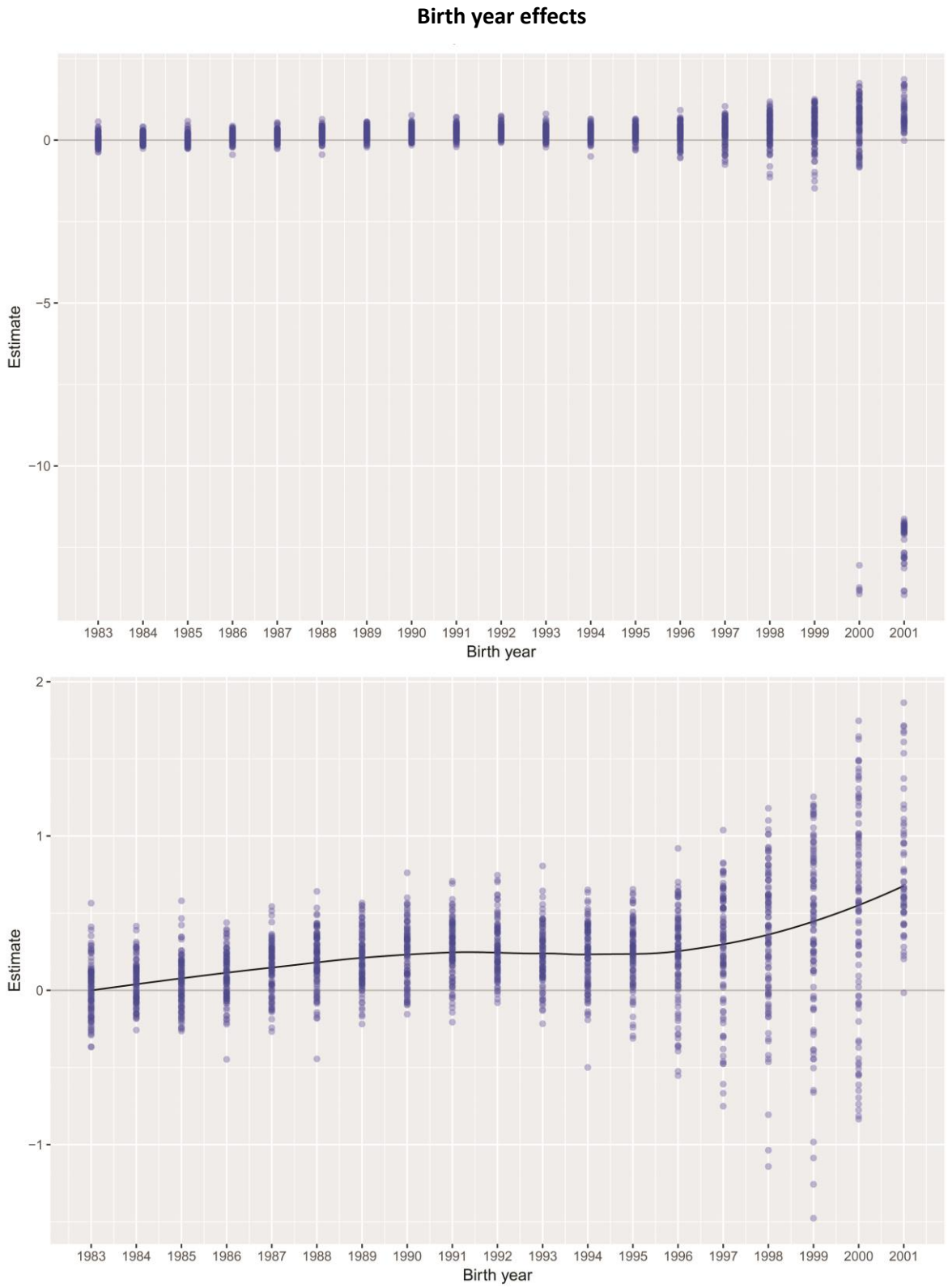


eFigure 6: Hazard ratio estimates from the 100 resampling/reestimates from the three models. Each dot represents a single resampling step.

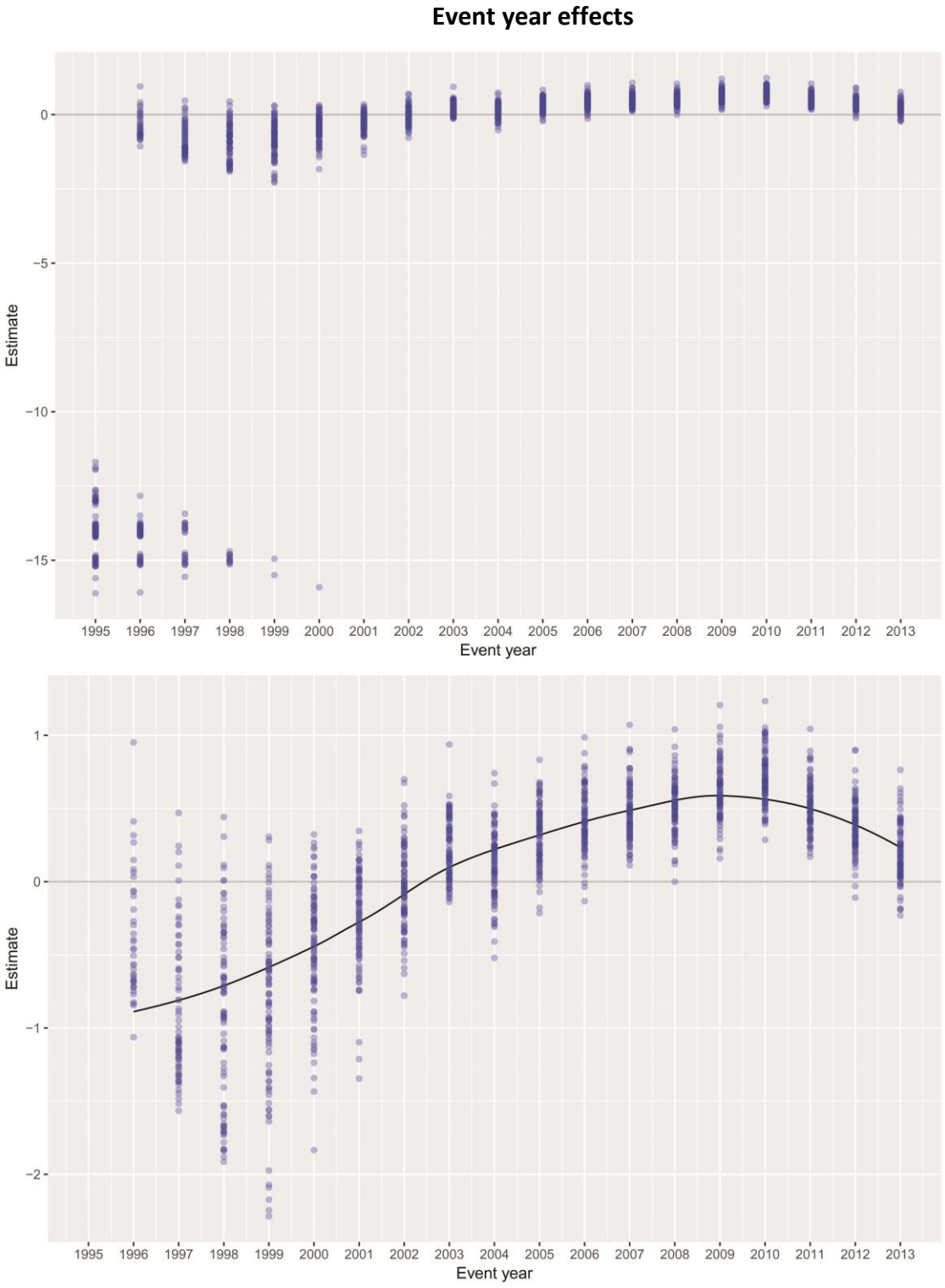
Estimated calendar time effects

Now, we look at the estimated time effects for the birth year- and event year models. They are illustrated in eFigures 7 (for birth year effects) and 8 (for event year effects). Each dot represents the parameter estimate from a single resampling/reestimation step. Note that due to the resampling technique and the rareness of the outcome, some datasets will not have sufficient information to validly estimate all event year and birth year effects, especially not those near the boundaries.

In eFigure 7 we see that there is a general increase in the risk of affective disorders with increasing birth year. Please note that the estimates are contrasts relative to the reference year 1982. We also find increasing variability with increasing birth years. However, this is most likely due to the rather short follow-up time available for the later generations. In eFigure 8 we present the estimated event year effects. Now the estimates are contrasts relative to the reference year 2014. We find very clear tendencies towards increased risk of an affective disorder event from 1996 to 2009 and a subsequent decrease. This corresponds quite well with the trends found in the Lexis heatmap. There thus exists a very strong event year (period) effect of affective disorder in Denmark during the study period, even though it does not seem to confound the relationship between cesarean delivery and affective disorder risk. This implies that future studies with affective disorder as the outcome should be cautious about potential period effect confounding.



eFigure 7: Parameter estimates of the birth year effects on their original scale. Each dot represents one resampling/reestimation step. In the upper panel, all resampling steps have been included, while resampling steps resulting in effect estimates less than -5 have been excluded in the lower panel. Here, a loess smoother has furthermore been added in black. Please note that the year 1982 serves as the reference category.



eFigure 8: Parameter estimates of the event year effects on their original scale. Each dot represents one resampling/reestimation step. In the upper panel, all resampling steps have been included, while resampling steps resulting in effect estimates less than -5 have been excluded in the lower panel. Here, a loess smoother has furthermore been added in black. Please note that the year 2014 serves as the reference category.

Sensitivity towards how the outcome variable is defined

In this section, we investigate the robustness of the results towards how the outcome variable is defined. Specifically, we compare the following three options for outcome variable definitions:

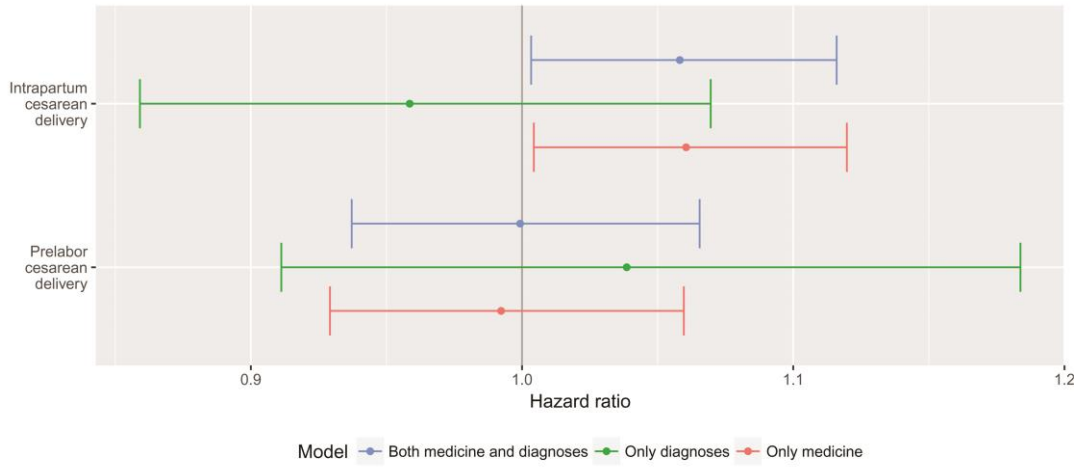
Both medicine and diagnoses: This is the definition used in the primary analyses. We model the time to either an affective disorder diagnosis (depression or bipolar disorder) or two redeemed prescriptions for medication (an antidepressive or lithium), whichever comes first. There are n_e^{ori} = 92,371 such events in the dataset.

Diagnoses: We only include information about diagnoses. There are n_e^{diag} = 22,067 such events in the dataset.

Medicine: We only include information about medicine. There are n_e^{med} = 88,245 such events in the dataset.

Please note that because there are a lot less diagnosis events than medicine events, the former outcome definition will result in less precise effect estimates. Note also that it is possible – and likely – to have both a diagnosis event and a medicine event, as evident from eTable 4. We fit a between-within model with confounder adjustment level 4 for each of these outcome definitions. The within-effect estimates for mode of delivery are presented in eFigure 9. We see rather broad confidence intervals for the *only diagnosis* model, as expected. The results using *only medicine* are very similar to those from the primary outcome definition, which is rather unsurprising as most events are medicine events. When the outcome is defined only using diagnosis events, we encounter a protective effect of being delivered by intrapatrum cesarean section, although not significantly so.

The overall conclusions regarding no effect of pre-labor cesarean delivery and only minimal effect of intrapartum cesarean delivery on the risk of affective disorders thus still stands, though the point estimates vary a bit when comparing medicine events and diagnosis events.



eFigure 9: Within-effect parameter estimates and 95 % confidence bands from the three between-within models with full confounder adjustment and varying definitions of the outcome variable. Only exposure effect estimates are displayed.

	Medicine: No	Medicine: Yes	Total
Diagnosis: No	917,073	70,304	987,377
Diagnosis: Yes	4,126	17,941	22,067
Total	921,199	88,245	1,009,444

eTable 4: The interplay between affective disorder diagnoses (depression or bipolar disorder) and -medicine (antidepressive or lithium). Please note that the number of events in the original outcome definition (diagnosis or medicine) can be calculated as the total number of medicine events (88,245) plus the total number of diagnosis events (22,067) minus the number of children with both events (17,941).

Sensitivity towards how mode of delivery is defined

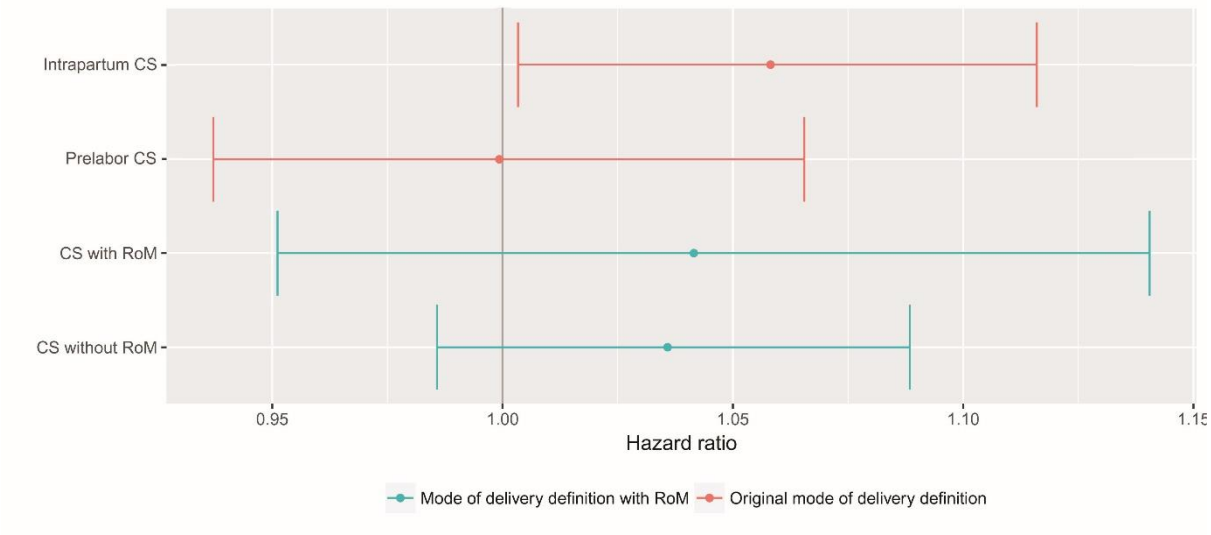
We now compare two definitions of mode of delivery, namely:

- Mode of delivery in three categories: Vaginal birth, pre-labor cesarean section, intrapartum cesarean section (as used in the main analyses).
- Mode of delivery in three categories: Vaginal birth, cesarean section before rupture of membranes (ROM), cesarean section after ROM.

In eTable 5 we present the interplay between the two variables. It should be noted that this table shows that the two variables are inconsistent: 31 children have been categorized as vaginal births in one variable and cesarean sections in the other. The new variable that uses information about ROM is expected to be less reliable than the original mode of delivery variable and thus these mistakes are likely to be due to errors in the new variable. In eFigure 10 we present within-effect estimates from two BW models with confounder adjustment level 4 that differ only by how their mode of delivery variables are defined. Only the exposure effects are included in the figure, but the remaining effect estimates produce virtually identical results across the two models. For mode of delivery, we reach the same overall conclusion in both models, but the point estimates differ a bit. In particular, when categorized according to ROM, both types of cesarean delivery result in increased risks of affective disorders, whereas the intrapartum/pre-labor definition finds an increased risk for intrapartum cesarean delivery. Moreover, the borderline-significant estimate found for intrapartum cesarean in the main analyses is not present for any of the cesarean deliveries using the ROM categorization. All in all, we conclude that there is little sensitivity towards the choice of defining mode of delivery relative to the timing of cesarean sections, rather than to the occurrence of rupture of membranes. The ROM categorization provides further evidence against the hypothesis of contact to vaginal microbiota being protective against affective disorder risk, as we have also interpreted the results from the main analyses to suggest.

New variables ⇨ / ***Original variables ⇩***	Vaginal birth	Cesarean section with ROM	Cesarean section without ROM	Total
Vaginal birth	882,584	0	31	882,615
Intrapartum cesarean	0	18,980	55,511	74,491
Pre-labor cesarean	0	968	51,370	52,338
Total	882,584	19,948	106,912	1,009,444

eTable 5: A cross-tabulation of the two variables for mode of delivery. Note that the data quality of the "new" variable is lower than the "original" variable, which might explain the inconsistencies found in this tabulation.



eFigure 10: Within-effect parameter estimates and 95% confidence bands from the two between-within models with full confounder adjustment and with varying definitions of the mode of delivery variable. Only exposure variable effects are shown. Note that in both models, the reference category is vaginal birth. *Cesarean section with ROM* and *Cesarean section without ROM* represent cesarean sections (CS) with and without rupture of membranes, respectively.