UNIVERSITY OF COPENHAGEN

Prenatal Exposure to Alcohol and Behavioural and Emotional Development at Age Seven Assessed by the Strengths and Difficulties Questionnaire (SDQ)

An investigation of the psychometric properties of the SDQ and empirical results from a large-scale cohort study

Niclasen, Janni

Publication date:
2013

Document version
Early version, also known as pre-print

Citation for published version (APA):
Niclasen, J. (2013). *Prenatal Exposure to Alcohol and Behavioural and Emotional Development at Age Seven Assessed by the Strengths and Difficulties Questionnaire (SDQ): An investigation of the psychometric properties of the SDQ and empirical results from a large-scale cohort study*.

Download date: 08. Apr. 2020

FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF COPENHAGEN

PhD thesis

Janni Niclasen

Prenatal Exposure to Alcohol and Behavioural and Emotional Development at Age Seven Assessed by the Strengths and Difficulties Questionnaire (SDQ):

An investigation of the psychometric properties of the SDQ and empirical results from a large-scale cohort study

Prenatal Exposure to Alcohol and Behavioural and Emotional Development at Age Seven Assessed by the Strengths and Difficulties Questionnaire (SDQ):

An investigation of the psychometric properties of the SDQ and empirical results from a large-scale cohort study

Janni Niclasen, August 2013

Department of Psychology, University of Copenhagen

Supervisors

Main supervisor	Associate professor Thomas W. Teasdale, Fil Dr., Dr. Med. Sci. Department of Psychology, University of Copenhagen, Denmark
Secondary supervisor	Associate professor Anne Mette Skovgaard, MD, Dr. Med. Sci., Child and Adolescent Psychiatric Centre Glostrup, Faculty of Health Science, University of Copenhagen, Denmark
Secondary supervisor	Professor Anne Marie Nybo Andersen, MD, PhD, Department of Public Health, University of Copenhagen, Denmark

Evaluation committee

Barbara Hoff Esbjørn	Associate professor, psychologist, PhD., Department of Psychology, University of Copenhagen, Denmark
Niels Bilenberg	Professor, MD, PhD., Department of Child and Adolescent Psychiatry, University of Southern Denmark, Denmark
Ron Gray	Consultant Clinical Epidemiologist, MB ChB, MPH, FFPH, FRCPsych, National Perinatal Epidemiology Unit (NPEU), University of Oxford, England

Table of Contents

List of papers

This thesis is based on the following papers:

Paper 1. Niclasen J, Teasdale TW, Andersen A-MN, Skovgaard AM, Elberling H, et al. (2012) Psychometric Properties of the Danish Strength and Difficulties Questionnaire: The SDQ Assessed for More than 70,000 Raters in Four Different Cohorts. *PLoS ONE* 7(2): e32025. doi:10.1371/journal.pone.0032025

Paper 2. Niclasen J, Skovgaard AM, Andersen AM, Sømhovd MJ, Obel C. A confirmatory approach to examining the factor structure of the Strengths and Difficulties Questionnaire (SDQ): a large scale cohort study. *Journal of Abnormal Child Psychology*, 2013 Apr;41(3):355-65. doi: 10.1007/s10802-012-9683-y

Paper 3. Niclasen, J., Andersen, AMN., Teasdale, TW., Strandberg-Larsen, K. Binge Drinking and Cumulated Alcohol Exposure in Pregnancy: Behavioural and Emotional Development at Age Seven (*submitted for publication*)

Paper 4. Niclasen, J., Andersen, AMN., Strandberg-Larsen, K., Teasdale, TW. Is Maternal Alcohol Binge Drinking in Early and Late Pregnancy Associations with Child Behavioural and Emotional Development at Age Seven? (*submitted for publication*)

Paper 5. Niclasen J. Drinking or not drinking in pregnancy: the multiplicity of confounding influences (*accepted for publication in Alcohol and Alcoholism*)

Thesis structure

This Ph.D. thesis consists of two parts. Part one is investigating the psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). Part two is concerned with prenatal exposure to alcohol and parent-rated SDQ scores at age seven. Although both are structured according to the IMRAD (introduction, method, results and discussion) structure, the composition of the two parts is somewhat different. The "SDQ" part is based on the results of Paper 1 and Paper 2 and additional analyses and results presented in Appendices A-I. The "alcohol" part of the thesis is based on the results of Paper 3, Paper 4, Paper 5 and Appendices J, K and L. The method sections of both parts of the thesis is somewhat more discussing than would be expected in scientific research articles. The result sections of the two parts of the thesis are very different. The result section of the "SDQ" part is very comprehensive and new results not presented in the articles are incorporated here. For the "alcohol" part of the thesis the results are briefly presented in bullet points. Some verbatim overlap inevitably appears in the "SDQ" section, but very minimally in the "alcohol" section.

Abbreviations

ABC:	Aarhus Birth Cohort
ADHD:	Attention Deficit Hyperactivity Disorder
ARBD:	Alcohol Related Birth Defects
ASD:	Autism Spectrum Disorder
AVE:	Average Variance Extracted
BAC:	Blood Alcohol Concentration
CBCL:	Child Behavioural Check-List
CCC2000:	Copenhagen Child Cohort 2000
CFA:	Confirmatory Factor Analysis
CR:	Composite Reliability
DAWBA:	Development And Well Being Assessment
DNBC:	Danish National Birth Cohort
EFA:	Exploratory Factor Analysis
FAS:	Fetal Alcohol Syndrome
FASD:	Fetal Alcohol Spectrum Disorder
GA:	Gestational Age
MAR:	Missing At Random
OCD:	Obsessive Compulsive Disorder
PAE:	Prenatal Alcohol Exposure
SD:	Standard Deviation
SDQ:	Strengths and Difficulties Questionnaire
SEM:	Structural Equation Modelling
SES:	Socio-Economic Status
SGA:	Small for Gestational Age

Acknowledgements

This Ph.D. project was supported by the Department of Psychology, University of Copenhagen. It was further supported financially by Ludvig og Sara Elsass Fond, Lundbeckfonden, Aase og Ejnar Danielsen's Fond, Fonden til lægevidenskabens fremme - The A.P. Møller Foundation for the Advancement of Medical Science, Børne- og Ungdomspsykiatrisk Selskab i Danmark, Carl J. Becker's Fond, Dagmar Marshalls Fond, Direktør Jakob Madsen og hustru Olga Madsens Fond. I am grateful to the Department of Psychology and all of these funding agencies and foundations.

I would like to thank my colleagues at the Department of Psychology who have supported me morally and academically throughout the process. These include Tina Birk Irner for invaluable discussions on prenatal exposure to alcohol and child development, and Mikael Julius Sømhovd and Christian Gaden Jensen for great discussions on structural equation modelling. Further, I am grateful to Ingrid Leth for being so open-minded and generous and inviting me to take part in exciting research projects. Also, a special thanks to other colleagues in the Copenhagen Child Anxiety Project (CCAP) for enriching discussions on child anxiety and developmental psychopathology. A very special thanks to my dear friend and good colleague Silja Henderson for precious lunch-company and for making my Ph.D. journey less lonely.

I am also truly indebted to Carsten Obel, Department of Public Health, AU for deciding to work with me. I am very grateful for having had the opportunity to work with Carsten, and thankful that he has shared his thorough knowledge on child and adolescent psychopathology in epidemiology with me.

Also, I would like to thank my two external supervisors, Anne Mette Skovgaard and Anne Marie Nybo Andersen. I am very grateful to have had the opportunity to work with Anne Mette Skovgaard and all of her colleagues at the Child and Adolescent Psychiatric Centre Glostrup. Thanks for sharing knowledge within the field of child- and adolescent psychiatry. Also thanks to Anne Marie Nybo Andersen and her colleagues in the POPE research group at the Department of Social Medicine, Institute of Public Health, UCPH. Thanks to Anne Marie for letting me stay one year at her department, and for sharing knowledge within the field of mother-child interactions in epidemiology.

I would particularly like to thank my main supervisor, Tom Teasdale for always being patient with me, and for having such a great sense of humour. I would also like to thank Tom for always responding quickly to my (many) emails - even one Christmas day! It has been great working with Tom who has supported me throughout the process when I needed advice on diverse issues. Tom has taught me much about statistics, and has encouraged me to continue my work when for one reason or another I was discouraged – he always managed to get me back on track with his emphatic and humorous personality.

A special thanks to my family, my parents Anna and Finn Niclasen for helping me out with my daughters, picking them up from day-care and looking after them when they were ill. The same goes for my aunt, Sys Kastrup, with invaluable support throughout the project. Also, a very big

thanks to Torben Baslund for being a good and caring father for our two lovely daughters and helping me out when the work load was particular demanding.

Finally, I am mostly grateful to my two fantastic daughters themselves, Liva Ingrid and Vega Tilla. Thanks to both of you for your powerful and unconditional love and thank you for keep reminding me that life is so much more than academia.

Janni Niclasen, August 2013

Abstract

Untreated childhood psychopathology may often develop into adult psychiatric disorders [1-3]. Because this is associated with great direct and indirect costs for the individual and for society, studies identifying the role of potential risk factors are needed. Risk factors for the development of psychopathology are numerous and comprise both pre- and post-natal factors. Prenatal risk factors include exposure to smoking [4-6], alcohol [7], malnutrition [8] and coffee [6, 9], maternal pre-pregnancy adiposity [10], season of birth [11], maternal stress and anxiety in pregnancy [12], and low birth weight [12]. Post-natal risk factors associated with later psychopathological development includes psychopathology on the maternal side [8], criminal behaviour on the paternal side [8], low income and little education [8], personality, IQ, a violent and abusive home environment [8, 13, 14] and a non-secure parental attachment style [8, 13, 14].

Whereas there is a long tradition within epidemiology to focus on prenatal factors, epidemiologists have generally paid very little attention to post-natal factors [15]. Psychologists have, conversely, long recognised the importance of the post-natal environment for the development of psychopathology, but have paid very little attention to pregnancy related risk factors [15]. The starting point of this Ph.D. project is an integration of knowledge from each of these two disciplines. Specifically, the aims are to: 1. investigate the psychometric properties of the Danish version of the Strengths and Difficulties Questionnaire (SDQ); and 2. investigate the potential associations between prenatal exposure to alcohol and behavioural and emotional development assessed by the SDQ at age seven.

The alcohol related studies are important as it has been claimed that that mother's drinking during pregnancy may affect the neurodevelopment of around 1 % of all children [16]. *Fetal Alcohol Spectrum Disorders* (FASD) is the umbrella term used to classify children exposed to alcohol prenatally from *Fetal Alcohol Syndrome* (FAS) at the one end of the disorder spectrum to *Alcohol-Related Birth Defects* (ARBD) at the other end. The term FAS was coined in the early 1970's to describe those children most heavily exposed to alcohol and who exhibit a specific pattern of growth retardation, dysmorphic facial features and Central Nervous System (CNS) dysfunctions. The emphasis in the early years was to identify the effects on mental health of being exposed to very large doses of alcohol. Gradually, the focus has shifted toward investigations looking at exposures to much lower doses of alcohol, typically < 1 unit/week. The "hot-topic" today is therefore to identify whether there is a safe, lower level below which drinking is not associated with any harm to the developing foetus.

The findings from observational studies are somewhat contradictory. Some studies have indeed found associations between prenatal exposure to lower doses of alcohol and neurodevelopmental outcomes in childhood [17-24], whereas other studies have observed no such associations [25-29]. Many studies have even reported a J-shaped association, such that exposure to low doses of alcohol has an apparently protective effect on the foetus [30-32].

Apart from exposure to very low doses of alcohol researchers have also paid separate attention to prenatal exposure to *binge drinking*, most often defined as an intake of a minimum five units of alcohol on a single occasion. The rationale for this distinction is that exposure to binge drinking is thought to be more devastating for the developing Central Nervous System (CNS) because it is the peak blood alcohol (BAC) concentration that determines the level of the damage on part of the child [33, 34]. Compared to the literature investigating exposure to lower average doses of alcohol, there seem to be somewhat more evidence for a serious effect of being exposed to binge drinking prenatally. One review [35] concluded that children exposed to binge drinking consistently showed poorer neurodevelopmental outcomes in childhood [23, 34, 36-40]. However, the literature is not entirely consistent and other studies have reported no associations with such outcomes [38, 41-44].

In the observational literature described above different outcome measures are used to assess neurodevelopment in childhood. One such very often used instrument is the Strengths and Difficulties Questionnaire (SDQ) which is a screening tool developed to assess behaviours, emotions and interpersonal relationships in young children and adolescents. The SDQ contains 25 questions that ask about different positive and negative aspects of the child's behaviour. The items are scored *"not true"*, "*somewhat true"* and "*certainly true"*, and are divided into five scales (*Hyperactivity/ inattention, Emotional, Conduct, Peer-problems* and *Prosocial*) comprising five items each [45]. The SDQ has been used internationally in clinical as well as research settings. Although Denmark might be the place in the world where the SDQ has been used most extensively the psychometric properties of the Danish version of the SDQ have hitherto not been investigated.

The present thesis consists of two parts. The purpose of the first part was to thoroughly investigate the psychometric properties of the SDQ from an exploratory (Paper 1) as well as from a confirmatory (Paper 2) factor analytic perspective. The aim was further to develop age and gender specific norms (Appendices C-H) for the Danish version of the SDQ based on more than 70.000 parent and teacher raters. Data for the studies derived from four large scale Danish birth cohorts: the Danish National Birth Cohort (DNBC), the Copenhagen Child Cohort (CCC2000), the Aarhus Birth Cohort (ABC) and SFI forløbsundersøgelse (SFI). The results from these studies revealed that the factor structure was manifested, and good scale reliability particularly for the *Hyperactivity scale*, and satisfactory validity was observed. It was concluded that the Danish version of the SDQ works well psychometrically, particularly so for older children rated as by teachers, and less so, but still acceptably, for younger children as rated by their parents.

The scope of the second part of the thesis was to investigate associations between exposure to lower doses of alcohol (Paper 3) and binge drinking assessed in full pregnancy (Paper 3) and in first and third part of pregnancy (Paper 4) on the one hand and parent-rated SDQ-scores at age seven on the other hand. The aim was further to describe the characteristics of women who drink and women who do not drink alcohol in pregnancy and discuss the methodological implications of these findings (Paper 5 and Appendices J, K and L). Data for these studies derived from the DNBC that contains information on more than 100.000 pregnancies.

The analyses revealed no statistically significant associations between prenatal exposure to low doses of alcohol and behavioural and emotional development at age seven. In fact, the most favourable outcomes were observed for the high exposure group, whereas the least favourable outcomes were observed for the children in the abstaining group. Exposure to binge drinking in full pregnancy (Paper 3) was found to be negatively associated with parent-rated *Externalising, Internalising,* and *Conduct* scores in boys, but not in girls. Exposure to binge drinking in early as well as late pregnancy (Paper 4) was found to be negatively associated with *Externalising* scores at age seven. The associations were found to be higher for late pregnancy exposure (that is, worse outcomes), compared to early exposure. Regarding the findings from the study investigating background characteristics of women who drink and who do not drink alcohol in pregnancy, highly statistically significant differences were observed between exposure groups on most variables. That is, very large differences were observed on most potential confounding factors.

The thesis end up discussing potential explanations for the lack of consistency in the observational literature investigating associations between exposure to low doses of alcohol and binge drinking on the one hand, and neurodevelopmental outcomes one the other hand. The focus of this discussion is that the lack of consistency in the literature is considered a consequence of methodological limitations that prevails the literature namely: 1. confounding factors that are insufficiently controlled for in the statistical analyses; 2. mediating factors that are insufficiently controlled for in the statistical analyses; 3. poorly defined alcohol exposure categories that do not sufficiently incorporate *"dose", "pattern"* and *"timing"*; 4. other issues concerned with the definitions of the alcohol exposure categories; 5. the use of outcome measures, like the SDQ, that may not be sensitive enough to detect potential harmful effects; and 6. the children may often be assessed at too early ages when an effect may not yet have manifested itself.

On the basis of the general literature and the studies conducted for the purpose of this thesis, no firm conclusions can yet be drawn. Most convincing is the evidence from the binge studies that indicate that being exposed to just one episode of binge drinking, particular in the last part of pregnancy, does appear to have an effect on behavioural development at age seven. Less conclusive are the findings from the studies investigating exposure to lower doses of alcohol. It cannot be concluded that prenatal exposure to low doses of alcohol is negatively associated with neurodevelopment in childhood. However, because of methodological limitations it also cannot be concluded either that prenatal exposure to alcohol is *not* negatively associated with neurodevelopmental outcomes.

Although we currently do not have evidence that exposure to low doses of alcohol is negatively associated with neurodevelopmental outcomes in childhood I recommend that pregnant women abstain from drinking. The main argument for this is that we currently do not have any evidence that being exposed to alcohol in any ways does anything beneficial for the developing foetus. In the words of Garcia-Algar and colleagues: *"no evidence of harm does not mean evidence of no harm"* [46]. Because our research designs are so full of methodological faults and limitations the wisest course is to recommend abstinence – for the sake of the unborn children.

Danish summary

Det er i dag et veletableret faktum, at børn og unge, der udvikler mentale sundhedsproblemer og ikke kommer i behandling, har en øget risiko for at være mærket af psykiske vanskeligheder i puberteten og ind i voksenlivet [1-3]. Fordi dette er associeret med store direkte og indirekte omkostninger for den enkelte og for samfundet, har vi brug for undersøgelser, der identificerer potentielle risikofaktorers betydning for udvikling af psykopatologi hos børn og unge. Disse er mange og inkluderer både prænatale og postnatale faktorer. De prænatale risikofaktorer inkluderer eksponering til rygning [4-6], alkohol [29], kaffe [6, 9] fejlernæring i graviditeten [8], maternel pre-graviditets fedme [10], maternel stress og angst i graviditeten [8, 12], og lav fødselsvægt [8, 12]. Postnatale risikofaktorer associeret med senere udvikling af psykopatologi inkluderer psykiske problemer hos især moderen [8], kriminel adfærd hos især faderen [8], lav indkomst og dårlig uddannelse [8], IQ [8], vold og misrøgt i familien [8] og usikker tilknytning til forældrene [13, 14].

Mens der indenfor epidemiologien er en lang tradition for at fokusere på de prænatale faktorer, har epidemiologer generelt set haft meget lidt fokus på de postnatale faktorer. Omvendt har psykologer selvsagt en lang tradition for at fokusere på postnatale opvækstfaktorer og disses betydning for udvikling af psykopatologi, mens de har haft et meget begrænset fokus på prænatale graviditetsafhængige faktorer. Udgangspunktet for dette Ph.d. projekt er en integration af viden fra disse to discipliner.

Studier, der undersøger effekterne af prenatal eksponering til alkohol, er vigtige, idet det hævdes, at maternel indtag af alkohol i graviditeten påvirker udviklingen hos op til 1 % af alle børn. *"Fetal Alcohol Spectrum Disorder"* (FASD) er den paraplyterm, der bruges om børn, der har været eksponeret til alkohol i fostertilværelsen, og den inkluderer børn med *"Fetal Alcohol Disorder"* (FAS) i den ene ende af skalaen til *"Alcohol Related Birth Defects"* (ARBD) i den anden ende. FAS-termen har været anvendt siden 1970'erne og beskriver de børn, der har været eksponeret til de største mængder af alkohol i graviditeten og udviser et specifikt mønster af symptomer, inklusiv hæmmet vækst, dysmorphic ansigtstræk og et dysfunktionelt centralnervesystem. Siden da er der sket et gradvist skift i fokus, så der i dag er en øget interesse for, hvorvidt prenatal eksponering til meget små mængder alkohol, ned til en genstand om ugen, er negativt associeret med kognitive og mentale udfald i barndommen. Det centrale spørgsmål for forskere indenfor feltet i dag er derfor, hvorvidt der findes en sikker nedre grænse under hvilken eksponering til alkohol ikke er forbundet med dårligere kognitive og mentale udfald i barndommen. Hidtil har forskere ikke været i stand til at konkludere entydigt. Mens nogle studier har påvist en sammenhæng mellem eksponering til meget små mængder af alkohol og kognitiv og mental udvikling i barndommen [17-20, 22-24, 47], har andre forskere ikke været i stand til at påvise en sådan sammenhæng [28, 29, 42-44, 48]. Endnu andre studier har rapporteret en J-formet kurve mellem prænatal alkohol eksponering og kognitiv og mental udvikling i barndommen [30-32]. En sådan kurve indikerer, at eksponering til små mængder af alkohol tilsyneladende har en beskyttende effekt på barnet.

Udover fokus på eksponering til små mængder alkohol har forskere også haft fokus på *"binge drinking"* [33, 35, 41, 47, 48], hvilket i litteraturen oftest er defineret som et indtag af minimum fem

genstande ved en enkelt lejlighed. Rationalet bag denne opdeling er baseret på, at man formoder, at eksponering til *"binge drinking"* er værre for centralnervesystemets udvikling end eksponering til den samme mængde fordelt over flere dage eller uger, fordi det er koncentrationen af alkohol i blodet, der bestemmer omfanget af skade hos fostret [33, 35]. Sammenligner man den del af litteraturen, der fokuserer på eksponering til små mængder af alkohol med den del, der har fokus på *"binge drinking",* ser det da også ud til, at der er mere evidens for, at *"binge drinking"* har en mere negativ indflydelse på barnets udvikling. Et omfattende review-studie [35]konkluderede, at børn, der har været eksponeret til *"binge drinking"* i graviditeten, vedvarende udviste dårligere kognitiv og mental udvikling i barndommen [23, 35-40].

I litteraturen beskrevet ovenfor er der anvendt forskellige psykologisk orienterede spørgeskemaer og neuropsykologiske tests til vurdering af børnenes kognitive og mentale udvikling ved followup tidspunktet. Et af de instrumenter, der har været oftest anvendt, er spørgeskemaet Strengths and Difficulties Questionnaire (SDQ), der er et kort screening redskab til vurdering af adfærd, emotioner og forholdet til venner hos børn og unge. SDQ-skemaet består af 25 positivt eller negativt formulerede spørgsmål, og scores *"passer ikke", "passer delvist"* og *"passer godt"*. Skemaet dækker fire problemområder, nemlig hyperaktivitet/ uopmærksomhed, emotionelle problemer, adfærdsproblemer og problemer i forholdet til jævnaldrene [45]. Derudover vurderes barnet på et socialt styrkeområde. Hvert af disse områder dækkes med fem spørgsmål. SDQ er oversat til mere end 70 sprog og har fundet bred anvendelse i kliniske såvel som ikke-kliniske sammenhænge. Selvom Danmark måske er det land i verden, hvor SDQ er blevet anvendt allermest, har ingen forskere til dato undersøgt de psykometriske egenskaber ved den danske version af SDQ.

Ph.d. afhandlingen består overordnet set af to dele. Formålet med første del er at se på de psykometriske egenskaber ved SDQ i en dansk sammenhæng fra både et eksplorativt (Artikel 1) og et konfirmatorisk (Artikel 2) faktor analytisk perspektiv. Derudover er formålet at udvikle alders- samt kønsspecifikke normer (Appendiks C-H) for den danske forældre og lærer version af SDQ. Disse psykometristudier baserer sig på mere end 70.000 spørgeskemaer, udfyldt af forældre og lærere i forbindelse med fire store danske fødselskohorter: Bedre Sundhed for Mor og Barn (BSMB), Copenhagen Child Cohort (CCC2000), Aarhus Birth Cohort (ABC) samt SFI's forløbsundersøgelse (SFI). Det blev konkluderet i disse studier, at faktorstrukturen er god, at der er god skala reliabilitet især for *Hyperaktivitetsskalaen*, og endvidere god validitet. Det blev konkluderet at den danske version af SDQ har gode psykometriske egenskaber, specielt for ældre børn der vurderes af deres lærere, og knapt så gode (om end stadig acceptable) for yngre børn der vurderes af deres forældre.

Formålet med anden del af afhandlingen var at belyse sammenhængen mellem prenatal eksponering til små mængder af alkohol (Artikel 3), *"binge drinking"* målt i hele graviditeten (Artikel 3), og *"binge drinking"* målt i første og sidste del af graviditeten specifikt (Artikel 4) på den ene side, og forældre SDQ-scorer på den anden side. Formålet var endvidere at beskrive karakteristika ved kvinder, der drikker, og kvinder der ikke drikker alkohol i graviditeten, og

diskutere de metodiske implikationer af sådanne potentielle forskelle mellem eksponeringsgrupperne (Artikel 5 samt Appendix J, K og L).

Resultaterne fra disse alkoholstudier viste, at der ikke kunne findes nogle statistisk signifikante associationer mellem prenatal eksponering til små mængder alkohol på den ene side og adfærds- og emotionel udvikling på den anden. Rent faktisk viste det sig, at de mest gunstige udfald blev observeret for høj-eksponeringsgruppen, mens de dårligste udfald blev observeret for gruppen af børn, hvis mødre ikke havde indtaget noget alkohol i graviditeten. I forhold til *"binge drinking"* målt i fuld graviditet viste det sig at være negativt associeret med forældre-vurderet *Eksternalisering, Internalisering* og *Adfærdsproblemscorer* ved syvårsalderen hos drenge, men ikke hos piger. Eksponering til *binge drinking* i første såvel som sidste del af graviditeten viste sig endvidere at være negativt associeret med *Externaliserende* scorer ved syvårsalderen. Associationerne var højere, og altså dårligere, sidst i graviditeten sammenlignet med først i graviditeten. Vedrørende undersøgelserne, der så på baggrundskarakteristika hos kvinder, der drikker og ikke drikker alkohol i graviditeten, viste det sig, at der var statistisk signifikante forskelle eksponeringsgrupperne imellem på stort set samtlige af de undersøgte variable. Med andre ord, store forskelle blev observeret på alle undersøgte potentielle confounder variable.

Ph.d. afhandlingen diskuterer afslutningsvis mulige forklaringer på de uoverensstemmelser, der generelt set findes i litteraturen mellem prenatal eksponering til små mængder alkohol og *"binge drinking"* på den ene side og kognitive og mentale udfald på den anden side. Det konkluderes, at årsagen til disse uoverensstemmelser skyldes en lang række metodiske begrænsninger ved vores nuværende måde at opstille undersøgelser på, samt måden hvorpå vi designer vores studier. Den første årsag, der diskuteres, er problemer med residuale confounding og residuale medierende faktorer – altså præ- og postnatale baggrundsvariable, der ikke i tilstrækkelig grad er kontrolleret for i de statistiske analyser. Dernæst diskuteres problemer i forhold til utilstrækkelige definitioner og dårligt afgrænsede alkoholeksponeringskategorier. Slutteligt diskuteres forskellige problemer i forbindelse med udfaldsmålene, altså de psykologiske spørgeskemaer og neuropsykologiske tests, der anvendes. Blandt andet diskuteres det, at der kan være problemer med mangel på sensitivitet i de anvendte psykologiske spørgeskemaer eller tests. Endvidere at der kan være problemer med, at børnene testes på forkerte alderstrin – alderstrin, der i epidemiologiske undersøgelser oftest er valgt ud fra praktiske og logistiske årsager snarere end ud fra teoretiske overvejelser.

På baggrund af litteraturen som helhed og de videnskabelige arbejder, der danner basis for denne Ph.d. afhandling, sluttes det, at der ikke kan drages nogen endegyldig konklusion. Evidensen fra "*binge drinking"* studierne er dog mest overbevisende, og disse syntes at indikere, at eksponering til en enkelt episode af "*binge drinking"* nok er forbundet med øget *Eksternaliserende* scorer ved syvårsalderen – specielt hvis eksponeringen har fundet sted i sidste del af graviditeten. Mindre entydige er resultaterne fra studierne der ser på eksponering til små mængder af alkohol. Det kan på baggrund af litteraturen som helhed samt resultaterne fra indeværende Ph.d. projekt ikke konkluderes, at eksponering til små mængder af alkohol er negativt associeret med kognitiv og mental udvikling i barndommen. Omvendt kan det heller ikke konkluderes, at prenatal

eksponering til alkohol *ikke* er negativt associeret med kognitiv og mental udvikling i barndommen.

Selvom vi på nuværende tidspunkt ikke har evidens for at sige, at eksponering til små mængder af alkohol er negativt associeret med kognitiv og mental udvikling i barndommen mener jeg, at den Danske Sundhedsstyrelse bør bibeholde deres nuværende nultolerance anbefalinger. Argumentet herfor er, at vi i hvert fald ikke har evidens for at eksponering til alkohol på nogen måde er positivt associeret med psykologisk-orienteret udfald. Med et citat fra Garcia-Algar og kolleger [46] *"ingen evidens for en skadesvirkning betyder ikke, at vi har evidens for ingen skadesvirkning"* (min oversættelse). Derfor, og fordi vores forskningsdesign er tydeligt fulde af metodiske fejl og begrænsninger, er den eneste fornuftige ting for barnets skyld at anbefale afholdenhed til gravide og kvinder, der planlægger at blive gravide.

INTRODUCTION: The Danish parent and teacher versions of the Strengths and Difficulties Questionnaire: psychometric properties and clinical cut-offs

The Strengths and Difficulties Questionnaire (SDQ) is an screening tool developed to assess behaviours, emotions and relationships in young children and adolescents. It is one of the most widely used, brief screening instruments for assessing mental health in children and adolescents, and it is used internationally in both developed and developing countries [49-67]. The primary aim of the questionnaire is to identify children who are at high risk of psychiatric disorders and who therefore warrant further assessment [45, 68]. The SDQ consist of 25 items that have been constructed on the basis of nosological concepts as well as factor analyses [69]. It contains five scales (*Hyperactivity-inattention* (hereafter *Hyperactivity), Conduct disorder, Emotional problems*, *Peer-problems* and *Prosocial)* of five items each. It was developed by Goodman in the early 1990s, and is based on the much longer Rutter questionnaire [49, 68]. Similarly, the Child Behavioural Checklist (CBCL) developed by Achenbach in 1991 is another screening tool that contains 118 questions that ask about problematic behaviours among 4-16-year-old children and adolescents [49, 70]. The goal of the SDQ was to meet the needs of educationalists, clinicians and researchers and it is used in clinical as well as non-clinical settings. In research settings with clinical samples, it has generally been found that diagnostic predictions made by the SDQ agree well with clinical diagnoses [71-74]. The advantage of the SDQ compared to, for example, the Rutter questionnaire and the CBCL is that it is much shorter and therefore more suitable for large-scale cohort-based research purposes. Furthermore, items on strengths on the part of the child are included rather than an exclusive focus on deficits as is the case of the Rutter questionnaire and the CBCL. This also makes it more suitable for use in low risk epidemiological settings and within educational environments. However, inclusion of the items on strengths as well as positively worded, so-called *"reversed"* items have also been found to be a major psychometric challenge [56].

Factor analysis is a statistical method that is used to describe the covariability among a number of observed variables in terms of a potentially lower number of unobserved, latent traits [75]. Two types of factor analytical methods have been deployed in the literature: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Whereas the goal of EFA is to identify factors based on data and to maximize the total amount of variance explained, the aim of CFA is to evaluate a priori hypotheses that are based on theory. Strictly speaking, EFA is a method that should be applied only when 1. there is no a priori hypothesis about how the measured variables relate to one another; 2. a new questionnaire that aims to measure underlying variables is constructed; or 3. a large data set needs to be reduced to a more manageable size while retaining as much of the original information as possible [75, 76]. Despite of this several studies have looked at the factor structure of the SDQ utilizing EFA methodology and most of these have been able to confirm a five factor structure [52, 55, 57, 64, 66, 77-79]. However, some problems have been observed for the conduct and peer problem scales [55].

Whereas EFA can be understood as a descriptive approach to factor analysis, the aim of CFA is to test how well data fit a hypothesised, a priori, theory-based measurement model [80]. It takes a

structural analytic approach and constitutes the measurement part of structural equation modeling (SEM). It is thus a technique that is used when the researcher holds some prior expectation about the structure of the latent factors, and aims to test how well data fit one or more theoretical derived models. The results of the CFA studies have varied. Some studies have found support for the originally proposed five factor model [56, 81, 82], others for a three factor solution that adds two second order *internalizing/ externalizing factors* to the model [56, 82, 83]. Other studies again have found support for a model that includes a hypothesized *positive construct factor* in addition to the four problem scales [84].

SDQ mean scores and standard deviations (SD) for non-clinical samples have been found to vary between different European settings [67, 69, 85]. British published mean scores tend, to be higher than Northern European means [67, 69, 79, 85, 86], but similar to or lower than the mean scores reported for Southern European countries [87, 88]. In other words, differences in means scores and cut-offs have been found to reflect variation on a North to South European gradient with higher scores observed in the South. Few studies have compared scores between age groups within the same cultural setting. However, one study did find that older children scored lower than younger ones on the *hyperactivity scale* indicating that younger ones exhibited more behavioural problems than older ones [85].

Despite its small size, Denmark is, as stated, probably the country in the world where the SDQ has been used most often. It is included in virtually all of the follow-ups of the large scale Danish birth cohort studies. It is also used as part of "skolesundhed.dk", a program that collects information on school-aged children's health and development that, among other things, screens for ADHD among children starting school. Despite this very wide use, no one had hitherto investigated the psychometric properties of the Danish version of the SDQ. Furthermore, norms and clinical cut-offs have not previously been developed. Because of its wide use in clinical and well as research settings it was believed that there was a great need for studies investigating issues of norming.

AIMS: Psychometric properties of the SDQ

The overall aim of these studies was to thoroughly investigate the psychometric properties of the Danish parent and teacher versions of the SDQ in a non-clinical sample created by merging data from four large-scale Danish cohorts. Specifically, the aims were to:

1. Thoroughly describe the Danish version of the parent and teacher forms of the SDQ and investigate the psychometric properties including the factor structure from an EFA perspective (paper 1).

2. Investigate the psychometric properties of the Danish version of the parent and teacher forms of the SDQ from a CFA perspective (paper 2).

3. Develop gender and age specific norms for the Danish version of the parent and teacher forms of the SDQ (now available at http://www.sdqinfo.com/DanishNorms/DanishNorms.html and Appendices C-H).

METHODS: Psychometric properties of the SDQ

Samples

Data for the factor analytic studies were derived from four large-scale birth cohorts, namely the Danish National Institute of Social Research (SFI), the Copenhagen Child Cohort (CCC2000), the Danish National Birth Cohort (DNBC), and the Aarhus Birth Cohort (ABC). The methodologies of the individual cohorts have been described in more detail elsewhere [89-92].

The SFI longitudinal project is a birth cohort initiated in 1995 [90]. A simple randomly selected community sample of 5,998 children born in the autumn of 1995 and their parents was initially contacted of whom 90.5 % of the parents agreed to participate in the study. For the 7-year follow-up a total of 4,971 parents participated in the study.

The CCC2000 is a birth cohort of children born in the year 2000 within the Copenhagen County and includes information on 6,090 children. Of the 5,898 eligible for 5-year follow up a total of 3,501 parents and teachers were included in the studies [93].

The DNBC includes information on 101,042 pregnancies and data were initially collected between 1996 and 2002 [92]. Of the 83,315 qualified for the 7-year follow-up in October 2009 (when the data were drawn) a total of 48,544 parents had filled in the questionnaire.

The ABC approached all pregnant women receiving prenatal care in Aarhus between 1989 and 1996. The ABC comprises information on 26,324 women who gave birth between 1990 and 1992 and among whom a total of 8,422 participated in the 10-12-year follow-up. Of the total number of teachers eligible to follow-up 55 % completed the SDQ [94-97]. In total, 77,005 raters were included in the four studies.

Materials

The SDQ contains 25 questions and an *Impact supplement*. The 25 questions ask about different positive and negative aspects of the child's behaviour, and can be scored *"not true"*, "*somewhat true"* and "*certainly true"*. Of the 25 questions, 10 are generally thought of as strengths, 14 as difficulties and 1 as a neutral question. The items are divided into five scales (*Hyperactivity/ inattention, Emotional, Conduct, Peer problems* and *Prosocial*) comprising five items each [45]. The first four scales are summed to obtain a *Total difficulties score* whereas the *Prosocial scale* was included in order to enhance acceptability on the part of the rater [45]. The questions have been selected on the basis of contemporary nosological concepts as well as factor analytically derived

dimensions [45, 98]. An extra *Impact supplement* begins with one screening question asking whether the rater ''*overall thinks that the child has difficulties in one or more of the following areas: emotions, concentration, behaviour or being able to get on with other people*''. If the rater answers ''*yes*'' to this question further items inquire about the severity of these difficulties. The *Impact supplement* provides an important estimate of the burden of the problems which is an essential part of the diagnostic criteria in the current diagnostic classification systems, ICD-10 and DSM-V [45, 98]. The parent and teacher versions of the SDQ were translated in 2001, implementing standard back-translation procedures and using concepts and terms that were in keeping with the time [99]. Parallel parent, teacher and self-rating versions of the questionnaire exist.

Statistical analyses

All analyses for Paper 1 were carried out using the statistical package SPSS version 18. For Paper 2 the method of Confirmatory Factor Analysis (CFA) was chosen as the appropriate means to test three hypothesised models and these analyses were performed using the statistical package M+ version 6.12. As the 25 items all had skewed or very skewed distributions, all statistical group analyses for Paper 1 were carried out by means of Mann-Whitney's U-test and all analyses for Paper 2 were likewise treated on a categorical level. Principal Component Analyses (PCA) were carried out for Paper 1. It can be argued that PCA strictly speaking is not a method of EFA. Whereas PCA assumes that the sample used is the entire population, EFA methods assume that participants are randomly selected. Whereas PCA decomposes the original data into a set of linear variates, EFA derives a mathematical model from which factors are estimated. Whereas PCA is concerned with identifying which linear components exist within the data and how a particular variable might contribute to that component, EFA estimates the underlying factors on the basis of various mathematical assumptions [75]. However, despite these differences, and because a model with 25 items with commonalities > 0.70 has been found to differ little from EFA extracted results, the term EFA is used consistently to refer to PCA followed by a rotation procedure below like in most parts of the literature.

RESULTS: Psychometric properties of the SDQ

Missing data

The devisor of the SDQ, Robert Goodman recommends a case-wise deletion, i.e. that cases are included only when a minimum of three of the five items are responded to on any single scale [100]. Kline, on the other hand, suggests a list-wise deletion of cases, if less than 5 % of data are missing on a single variable [101]. In practice, few researchers apply Goodman's recommendations [67] and in the literature any missing values most often result in a list-wise deletion of cases [85, 102, 103]. In the present studies missing values were considered missing at random (MAR), and since they constituted less than 0.05 % of all data, they resulted in a list-wise deletion of cases.

Response frequencies

As stated, all of the SDQ items were found to be skewed or very skewed. Because this skewness is of great importance for the way the data are treated in the remaining part of the thesis, the response frequencies for each of the 25 items for the 5-7-year-old parent ratings are included in appendix A. It appears that all items are non-normally distributed, especially the conduct and peer problem items. Particularly skewed are the two conduct items *"often fights with other children or bullies them"* and *"steals from home, school or elsewhere"* with only 0.6 and 0.3% of responders agreeing the item to be *"certainly true"* and 95.6% and 98.1% declaring it *"not true"*.

Factor analyses

In order to determine what number of factors to extract for Paper 1 the Scree plot as well as the number of factors with an initial Eigenvalue > 1 were evaluated. For all of the tested samples (different cohorts, younger and older children, boys and girls and parent and teacher raters) the optimal solution proved to be a replication of Goodman's originally proposed five factor solution [98].

It was decided to report estimates from the Promax rotation (using Kappa setting by 0.40) because this rotation includes results from *Orthogonal* as well as *Oblique rotations*. The *Structure Matrix* reports the estimated factor scores from the *Orthogonal rotation*, which is a rotation method that assumes independence between the underlying factors. The *Pattern Matrix* on the other hand displays the estimates from the *Oblique rotation*. As oppose to the *Structure Matrix*, this rotation method does not assume the underlying factors to be independent of each other. Rather, the method allows the underlying factors to be related to one another.

Overall, the results of the EFAs revealed that virtually all of the 25 items showed the highest loadings on their respective scales. Higher factor loadings were generally found for teacher ratings than for parent ratings. The values of the *Structure matrices* for both parents and teachers showed unequivocally high loadings on their intended scales. The picture for the *Pattern matrices* on the other hand revealed a somewhat more ambiguous picture. For parents, some *conduct items* showed high loadings on the other scales and conversely, *non-conduct items* loaded highly onto the *Conduct scale*. Positively worded items further tended to load on to the *Prosocial scale*. This picture was even more pronounced for teachers' ratings. High factor loadings were revealed for all five positively worded items on the *Prosocial scale* and four of the five *conduct items* loaded highly onto the *Hyperactivity scale*. Additionally, high cross-loadings were observed for some *peer-problem items* on the *Emotional* and *Conduct scales*, and conversely, *emotional* and *conduct items* did tend to load highly on the *Peer problem scale*.

In CFA, three different approaches to testing structural equation models can be applied [80, 104]: a *"strictly confirmatory"* approach in which it is tested how well a single model based on theory fits

the data. The model is either confirmed or rejected, and no further modifications are made to the model. In the second *"alternative models"* approach, several theoretical models are proposed and one model is selected as the most appropriate in representing the sample data. The final *"model generating"* approach represents the case where a theoretically-based model has been rejected, and on this basis one proceeds in an exploratory manner where the model is modified and re-estimated [104].

In Paper 2 it was decided to investigate how well data fitted three "*alternative"* theoretically-based, hypothesized measurement models (please see Figure 1). In addition, it was decided to continue with a *"model generating"* approach, and allow for a minimum of model modifications. These were only allowed provided that they made theoretical sense. The three models were initially evaluated on the basis of four different overall model fits: the Chi-square test statistics, Root Mean Square Error of Approximation (RMSEA), Bentler Comparative Fit Index (CFI) and Tucker-Lewis Fit Index (TLI). It is important to bear in mind that the fit indices do not yield much information themselves. Rather, they indicate how well the model overall fits the data, and indicate whether something is overall wrong with the model.

For the unadjusted models 1, 2 and 3 the Chi-square model fit, CFI and TLI were consistently found to be unacceptable for the parent samples. For the teacher samples unacceptably high Chi-Square model fits and just acceptable CFI and TLI model fits were observed. To achieve better model fits it was decided to opt for the "*model generating"* approach and allow for a minimum of theoretically meaningful modifications. A model that allowed for the following modifications was decided on: cross-loadings between items 22 and 18 (two *conduct items*), items 10 and 2 (two *hyperactivity items*) and items 20 and 9 (two *prosocial items*) as well as cross-loadings between the *prosocial scale* and the positively-worded, reversed items 21 and 14. These modifications significantly improved the model fits for Model 1 and Model 2 for all of the samples. Despite the modifications the value of the Chi-square model fit consistently remained extremely high. However, it is a well-known problem in SEM that large sample sizes do cause problems for the Chi Square model fit.

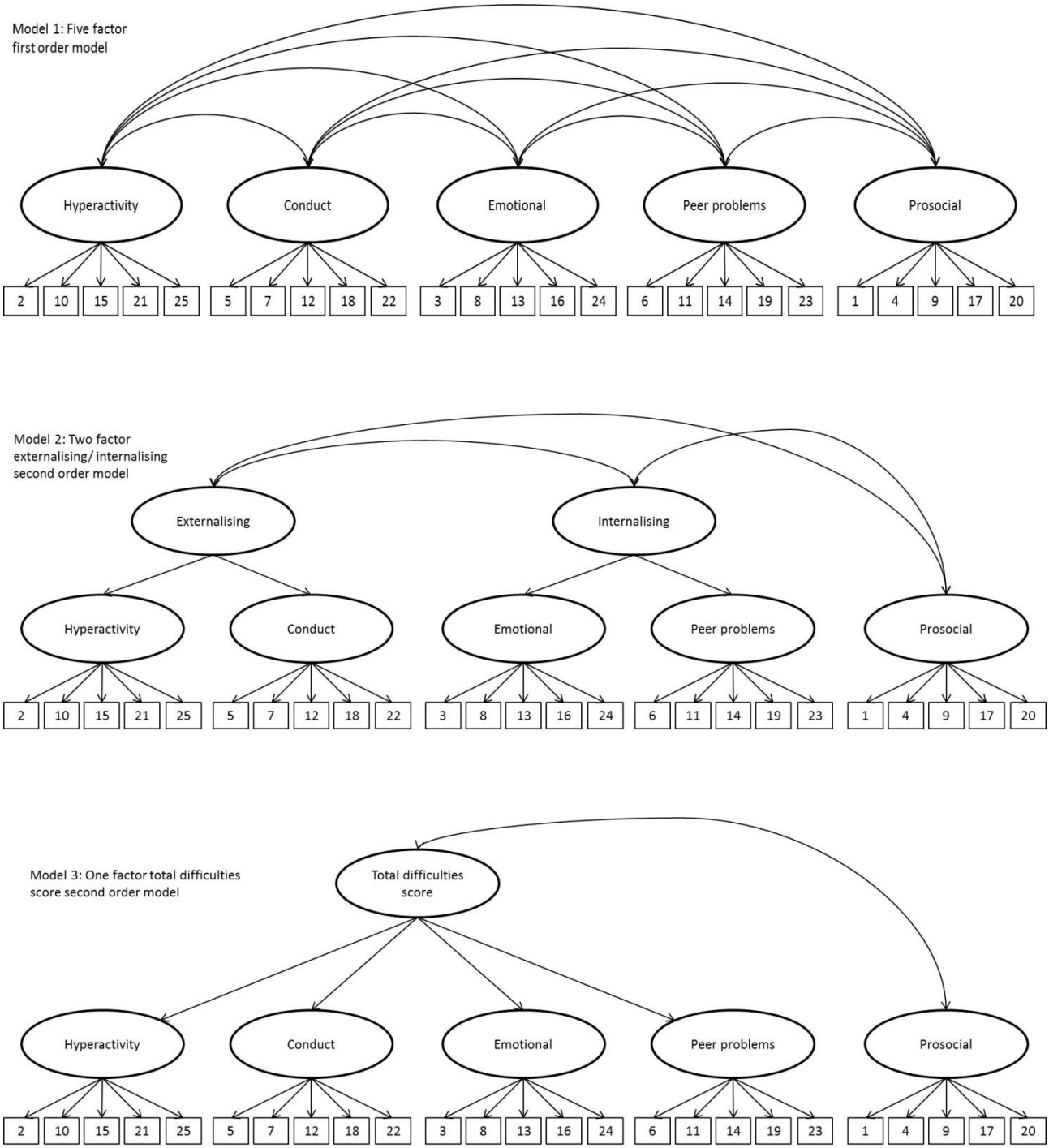


Figure 1: the three theoretical models tested in the CFA study (Paper 2)

Secondly, the large Chi Square model fit is considered a result of the misfit between the data and the models. Because the Chi Square is very sensitivity to the large samples sizes it was decided to report the RMSEA, CFI and TLI fit statistics as well. The RMSEA is an *"absolute fit index"* and estimates how much misspecification there is in the model per degree of freedom [80, 101]. The starting point is an assumption that there is misspecification in the model and it measures the size of this. It takes sample size and complexity into account, and because the size of at least some of

the included samples were very large much of the interpretations in the present study should rely on this statistic. Although the RMSEA model fits were acceptable to good for parents and good for teachers for the original, non-modified Model 1 and Model 2, the fits significantly improved with the modifications.

The CFI and TLI are both *"incremental fit indices"* and these rely on the proportionate improvement in fit of a hypothesised model compared to a more restricted, nested baseline model [80, 101]. The values of CFI and TLI for the non-modified Model 1 and Model 2 for parents were < 0.90 indicating poor fits and > 0.90 - < 0.95 for teachers indicating acceptable fits. For the modified Model 1 and Model 2 the fits for the parent samples ranged from < 0.90 to < 0.95 indicating poor to acceptable model fits, but for the teacher samples almost all were > 0.95 reflecting good model fits. The CFI and TLIs for Model 3 were generally poorer than for Model 1 and Model 2. On the basis of the model fits it was concluded that the SDQ overall works better for older children compared to younger ones, better for girls than for boys and better for teacher raters than for parent raters. Furthermore, because Model 1 (the five factor first order model) and Model 2 (attaching two second order *Internalizing/ Externalizing factors* to Model 1) fitted data equally well a future use of these two models is recommended. The adjusted Model 3's (including a *Total difficulties second order factor*) was non-identified and could not be computed. However, as this model was consistently found to have the poorest fits for all subsamples these issues were not investigated further and the use of this model cannot be recommended.

Reliability of the SDQ

Having confirmed the factor structure, the next step was to investigate the reliability of the SDQ. Reliability concerns the degree to which the scores are free from random measurement error, and estimates the proportion of total variance not due to random error [75]. One particular type of reliability is internal consistency that was measured by means of *Cronbach's Alpha* (Paper 1), *Average Variance Extracted (AVE)* (Paper 2) and *Composite Reliability (CR)* (Paper 2).

Cronbach's Alpha is the most widely used estimate of internal consistency and measures how closely related a set of items are as a group. The assumption behind *Cronbach's Alpha* is that the unique variance within items should be small compared to the covariance between scale items [75]. A high *Cronbach's Alpha* (most often defined as >0.70 or >0.80, however lower for research purposes) is most often seen as evidence that the items measure a latent construct. However, this should be interpreted with caution since the magnitude of *Cronbach's Alpha* is also positively related to the number of included items. Thus a higher Alpha will be obtained for a scale that comprises many items rather than few items [75].

Notwithstanding the fact that SDQ subscales only include five items, the coefficients of *Cronbach's Alpha* were generally considered high. Highest estimates were found for the *Hyperactivity scale* (*Cronbach's Alpha*: 0.73-0.86 for the eight subsamples) and for the 20 item *Total difficulties scale* (*Cronbach's Alpha's*: 0.75 – 0.88 for the eight subsamples). However, the lowest estimates were

observed for the *Conduct scale* (*Cronbach's Alpha's*: 0.44 – 0.73 for the eight subsamples), and estimates within this range are generally considered poor to, at best, acceptable. The estimates were generally found to be higher for boys than for girls and typically higher for teacher ratings compared to parent ratings for the individual subscales and *Total difficulties score*, but lower so for the *Impact scores*. These somewhat lower reliability estimates for the *Impact score* may broadly be a result of the fact that teacher *impact* estimates are calculated on the basis of only three items whereas parent estimates are based on five items.

CR is another measure that assesses the internal consistency of a scale. It is calculated on the basis of the standardised factor loadings for each item and the corresponding error terms, and resembles *Cronbach's Alpha* in many ways. A good *CR* should be > 0.70. In Paper 2 all *CR's* were found to be > 0.7 for all scales for all subsamples and thus considered good (Paper 2). Note however, that the lowest values of *CR* were found for younger children with parent raters and highest values were found for older children with teacher raters. No substantial differences were found between boys and girls.

The *AVE* is yet another measure of scale internal consistency and measures the amount of variance that is captured by the latent variable in relation to the amount of variance due to its measurement error. It is thus a measure of the error-free variance of a set of items [105]. If an item is overall poor for its scale it will result in a low *AVE*. The *AVEs* revealed that all factors worked well for older children rated by teachers and also that no items from the *Hyperactivity subscale* were problematic for any of the subsamples. Single items on the *Emotional, Conduct, Peer-problems* and *Prosocial scales*, however, did tend to create problems for these scales for younger children rated by parents, resulting in poor values of *AVE*. This is not surprising as 14 items and 16 items out of 25 explained < 0.50 % of the total variance for these samples for boys and girls, respectively.

Validity of the SDQ

Validity can be defined as the agreement between a test score or measure and the quality it is intended to measure [106]. Different types of validity are recognised, two of which are concerned with the measured construct, and these are known as *Discriminant* and *Convergent validities*. *Discriminant validity* concerns whether concepts or measurements that are supposed to be unrelated are, in fact, unrelated [107]. By contrast, *Convergent validity* is established if two measures (scales or items) of a constructs that theoretically should be related, are in fact related [106-108]. No single definitive test of *Convergent* and *Discriminant validities* exists.

Convergent validity can be established if correlations among variables believed to measure the same construct are at least moderate in magnitude. Correspondingly, if variables believed to measure different constructs show sufficiently low correlations, *Discriminant validity* has been established. The highest correlations in the present data were indeed observed among items within rather than between scales indicating good *Discriminant* and *Convergent item validity*.

Discriminant validity can also be established if the estimated correlations between the individual factors are not excessively high (> 0.85) [109]. Because this is considered important in order to identify where possible problems are hidden within the SDQ these data are presented in Appendix B. Problematically high correlations were observed between the *Internalising/ Peer-problems, Externalising/ Conduct, and Total difficulties/ Conduct scales* (> 0.85). This indicated that these pairs of scales shares too much common variance between them, and consequently show poor *Discriminant validity*. This indicated that the first order factors (i.e. *Conduct* and *Peer-problems*) explain too much of the variance of the second order factors (i.e. *Externalising* and *Internalising factors*). One way of overcoming this problem could be to test a more parsimonious model with three first order factors, i.e. *Internalising*, *Externalising* along with the *Prosocial scale*. Allowing 10 items to load on to each of the *Internalising* and *Externalising factors* did not result in better model fits indicating that such models does not work very well.

Discriminant validity can also be established if at least 50 % of the variance of every indicator can be explained by the model. Some problems of establishing this type of *Discriminant validity* were found for the *Emotional, Peer-problems* and *Conduct scales* as some items (particularly items 3, 5, 6, 11 and 22) showed particularly low loadings on their respective scales. The problems of these items is that they explain relatively little of the total variance. At least 50 % of the total variance of every indicator should be explained by the model. The value of R-square indicated how much of the variance is attributable to the test item itself with the remaining unexplained parts of the variance being attributable to other, residual factors. Values of R-square < 0.50 are considered critically low as more than 50 % of the variance is explained by other factors than the test item itself. The value of R-square for item 3 (*"Often complains of headaches, stomach-aches or sickness")* was lowest for all subsamples (Paper 2, table 3). For parent raters and younger children this test item was consistently and critically low, explaining < 0.20 of the total variance (e.g. 0.38^2 = 14.4 % of the total variance for young boys rated by their parents leaving 85.6 % unexplained). For older children and teacher raters the R-squares for items 3 and 11 showed that they were the only items explaining less than 50 % of the total variance (but with values only a little below 0.50). For younger children being rated by their parents as many as 16 of the 25 items explained < 0.50 of the total variance indicating severe problems in several test items for this age groups with parent raters. For older children and teacher raters the factor loadings were considered very good (Paper 2).

Danish norms and clinical cut-offs

The cut-off scores, means, SDs and frequency distributions are presented in Appendices C-H for the samples of 5-7- and 10-12-year-olds, separately [110]. The cut-offs are presented for the full sample and for boys and girls separately, whereas the means, SDs and frequency distributions only are presented for boys and girls separately. The cut-off scores are banded according to Goodman's recommendations. Thus approximately 80 % of the children and adolescence are clustered in the "*normal*" banding, with 10 % in the "*borderline*" banding and the remaining 10 % grouped in the

abnormal or "*clinical*" banding [111]. When the distribution of scores on the five sub-scales did not permit a precise cut-off at the 90th percentile the score *above* this percentiles was chosen. This was done in order to yield slightly lower percentages of scores in the "*clinical*" banding in order to limit the total number of false positives. This principle needed only to be applied for the individual subscales because of the limited number of discrete values (0-10) attainable on these scales. As anticipated on the basis of the presented mean scores, girls were generally rated as having fewer difficulties than boys, thus contributing to the broader range of scores for girls in the "*clinical"* banding. This difference was particularly noticeable on the *Hyperactivity scale* which also contributes to the differences in total difficulties score between boys and girls. Girls on the other hand, had a narrower band of scores in the prosocial *"normal"* banding indicating higher prosocial ratings among girls. Comparing teacher with parent ratings the differences in scores on the *Hyperactivity scale* were even more marked, indicating that teachers are more likely to rate boys and girls differently on this scale. The "*normal*" bandings for teacher ratings for 5-7 and 10-12-year-old boys on the *impact scores* should also be noted. These figures indicate that teacher ratings of boys are the most likely to report on an impact of the observed behaviours.

Because the attrition rates were found to differ between the samples (the DNBC, ABC and CCC2000 on the one hand and SFI on the other hand) it was decided to compare the mean scores of the DNBC and SFI 7-year samples by means of Cohen's D [112]. An effect size of Cohen's D = 0.2 to 0.3 can be considered a small effect, around 0.5 a medium effect and 0.8 to infinity, a large effect. The sizes of Cohen's D were all considered small as they were all found to be < 0.30, and the only effect sizes which was > 0.20 was those for the *Prosocial scale* (Appendix I). Interestingly, the highest mean scores were actually observed for the SFI sample (i.e. more *prosocial behaviour*), a finding that may be contrary to what would be expected.

DISCUSSION: Psychometric properties of the SDQ

Overall, the EFA study supported the five faceted factor structure of the SDQ (Paper 1) [64]. Furthermore, a three factor and a five factor model were found to have equally good fits in the CFA study (Paper 2) [63]. In Paper 1, the *Orthogonal rotation* of the *Structure Matrix* of the EFA replicated Goodman's five factor structure for parents and for teachers [98]. However, this is somewhat not surprising as Goodman made use of *Varimax rotation*, one particular type of *Orthogonal rotation* [75]. As mentioned, the correlation coefficients of the *Structure Matrix* do not assume that the underlying factors are related. However, from a psychological perspective this makes little sense. Within the field of child and adolescent psychiatry comorbidity is commonly observed in children with mental health problems [1-3]. Psychological factors are indeed related to one another and it therefore makes little sense to assume these to be independent of one another. For the SDQ this means that it is assumed that the *Hyperactivity* and *Conduct scales* are entirely independent constructs – something that most professionals would probably disagree with. The *Structure Matrix* in Paper 1 was therefore primarily reported in order for the results to

be comparable to the reported findings from the many EFA studies reporting the results from *Varimax rotation* only.

The *Pattern Matrix* on the other hand allows factors to be related or correlated with each other [75]. For example, the factors of *Hyperactivity* and *Conduct disorder*, were allowed to correlate. For parents, the *Pattern Matrix* revealed that some *Conduct items* actually did tend to load onto other scales, whereas *non-Conduct items* showed high loadings on the *Conduct scale*. Furthermore, some positive, "reversed" items did load highly onto the *prosocial factor*. For teachers, the picture was even more "mis-matched" and there was an even greater tendency for items to show high factor loadings on more than one scale. All five "reversed" items loaded highly onto the *Prosocial scale*, indicating that teachers are more prone to experience all positively worded items as one construct (i.e., the five reversed items as well as the five *Prosocial items*). For teachers, four of the five *Conduct items* showed high loadings on the *Hyperactivity scale*. By contrast, high loadings were revealed for two *Hyperactivity items* on the *Conduct scale*. Regarding the high loadings of the *Conduct items* on the other scales it seems that these items are as much part of a *Hyperactivity/ inattention construct* as part of a notion of *Conduct* for teachers. These findings indicate that teachers are more prone to view *Conduct* and *Hyperactivity* as one construct. If a teacher reports a child as exhibiting *Conduct symptoms*, he or she will also be very likely to rate the child as having symptoms of *Hyperactivity*. On the other hand, if a child is rated as exhibiting *Hyperactivity/inattention symptoms* by its teacher, there will be some tendency for the teacher to rate the child as having *Conduct problems* as well but not to as large an extent as the reversed.

It was concluded in Paper 2 that Model 1 and Model 2 showed equally acceptable to good overall model fits for all subsamples. This implies that both models work equally well and suggests that these two models can be applied equally successfully for clinical as well as research purposes. However, the limitations of the overall models fits need mentioning.

The fit indices indicate only how well the data on average or overall fit the theoretically-based model. This implies that some parts of the model may poorly fit the data even if the value of a particular index seems favourable [80, 101]. Also, because each of the model fits reflects only one particular aspect of the fit, a favourable value of one model fit does not alone indicate a good overall fit. Although the fits presented in Paper 2 for the modified model proved acceptable to good, some problems were observed for specific parts of the model. The relatively low values of the AVE on the *Emotional, Peer-problems, Conduct* and *Prosocial scales* for younger children rated by their parents indicate specific problems with certain parts of the model, namely with single items on these scales. This implies that clinically too much emphasis should not be put on any individual items, particularly not in the case of younger children. It highlights the importance of emphasising that the SDQ is a screening tool, that can and should only be used as such.

Furthermore, by taking a closer look at the covariances between the factors did reveal some severe problems with the *Discriminant validity* between *Internalising/ Peer-problems, Externalising/ Conduct,* and *Total difficulties/ Conduct scales*. That there might be a greater overlap between these sets of scales than what would be preferred was also hinted from the

Pattern Matrix in Paper 1. This certainly indicates that there is a great overlap and comorbidity in the behavioural manifestation of the mental disorder that these scales are supposed to measure. The EFAs revealed that the *Hyperactivity scale* and *Hyperactivity items* were the ones with the least overlap with other items and scales. This was similarly observed in the *AVE* reliability estimates. To estimate the *AVE* the values of the of R-square values of the individual items are used. If a single item cross-loads between more factors it will result in a low value of *AVE*. The highest estimates were observed for the *Hyperactivity scale*.

Finally, it was found that a model with two broader *Externalising* and *Internalising scales* worked well. This makes theoretical sense as the SDQ was originally developed from the Rutter questionnaire the purpose of which is exactly to tap into *Emotional* (*Internalising*) and *Behavioural* (*Externalising*) *problems* [68]. Goodman's original model with a *Total difficulties second order factor* showed the poorest fits and the standard errors and related estimates could not be computed for the modified model. There may be two plausible reasons for this. First, the model is too complex with too many parameters and this results in a non-identified model. However, since Model 2 is more complex than Model 3 this seems unlikely. Secondly, the model is non-convergent, indicating that there is something wrong with the model causing problems for the estimations of the best fits for the parameters. If this is the cause of the problems then it may in turn be due to the skewness of some of the items leaving too little information on some of the parameters. This might have been solved by removing the most skewed items. However, this did not result in better overall model fits. Alternatively, the answer categories could be dichotomised. In total, on this basis it is recommended that the *Total difficulties scale* should not be used. Instead, the use of the broader *Externalising/ Internalising scales* is recommended for research purposes.

It could be argued that there was no need for yet another study investigating the psychometric properties of the SDQ from an EFA perspective. This is in many respects a fair criticism. It has been proposed that EFA has three main uses [75]: 1. to understand the structure of a set of variables; 2. to construct a questionnaire that measures one or more underlying factors; and 3. to reduce a dataset to a more manageable size while retaining as much of the original information as possible. A fair claim would be that the purpose of the EFA study was none of these. The structure of the items was, at least to some degree, already known. The questionnaire was already constructed, and is indeed very short and not in need of further reduction, i.e. a *'short version'* of the SDQ is not required. Rather, it could be argued that the SDQ already forms a short version of the Rutter questionnaire. Despite this, one reason for doing the EFA study arose because of the very large size of the sample. No one has to date investigated the factor structure of the SDQ with such a large sample, which here has allowed for highly relevant gender, age and rater specific analyses without a noticeable lack of power. Thus, all analyses were carried out separately for boys and girls, separately for younger and older children and separately for parent and teacher raters. Gender specific analyses had not previously been carried out and although no differences were observed between boys and girls on the factor structure this is in itself a very important finding. Differences in mean scores were, for example, observed for boys and girls and although factor structure is an entirely different thing, it could very well have been that the factor structure would

work better for boys than for girls, or vice versa. However, this was not the case. Paper 1 also reported the findings from the *Structure Matrix* as well as from the *Pattern Matrix*. Both of these are in the literature generally not presented together, rather one or the other is reported. However, the reporting of both allowed for some very important comparisons with other studies. For example, the original article by Goodman [98] only included results from the *Structure Matrix* (using *Varimax rotation*). It may very well be that if Goodman had decided on an *Oblique rotation* instead and reported the results of the *Pattern Matrix* in his original work the SDQ would look different today – different scales or different questions might have been included.

The means and cut-off scores presented in Appendices D and G are in line with those reported for other Scandinavian studies and somewhat lower at least on the *Hyperactivity, Peer-problems* and *Total difficulties scales* than those found in other European and non-European studies [55, 61, 85, 99, 113-119]. Goodman [111] recommends that cut-off scores be adjusted according to age and gender, chosen according to the likely disorder rate in the sample being studied and according to the relative importance of false positives and false negatives. In a general population it seems more appropriate to include too few clinical cases rather than too many, i.e. using higher cut-off scores. It was for this reason decided to select appropriate "*clinical*" cut-offs, *above* rather than closest to, the 90th percentile. Another way to overcome the problem of including too many false positives could be to use 90/ 97.50 percentiles rather than the 80/ 90 percentiles as recommended by Goodman [74, 98]. This would results in even fewer false positives but probably also in more false negatives.

The proposed *Total difficulties cut-off scores* were found to be between 11 and 14 for parent ratings and between 12 and 18 for teacher ratings. These parent cut-offs are somewhat lower than the British recommendation of 17 [98], German of 16 [85] and Swedish of 14 [119] and indicate that children of all the included age groups are rated as exhibiting fewer emotional and behavioural problems compared to other samples. Different explanations for these differences can be given. Firstly, they may indicate that Danish parents and teachers rate children and adolescence more positively than do British parents and teachers. However, it is difficult to see why this should be the case. Secondly, it may be that the included samples are more selective and therefore less representative of the general population compared to the samples included in other studies. The data for the present study derived from some of the large-scale cohorts that to a degree are characterised by fewer mothers outside the workforce and with no further education beyond compulsory school, fewer single parents and fewer parents from the lowest income groups. This was particularly so for the large DNBC cohort and since data were included into the analyses in an unweighted manner this may have introduced a potential source of bias to the analyses [120]. Thirdly, it may reflect actual behavioural and emotional differences in the Scandinavian countries – countries that are characterised by better social security, low poverty, high living standards and less economic and social inequality. Meltzer et al. [121] have demonstrated that children with mental disorder are more likely to live in lower income households, with a single parent and in social sector housing. Denmark is characterised by a relatively homogenous population with a high level

of social security and these circumstances may very likely cause the higher cut-offs indicating fewer behavioural and emotional problems in the general Danish population.

STRENGTHS AND LIMITATIONS: Psychometric properties of the SDQ

The studies for Paper 1 and Paper 2 are based on data from four of the large-scale Danish cohorts. At least three of these (DNBC, ABC and CCC2000) are known not to be fully representative of the background population, and relatively low participation rates have been reported (< 50 % in the first waves) [91, 122]. The SFI sample on the other hand did have a much higher participation rate (≈ 90 % in the first wave) (Paper 1) [90]. To test whether the different attrition rates for the different cohorts had an effect on the observed mean SDQ scores between the samples the mean SDQ scores between the 7-year SFI and the 7-year DNBC samples were compared by means of Cohen's D [112]. An effect size of Cohen's D of 0.2 to 0.3 can be considered a small effect, around 0.5 a medium effect and 0.8 to infinity, a large effect. The sizes of Cohen's D were all considered small as they were all found to be > 0.30, and the only effect sizes that were > 0.20 was those for the *Prosocial scale* (Appendix I). Interestingly, the highest mean scores were actually observed for the SFI sample, a finding that may be contrary to what would be expected. Because of these low effect sizes it can be argued that the identified psychometric properties and norms and clinical cut-offs do resemble what would have been observed if the study had been more representative of the background population, and therefore is applicable to the general population.

The inclusion of a clinical, high risk sample could have been an advantageous in the present study. As one study looking at the factor structure of the SDQ from an EFA perspective concludes "*there are advantages to using the broader internalising and externalising SDQ subscales for analyses in low-risk samples, while retaining all five subscales when screening for disorders*" [82]. In Paper 2, on the other hand it was concluded that the two models showed equally good model fits in the non-clinical sample included in the present studies. It would be very relevant to investigate whether the findings from the study by Goodman or the findings from the Paper 2 could be replicated in a Danish high-risk sample. Such findings would be applicable in the many clinical studies using the SDQ.

The reported response frequencies (Appendix A) revealed skewed or very skewed distribution of scores on most items. Despite these obvious non-normal distribution of scores, means and SDs were presented in Paper 1. In fact, it would arguably have been more appropriately to report medians and inter-quartile ranges (i.e. the interval between the 25th and 75th fractiles) rather than means and SDs. This is recommended for future studies presenting the descriptive statistics of the SDQ.

FUTURE STUDIES: Psychometric properties of the SDQ

The *Pattern Matrix* of the EFA revealed some overlap between the *Conduct items* and the *non-Conduct scales* (particularly the *Peer-problem* and *Hyperactivity scales*), *non-Conduct items* (particular *Peer-problems* and *Hyperactivity items*) and the *Conduct scale*. A similar tendency was observed for the *Peer-problem scale* and *Peer-problem items*. Furthermore, *Discriminant validity* could not be fully established between the *Internalising/ Peer-problems, Externalising/ Conduct,* and *Total difficulties/ Conduct scales*. This is of great clinical importance. The reason that *Peer-problems* is at all considered part of an *Internalising construct* is that withdrawal from social life is indeed a very good predictor of future development of anxiety and depression. However, the great overlap with the externalising items and scales also indicate that *Peer-problem items* are very good indicators of the *Externalising constructs.* Future research should investigate these matters further. It is currently not known for example whether the SDQ can actually predict the development of other mental disorders by compiling the existing questions into new constellations. Currently, the SDQ covers *Hyperactivity, Conduct disorder, Emotional and Peer-problems*. However, other disorders like Obsessive compulsive disorder (OCD), Tourette syndrome, Autism spectrum disorder (ASD), and eating disorders, are not covered by the SDQ. It may be that the existing questions in new constellations actually do have the potential of predicting future diagnoses of some of these disorders. In other words, a possible (hidden) potential of the SDQ deserves to be unravelled. The aim of such study would be to investigate whether the current SDQ questions can predict future diagnoses of other prevalent childhood mental health disorders, not currently covered by the SDQ.

The predictive validity of the SDQ over time has to date not been investigated. This seems somewhat problematic as we do not know how well the SDQ actually predicts future diagnoses of, for instance, ADHD, Conduct disorder, anxiety and depression. It is recommended that future studies investigate the screening properties of the SDQ, i.e. how well the SDQ predicts future diagnoses of these child mental health disorders over time.

The Danish version of the self-rate SDQ has not been used as extensively in the large-scale cohorts as the parent and teacher versions. When such data have been collected it is recommended that psychometric properties are investigated and norms and clinical cut-offs are developed for the Danish self-rate version of the SDQ. This is particularly important considering the relatively large differences observed between the Danish parent and teacher cut-offs compared to the British developed cut-offs.

INTRODUCTION: prenatal exposure to alcohol and child behavioural and emotional development at age seven

Fetal Alcohol Syndrome (FAS): a historical perspective

Fetal Alcohol Syndrome (FAS) was coined by Jones and Smith in their now legendary article from 1973, *"Recognition of the fetal alcohol syndrome"* [123]. In this and two later articles they systematically delineate the association between maternal alcohol abuse in pregnancy and a specific pattern of growth retardation, dysmorphic facial features and CNS dysfunctions in the child [123-125]. These were the first scientific articles in English that had been published in many years on the deleterious effects of alcohol on the developing foetus [126-129].

However, the suspicion of alcohol as a culprit of dysfunctions and abnormalities related to human reproduction and child development was not new. In the first half of the 1700s, during the *"Gin epidemic"* in England several reports documented the adverse effects on the developing foetus of maternal drinking in pregnancy [126, 128-130]. In 1725 James Sedgewick, a London apothecary, noticed that there was a relation between the *"mothers ill-spent life during her pregnancy and consequences on infants"* [129]. A year later the College of Physicians petitioned the parliament to control the distilling trade and called gin *"a cause of weak, feasible distempered children … born weak and silly … shriveled and old, as though they had numbered many years"* [126]. Novelist and anti-gin campaigner Henry Fielding blamed gin-consumption for the *"increased crime and increased ill-health among children"* [126, 128-130]. However, because alcohol along with opium was the only anaesthetic available in the 18th century it could not be prohibited by the obstetricians [129].

The first epidemiological study of women consuming alcohol in pregnancy was carried out by William Sullivan, a deputy medical officer of the convict prison in Parkhurst, England [126, 128, 129]. In 1899 he followed 600 children born to 120 imprisoned, alcoholic women and 28 non-drinking relatives as controls. Among the alcoholic women he observed twice as many infant mortalities, he observed that 80 women had three or more infant deaths, that 55.8 % died at birth or before the age of two and that the children of the alcoholic women not were reproductive members of society later in life [126, 128, 129]. Around the same time, in 1905, in the USA, MacNicholl surveyed alcohol as a cause of mental retardation among school children in the city of New York. Among the 6624 children of drinking parents he found 53 % to be *"dullards"*, compared to 10 % *"dullards"* among the 13,523 children of abstainers [131]. It was also around this time several researchers started using animal models to demonstrate the deleterious effects of prenatal alcohol exposure (PAE) and showed that offspring of alcohol-exposed parents often had physical defects [128].

From the beginning of the 1900s to the 1960s, interest on the topic virtually disappeared. There was a general paradigm shift from a focus on prenatal and hereditary factors, to an approach emphasizing the importance of early childhood factors as important for child development [129]. In the early 1970s interest in the adverse effects of alcohol was renewed [123, 130, 132]. In particular, the introduction of the FAS term made the topic of prenatal alcohol exposure (PAE) and neurodevelopment find its way back onto the agenda. Whereas the early studies mostly described

single cases severely affected by FAS, researchers gradually began to investigate the effects of much lower doses of alcohol. As a consequence, the new "hot-topic" was to identify whether there exists a safe, lower level below which drinking is not associated with any harm to the developing foetus [44, 48, 133, 134].

The diagnostic criteria of Fetal Alcohol Syndrome (FAS) and Fetal Alcohol Spectrum Disorder (FASD)

Today, it is well agreed that exposure to large doses of alcohol act as a teratogen [128] and that it can have a wide range of deleterious effects on children's cognitive, behavioural and physical development [19, 124, 135, 136]. Fetal Alcohol Spectrum Disorders (FASD) is the umbrella term used to classify children exposed to alcohol prenatally from FAS in the one end of the spectrum to Alcohol Related Birth Defects (ARBD) in the other end. It has been claimed that mothers drinking during pregnancy may affect the neurodevelopment of around 1 % of all children [16].

The FAS-term has been used persistently since Jones and Smith's article from 1973 to describe those children most heavily exposed to alcohol and who exhibit the triad of symptoms described above [136, 137]. The FAS diagnosis is also the only expression of FASD that has garnered consensus among experts and is included in WHO's internationally used diagnostic manual ICD-10-CM (the "Clinical Modification"-version) under Q86.0 Fetal Alcohol Syndrome [138]. The American Psychiatric Association (APA) has not included a diagnosis within the FASD spectrum in the DSM-IV [139]. Instead, the Institute of Medicine's (IoM) "five diagnostic categories" and the University of Washington's "4-digit diagnostic code" have been widely applied since the mid-1990s. In 1996, the IoM recommended the use of five different diagnoses under the umbrella term fetal alcohol spectrum disorder (FASD) [136]. These include: 1. Fetal alcohol syndrome with confirmed maternal alcohol exposure; 2. Fetal alcohol syndrome without confirmed maternal alcohol exposure; 3. Partial FAS with confirmed maternal alcohol exposure; 4. Alcohol-related birth defects (ARBD); and 5. Alcohol-related neurodevelopmental disorder (ARND) [136]. The FASD 4-digit diagnostic code was developed by the Washington State FAS diagnostic and prevention network in 1997 and it is a simple, evidence-based method for diagnosing FASD on the basis of the following four features: growth deficiency, FAS facial features, CNS structural and functional abnormalities and prenatal alcohol exposure [140].

In the newly released DSM-V the term *"neurobehavioural disorder associated with prenatal alcohol exposure"* is included in the section *"conditions for further study"* [141]. This section contains conditions on which future research is encouraged. It is included to provide a common language for researchers and clinicians who are interested in studying the disorders but is *not* intended for clinical use. The proposed criteria includes: A. "more than minimal exposure to alcohol during gestation, including prior to pregnancy recognition. Confirmation of gestational exposure to alcohol may be obtained from maternal self-report of alcohol use in pregnancy, medical or other records, or clinical observations"; B. impaired neurocognitive functioning; C. impaired self-regulation; D. impairment in adaptive functioning; E. onset of the disorder (symptoms in criteria B,

C, and D) occurs in childhood; F. the disturbance causes clinically significant distress or impairment in social, academic, occupational, or other important areas of functioning; and G. the disorder is not better explained by the direct physiological effects associated with post-natal use of a substance (e.g. a medication, alcohol or drugs), a general medical condition (e.g. traumatic brain injury, delirium, dementia), another known teratogen (e.g. fetal hydantoin syndrome), a genetic condition (e.g. William syndrome, Down syndrome, Cornelia de Lange syndrome), or environmental neglect.

Prenatal Alcohol Exposure (PAE) and neurodevelopment: what is known and what remains uncertain? [16]

The absence of diagnostic agreement presented above probably reflects a general lack of consensus that is observed in the scientific literature investigating PAE and neurodevelopmental outcomes. Whereas few would disagree that prenatal exposure to larger average doses of alcohol causes irreversible brain damages [19, 137], including structural damages to the corpus callosum, cerebellum and hippocampal areas [142], an on-going debate prevails as to whether exposure to lower doses of alcohol is damaging for the developing foetus.

Since the early 1980s it has been hypothesised that the larger a *"dose"* a mother drinks in pregnancy the more CNS deficits in the child. In the high end of the spectrum it has consistently been found that prenatal exposure to large doses of alcohol is negatively associated with neurodevelopment in childhood, including problems with IQ [19, 143-146], executive functioning [144, 147, 148], motor development [19, 143, 146], learning and memory [19, 144, 149-151], speech and communicative skills [146] and behavioural outcomes [144, 146, 149, 152, 153]. The secondary disabilities comprise of difficulties understanding the consequences of their actions and learning from past mistakes [154], problems with adaptive functioning leading to difficulties with independent living and employment [146, 154], and increased rates of mental disorders [19, 146, 155-158].

Studies investigating PAE to low-moderate average doses of alcohol and neurodevelopmental outcomes in childhood has on the other hand been far less convincing and no dose-response associations have been established [16, 34, 146, 159, 160]. Some studies have found negative associations in childhood with externalising and aggressive behaviour [22], mental health [17, 21, 24], IQ [18, 23], hyperactivity [19, 20], impulsivity [19], attention [151, 161], learning difficulties [19, 20], memory [19], coordination [19], executive functioning[19] and social abilities [19]. A recent study has reported that such negative behavioural effects may even persist into adulthood [162]. Other studies have reported no such associations in childhood with mental health [134], IQ [163], academic achievement [134], hyperactivity/ inattention [29], language delay [28], attention [44] or executive functions [43]. Even other studies have reported on a J-alcohol shape, indicating that exposure to a little alcohol apparently act as a protective factor for the developing foetus. One such study found that the worst mental health and cognitive outcomes at age three were apparent in offspring of abstainers and heavy drinkers [30]. The same sample was followed at age five and favourable outcomes for boys exposed to light drinking in pregnancy were reported [134]. Another study showed that light

and moderate drinking in the first three months of pregnancy was positively associated with mental health scores at age 14 [164].

One thorough systematic review concluded that there is no convincing evidence that PAE to low-moderate average doses of alcohol is negatively associated with neurobehavioural outcomes. However, the authors did acknowledge that many of the studies had methodological weaknesses, mostly concerning confounding factors [159]. Many of the reported studies did not control for confounders at all, others controlled insufficiently or inappropriately. They also concluded that the J-alcohol shape reported by many of the studies, probably reflected a *"healthy drinker effect"* in which women with a poor obstetric history were more likely to abstain from drinking [159]. Similarly, another review found no consistent evidence that PAE to low-moderate amounts of alcohol was negatively associated with neurodevelopment [34].

The literature presented above distinguishes between exposure to different average *"doses"* of alcohol (i.e. low, moderate or high doses of alcohol). However, the most recent literature further recognised the importance of considering the *"pattern"* of the exposure, i.e. the quantity consumed on a typical occasion [165]. Most studies today therefore distinguish between exposure to average (lower) doses of alcohol and binge drinking [28, 33, 34, 43, 44, 159, 166-168]. Binge drinking is in the literature most often defined as an intake of a minimum of five alcohol containing units on a single occasion [33, 34]. The rationale for this distinction is that exposure to binge drinking is more devastating for the developing CNS because it is the peak blood alcohol concentration (BAC) that determines the level of the damage [33, 34]. In other words, binge drinking causes greater harm than exposure to a comparable amount spread over several days, weeks or months [33, 169].

Compared to the literature investigating exposure to lower average doses of alcohol, there seem to be somewhat more evidence for a devastating effect of being exposed to binge drinking. One review concluded [35] that children exposed to binge drinking consistently showed poorer neurodevelopmental outcomes in childhood [23, 34, 36-40]. A similar conclusion was drawn in a recent meta-analytic study [168]. Studies have generally reported negative associations with disinhibited behaviour [37], mental health problems [41, 170], IQ scores [36, 171], delinquent behaviour [36], academic achievement [171], antisocial behaviour [20, 41], learning problems [20, 23], classroom behaviour [171], attention [171], and behavioural problems [20, 41, 172]. However, the literature is not entirely conclusive and other studies have reported no associations with IQ [38, 41], attention [44] or executive functions [43].

Apart from *"dose"* and *"pattern"*, a third factor, *"timing"* seem to be of particular importance if one is to understand the effects of PAE on neurodevelopmental outcomes. Most studies have focused upon alcohol exposure during early pregnancy only [21, 41, 43, 44, 48, 134, 164] despite the fact that there seem to be two critical periods in human when the brain is especially vulnerable to insult [142]. The first period occur during the first trimester, from gestational weeks 12-20, and is characterised by a rapid rate of nerve-cell proliferation. The second period occur during the third trimester and does not end until age 18 month of the child and is characterised by a brain growth spurt. Unfortunately, very few human studies have attempted to investigate possible effects of

timing on such outcomes. One study that did so did find that exposure to moderate-high levels of alcohol in the third trimester or binge drinking in the second or third trimester was negatively associated with language delay at age two [28]. Another study found that binge drinking in early pregnancy was the best predictor of behavioural problems and performance in school-aged children [171]. A very recent study found that binge drinking in the first four weeks after conception had a very strong and predictive effect on SDQ scores at age five [172]. Yet another study concluded that binge drinking in the second or third trimester was associated with mental health problems in the children [41], whereas mid-pregnancy PAE was found to be significantly related to poorer habituation and increased low arousal in new-born infants [39].

As should be apparent from the literature presented above, the findings are generally very inconsistent and inconclusive. The disparity is greater in the part of the literature investigating exposure to lower doses of alcohol than to binge drinking, and there is no consistency in the *"timing"* literature as to whether exposure in the first, second or third trimester is more negatively associated with neurodevelopmental outcomes in childhood. This inconclusiveness may in part be due to a large number of methodological limitations. First, the effects of alcohol seem to be at least *"dose"*, *"pattern"* and *"timing"* dependent [16, 34], and the scientific literature has by and large failed to focus on more than one, or at best two, of these three factors. This may be at least part of the reason that the literature for example has not been able to establish a dose-response relationship [21, 30, 43, 44, 48, 134, 159, 163-165, 173]. Secondly, it has been suggested that the lack of consistency is due to measurement error, bias and confounding in epidemiological studies [174]. Thirdly, there is no standardisation on a quantitative definition of what is meant by a low, moderate and high level of alcohol between studies [34, 175]. This may also confuse the understanding of the literature. One comprehensive review concluded that *"… we may have reached the limits of what we can determine from the standard case control and cohort designs"* [34] (p. 21). I will therefore now turn to animal studies.

Biochemistry and animal models

Because human observational studies are prone to methodological limitations including measurement error, bias and confounding, animal studies are needed to strengthening the case that PAE act as a teratogen on the developing foetus [128]. If indications from the human studies can be replicated in animal studies it supports the idea that alcohol does indeed act as a human teratogen. The advantages of animal studies are that they can experimentally be designed to disentangle the effects of *"dose", "pattern"* and *"timing"*, and they are to some extent free from issues concerning measurement error, bias and confounding.

Experimental animal studies have indeed demonstrated that the neuroteratogenic effects depends on *"dose"*, *"pattern"* and *"timing"* [176, 177]. Regarding *"dose",* animal studies have concluded that exposure to low-moderate doses of alcohol is negatively associated with the development of the forebrain in rats [178], and attention and neuromotor functioning in rhesus monkeys [179]. Dose-response associations, that in the human has virtually been absent, have been

demonstrated in numerous animal studies investigating spontaneous alternation, learning and conditioned taste aversion learning tasks [16, 34, 136, 180].

Regarding *"patterns"*, animal studies have consistently shown negative effects on offspring exposed to binge-like patterns on discrimination tasks and association learning (as measured in mazes), motor skills (as measured by poor reflexes, coordination and balances), hyperactivity/ inattention (as measured by preservation on reversal tasks and heightened exploratory behaviour), and social ability (as measured by play tasks, mating abilities and corporation) [16, 34, 181].

The *"Timing"* aspect has in the animal literature received more attention compared to the human literature [16, 34]. What corresponds to the first and third trimesters in humans appears to be particularly sensitive periods for inducing CNS abnormalities [146, 182]. PAE in the first trimester has in rhesus monkeys been found to significantly decrease scores on infant neurobehavioural tests, whereas mid- to late gestation exposure resulted in reduced motor maturity [179]. Third trimester human brain development is the period of greatest brain growths and is equivalent to the early post-natal periods in rats. Numerous rat studies have reported that heavy alcohol exposure in this period is associated with reduced brain weight and volume, particularly in the forebrain, brainstem, cerebellum and corpus callosum [178, 183, 184]. A rhesus monkey study reported early gestation exposure as being negatively associated with infant neurobehavioural test scores whereas mid- to late exposure resulted in reduced motor maturity [179]. A study investigating binge-like drinking during the third trimester in vervet monkeys found significantly fewer neurons in the frontal cortex of the exposed offspring. The authors concluded that the entire dorsolateral prefrontal cortex was affected and that this provides an anatomical basis for the behavioural and cognitive deficits observed in children exposed to alcohol prenatally [185].

The advantages of studying FASD by means of animal models is first and foremost that many studies could for obvious ethical reasons not be carried out in humans, and secondly that confounding factors can be controlled for in a more rigorous manner [176]. Thirdly, less measurement error and bias is introduced into the statistical analyses. Although animal studies have contributed greatly to the understanding of the effects of PAE on neurodevelopmental outcomes they do have their limitations. The complex pre- and postnatal environments in humans are poorly approximated in animal models. Further, the complexity of human social behaviour in interaction with others, and the higher cognitive functioning in humans is very poorly resembled in animals. On the biological side, the rapid third trimester brain development in humans does not take place until the early post-partum period in rodents. These factors may all limit the applicability of the results from the animal studies [176]. In order to understand the importance of the pre- and post-natal environments, early childhood environment and complexity of human social behaviour on CNS development, I will now return to human research and introduce a theoretical model. The model can be used to understand the complexity of the early child environment and to understand the factors that are known to influences the development of cognitive and mental health development in children.

Understanding the associations between prenatal exposure to alcohol and child development: a theoretical framework

It was demonstrated above that exposure to at least substantial amounts of alcohol most likely is negatively associated with child neurodevelopment. Less conclusive was the investigations of exposure to lower doses of alcohol. Although exposure to alcohol in the intrauterine environment may at a first glance seem like a fairly "easy-to-understand" biological process the subject may actually be rather more complex.

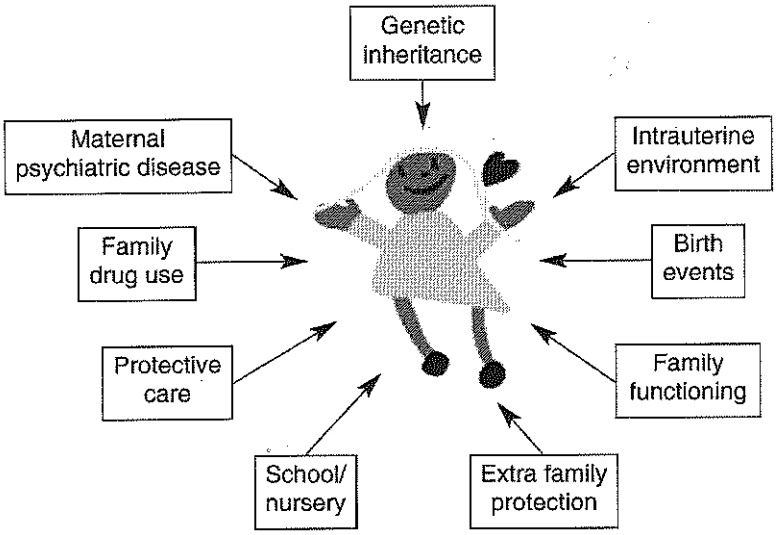


Figure 2: Cumulative effect on infant development (adapted with permission from Dr. Philip M. Preece, personal communication 9th July 2013)

Child behavioural and emotional development are intimately related to intrauterine circumstances as well as factors at birth and in childhood. Figure 2 and Figure 3 illustrate the complexity of the subject by showing the multitude of factors that are known to influence infant and child development [186]. Whereas Figure 2 demonstrates the factors influencing infant and child development, Figure 3 more specifically illustrates that factors from conception, in utero, at birth, infant and childhood all play part in the development of behaviours and emotions. The models have specifically been developed to understand children with alcohol and drug *abusing* parents, which is best illustrated by the inclusion of "foster care" in Figure 3. Despite this, the models are still applied as they are believed to be useful tools to understand the finding from the studies investigating PAE to lower doses of alcohol and behavioural and emotional development in a non-abusive, stable home environment.

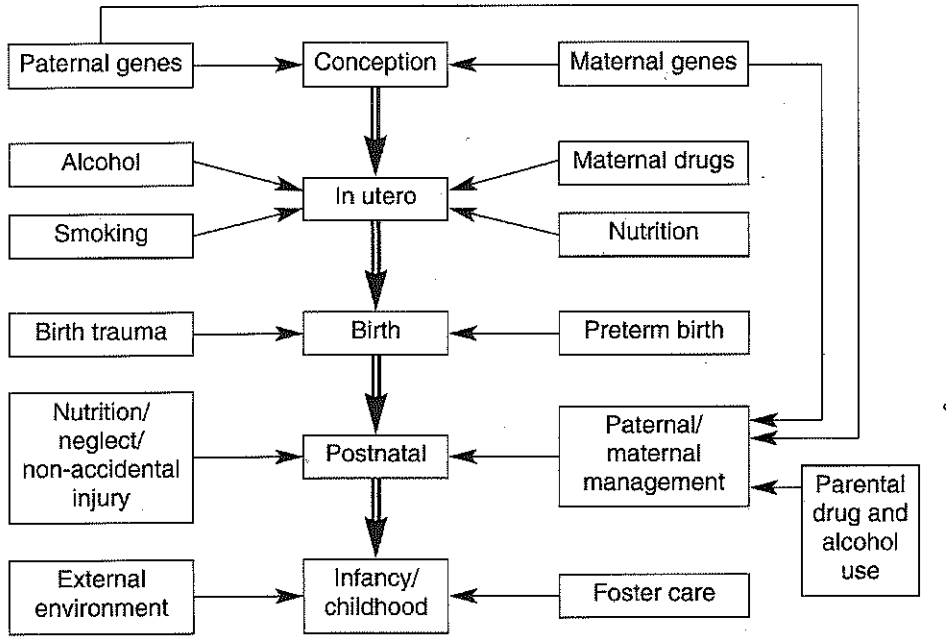


Figure 3: Factors influencing infant development (adapted with permission from Dr. Philip M. Preece, personal communication 9th July 2013)

What is important to understand is that most (if not all) of the factors in the model act through a number of different pathways and on a number of different levels. The effect of PAE is modified and influenced by a wide range of pre- and post-natal factors [186]. For example, prenatal exposure to alcohol was in the scientific literature above presented as being associated with behavioural development in childhood. However, apart from exposure to alcohol in the "intrauterine environment", a women drinking alcohol in pregnancy may also transfer a "genetic inheritance", or disposition for behavioural problems to her child, the mothers may be more likely to suffer from a "psychiatric disease", the "family functioning" may be poorer than average, there may be less "extra family protection" than in an average family, and the child may be less likely to attend a resource-full "school/ nursery". The same may be evident on the paternal side.

For example, a parent with behavioural problems may, apart from being more likely to drink alcohol, also be more likely to expose the foetus to psychotropic medication, cigarette smoking, illicit drugs and poorer nutrition in pregnancy. These factors all influence the intrauterine growth environment and hence the child – factors that may all make the child more prone to exhibiting behavioural problems in childhood [186]. Figure 2 and Figure 3 illustrate how difficult, if not impossible, it is to disentangle the effects of alcohol from other intrauterine exposures, genetic predisposition and subsequent lifestyle, family and socio-demographic factors [186].

Now that the complexity of the problem has been illustrated and a theoretical model has been introduced I will move on to presenting the aims, methodologies and results of Paper 3, Paper 4 and Paper 5, and Appendices J, K and L.

AIMS: Prenatal alcohol exposure and child development

The primary aim of the studies was to contribute to the observational literature investigating potential associations between PAE on the one hand and neurodevelopment on the other. The secondary aim was to thoroughly describe women with different habits of alcohol intake in pregnancy on a large number of confounding factors, and to discuss the influence of these on the results from the observational literature. Specifically, the aims were to:

1. Investigate the association between prenatal exposure to low-moderate doses of alcohol and binge drinking on the one hand and behavioural and emotional development as measured by parent-rated SDQ scores at age seven on the other hand (Paper 3).

2. Investigate the association between exposure to binge drinking in early and late pregnancy on the one hand and behavioural and emotional development as measured by parent-rated SDQ scores at age seven on the other hand (Paper 4).

3. Describe the characteristics of women who drink, and who do not drink, alcohol in pregnancy, as measured by a full pregnancy cumulated alcohol measure, and discuss the methodological implications of these findings (Paper 5).

4. Describe the characteristics of women who drink, and who do not drink, alcohol in pregnancy, as measured by the reported average intake in the first, second or third part of pregnancy, respectively, and discuss the methodological implications of these findings (Appendices J, K and L).

METHODS: Prenatal alcohol exposure and child development

Sample

The data for Paper 3, Paper 4, Paper 5 and Appendices J, K and L were derived from the DNBC that includes information on 101.042 pregnancies [187]. Between 1996 and 2002 pregnant women were nationwide invited to participate in the study. The aim of the DNBC was to provide information about the period from conception to early childhood and to permit for studies investigating how this period influences health conditions and development later in life. In particular, the aim was to study side effects of medications and infections. The participating women were interviewed twice prenatally, in approximately weeks 15 and 30, regarding their lifestyle during the first and second part of their pregnancy, and again at age six months of the child about their lifestyle in the final part of their pregnancy. All interviews include questions on maternal health and illness, lifestyle and socio-demographic characteristics. At age seven a questionnaire concerning the child's health and illness, behaviour, lifestyle and socioeconomic characteristics was sent to the child's mother.

Restriction of the sample

For Paper 3, Paper 4, Paper 5 and Appendices J, K and L complete case analyses were decided on. Specifically, for Paper 3, Appendices J, K and L the sample was restricted to women with full information on key alcohol variables (average alcohol intake and binge drinking) and full information on the four SDQ difficulty scales (*Hyperactivity*, *Conduct*, *Emotional* and *Peer-problems*). The sample was further restricted to live-born, term singletons, i.e. to singletons with a gestational age of ≥ 37 completed weeks. This left a total of 37,152 mother-child dyads in these studies. The sample restrictions for Paper 4 were virtually identical to those of Paper 3, however only restricting women with missing data on the binge drinking variable, not on the average alcohol variables. This left at total of 37,315 mother-child dyads in Paper 4. Inclusion criteria for Paper 5 was full information on average alcohol intake variables from the first three interviews leaving a total of 63,464 women in the study.

Measures of average alcohol intake and binge drinking

A standard drink in Denmark is defined by the National Board of Health as 12 grams of absolute alcohol [188]. In all three questionnaires the interviewees were asked separate questions concerning average alcohol intake and binge drinking (the interviews are available in English at www.dnbc.dk → data available). This *"pattern"* dependent subdivision of the exposure categories (i.e. low-moderate doses and binge drinking) was applied in Paper 3, Paper 4, Paper 5 and Appendices J, K and L.

Regarding average alcohol intake in the first and second questionnaires the pregnant women were asked about their intake of *"different kinds of beverages"*. After a few questions concerning their intake of tea and coffee, they were asked separate questions regarding their intake of beer, wine and spirits: *"how many normal beers/ glasses of wine/ glasses of spirits do you drink per week"* (Table 1). In the first interview the questions were followed by similar questions regarding their pre-pregnancy weekly average alcohol intake.

How many glasses of beer/ wine/ distilled alcohol do you drink per week now?
If you think about the entire period of pregnancy – also the very beginning – how many times did you then have 5 drinks or more in one night/ event?
- **What week(s) of gestation were you in the 1st, 2nd, 3rd etc. time?**

Table 1: Questions regarding average alcohol intake and binge drinking in the DNBC

In the third interview taking place six month post-partum the women were told: *"now follows a few questions about diet and different life style habits. We are still talking about the part of the pregnancy from the last interview until birth".* This section started with a number of questions regarding intake of vitamins and other supplements, food, smoking, coffee and tea followed by the questions on alcohol intake: *"how many normal beers/ glasses of wine/ glasses of spirits did you drink per week?"*.

In each of the three interviews the questions on average doses of alcohol were immediate followed by questions enquiring about episodes of binge drinking. In the first interview the women were asked: *"If you think about the entire period of pregnancy, also the very beginning, how many times did you then have 5 drinks or more in one night/event"?* If a woman answered yes to this question she was further asked about *"the number times"* and at *"what week(s) of gestation she was in the 1st, 2nd, 3rd etc. time"?* The phrasing of the binge drinking questions was similar in the second interviews, i.e. enquiring about the number of episodes from the beginning of pregnancy. In the third interview the women were asked about binge episodes from week 30 and until birth. The procedures applied in the DNBC to gather information about alcohol consumption have been shown to yield reliable information among pregnant Danish women with a low to moderate alcohol intake [188, 189].

Outcome measure: parent-rated SDQ scores at age seven

The outcome measure consisted of parent-rated SDQ scores at age seven (please see the SDQ section of this thesis for a general introduction to the SDQ). The four difficulty scales were used as outcome measure. Because it was found in Paper 2 that Model 1 and Model 2 had equally good model fits, the difficulty scales were used both as four separate scale models (i.e. *Hyperactivity*, *Conduct*, *Emotional* and *Peer-problems*) and as a broader model including the *Externalising* and *Internalising scales* [63]. For Paper 3 the four scale model used the 10 % *"clinical"* bandings available at http://www.sdqinfo.org/DanishNorms/DanishNorms.html [190] and presented in Appendix. Because of the small size of the late exposure group (N = 94) in Paper 4 it was decided to use the 20 % *"borderline"* bandings in order to include a higher number of potential cases.

Confounding factors

Confounding is defined as the mixing together of the effect of an exposure with a factor that is statistically associated with the exposure and causally associated with the outcome (Figure 4) [191]. For example, PAE is hypothesized to be causally associated with neurodevelopment in the foetus. Smoking on the other hand is also hypothesized to be causally associated with the outcome (foetal neurodevelopment) and also to be statistically associated with alcohol intake. Therefore, smoking is considered a confounder that should be controlled for in the statistical analyses. As a result, controlling for smoking will remove some of the apparent association between alcohol and neurodevelopment.

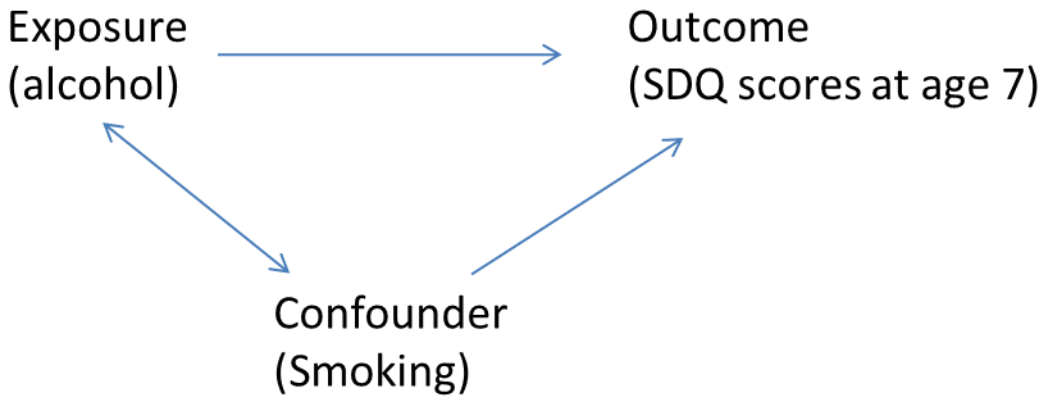


Figure 4: Model showing the association between the exposure (alcohol) and outcome (SDQ) and the influence of a confounder (smoking)

Residual confounding refers to the confounding that remains after attempting to adjust completely for confounding [191, 192]. Residual confounding remains if important confounders have not been controlled for or if a confounder has not been classified correctly and hence leading to misclassification. The effects of residual confounding can either mask a true association or create a spurious association, i.e. a false or non-causal association owing to chance, bias or confounding [193].

A wide range of potential confounders could be controlled for including prenatal exposure to smoking, caffeine, marijuana and other illicit drugs, psychotropic medication, vitamin, fish and fish oil, maternal physical activity in pregnancy, parental education, income and other measures of socioeconomic status (SES), having a partner in pregnancy, social support in pregnancy, parental psychiatric problems, parental criminal behaviour, parental personalities, IQ and age. These factors can all be considered potential confounders as they have all been found to be causally associated with neurodevelopment in the child and statistically associated with maternal alcohol intake in pregnancy [4-6, 8-12, 29].

In Paper 3 it was decided to control for parental smoking, education, past histories of psychiatric diagnosis, and maternal well-being in pregnancy. Information on past psychiatric history came from the Danish Psychiatric Central Register [194], information on educational attainments from the Integrated Database for Longitudinal Labour Market Research (IDA) [195], and smoking and maternal well-being in pregnancy from the structured DNBC interviews. In Paper 4 it was decided only to control for confounders on the maternal side including education, psychiatric diagnosis up to the age of seven of the child, age and smoking in pregnancy. This was decided on because of lack of power due to the small sample size of the late exposure group (N = 94). Information on past psychiatric history, education and age came from registries, the smoking variable from self-reports.

Mediating factors

A mediator or intermediate factor can be defined as *"a variable that occurs in a causal pathway from a causal (independent) variable to an outcome (dependent) variable. It causes variation in the outcome variable and itself is caused to vary by the original causal variable. Such a variable will be associated with both the causal and the outcome variable"* [193] (p. 131) (See Figure 5). In other words, mediators explain how external physical events take on internal psychological significance [15] (p. 1176). Mediation is said to have occurred if the strength of the relationship between the exposure variables and the outcome is reduced by including a mediator. Perfect mediation occurs when the relationship between the exposure variable and the outcome is completely wiped out by including a mediator in the model [196]. For example, attachment in childhood causes variation in SDQ scores at age seven, but is itself also caused to vary with the level of alcohol intake in pregnancy.

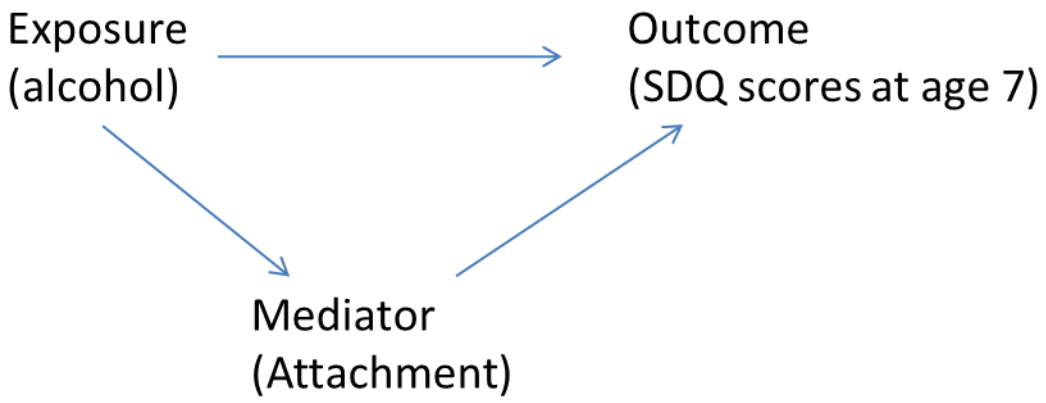


Figure 5: Model showing the association between the exposure (alcohol) and outcome (SDQ) and the influence of a mediating factor (attachment in childhood)

A wide range of mediating factors could be relevant to control for including attachment, parental and family functioning, parent-infant interaction, maternal post-natal stress and depression, colic, child motor and cognitive development, SES, and social support. These factors can all be considered potential mediating factors as they causes variability in the outcome measure (SDQ scores at age seven), but are themselves caused to vary by the exposure variable (alcohol).

Because the particular aim of the DNBC was to study medications and infections, no or only poor information was available on the potential psychologically-oriented mediating factors. For example, it is well known that home environment and attachment style affect the developing brain [13, 197, 198]. Specifically, attachment to caregiver and home environment during the first two years of life, when the brain is particularly plastic, is known to significantly influence child's development. The questions on "attachment" and "home environment" included in the DNBC were not derived from standardised questionnaires, but constituted a number of non-validated,

non-standardised questions. Because such variables would very likely introduce misclassification and measurement error into the statistical analyses they were not included in any of the analyses.

Paper 3: rationale behind the definitions of low-moderate doses of alcohol

As mentioned in the introductory section above there is no uniformly accepted definition in the literature as to what is meant by a low, moderate and high average *"dose"* of alcohol [34]. The subdivision of the alcohol categories for Paper 3 was thus somewhat arbitrary. However, the categories were decided on, bearing the aim of the study in mind, i.e. to investigate the importance of *"dose"* and *"pattern"*.

1. Alcohol exposure from full pregnancy vs. early pregnancy:

It was concluded in the introductory section above that there is no evidence that exposure to alcohol in the early part of pregnancy is more negatively associated with neurodevelopmental outcomes that exposure later in pregnancy. Despite this, there is a general tendency in the literature to investigate exposure to alcohol in the early part of pregnancy. Because it was hypothesised that "total exposure" would be a better predictor of CNS damage in the child than "early exposure" it was decided to use information on alcohol intake from full pregnancy rather than from early pregnancy for Paper 3.

2. *Exposure groups in the very low end of the spectrum:*

Out of the over 100,000 pregnant women in the DNBC, very few reported on what can be considered a high or even a moderate intake of alcohol. Therefore, the strength of the DNBC is the extensive information on women with a low intake of alcohol. It was therefore decided to thoroughly investigate exposure down to very low doses of alcohol, namely, to as little as 0, >0-5, >5-15, >15-45, >45-90 and >90 units of alcohol in full pregnancy. In other words, the "high" exposure (>90 units) group consisted of women with an average intake of a little more than two units/ week. This classification would permit for analyses of a dose-response association in the very low end of the spectrum.

3. *Cumulated alcohol exposure vs. exposure to average doses of alcohol:*

In the literature most researchers investigate exposure to average doses of alcohol, i.e. the intake is divided with number of weeks. However, the use of an average measure impedes for the inclusion of exposure categories in the very low end of the spectrum. Because the aim was to thoroughly investigate exposure to very low doses of alcohol it was decided to define the exposure categories on the basis of a cumulated alcohol exposure measure across pregnancy instead of an

average measure. This permitted for exposure categories in the very low end of the spectrum, i.e. exposure to as little as > 0-5 units of alcohol in full pregnancy.

4. Exposure from the early unrecognised part of pregnancy:

It is well recognised that most women cease drinking or minimise their intake of alcohol once they recognise their pregnancy. The percentage of women ceasing to drink in pregnancy differs between socio-economic groups [199]. Although very early exposure seem to be a critical period and most studies investigate effects of early pregnancy exposure most studies do not include pre-pregnancy intake as part of the exposure measure. However, because of the discrepancy between alcohol intake in-pregnancy and prior-to-pregnancy it was decided to include information on average alcohol intake as a measure of exposure in the very early, unrecognised part of pregnancy.

Paper 3 and Paper 4: rationale behind the definitions of binge drinking categories

Apart from exposure to lower doses of alcohol the aim of Paper 3 was to investigate the association between binge drinking and parent-rated SDQ scores at age seven. The exposure categories were decided on, bearing the importance of *"dose"* and *"timing"* in mind. The categories were defined on the basis of total number of reported episodes in full pregnancy (including episodes in the early, unrecognised part of pregnancy) rather than early pregnancy only, as otherwise seems to be the methodology applied by most researchers [43, 44, 167]. Again, the rationale for doing so was that it was hypothesised that the "total exposure" would be a better predictor of the CNS damage on part of the child than "early exposure".

In Paper 3, small associations were observed between prenatal exposure to binge drinking and parent-rated *Internalising*, *Externalising* and *Conduct* scores in boys at age seven. Because Paper 3 found such small associations, but only investigated the *"dose"* and *"pattern"* parameters, it was decided to specifically investigate the *"timing"* parameter in Paper 4. The reasons for specifically investigating early and late pregnancy were twofold: 1. It has been suggested in the literature that exposure to binge drinking in the first and third trimester is most devastating to CNS development [34, 146]; 2. The women in the DNBC were in the second interview asked about the number of binge episodes in the entire pregnancy (rather than, for example, in the period from the first interview until the second interview). The inclusion of a "second part of pregnancy"-exposure group would therefore be subjected to some uncertainty and misclassification and it was decided to include information from first and third part of pregnancy only.

Paper 5: rationale behind the included outcome measures

In Paper 3, no statistically significant associations were observed between exposure to low-moderate doses of alcohol and any of the SDQ outcomes at age seven, despite the fact that (what was at the time considered) the most important confounders were controlled for. However, because large differences were observed (in Table 1 in Paper 3) between exposure groups on the reported background characteristics, it was decided to investigate these matters further. The rationale for doing so was that if large inter-group differences were observed on other potentially important confounders, and these were not controlled for, residual confounding could in turn mask true associations or create spurious associations between exposure to low doses of alcohol and SDQ scores at age seven. The aim of Paper 5 was therefore to thoroughly describe women who drink and who do not drink alcohol in pregnancy on a very large number of background characteristics, i.e. potential confounding factors. The list of included variables should not be considered exhaustive. They were included because data was available on these variables. Other confounding variables including IQ, parental personality and criminal behaviour would have been at least as relevant to control for. However, no information was available on these variables.

Appendices J, K and L: rationale for including distributions of background characteristics for women subdivided on the basis of average alcohol intake in the first, second and third part of pregnancy, respectively

The Tables presented in Appendices J, K and L contain information on sampling characteristics including gestational age in the first and the second interview, child age the time of the third interview, unplanned pregnancies, time to pregnancy, and fertility treatment. Data on family characteristics includes maternal and paternal age, marriage, maternal and paternal education, a number of variables concerned with maternal and paternal psychiatric diagnoses/ psychological problems, a number of variables on maternal pre- and in-pregnancy alcohol intake, maternal and paternal smoking , use of nicotine substitutes, intake of vitamins, iron supplement and fish oil, habits of fish eating, pre-pregnancy BMI outside the normal range, intake of cola, coffee and painkillers, diabetes, asthma and anaemia in pregnancy, television watching, being a tenant, exercise in pregnancy. Finally, information on child characteristics includes sex of the child, child not growing in the last part of pregnancy, birth weight, small for gestational age (SGA), gestational age (GA), APGAR scores and child psychiatric diagnosis in the first seven years of life.

The appendices present the means and percentages for all of the background characteristics between exposure groups defined on the basis of information on average alcohol exposure and binge drinking from the first, second and third interviews, respectively. The average alcohol exposure groups were defined as follows: 0, >0-2, >2-4 and >4 units of alcohol/ week, whereas the binge drinking categories were dichotomised (yes/ no). These are somewhat arbitrary subdivisions but they resemble the exposure categories that are used in many of the large-scale cohort studies in the literature.

The first aim of Appendices J, K and L was to illustrate that the significant differences observed between intake groups in Paper 5 are also present when exposure groups are defined on the basis of average doses of alcohol in the first, second or third part of pregnancy. The second aim of Appendices J, K and L was to illustrate that the observed trends between exposure groups actually dependent on the applied exposure categories. In other words, the trends observed between exposure groups defined on the basis of average alcohol exposure look somewhat different from the trends observed between exposure groups defined on the basis of cumulated alcohol exposure. For example, linear trends were observed for the educational variables in Paper 5 (full pregnancy cumulated alcohol exposure) whereas U-shaped trends were observed in Appendices J, K and L (average alcohol intake in the first, second and third part of pregnancy). Similar differences in trends were observed for many of the other variables.

Having presented the rationales of Paper 3, Paper 4 and Paper 5, and Appendices J, K and L I will move on to summarise the results of each of the papers and Appendices J, K and L.

RESULTS: Prenatal alcohol exposure and child development

The main findings of the alcohol studies were:

Paper 3

- Prenatal exposure to low-moderate doses of alcohol in pregnancy was not found to be associated with any of the parent-rated behavioural and emotional outcomes in boys or girls measured at age seven.
- Some statistically significant results, the least favourable outcomes, were observed for the abstaining group between prenatal exposure to low-moderate doses of alcohol and *Conduct* and *peer-problems* at age seven.
- Some statistically significant results, the most favourable outcomes, were observed for the high exposure (> 90 units) group between prenatal exposure to low-moderate doses of alcohol and *Externalising*, *Internalising*, *Hyperactivity*, and *Emotional* outcomes at age seven.
- Exposure to binge drinking in pregnancy was found to be negatively associated with parent-rated *Externalising, Internalising* and *Conduct* scores in boys measured at age seven.
- Exposure to binge drinking in pregnancy was not found to be associated with any of the parent-rated *Behavioural* and *Emotional* outcomes in girls measured at age seven.

Paper 4

- Compared to the no-binge exposure group, binge drinking in early pregnancy was found to be weakly, but statistically significantly associated with parent-rated *Externalising* scores at age seven.
- Compared to the no-binge exposure group, binge drinking in late pregnancy was found to be negatively associated with parent-rated *Externalising* scores at age seven.
- Being exposed to binge drinking in last part of pregnancy was found to be significantly more negatively associated with parent-rated *Externalising* scores at age seven than exposure to binge drinking in early part of pregnancy.

Paper 5

- Highly statistically significant differences were observed on virtually all of the background characteristics between women with different reported intakes of cumulated low-moderate doses of alcohol in pregnancy. Linear or J-shaped associations were observed on all virtually all of the variables.
- The included variables were all considered potential confounding factors to a smaller or larger extent in observational studies investigating PAE and neurodevelopmental outcomes in childhood. Because observational studies only control for a limited number of variables and (obviously) only control for the confounders that are actually available, residual confounding may mask potential associations between exposure to low-moderate doses of alcohol and neurodevelopmental outcomes in childhood.
- The list of included variables is not complete, sufficient nor exhaustive. Rather, the included variables were those available and those known from the literature to a smaller or larger extent to be associated with neurodevelopment in childhood. They were included to illustrate that large variations also exists on other variables than those usually considered key confounding variables.
- Mediating variables were not included in Paper 5 as no information from standardised tests or questionnaires was available on such variables. The lack of mediation analyses in the observational literature was discussed as a potential source of bias.
- It was concluded that future observational studies should control for factor scores rather than for the observed variables as is practice today. This may minimise the total amount of residual confounding in observational studies.

Appendices J, K and L

- Differences were observed between exposure groups defined on the basis of reported average doses of alcohol in the first, second or third part of pregnancy. The trends observed when the exposure groups were classified on the basis of information from the

first interview was very similar to the trends observed when the women were classified on the basis of information from the second or third interview.

- The trends observed between exposure groups defined on the basis of "full pregnancy cumulated alcohol exposure" were different from the trends observed when exposure groups were defined according to an "average measure" in the first, second or third part of pregnancy. For example a linear trend was observed for the education variables according to the cumulated measure, but a U-shape relation was observed when exposure groups were defined according to an average measure. The results from observational studies may reflect these different trends in background characteristics rather than effects of PAE.
- When an average measure was applied the abstaining group and the high intake group looked very similar on most background characteristics. The subdivision of the abstaining and high intake groups in Paper 5 on the other hand revealed the impact the adapted classification method had on the distribution of confounding variables. It reveals that the trends observed for the confounders are very sensitive to the subdivision of alcohol categories and much information is lost when the exposure categories are treated on a categorical level.

DISCUSSION: Prenatal alcohol exposure and child development

For centuries a debate has prevailed as to whether, and to what extent, prenatal exposure to alcohol is negatively associated with neurodevelopmental outcomes in childhood. There seems to be evidence that exposure to very large doses of alcohol is negatively associated with neurodevelopmental outcomes in childhood. However, the research question that, in some respect, was the main focus of this thesis, namely whether exposure to low doses of alcohol is negatively associated with neurodevelopmental outcomes in childhood, remains unresolved. Regarding binge drinking, the literature as well as the findings from Paper 3 and Paper 4 seem to suggest that exposure to binge drinking is indeed negatively, albeit weakly, associated with neurodevelopmental outcomes in childhood. Exposure in late pregnancy is more negatively associated with *Externalising* behaviour than early exposure.

In order to understand the lack of negative association with PAE to low doses of alcohol we need to draw on knowledge from epidemiology as well as psychology. The starting point of Paper 3 is an extensive amounts of epidemiological data from the DNBC. However, in order to understand the results an unambiguous focus on prenatal factors, as has otherwise been the norm in the literature, is insufficient. Instead, I will introduce and discuss the role of post-natal factors as well, factors that have generally received very little attention in the literature. The point I will try to make is that such psychologically-oriented early childhood factors should seriously be taken into account when interpreting the results of the epidemiological studies.

I will discuss the literature by introducing a number of methodological limitations. These include: 1. *an issue of confounding*, i.e. the lack of negative association may be due to confounding factors

insufficiently controlled for; 2. *an issue of mediation*, i.e. the lack of negative association may be due to mediating factors insufficiently controlled for or not controlled for altogether; 3. *an issue of "dose", "pattern" and "timing"*, i.e. the lack of negative association may be due to the fact that these parameters have not been sufficiently incorporated into the definition of the alcohol exposure variable; 4. *an issue of defining the alcohol exposure categories*, i.e. the lack of negative association is in part associated with problems of the information on which the alcohol categories have been defined; 5. *an issue of the use of the SDQ*, i.e. is the SDQ and other outcome measures used in other observational studies may not be sensitive enough to detect potential (small) effects of exposure to PAE; and 6. *an issue of age at which the child was assessed*, i.e. the child may not be assessed at an appropriate age at which a potential effect is present? Below, these six "issues" will be discussed one at a time mainly in relation to the findings from Paper 3, Paper 4, Paper 5, and Appendices J, K and L, but also drawing on the findings from the SDQ part of the thesis.

The issue of confounding

Above, a confounder was defined as the mixing together of the effect of an exposure with a factor that is statistically associated with the exposure and causally associated with the outcome (Figure 4) [191]. Two matters should be mentioned in relation to this. First, if one fails to control for a confounding factor, for example smoking, *residual confounding* remains. Second, if one fails to measure the confounder accurately it will lead to *misclassification*. Both of these factors can either create a spurious (i.e. false) association or mask a true association [34].

Regarding *residual confounding* factors, it was in Paper 3 decided to control for parental smoking, education and past psychiatric history as well as maternal well-being in pregnancy. The multivariate statistical analyses revealed that these were indeed important confounders. However, other factors including parental age, IQ, personality, temperament, socio-economic status, social support, stress and lifestyle factors could advantageously have been controlled for. Information was available on age, but was unfortunately considered unimportant at the time of designing the study, and hence not controlled for. No information was available on the other factors. The failure to control for these may likely have led to the remaining of residual confounding, and in turn have masked a true association for the low exposure groups [160]. In Paper 3 an apparent protective effect was observed for the high exposure group on many of the outcomes. These associations are likely to be spurious. Abel [146] noted that moderate drinking is correlated with higher parental education and many other social advantages. He therefore declares that it is not surprising that when middle- and upper-income families are studied (as in the DNBC), infants exposed to higher quantities of alcohol perform better compared to children of (less well educated) abstainers. This is backed up by Paper 5 in which it was shown that very large intergroup differences existed on most variables, and that the most favourable outcomes were indeed observed for the high exposure group.

Regarding psychological confounders such as maternal and paternal IQ, temperament and personality no data was available. However, it is speculated that similar trends may have been

observed on such variables had data been available. The aim of the DNBC was in particular to study the side effects of medication and infections [187], and information on psychological factors in the DNBC, and most other birth cohorts for that matter, is generally sparse. The literature therefore generally fails to control for such psychological-oriented confounders, and this may indeed in part explain the lack of consistency in the literature and the lack of negative association in Paper 3.

The aim of Paper 5 was to illustrate differences in background characteristics of women with different reported intakes of alcohol in pregnancy. Such a thorough descriptive study is important as it shed light on the problem of residual confounding apparent in observational studies. The significant differences and linear trends observed on most variables suggests that the problem of residual confounding may indeed be present. The Paper illustrates that differences exist on most variables, and most likely also on psychologically-oriented variables that were unfortunately not included.

It can be argued that at least the lifestyle factors only introduce little "noise" to the statistical analyses and therefore do not need to be controlled for. However, the results of the EFAs presented in Paper 5 did actually identify a number of latent variables of which the second was a "stimulant" factor and the sixth an "educational-related lifestyle" factor. Considering that the factor scores grasp the underlying variability of all the variables constituting any single factors it is indeed meaningful to control for such factors. The conclusion of Paper 5 was therefore that epidemiological studies should be aware of this when designing studies of PAE and neurodevelopmental outcomes.

Regarding *misclassification*, if included confounders have not been classified correctly it may introduce bias to the results. Alternatively, if a confounder has not been measured accurately it will also lead to misclassification, that in turn will introduce bias to the result. This issue of misclassification is a particular problem when the aim is to control for psychosocial factors [200]. The "past psychiatric history"-variable was register-based, and therefore as such did not introduce reporting bias into the statistical analyses. However, it was decided to dichotomise the variable into those with and those without a psychiatric diagnosis (rather than for example classify the women on the basis of type of diagnosis). Because some psychiatric diagnoses may be more of a risk factor that others (for example maternal depression in pregnancy), the dichotomised classification may be too broad and lead to bias.

The "self-reported psychological well-being in pregnancy"- variable was controlled for in Paper 3. However, this composite variable was constituted from a broad range of self-reported measure of well-being in pregnancy. These were then summed and the sum-scores were then treated on a categorical 0/ 1/ 2 level. This variable most likely have introduced misclassification into the statistical analyses. Such variables, "invented for the purpose of the cohort" probably introduce more bias into the statistical analyses than what is gained by attempting to control for such factors. Standardised questionnaires should have been used instead.

When the aim of a study is to investigate exposure to very small doses of alcohol a small effect (at most) is hypothesised. It is obvious that when a small effect is hypothesised it is even more important to accurately control for major confounding factors as these will otherwise mask a true association. The conclusion of Paper 3 in relation to low doses of alcohol and confounders is that such factors were controlled for insufficiently and this in part may explain the lack of association between exposure to low doses of alcohol and SDQ scores as age seven. The most favourable outcomes observed for the high exposure group, and the least favourable outcomes for the abstaining group is probably in part a product of residual confounding. A similar conclusion can be drawn for binge drinking in Paper 3 and Paper 4. However, because the effect of binge drinking is expected to be larger it takes more to fully mask an association. Another factor that is at least as important is the issue of mediation that is virtually absent in the epidemiological literature.

The issue of mediation

Above, a mediator was referred to as an intermediate variable on the causal pathway from exposure to outcome (Figure 5). First, if one fails to control for mediating factors, for example attachment, *residual mediation* remains. Second, if one fails to measure the mediator accurately it will lead to *misclassification*. As with the confounders both of these can either create spurious (i.e. false) associations or mask true associations [34].

In Paper 3, Paper 4 and Paper 5 and Appendices J, K and L mediators were not included as information from standardised questionnaires was not available on such variables. It is therefore beyond doubt that *residual mediation* remains in the statistical analyses in Paper 3 and Paper 4. It was discussed above that the aim of the DNBC was to study side effects of medications and infections [187] which is probably the reason for the lack of information on psychologically-oriented variables. As discussed above, epidemiologists generally have a long tradition and a very thorough focus on confounding factors. However, they generally pay very little attention to mediation [15]. Psychologists on the other hand, have long recognised the importance of psychologically-oriented mediating variables but have focused little on pregnancy related confounders [15]. It has been recognised in the literature that there has been too little focus on the post-natal environment [160]. Abel noted that *"that a child's postnatal environment affects his/ her behaviour is hardly surprising, but it has not received the attention it deserves in the present context"* [146] (p. 127). Further, a comprehensive review concluded that *"the two most important types of confounding of effects on neurodevelopmental outcome are failure to control for the postnatal environment and failure to control for factors which are strongly genetically influenced and which may be related to both prenatal alcohol exposure and the outcome"* (p. 20) [34]. A combined approach of epidemiologists thorough knowledge on confounding and psychologists in-depth focus on childhood mediators is recommended for future studies.

In Figure 5 "attachment" was given as an example of a psychologically-oriented "postnatal environmental"-variable. Other important mediators include IQ on part of the child [158], parental mental health [160], home environment [146, 158], quality of parenting [160], changes in living

arrangement [158], violence [158], social support [146] and early motor and cognitive development. These factors are particularly important in the first two years of life when the child's brain is particularly plastic [146]. Therefore, a child's behaviour and cognitive functioning at for example age seven is a product not only of prenatal factors (like alcohol) but also of postnatal exposures, as was illustrated in Figure 2 and Figure 3. In other words, the impact of PAE cannot be separated from the impact of the developmental processes in childhood.

As already mentioned Abel [146] points out that the factor *"home environment"* has not received the scrutiny it deserves. The analyses in Paper 3 revealed that the most favourable outcomes were found for the children of the high intakers, and the least favourable outcomes for the abstainers. The high intakers (>90 units) were older, more likely to have a university degree, to eat fish, and have to a pre-pregnancy BMI above or below normal range. They were on the other hand less likely to watch TV and drink cola. The abstainers (0 units) were younger, more likely to have mandatory education only, drink cola, watch TV, smoke cigarettes, live alone and to have psychiatric problems. They were less likely to do exercise and eat fish. These characteristics, along with unmeasured mediating variables such as IQ, attachment style and personality could very likely be mentally protective for the high exposed children, but disadvantageous for the unexposed children.

In the 1950's Bowlby was the first to demonstrate the lasting consequences that the quality of the mother-child relationship has for a wide range of developmental cognitive and mental health outcomes [201]. He concluded that infants who develop a secure attachment style are those with a history of sensitive and responsive maternal care and this style is later associated with better emotional regulation, higher self-esteem, and more develop coping skills. This in turn makes the children better able to handle stressful or challenging situations and reversely lowers the risk for poorer mental health outcomes later in life. On the other hand, children with an insecure attachment are at greater risk for poor mental health outcomes [202, 203]. It is known that a secure attachment is associated with better academic performance in adolescents and better cognitive performance in childhood [204, 205]. Further, if there is a negative effect of being exposed to alcohol prenatally, the effect is most likely very small. "Attachment" and "home environment" on the other hand has a substantial effect on mental health development. Therefore, the negative effect of alcohol will most likely be masked by the relatively larger effect of these variables.

Regarding *misclassification* it is obvious that because no mediators were included in the statistical analyses in Paper 3 and Paper 4 no mediators have been misclassified and thus introducing bias to the results altogether. This also applies to the literature in general where it is rare that researcher control for mediators all together, and even more rare that they control for such factors by means of standardised questionnaires [146]. In the DNBC information was actually available on different psychologically-oriented mediators from the first few years of the child's life such as "mother-child relation", "mother-partner relation", "environmental impacts on part of the mother", and "mother psychological well-being". However, none of these were derived from standardised questionnaires. In other words, the questions may be face valid but it is difficult to know exactly

what the questions enquire about. Introducing such variables into the statistical analyses will at best do nothing to the results but at worst introduce bias to the analyses. This will in turn over- or underestimate the effects of PAE on the neurodevelopmental outcomes.

The role of mediators also deserves mentioning when one aims to understand the results of epidemiological studies versus clinical studies. It was concluded above that there is no consistency in the epidemiological literature that investigates exposure to lower doses of alcohol. Some studies, including Paper 3, find no associations with neurodevelopmental outcomes [28, 29, 43, 44, 134, 163], others have reported on negative associations [17-19, 21, 22, 24], whereas others have even reported on a J-alcohol shape association [30, 164]. Clinical samples have on the other hand more consistently been able to report negative associations with neurodevelopmental outcomes in childhood [39, 40, 145, 149, 153, 154, 161, 206-209]. These results suggest that exposure to large doses of alcohol does indeed have a negative effect on child neurodevelopment. However, as Figure 2 and Figure 3 illustrate it may not be all that straightforward. Apart from being exposed to alcohol prenatally, these children may also be exposed to illicit drug and poor nutrition in the intrauterine environment, the mothers may suffer from psychiatric disorders, there may be little extra family protection, the children may have experienced violence and neglect, they may have been in foster care and be non-securely attached to their caregivers.

Although this thesis only focuses on exposure to lower doses of alcohol in non-clinical, epidemiological samples this is still important. If we assume that the poor cognitive and behavioural outcomes for the high exposed children solely is an effect of the prenatal exposure to alcohol we may induce that exposure to lower doses is associated with a smaller effect on part of the child. It will lead us to understand that exposure to alcohol as an easy-to-understand biological process. If, on the other hand, the association for the high exposed group in part is spurious and due to residual confounding and mediating factors (and therefore not exclusively biological), it will still influence our understanding of the findings from the epidemiological, birth cohort studies. If we do not expect negative findings in studies with high exposed groups, then negative (smaller) effects from cohort studies looking at exposure to much low doses of alcohol would not be expected. If this is so, these studies in turn would probably not receive the amount of attention they currently do.

The issue of "dose", "pattern" and "timing"

Whereas the issues of confounding and mediation have to do with "noise" in the statistical analyses the issues of "dose", "pattern" and "timing" is concerned with "noise" in the definitions of the exposure categories. It was described above how the early literature investigated exposure to daily consumptions, that is, very large *"doses"* of alcohol [*130, 145, 153-156*], whereas focus gradually has shifted towards investigations of weekly consumptions[22, 43, 44, 163, 202], that is more moderate *"doses"* of alcohol. In Paper 3 the aim was to investigate associations with even lower *"doses"* of alcohol and a cumulated measure rather than an average measure was therefore decided on.

Regarding *"dose"*, the use of a cumulated measure allowed for some more detailed comparisons of exposure groups within the low spectrum, and the linear trends observed in Paper 5 somehow justified this new approach to classification. In that respect it seems like a solid way to subdivide the women. Having said that, the idea of investigating exposure to very low doses of alcohol (down to as little as >0-5 units in pregnancy) taking into account the issues of confounding and mediation described above it makes little sense. The effects of being exposed to very low doses of alcohol will, all other things being equal, be expected to be very small. The issues of residual confounding and mediating factors will have a much greater influence on the statistical analyses. Therefore, even if there is an effect of being exposed to very low doses of alcohol (i.e. >0-5 units) it would most likely not be detectable in the statistical analyses. With hindsight, such small exposure categories should probably not have been included.

Regarding *"pattern"* it was decided to subdivide the categories into exposure to lower doses of alcohol and binge drinking as is standard in the most recent epidemiological literature [43, 44, 163]. On the one hand this subdivision makes sense if it is the level of BAC that determines the level of damage. On the other hand, this subdivision makes little sense as considering the definition of binge drinking (i.e. five units of alcohol), seem like a somewhat arbitrary one. It is not based on empirical evidence that a biological threshold (i.e. five units) exists under which drinking is not associated with any damages to the developing foetus. Even if such a threshold exist it will, apart from the consumed amount of alcohol, also depend on numerous genetic, biological and timing factors.

The use of an average measure has also been criticised for not being sensitive to the fact that women rarely drinks every day [210] and therefore does not take "intensity" into account. For example it does not allow for distinctions between women who drink three drinks on a Saturday night and women who drink three drinks on three consecutive days. O'Leary and colleagues recommend the use of a "composite method" in which "dose", "pattern" and "timing" parameters are combined into a composite measure [165]. Using this method the women are asked to indicate 1. how often they drink alcohol, 2. the quantity consumed on a typical occasion, and 3. these questions were asked separate for each trimester. This method is probably better and more precise than the standard "exposure to an average dose of alcohol"-method. It takes into account "dose", "timing" and "pattern", but also "intensity". If it is the BAC that determines the level of damage it is indeed of central importance to inquire about intensity, i.e. "quantity consumed per occasion" rather than just "glasses per week". However, the method does hold its limitations. It does not truly overcome the problem of using an average measure, it does not overcome the problem that exposure in some parts of pregnancy may be more devastating that exposure in other parts of pregnancy (please see below), it does not overcome the problem of classifying the women into very broad categories (i.e. low, moderate, binge and heavy), it does not overcome the problem of reporting bias (please see below), and it does not overcome the fundamental problem of comparing groups of women who (might) differ on confounding and mediating factors al together. For example the low category was defined by O'Leary and colleagues as women who

"over a week, should have less than seven standard drinks, and, on any one day, no more than 1-2 standard drinks" [165] (p. 957).

In the DNBC the women were initially asked about their weekly average intake. Hereafter they were asked about potential episodes of binge drinking. It is very likely that some of the women have averaged the number of drinks from binge episodes and hereafter have reported no binge episodes. Other women may have reported their "binge intake" both as an average measure and as a binge episode per se. Even others will only have reported the binge episodes as such. In the former two cases it will induce bias to the statistical analyses. Therefore the subdivision of questions in the DNBC into average alcohol intake and binge drinking seem insufficient. Although limited, the "composite method" in many respects seem superior. Considering no information on intensity was available in the DNBC it can be concluded that the analyses of lower doses of alcohol and binge drinking in Paper 3 instead should have been combined.

Regarding *"timing"* it was decided to compose the exposure for Paper 3 from full pregnancy rather than early pregnancy information. The rationale for doing so was that it was hypothesised that "total exposure" would be a better predictor of behavioural development than "early exposure". As described above it is been reported in the literature that early as well as late pregnancy exposure is most negatively associated with neurodevelopment [146] and Paper 4 concluded that exposure in the last part of pregnancy was more negatively associated with externalising behaviours. On this ground, because the cumulated measure does not per se take into account the effect of timing and because it is not sensitive to the differentiated effect of being exposed in different part of pregnancy the measure does hold limitations.

The specific aim of Paper 4 was to investigate the effects of timing exactly this because human studies have primarily focused on early pregnancy exposure [21, 31, 43, 44, 48]. The effect of exposure later in pregnancy have largely been ignored in epidemiological studies. Most studies use information on early pregnancy exposure only despite the fact that animal studies suggests that late exposure is at least as devastating for the developing foetus. One possible explanation could be that cohorts only have information on early pregnancy exposure. Another plausible explanation is that there is a long tradition in epidemiology to investigate early pregnancy exposures (not just to alcohol) and associations with still birth, spontaneous abortion, congenital malformation and other physical outcomes (outcomes that have not otherwise been touched upon in this thesis). Studying such outcomes it probably makes very good sense to investigate early pregnancy exposure. However, when the outcome is neurodevelopment, and hence CNS dependent, it no longer make sense to have a sole focus on early exposure.

The finding from Paper 4 confirmed this as it was found that late exposure was more negatively associated with externalising scores than no exposure or early exposure. This findings deserves a few comments. It is an important study because it is among the first human studies to investigate the effects of timing using an 'animal-like' design. The finding is also important as the significant association from early pregnancy is even more strongly associated with late pregnancy exposure. The finding indicates that late binge exposure may be worse for the developing CNS – a finding

that makes sense if it is the BAC that determines the level of damage to the CNS. Brain development occurs in multiple stages and different brain regions that subserve different observable behaviours are sensitive to alcohol exposure at different points in time. Broadly speaking, first trimester is mostly concerned with cell proliferation and migration of cells [211], the second trimester with neuronal and synapse formation, axonal and dendritic outgrowth and programmed cell death [212], whereas in the third trimester the brain is growing larger with synapse formation and myelination taking place [212]. From this perspective exposure to alcohol in the early part of pregnancy could result in interrupted cell proliferation and migration of cells to all parts of the central nervous system. Late pregnancy exposure, on the other hand would be expected to lead to an interruption of brain development in specific parts of the brain.

The finding also makes sense in relation to a study investigating neuronal reduction in the frontal cortex of vervet monkeys after PAE [185]. The vervet monkeys were allowed to drink the equivalent of 3-5 standard drinks (i.e. binge drinking) during the third trimester. The offspring were perfused at birth and the authors found significantly fewer cells in the frontal lobes of the exposed offspring as well as increased density of interstitial white matter neurons. The authors concluded that these changes are consistent with the behavioural and cognitive changes observed in FASD [185]. Because the third trimester in humans is the period of greatest brain growth [212, 213] it may be that exposure to one binge episode in late pregnancy is enough to lead to a lower number of frontal cortex neurons and higher number of interstitial neurons in frontal white matter [185].

The findings from Paper 4 should on the other hand also be viewed with caution. It appears from the background characteristics that fewer women in the late-binge group had >13 years of education (38.3 %) compared to the no-binge group (52.3 %) and the early-binge (54.2%), and more mothers had a psychiatric diagnosis (10.6 %) compared to the no-binge (4.8 %) and the early-binge (5.2 %). Interestingly, fewer fathers in the no-binge category have a psychiatric diagnosis (1.1 %) compared to the maternal no-binge (3.4 %) and early-binge (3.7 %) groups. The early-binge and late-binge groups also consisted of more maternal smokers (30.6 % and 32.3 %, respectively) compared to the no-binge group (18.9 %). Therefore, although the effect after binge drinking will be expected to be larger than after exposure to lower doses of alcohol, the issues of residual confounding and mediation are still operative and the study therefore needs to be replicated.

The issue of defining the alcohol exposure categories

As mentioned above there is no uniformly accepted definition of what is meant by low, moderate and high doses of alcohol [35, 136, 175]. What some researchers consider a “moderate” intake is by others classified as a “low” or a “high” intake. In Paper 3 the “high” exposure group was defined as a cumulated intake of >90 unit in pregnancy, i.e. a little over two units per week. Two units per week would by most other researchers be defined as a low intake. This lack of standard of how to classify the exposure categories may in part explain the inconsistency of results between studies,

and has led the Centre for Disease Control and prevention (CDC) to use the term “risky drinking” rather than the imprecise term “moderate drinking” [35].

Apart from the issue of what is actually meant by exposure to a low, moderate or high level of alcohol there are three more serious issues regarding how to define the alcohol exposure categories and potential impacts this may have on the results. These include: 1. on the basis of what information has the alcohol categories been defined?; 2. on the basis of what information are the abstainers and the high intakers categorised? 3. is a substantial number of women misclassified into wrong exposure categories?

Firstly, Appendix J, K and L were included to illustrate the degree of impact the classification has on the distribution of covariates. The intake categories in the Appendices were defined according to average intake in the first, second and third part of pregnancy, whereas the classification in Paper 5 was based on the cumulated measure. Comparison of these classification methods revealed different distribution of scores on potential covariates (for example education). Linear trends were observed for the cumulated measures whereas U-shaped associations were observed for the average measures. Considering the substantial impact residual confounding and mediation has on the results, the variability of results between studies can likely be explained by such different distributions of covariates presented in Appendices J, K and L and Paper 5.

Secondly, the subdivision of the abstainers and the high-intakers in Paper 5 also illustrates how two ‘subgroups’ of women are embedded within each of these two groups. The two sub-groups embedded within the abstaining and high intake groups differed significantly on most covariates (as is discussed above). In a review by Henderson et al it is concluded that the J-shape reflects a *“healthy drinkers effects”* in which women with a poor obstetric history are more likely to abstain from drinking [35]. However, Paper 5 shows that it is not only the women with a poor obstetric history that abstain from drinking. The abstaining group consists of a pregnancy-abstaining group and an all-time abstaining group. The pregnancy-abstaining group could indeed be referred to as a *“healthy drinkers group”*. However, it was revealed that the all-time abstainers include women with poor education, unhealthy lifestyle habits and diet. It would probably be more appropriate to refer to this subgroup as an *“unhealthy, non-drinker group”.* Because these two subgroups are bi-modally embedded into one it is recommended that these groups are explicitly subdivided in future studies. Both because the data is not normally distributed but also because the results from the statistical analyses are dependent on residual confounding factors.

Thirdly, there may be an issue of reporting bias in which women with a high intake report on a zero or a very low intake and thus are misclassified into the abstaining or low exposure groups [146]. It has been demonstrated that mothers who drink deny [46], that mothers who admit do not remember correctly [46], and that other women even purposely under-report their alcohol intake [188, 189]. Particularly women with a high intake tend to under-report alcohol consumption during pregnancy – especially so if the information is not collected in a careful and sensitive manner [169]. Because the under-reporting probably varies between groups it will lead to differential misclassification, i.e. misclassification that varies between intake groups. That there may be some

misclassification is backed up by the finding from Paper 5 that showed linear trends for the middle exposure groups but curves for the abstaining and high exposure groups. This means that in Paper 3 the little higher, and hence poorer, relative changes in means observed for the abstaining groups may be an effect attributable to misclassified women with high consumptions.

The issue of the use of the SDQ

In the first section of this thesis the psychometric properties of the SDQ were investigated very thoroughly. It was concluded that the SDQ works particularly well for older children rated by teachers and particularly well for the hyperactivity scale. It was further concluded that Model 1 (a five factor model) and Model 2 (adding two second order factors to Model 1) worked equally well.

Because of lack of information on psychologically-oriented outcomes in the DNBC, the only measure applied in Paper 3 and Paper 4 was the parent-rated SDQ. This is a limitation of Paper 3 and Paper 4. Further, the SDQ was found to have the poorest psychometric model fits for young (5-7-year-old) children rated by their parents. However, the model fits were found to be acceptable to good so this should be of little concern. Three other factors deserves mentioning.

Firstly, because of its brevity the SDQ is probably the most often included screening tool in the large-scale cohorts including the DNBC, CCC2000, ABC, Healthy Habits for two, Avon Longitudinal Study of Parents and Children (ALSPAC), and the UK Millennium cohort. It has therefore been applied in several studies investigating the effects of PAE rendering for comparison of results across studies. This is an obvious advantage. However, most studies include other outcomes measures as well [30, 31, 41, 48] and often include information on the SDQ from teachers as well as parents [42]. In Paper 4 in particular where an association was observed with *Externalising scores* it would have strengthened the validity of the results if another outcome measure had been included and similar findings had showed. With just one outcome measure (the SDQ) it is only possible to conclude very cautiously.

Secondly, it has been reported that people with PAE exhibit impairments on the performance of relatively complex and novel tasks [147, 214], and the SDQ may simply not be sensitive enough to grasp potential damages from PAE. Each of the SDQ scales consist of five items rated on a 0-2 Likert scale allowing for sum-scores from 0-10. Even though more questions by no means in themselves would be an advantage the inclusion of more *reliable* questions would. If each scale consisted of more *reliable* questions it would allow for more variability (i.e. a broader range of sum-scores) and this would likely be better at grasping subtle differences between exposure groups. Thus, complex, neuropsychological tasks may be needed instead [215].

Thirdly, as briefly mentioned information was only available from parent raters although it is known from the literature that teachers are actually better at identifying externalising symptoms. Parent ratings are known to give a conservative estimate of behavioural problems and parent scores are generally lower than teacher scores [99, 190]. Appendix C shows that the teacher cut-offs generally include a wider range of scores in the *"normal"* banding, indicating that a high sum-score

is needed in order to be placed in the *"borderline"* or *"clinical"* bandings. Therefore the inclusion of teacher ratings would have heightened the usefulness of Paper 3 and Paper 4.

Fourthly, from age of 11 the inclusion of both parent, teacher and self-rate scores can be applied. This is a great advantage because when the SDQ is available from multiple informants, that is parents, teachers and self-rating a computerised algorithm can be applied. This algorithm is based on the symptom and impact scores derived from the SDQ and the predictive algorithm generates *"unlikely", "possible"* or *"probable"* ratings for three groups of disorders, namely, conduct-oppositional disorders, hyperactivity-inattention disorders and anxiety-depressive disorders [74, 216]. Predictions of these three groups of disorders are then combined to generate an overall prediction about the presence or absence of any psychiatric disorder [74]. The algorithm has been found to identify two-thirds of the children with a psychiatric disorder in a community sample [74]. However the screening efficiency depends on the diagnosis, and has been found to be good for conduct-oppositional disorders, hyperactivity-inattention, depression, pervasive developmental disorders and some anxiety disorders. It has on the other hand been found to be poor for specific phobias, panic disorders, eating disorders and separation anxiety. It is recommended that future studies with information from this triad of informants should make use of the computerised algorithm as this is probably the best predictor of mental health problems in the child.

Fifthly, a final issue concerns the impact part of the SDQ. One (unpublished) study found that impact scores in preschool years was the best predictor of ADHD diagnoses later in childhood [217]. On this basis impact scores should have been included as a separate outcome measure in Paper 3 and Paper 4. It is recommended that the impact part should be used in future studies using the SDQ where information from only one informant is available.

The issue of age at which the child was assessed

The final issue is concerned with the age at which the child was assessed. In the present study the children were assessed at age seven. The main reason that this age span was decided on was a pragmatic one, that is, that was when the follow-up was carried out. It was not decided on for any theoretical reason. At least when it comes to neurodevelopmental outcomes the ages at which follow-ups take place should be theory-driven rather than practicality-driven. Other cohort studies have been criticised for assessing the children at inappropriate ages [215, 218, 219], a similar argument can be used for Paper 3 and Paper 4. The argument is that behavioural effects may not be present until later in childhood, and age seven is thus too early to assess a child exposed to alcohol prenatally for behavioural and emotional difficulties. For example, one study found that among 2600 children with FAS 50 % were found to have normal developmental scores at pre-school whereas all had severe brain dysfunction at age ten. In the same group of children 10 % were found to have attention problems at age five, whereas 60 % had attention problems at age ten [220]. The lack of consistency in the literature, and the lack of negative associations in Paper 3 may also in part be a consequence of this – that the children were assessed too early, at an inappropriate age that was adapted for logistic and practical reasons.

CONCLUSION: Prenatal alcohol exposure and child development

On the basis of the discussion above on issues concerning confounding and mediation, issues dealing with *"dose", "timing"* and *"pattern"*, issues of defining the alcohol exposure categories, and issues related to the outcome measure, no firm conclusion can be drawn.

Most convincing is the evidence from binge studies. These findings indicate that being exposed to just one episode of binge drinking, particular in the last part of pregnancy, probably has an effect on behavioural development at age seven.

Less conclusive are the findings from the studies investigating exposure to lower doses of alcohol. On the basis of the literature and the results of Paper 3 it cannot be concluded that prenatal exposure to low doses of alcohol is negatively associated with neurodevelopmental outcomes in childhood. However it cannot either be concluded that prenatal exposure to alcohol is *not* negatively associated with neurodevelopmental outcomes in childhood. The reason for this is that the research methods currently applied are insufficient and all infected by the six *issues* or methodological limitations presented above.

Although we currently do not have evidence that exposure to low doses of alcohol is negatively associated with neurodevelopmental outcomes I strongly believe that the National Board of Health should continue to recommend abstinence. The main argument I would use for this is that we currently do not have any evidence that being exposed to alcohol in any ways does anything good for the developing foetus. Quoting Garcia-Algar and colleagues: *"no evidence of harm does not mean evidence of no harm"* [46]. As long as our research designs are so obviously full of methodological faults and limitations the only reasonable thing to do is to recommend abstinence –for the sake of the unborn children.

STRENGTHS AND LIMITATIONS: Prenatal alcohol exposure and child development

Apart from the strengths and (not least) THE limitations discussed in the Discussion section above few other issues deserved mentioning. The present Ph.D. project is based on prospectively collected data that has already been collected as part of existing cohorts. This is considered a strength as well as a limitation. The considerable size of the samples applied in all of the five articles and all of the appendices is an obvious strengths. It goes without saying that such large amounts of data could not have been collected within the scope of a three year Ph.D. On the negative side however is the limited access to psychologically oriented data in all of the follow-up sessions. The main aim of the DNBC was to study side effects of infections and medication and the amount of psychological data was therefore sparse. Ideally, other questionnaires or diagnostic interviews like for example the Development and Well-Being Assessment (DAWBA) to examine

mental health development more thoroughly would have been included. Further it was not possible to change the age at which the child was assessed.

It is also a strength that information on alcohol intake was assessed during pregnancy for the first two interviews. However, it is a limitation that data for the third interview was collected six months post-partum. This may have introduced recall bias. A mother who has given birth to a child with a *"difficult"* temperament may for example be more likely to report a higher intake of alcohol that a woman with a child with an *"easy"* temperament.

Another obvious strength of the study is the inclusion of register-based data from Statistics Denmark. In Denmark all citizens carry a personal identity number from birth until death. Through this number it is relatively easy to tract people over time. The loss to follow-up is therefore relatively small by usual longitudinal research standards making such data immensely reliable and not subjected to reporting bias.

FUTURE STUDIES: Prenatal alcohol exposure and child development

Because of the limitations discussed above there is no need for yet more studies comparing groups with different average intakes. I agree with Abel who concludes that *"until research progresses beyond simple group comparisons, our understanding of the nature of the cognitive deficits associated with the uniqueness of FAS/ ARBD will never progress beyond superficiality"* [146] (p. 134). I further agree with Gray and Henderson who concludes that: *"… we may have reached the limits of what we can determine from the standard case control and cohort designs. Teasing apart the relative contributions of exposure and confounding variables and trying to adjust for genetic influences is likely to require the application of study designs that are new for this particular research area"* (p. 21) [34].

One way of overcoming the methodological problems (at least in part) is to include a whole range of natural experiments and research designs because none provide an adequate solution on its own. Because each study design will have its own strengths, limitations and biases it is concluded that a combination of research strategies are needed [16, 221]. Such studies will by no means solve all the problems but is the best we have at the moment and will be one step in the right direction.

One way is to compare data from different cultural settings in which drinking in pregnancy is observed within different socio-economic groups. Therefore, studies that compare results from different countries are welcomed. For example, one study compared results from Denmark where it is the affluent that drink in pregnancy and Finland where women with low SES drinks in pregnancy [29]. This study did not find low doses of alcohol to be related to child inattention/ hyperactivity once social adversity and smoking were taken into account.

Another way is by means of a sibling design in which mothers who changed their drinking habits between pregnancies are included. That is, only mothers who drank heavily in one but not in

another pregnancy are included. This way the effects of bias, confounding and mediation are minimized, but not removed all together. One study that did so did find that high exposed siblings exhibited more conduct problems suggesting a causal association. However, they did not find any associations with attention and impulsivity [222]. Studies using twins families, and numerous mothers and children from extended families and adopted children can also be recommended.

A third way would be to compare neurodevelopmental outcomes in children whose fathers drank in pregnancy but whose mothers did not drink in pregnancy, with children whose fathers did not drink but whose mothers did. If it is assumed that mothers drinking act as a biological teratogen, then similar effects in the two groups would imply that maternal drinking is not causative [16].

Finally, it is recommended that Paper 5 is replicated in other cohorts including potential confounding factors, but even more importantly includes psychologically-oriented mediating factors. Such study would highlight the importance of controlling for mediators in future studies investigating exposure to alcohol in pregnancy and neurodevelopmental outcomes in childhood.

REFERENCE LIST

(1) Costello EJ, Egger H, Angold A. 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *J Am Acad Child Adolesc Psychiatry* 2005;44(10):972-986.

(2) Ford T, Goodman R, Meltzer H. The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. *J Am Acad Child Adolesc Psychiatry* 2003;42(10):1203-1211.

(3) Costello EJ, Foley DL, Angold A. 10-year research update review: the epidemiology of child and adolescent psychiatric disorders: II. Developmental epidemiology. *J Am Acad Child Adolesc Psychiatry* 2006;45(1):8-25.

(4) Obel C, Linnet KM, Henriksen TB et al. Smoking during pregnancy and hyperactivity-inattention in the offspring - comparing results from three Nordic cohorts. *Int J Epidemiol* 2008.

(5) Braun JM, Kahn RS, Froehlich T, Auinger P, Lanphear BP. Exposures to environmental toxicants and attention deficit hyperactivity disorder in U.S. children. *Environ Health Perspect* 2006;114(12):1904-1909.

(6) Linnet KM, Dalsgaard S, Obel C et al. Maternal lifestyle factors in pregnancy risk of attention deficit hyperactivity disorder and associated behaviors: review of the current evidence. *Am J Psychiatry* 2003;160(6):1028-1040.

(7) Rodriguez A, Olsen J, Kotimaa AJ et al. Is prenatal alcohol exposure related to inattention and hyperactivity symptoms in children? Disentangling the effects of social adversity. *J Child Psychol Psychiatry* 2009;50(9):1073-1083.

(8) Thomsen PH, Skovgaard AM. *Børne- og Ungdoms psykiatri. En lærebog om børn og unges psykiske sygdomme*. FADL Forlag; 2007.

(9) Linnet KM, Wisborg K, Secher NJ et al. Coffee consumption during pregnancy and the risk of hyperkinetic disorder and ADHD: a prospective cohort study. *Acta Paediatr* 2009;98(1):173-179.

(10) Rodriguez A, Miettunen J, Henriksen TB et al. Maternal adiposity prior to pregnancy is associated with ADHD symptoms in offspring: evidence from three prospective pregnancy cohorts. *Int J Obes (Lond)* 2008;32(3):550-557.

(11) Mick E, Biederman J, Faraone SV. Is season of birth a risk factor for attention-deficit hyperactivity disorder? *J Am Acad Child Adolesc Psychiatry* 1996;35(11):1470-1476.

(12) Latimer K, Wilson P, Kemp J et al. Disruptive behaviour disorders: a systematic review of environmental antenatal and early years risk factors. *Child Care Health Dev* 2012;38(5):611-628.

(13) Cassidy J, Shaver PR. *Handbook of attachment, theory, research and clinical applications*. 2nd ed. New York: Guilford Press; 2008.

(14) Sameroff AJ, Lewis M, Miller SM. *Handbook of developmental psychopathology*. Kluwer academic; 2000.

(15) Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;51(6):1173-1182.

(16) Gray R, Mukherjee RA, Rutter M. Alcohol consumption during pregnancy and its effects on neurodevelopment: what is known and what remains uncertain. *Addiction* 2009;104(8):1270-1273.

(17) Fried PA, Watkinson B. 12- and 24-month neurobehavioural follow-up of children prenatally exposed to marihuana, cigarettes and alcohol. *Neurotoxicol Teratol* 1988;10(4):305-313.

(18) Larroque B, Kaminski M. Prenatal alcohol exposure and development at preschool age: main results of a French study. *Alcohol Clin Exp Res* 1998;22(2):295-303.

(19) NIAAA: National Institute of Alcohol Abuse and Alcoholism. 10th special report to the US congress on alcohol and health. Highlights from current research. Bethesda: MD: Department of health and human services; 2000. Report No.: 00-1583.

(20) Olson HC, Streissguth AP, Sampson PD, Barr HM, Bookstein FL, Thiede K. Association of prenatal alcohol exposure with behavioral and learning problems in early adolescence. *J Am Acad Child Adolesc Psychiatry* 1997;36(9):1187-1194.

(21) Sayal K, Heron J, Golding J, Emond A. Prenatal alcohol exposure and gender differences in childhood mental health problems: a longitudinal population-based study. *PEDIATRICS* 2007;119(2):e426-e434.

(22) Sood B, aney-Black V, Covington C et al. Prenatal alcohol exposure and childhood behavior at age 6 to 7 years: I. dose-response effect. *Pediatrics* 2001;108(2):E34.

(23) Streissguth AP, Barr HM, Sampson PD. Moderate prenatal alcohol exposure: Effects on child IQ and learning problems at age 7½ years. *Alcoholism: Clinical and Experimental Research* 1990;14(5):662-669.

(24) Testa M, Quigley BM, Eiden RD. The effects of prenatal alcohol exposure on infant mental development: a meta-analytical review. *Alcohol Alcohol* 2007;38(4):295-304.

(25) Sayal K, Draper ES, Fraser R, Barrow M, Davey SG, Gray R. Light drinking in pregnancy and mid-childhood mental health and learning outcomes. *Arch Dis Child* 2013;98(2):107-111.

(26) Skogerbo A, Kesmodel US, Wimberley T et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on executive function in 5-year-old children. *BJOG* 2012;119(10):1201-1210.

(27) Underbjerg M, Kesmodel US, Landro NI, et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on selective and sustained attention in 5-year-old children. *BJOG* 2012;119(10):1211-1221.

(28) O'Leary CM, Zubrick SR, Taylor CL, Dixon G, Bower C. Prenatal alcohol exposure and language delay in 2-year-old children: the importance of dose and timing on risk. *PEDIATRICS* 2009;123(2):547-554.

(29) Rodriguez A, Olsen J, Kotimaa AJ et al. Is prenatal alcohol exposure related to inattention and hyperactivity symptoms in children? Disentangling the effects of social adversity. *J Child Psychol Psychiatry* 2009;50(9):1073-1083.

(30) Kelly Y, Sacker A, Gray R, Kelly J, Wolke D, Quigley MA. Light drinking in pregnancy, a risk for behavioural problems and cognitive deficits at 3 years of age? *Int J Epidemiol* 2009;38(1):129-140.

(31) Kelly YJ, Sacker A, Gray R et al. Light drinking during pregnancy: still no increased risk for socioemotional difficulties or cognitive deficits at 5 years of age? *J Epidemiol Community Health* 2012;66:41-48.

(32) Robinson M, Oddy WH, McLean NJ et al. Low-moderate prenatal alcohol exposure and risk to child behavioural development: a prospective cohort study. *BJOG* 2010;117(9):1139-1150.

(33) Henderson J, Kesmodel U, Gray R. Systematic review of the fetal effects of prenatal binge drinking. *J Epidemiol Community Health* 2007;61:1069-1073.

(34) Gray R, Henderson J. Rewiev of the Fetal Effects of Prenatal Alcohol Exposure. National Perinatal Epidemiology Unit, University of Oxford; 2007 May.

(35) Gray R, Henderson J. Review of the Fetal Effects of Prenatal Alcohol Exposure. National Perinatal Epidemiology Unit, University of Oxford; 2007 May.

(36) Bailey BN, aney-Black V, Covington CY et al. Prenatal exposure to binge drinking and cognitive and behavioral outcomes at age 7 years. *Am J Obstet Gynecol* 2004;191(3):1037-1043.

(37) Nulman I, Rovet J, Kennedy D et al. Binge alcohol consumption by non-alcohol-dependent women during pregnancy affects child behaviour, but not general intellectual functioning; a prospective controlled study. *Arch Womens Ment Health* 2004;7(3):173-181.

(38) Olsen J. Effects of moderate alcohol consumption during pregnancy on child development at 18 and 42 months. *Alcohol Clin Exp Res* 1994;18(5):1109-1113.

(39) Streissguth AP, Barr HM, Martin DC. Maternal alcohol use and neonatal habituation assessed with the Brazelton scale. *Child Dev* 1983;54(5):1109-1118.

(40) Streissguth AP, Bookstein FL, Sampson PD, Barr HM. Neurobehavioral effects of prenatal alcohol: Part III. PLS analyses of neuropsychologic tests. *Neurotoxicol Teratol* 1989;11(5):493-507.

(41) Sayal K, Heron J, Golding J et al. Binge pattern of alcohol consumption during pregnancy and childhood mental health outcomes: longitudinal population-based study. *PEDIATRICS* 2009;123(2):e289-e296.

(42) Skogerbo A, Kesmodel U, Denny C et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on behaviour in 5-year-old children: a prospective cohort study on 1628 children. *BJOG* 2013;120(9):1042-1050.

(43) Skogerbo A, Kesmodel US, Wimberley T et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on executive function in 5-year-old children. *BJOG* 2012;119(10):1201-1210.

(44) Underbjerg M, Kesmodel US, Landro NI et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on selective and sustained attention in 5-year-old children. *BJOG* 2012;119(10):1211-1221.

(45) Goodman R. The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry* 1997;38(5):581-586.

(46) Garcia-Algar O, Black D, Guerri C, Pichini S. The effect of different alcohol drinking patterns in early to mid-pregnancy. *BJOG* 2012;119(13):1670-1671.

(47) Sayal K, Heron J., Golding J., Emond A. Prenatal Alcohol Exposure and Gender Differences in Childhood Mental Health Problems: A Longitudinal Population-Based Study . *PEDIATRICS* 2007;119:426-434.

(48) Sayal K, Draper ES, Fraser R, Barrow M, Davey SG, Gray R. Light drinking in pregnancy and mid-childhood mental health and learning outcomes. *Arch Dis Child* 2013;98(2):107-111.

(49) Achenbach TM, Becker A, Dopfner M et al. Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. *J Child Psychol Psychiatry* 2008;49(3):251-275.

(50) Almaqrami MH, Shuwail AY. Validity of the self-report version of the strengths and difficulties questionnaire in Yemen. *Saudi Med J* 2004;25(5):592-601.

(51) Altendorfer-Kling U. [The self-report version of the Strengths and Difficulties Questionnaire in an Austrian field sample]. *Z Kinder Jugendpsychiatr Psychother* 2007;35(4):265-271.

(52) Alyahri A, Goodman R. Validation of the Arabic Strengths and Difficulties Questionnaire and the Development and Well-Being Assessment. *East Mediterr Health J* 2006;12 Suppl 2:S138-46.:S138-S146.

(53) Borg AM, Kaukonen P, Salmelin R, Joukamaa M, Tamminen T. Reliability of the Strengths and Difficulties Questionnaire among Finnish 4-9-year-old children. *Nord J Psychiatry* 2012.

(54) Cury CR, Golfeto JH. Strengths and difficulties questionnaire (SDQ): a study of school children in Ribeirao Preto. *Rev Bras Psiquiatr* 2003;25(3):139-145.

(55) Du Y, Kou J, Coghill D. The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. *Child Adolesc Psychiatry Ment Health* 2008;2(1):8.

(56) Essau CA, Olaya B, Anastassiou-Hadjicharalambous X et al. Psychometric properties of the Strength and Difficulties Questionnaire from five European countries. *Int J Methods Psychiatr Res* 2012;21(3):232-245.

(57) Hawes DJ, Dadds MR. Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Aust N Z J Psychiatry* 2004;38(8):644-651.

(58) Koskelainen M, Sourander A, Vauras M. Self-reported strengths and difficulties in a community sample of Finnish adolescents. *Eur Child Adolesc Psychiatry* 2001;10(3):180-185.

(59) Lai KY, Luk ES, Leung PW, Wong AS, Law L, Ho K. Validation of the Chinese version of the strengths and difficulties questionnaire in Hong Kong. *Soc Psychiatry Psychiatr Epidemiol* 2009.

(60) Lukumar P, Wijewardana K, Hermansson J, Lindmark G. Validity and reliability of Tamil version of Strengths and Difficulties Questionnaire self-report. *Ceylon Med J* 2008;53(2):48-52.

(61) Malmberg M, Rydell AM, Smedje H. Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nord J Psychiatry* 2003;57(5):357-363.

(62) Mellor D. Normative data for the Strengths and Difficulties Questionnaire in Australia. *Australian Psychologist Vol 40(3)* 2005;pp. 215-222:-222.

(63) Niclasen J, Skovgaard AM, Andersen AM, Somhovd MJ, Obel C. A Confirmatory Approach to Examining the Factor Structure of the Strengths and Difficulties Questionnaire (SDQ): A Large Scale Cohort Study. *J Abnorm Child Psychol* 2012.

(64) Niclasen J, Teasdale TW, Andersen AM, Skovgaard AM, Elberling H, Obel C. Psychometric properties of the Danish strength and difficulties questionnaire: The SDQ assessed for more than 70,000 raters in four different cohorts. *PLoS One* 2012;7(2):e32025.

(65) Ruchkin V, Koposov R, Vermeiren R, Schwab-Stone M. The Strength and Difficulties Questionnaire: Russian validation of the teacher version and comparison of teacher and student reports. *J Adolesc* 2011.

(66) Stone LL, Otten R, Engels RC, Vermulst AA, Janssens JM. Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-Olds: A Review. *Clin Child Fam Psychol Rev* 2010;13(3):254-274.

(67) van Widenfelt BM, Goedhart AW, Treffers PD, Goodman R. Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 2003;12(6):281-289.

(68) Goodman R. A modified version of the Rutter parent questionnaire including extra items on children's strengths: a research note. *Journal of Child Psychology and Psychiatry* 1994;35(8):1483-1494.

(69) Goodman R, Scott S. Comparing the strengths and difficulties questionnaire and the child behavior checklist: is small beautiful? *Journal of Abnormal Child Psychology* 1999;27(1):17-24.

(70) Achenbach TM. Manual for the Child Behavior Checklist/4-18. 1991. Burlington, University of Vermont, Department of Psychiatry.

(71) Anselmi L, Fleitlich-Bilyk B, Menezes AM, Araujo CL, Rohde LA. Prevalence of psychiatric disorders in a Brazilian birth cohort of 11-year-olds. *Soc Psychiatry Psychiatr Epidemiol* 2009.

(72) Becker A, Steinhausen HC, Baldursson G et al. Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. *Eur Child Adolesc Psychiatry* 2006;15 Suppl 1:i56-i62.:i56-i62.

(73) Goodman A, Goodman R. Strengths and difficulties questionnaire as a dimensional measure of child mental health. *J Am Acad Child Adolesc Psychiatry* 2009;48(4):400-403.

(74) Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry* 2000;177:534-9.:534-539.

(75) Field A. *Discovering statistics using SPSS*. 3rd ed. SAGE; 2009.

(76) Kaplan RM, Saccuzzo DP. First edition ed. Wadsworth; 2001.

(77) Gomez-Beneyto M, Nolasco A, Moncho J et al. Psychometric behaviour of the strengths and difficulties questionnaire (SDQ) in the Spanish national health survey 2006. *BMC Psychiatry* 2013;13(1):95.

(78) Hayes L. Problem behaviours in early primary school children: Australian normative data using the Strengths and Difficulties Questionnaire. *Aust N Z J Psychiatry* 2007;41(3):231-238.

(79) Rothenberger A, Becker A, Erhart M, Wille N, Ravens-Sieberer U. Psychometric properties of the parent strengths and difficulties questionnaire in the general population of German children and adolescents: results of the BELLA study. *Eur Child Adolesc Psychiatry* 2008;17 Suppl 1:99-105.

(80) Byrne B. *Structural equation modeling with MPlus*. Routledge; 2011.

(81) Van RB, Veenstra M, Clench-Aas J. Construct validity of the five-factor strengths and difficulties questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry* 2008;49(12):1304-1312.

(82) Goodman A, Lamping DL, Ploubidis GB. When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the strengths and difficulties questionnaire (SDQ): data from british parents, teachers and children. *Journal of Abnormal Child Psychology* 2010;38(8):1179-1191.

(83) Dickey WC, Blumberg SJ. Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. *Journal of the American Academy of Child & Adolescent Psychiatry* 2004;43(9):1159-1167.

(84) Palmieri PA, Smith GC. Examining the structural validity of the strengths and difficulties questionnaire (SDQ) in a U.S. sample of custodial grandmothers. *Psychological Assessment* 2007;19(2):189-198.

(85) Woerner W, Becker A, Rothenberger A. Normative data and scale properties of the German parent SDQ. *Eur Child Adolesc Psychiatry* 2004;13 Suppl 2:II3-10.:II3-10.

(86) Muris P, Meesters C, van den BF. The Strengths and Difficulties Questionnaire (SDQ)--further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *Eur Child Adolesc Psychiatry* 2003;12(1):1-8.

(87) Marzocchi GM, Capron C, Di PM et al. The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. *Eur Child Adolesc Psychiatry* 2004;13 Suppl 2:II40-6.:II40-II46.

(88) Shojaei T, Wazana A, Pitrou I, Kovess V. The strengths and difficulties questionnaire: validation study in French school-aged children and cross-cultural comparisons. *Soc Psychiatry Psychiatr Epidemiol* 2009;44(9):740-747.

(89) Aarhus Birth Cohort. Aarhus Birth Cohort, www.aarhus-born.dk. 2008.

(90) Christensen E. 7 års børneliv. Velfærd, sundhed og trivsel hos børn født i 1995. København: Socialforskningsinstituttet; 2004. Report No.: 04:13.

(91) Elberling H, Linneberg A, Olsen EM, Goodman R, Skovgaard AM. The prevalence of SDQ-measured mental health problems at age 5-7 years and identification of predictors from birth to preschool age in a Danish birth cohort: The Copenhagen Child Cohort 2000. *European Child & Adolescent Psychiatry* 2010;19(9):725-735.

(92) Olsen J, Melbye M, Olsen SF et al. The Danish National Birth Cohort--its background, structure and aim. *Scand J Public Health* 2001;29(4):300-307.

(93) Skovgaard AM, Olsen EM, Houmann T et al. The Copenhagen County child cohort: design of a longitudinal study of child mental health. *Scand J Public Health* 2005;33(3):197-202.

(94) Kirkegaard I, Obel C, Hedegaard M, Henriksen TB. Gestational age and birth weight in relation to school performance of 10-year-old children: a follow-up study of children born after 32 completed weeks. *PEDIATRICS* 2006;118(4):1600-1606.

(95) Obel C, Linnet KM, Henriksen TB et al. Smoking during pregnancy and hyperactivity-inattention in the offspring--comparing results from three Nordic cohorts. *Int J Epidemiol* 2008.

(96) Rodriguez A, Jarvelin MR, Obel C et al. Do inattention and hyperactivity symptoms equal scholastic impairment? Evidence from three European cohorts. *BMC Public Health* 2007;7(147):327.

(97) Vestergaard M, Wisborg K, Henriksen TB, Secher NJ, Ostergaard JR, Olsen J. Prenatal exposure to cigarettes, alcohol, and coffee and the risk for febrile seizures. *PEDIATRICS* 2005;116(5):1089-1094.

(98) Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry* 2001;40(11):1337-1345.

(99) Obel C, Heiervang E, Rodriguez A et al. The Strengths and Difficulties Questionnaire in the Nordic countries. *Eur Child Adolesc Psychiatry* 2004;13 Suppl 2:II32-9.:II32-II39.

(100) SDQ. SDQ. 2010.

(101) Kline RB. *Principles and practice of structural equation modeling*. Third ed. Guilford; 2011.

(102) Sanne B, Torsheim T, Heiervang E, Stormark KM. The Strengths and Difficulties Questionnaire in the Bergen Child Study: a conceptually and methodically motivated structural analysis. *Psychol Assess* 2009;21(3):352-364.

(103) Van RB, Veenstra M, Clench-Aas J. Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *J Child Psychol Psychiatry* 2008;49(12):1304-1312.

(104) Jöreskog KG. Testing structural equation models. In: Bollen KA, Long JS, editors. *Testing structural equation models*.Newsbury Park: CA: Sage; 1993. 294-316.

(105) Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 1981;18(1):39-50.

(106) Kaplan RM, Saccuzzo DP. Validity. *Psychological Testing. Principles, Applications, and Issues*. First edition ed. Wadsworth; 2001. 130-152.

(107) Campell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 1959;56:81-105.

(108) Nunnally JC. Validity. *Psychometric Theory*. Second edition ed. McGraw-Hill Book Company; 1978. 86-116.

(109) Nielsen S, Nielsen B. Introduction to Structural equation modeling (SEM). 2011.

(110) Niclasen J. SDQ Danish norms. 2012.

(111) Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 1997;38(5):581-586.

(112) Ellis PD. *The essential guide to effect sizes. statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press; 2010.

(113) Heiervang E, Stormark KM, Lundervold AJ et al. Psychiatric disorders in Norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. *J Am Acad Child Adolesc Psychiatry* 2007;46(4):438-447.

(114) Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz DS. The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *J Am Acad Child Adolesc Psychiatry* 2005;44(6):557-564.

(115) Goodman R, Neves dos SD, Robatto Nunes AP, Pereira de MD, Fleitlich-Bilyk B, Almeida FN. The Ilha de Mare study: a survey of child mental health problems in a predominantly African-Brazilian rural community. *Soc Psychiatry Psychiatr Epidemiol* 2005;40(1):11-17.

(116) Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *J Abnorm Child Psychol* 1999;27(1):17-24.

(117) Klasen H, Woerner W, Wolke D et al. Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *Eur Child Adolesc Psychiatry* 2000;9(4):271-276.

(118) Mellor D. Furthering the use of the strengths and difficulties questionnaire: reliability with younger child respondents. *Psychol Assess* 2004;16(4):396-401.

(119) Smedje H, Broman JE, Hetta J, von Knorring AL. Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". *Eur Child Adolesc Psychiatry* 1999;8(2):63-70.

(120) Jacobsen TN, Nohr EA, Frydenberg M. Selection by socioeconomic factors into the Danish National Birth Cohort. *European Journal of Epidemiology* 2010;25(5):349-355.

(121) Meltzer H, Gatward R, Goodman R, Ford T. Mental health of children and adolescents in Great Britain. *Int Rev Psychiatry* 2003;15(1-2):185-187.

(122) Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? *Epidemiology* 2006;17(4):413-418.

(123) Jones KL, Smith DW. Recognition of the fetal alcohol syndrome in early infancy. *Lancet* 1973;302(7836):999-1001.

(124) Jones KL, Smith DW, Ulleland CN, Streissguth P. Pattern of malformation in offspring of chronic alcoholic mothers. *Lancet* 1973;1(7815):1267-1271.

(125) Jones KL, Smith DW, Streissguth AP, Myrianthopoulos NC. Outcome in offspring of chronic alcoholic women. *Lancet* 1974;1(7866):1076-1078.

(126) Calhoun F, Warren K. Fetal alcohol syndrome: historical perspectives. *Neurosci Biobehav Rev* 2007;31(2):168-171.

(127) George MD. *London life in the eighteen century*. 1951.

(128) Randall CL. Alcohol and pregnancy: highlights from three decades of research. *J Stud Alcohol* 2001;62(5):554-561.

(129) Warner RH, Rosett HL. The effects of drinking on offspring: an historical survey of the American and British literature. *J Stud Alcohol* 1975;36(11):1395-1420.

(130) Mattson SN, Riley EP. A review of the neurobehavioral deficits in children with fetal alcohol syndrome or prenatal exposure to alcohol. *Alcohol Clin Exp Res* 1998;22(2):279-294.

(131) MacNicholl TA. A study of the effects of alcohol on schoold children. *Quarterly Journal of Inebriety* 1905;27:113-117.

(132) Jones KL. FASD from a historical perspective. Barcelona, Spain 2012.

(133) J Henderson, R Gray, P Brocklehurst. Systematic review of effects of low–moderate prenatal alcohol exposure on pregnancy outcome. 2007.

(134) Kelly YJ, Sacker A, Gray R et al. Light drinking during pregnancy: still no increased risk for socioemotional difficulties or cognitive deficits at 5 years of age? *J Epidemiol Community Health* 2010.

(135) O'Leary CM. Fetal alcohol syndrome: diagnosis, epidemiology, and developmental outcomes. *J Paediatr Child Health* 2004;40(1-2):2-7.

(136) Stratton K, Howe C, Battaglia F. *Fetal Alcohol Syndrome. Diagnosis, Epidemiology, Prevention and Treatment*. Washington, D. C.: National Academy Press; 1996.

(137) Rasmussen C, Andrew G, Zwaigenbaum L, Tough S. Neurobehavioural outcomes of children with fetal alcohol spectrum disorders: A Canadian perspective. *Paediatr Child Health* 2008;13(3):185-191.

(138) World Health Organisation. *The ICD-10 classification of mental and behavioural disorders diagnostic criteria for research*. 1993.

(139) American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, DSM-IV*. Washington DC: 1994.

(140) Astley SJ. Comparison of the 4-digit diagnostic code and the Hoyme diagnostic guidelines for fetal alcohol spectrum disorders. *PEDIATRICS* 2006;118(4):1532-1545.

(141) American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, DSM-V*. Washington DC: 2013.

(142) Abel EL. Neurodevelopmental abnormalities. Structural and sensory. *Fetal alcohol abuse syndrome*.New York: Plenum; 1998. 101-110.

(143) Autti-Ramo I, Korkman M, Hilakivi-Clarke L, Lehtonen M, Halmesmaki E, Granstrom ML. Mental development of 2-year-old children exposed to alcohol in utero. *J Pediatr* 1992;120(5):740-746.

(144) Irner TB. Substance exposure in utero and developmental consequences in adolescence: A systematic review. *Child Neuropsychol* 2011.

(145) Spohr HL, Willms J, Steinhausen HC. Prenatal alcohol exposure and long-term developmental consequences. *Lancet* 1993;341(8850):907-910.

(146) Abel EL. Neurodevelopmental abnormalities. Behavioural and cognitive. *Fetal alcohol abuse syndrome*.New York: Plenum; 1998. 111-138.

(147) Kodituwakku PW, Kalberg W, May PA. The effects of prenatal alcohol exposure on executive functioning. *Alcohol Res Health* 2001;25(3):192-198.

(148) Mattson SN, Goodman AM, Caine C, Delis DC, Riley EP. Executive functioning in children with heavy prenatal alcohol exposure. *Alcohol Clin Exp Res* 1999;23(11):1808-1815.

(149) Streissguth AP, Barr HM, Sampson PD, Bookstein FL. Prenatal alcohol and offspring development: the first fourteen years. *Drug Alcohol Depend* 1994;36(2):89-99.

(150) Autti-Ramo I. Twelve-year follow-up of children exposed to alcohol in utero. *Dev Med Child Neurol* 2000;42(6):406-411.

(151) Streissguth AP, Sampson PD, Olson HC et al. Maternal drinking during pregnancy: attention and short-term memory in 14-year-old offspring--a longitudinal prospective study. *Alcohol Clin Exp Res* 1994;18(1):202-218.

(152) Jacobson JL, Dodge NC, Burden MJ, Klorman R, Jacobson SW. Number processing in adolescents with prenatal alcohol exposure and ADHD: differences in the neurobehavioral phenotype. *Alcohol Clin Exp Res* 2011;35(3):431-442.

(153) Spohr HL, Steinhausen HC. Follow-up studies of children with fetal alcohol syndrome. *Neuropediatrics* 1987;18(1):13-17.

(154) Spohr HL, Willms J, Steinhausen HC. Fetal alcohol spectrum disorders in young adulthood. *J Pediatr* 2007;150(2):175-179.

(155) Streissguth AP, Martin DC, Martin JC, Barr HM. The Seattle longitudinal prospective study on alcohol and pregnancy. *Neurobehav Toxicol Teratol* 1981;3(2):223-233.

(156) Spohr HL, Willms J, Steinhausen HC. The fetal alcohol syndrome in adolescence. *Acta Paediatr Suppl* 1994;404:19-26.

(157) Streissguth AP, O'Malley K. Neuropsychiatric implications and long-term consequences of fetal alcohol spectrum disorders. *Semin Clin Neuropsychiatry* 2000;5(3):177-190.

(158) Streissguth AP, Barr HM, Kogan J, Bookstein FL. Understanding the occurence of secondary disabilities in clients with fetal alcohol syndrome (FAS) and fetal alcohol effcts (FAE): Final report. Seattle: WA: University of Washington school of medicine. Dept. Psychiatry and behavioural sciences.; 1996.

(159) Henderson J, Gray R, Brocklehurst P. Systematic review of effects of low-moderate prenatal alcohol exposure on pregnancy outcome. *BJOG* 2007;114(3):243-252.

(160) Sayal K. Alcohol consumption in pregnancy as a risk factor for later mental health problems. *Evid Based Ment Health* 2007;10(4):98-100.

(161) Streissguth AP, Barr HM, Sampson PD, Parrish-Johnson JC, Kirchner GL, Martin DC. Attention, distraction and reaction time at age 7 years and prenatal alcohol exposure. *Neurobehav Toxicol Teratol* 1986;8(6):717-725.

(162) Day NL, Helsel A, Sonon K, Goldschmidt L. The Association Between Prenatal Alcohol Exposure and Behavior at 22 Years of Age. *Alcohol Clin Exp Res* 2013.

(163) Eriksen HLF, Mortensen EL, Kilburn T et al. The effects of low to moderate prenatal alcohol exposure in early pregnancy on IQ in 5-year-old children. *BJOG* 2012;119(10):1191-1200.

(164) Robinson M, Oddy WH, McLean NJ et al. Low-moderate prenatal alcohol exposure and risk to child behavioural development: a prospective cohort study. *BJOG* 2010;117(9):1139-1150.

(165) O'Leary CM, Bower C, Zubrick SR, Geelhoed E, Kurinczuk JJ, Nassar N. A new method of prenatal alcohol classification accounting for dose, pattern and timing of exposure: improving our ability to examine fetal effects from low to moderate alcohol. *J Epidemiol Community Health* 2010;64(11):956-962.

(166) Falgreen Eriksen HL, Mortensen EL, Kilburn T et al. The effects of low to moderate prenatal alcohol exposure in early pregnancy on IQ in 5-year-old children. *BJOG* 2012;119(10):1191-1200.

(167) Kesmodel US, Eriksen HL, Underbjerg M et al. The effect of alcohol binge drinking in early pregnancy on general intelligence in children. *BJOG* 2012;119(10):1222-1231.

(168) Flak AL, Su S, Bertrand J, Denny CH, Kesmodel US, Cogswell ME. The Association of Mild, Moderate, and Binge Prenatal Alcohol Exposure and Child Neuropsychological Outcomes: A Meta-Analysis. *Alcohol Clin Exp Res* 2013.

(169) Kesmodel U. Binge drinking in pregnancy--frequency and methodology. *Am J Epidemiol* 2001;154(8):777-782.

(170) Barr HM, Bookstein FL, O'Malley KD, Connor PD, Huggins JE, Streissguth AP. Binge drinking during pregnancy as a predictor of psychiatric disorders on the Structured Clinical Interview for DSM-IV in young adult offspring. *Am J Psychiatry* 2006;163(6):1061-1065.

(171) Sampson PD, Streissguth AP, Barr HM, Bookstein FL. Neurobehavioral effects of prenatal alcohol: Part II. Partial least square analyses. *Neurotoxicol Teratol* 1989;11:477-491.

(172) Alvik A, Aalen OO, Lindemann R. Early Fetal Binge Alcohol Exposure Predicts High Behavioral Symptom Scores in 5.5-Year-Old Children. *Alcohol Clin Exp Res* 2013.

(173) O'Leary CM, Nassar N, Zubrick SR, Kurinczuk JJ, Stanley F, Bower C. Evidence of a complex association between dose, pattern and timing of prenatal alcohol exposure and child behaviour problems. *Addiction* 2010;105(1):74-86.

(174) O'Leary CM, Bower C. Guidelines for pregnancy: what's an acceptable risk, and how is the evidence (finally) shaping up? *Drug Alcohol Rev* 2012;31(2):170-183.

(175) Kalter H. Teratology in the 20th century: environmental causes of congenital malformations in humans and how they were established. *Neurotoxicol Teratol* 2003;25(2):131-282.

(176) Guerri C, Bazinet A, Riley EP. Foetal Alcohol Spectrum Disorders and alterations in brain and behaviour. *Alcohol Alcohol* 2009;44(2):108-114.

(177) Plant ML, Abel EL, Guerri C. Alcohol and pregnancy. In: MacDonald I, editor. *Health issues related to alcohol consumption*. 2nd ed. ILSI Europe; 1999. 181-214.

(178) Ikonomidou C, Bittigau P, Ishimaru MJ et al. Ethanol-induced apoptotic neurodegeneration and fetal alcohol syndrome. *Science* 2000;287(5455):1056-1060.

(179) Schneider ML, Moore CF, Becker EF. Timing of moderate alcohol exposure during pregnancy and neonatal outcome in rhesus monkeys (Macaca mulatta). *Alcohol Clin Exp Res* 2001;25(8):1238-1245.

(180) Riley EP, Lochry EA, Shapiro NR. Lack of response inhibition in rats prenatally exposed to alcohol. *Psychopharmacology (Berl)* 1979;62(1):47-52.

(181) Driscoll CD, Streissguth AP, Riley EP. Prenatal alcohol exposure: comparability of effects in humans and animal models. *Neurotoxicol Teratol* 1990;12(3):231-237.

(182) Abel EL. *Fetal alcohol abuse syndrome*. New York: Plenum; 1998.

(183) Chen WJ, Maier SE, Parnell SE, West JR. Alcohol and the developing brain: neuroanatomical studies. *Alcohol Res Health* 2003;27(2):174-180.

(184) West JR. Fetal alcohol-induced brain damage and the problem of determining temporal vulnerability: a review. *Alcohol Drug Res* 1987;7(5-6):423-441.

(185) Burke MW, Palmour RM, Ervin FR, Ptito M. Neuronal reduction in frontal cortex of primates after prenatal alcohol exposure. *Neuroreport* 2009;20(1):13-17.

(186) Preece PM. Introduction. In: Preece PM, Riley EP, editors. *Alcohol, drugs and medication in pregnancy: the long-term outcome for the child*. 1st ed. Mac Keith Press; 2010. 1-11.

(187) Olsen J, Melbye M, Olsen SF et al. The Danish national birth cohort--its background, structure and aim. *Scandinavian Journal of Public Health* 2001;29(4):300-307.

(188) Kesmodel U. Binge drinking in pregnancy--frequency and methodology. *Am J Epidemiol* 2001;154(8):777-782.

(189) Kesmodel U, Frydenberg M. Binge drinking during pregnancy--is it possible to obtain valid information on a weekly basis? *Am J Epidemiol* 2004;159(8):803-808.

(190) Niclasen J. SDQ Danish cut off scores: 5-7-year-olds. 18-12-2012.

(191) Juul S. *Epidemiologi og evidens*. 1st ed. Munksgaard; 2008.

(192) Rothman KJ. *Epidemiology. An intoduction*. Oxford University Press; 2002.

(193) Porta M. *A dictionary of epidemiology*. Oxford University Press; 2008.

(194) Munk-Jorgensen P, Mortensen PB. The Danish Psychiatric Central Register. *Dan Med Bull* 1997;44(1):82-84.

(195) Statistics Denmark. IDA: an integrated database for labour market research. 1991.

(196) Field A. Moderation, mediation and more regression. *Discovering statistics using IBM SPSS statistics*. 4th ed. SAGE; 2013. 392-428.

(197) Schechter DS, Willheim E. Disturbances of attachment and parental psychopathology in early childhood. *Child Adolesc Psychiatr Clin N Am* 2009;18(3):665-686.

(198) Svane V, Mogensen J. Mønsterbrud trods traumer i hjerne og sind. In: Bach AS, Ejrnæs M, et al, editors. *Mønsterbrud i opbrud*. 1 ed. Dafolo Forlag; 2010. 50-83.

(199) Strandberg-Larsen K, Rod NN, Nybo Andersen AM, Olsen J, Gronbaek M. Characteristics of women who binge drink before and after they become aware of their pregnancy. *Eur J Epidemiol* 2008;23(8):565-572.

(200) Rugulies R, Aust B, Syme SL. Epidemiology of health and illness: a socio-psycho-phisiological perspective. *The SAGE handbook of health psychology*. 2013. 27-68.

(201) Bowlby J. Research into the Origins of Delinquent Behaviour. *Br Med J* 1950;1(4653):570-573.

(202) Ainsworth MD, Bell SM. Attachment, exploration, and separation: illustrated by the behavior of one-year-olds in a strange situation. *Child Dev* 1970;41(1):49-67.

(203) Sroufe LA. Attachment and development: a prospective, longitudinal study from birth to adulthood. *Attach Hum Dev* 2005;7(4):349-367.

(204) Mikulincer M, Shaver PR. *Attachment in adulthood. Structure, dynamics and change*. 1st ed. The Guildford Press; 2007.

(205) Mikulincer M, Shaver PR. Attachment orientations, behavioural self-regulation, and personal growth. *Attachment in adulthood. Structure, dynamics and change*. 1st ed. The Guildford Press; 2007. 219-250.

(206) Streissguth AP, Barr HM, Martin DC. Alcohol exposure in utero and functional deficits in children during the first four years of life. *Ciba Found Symp* 1984;105:176-96.:176-196.

(207) Streissguth AP, Barr HM, Sampson PD, Darby BL. IQ at age 4 in relation to maternal alcohol use and smoking during pregnancy. *Developmental Psychology Vol 25(1)* 1989;pp. 3-11:-11.

(208) Irner TB, Teasdale TW, Olofsson M. Cognitive and social development in preschool children born to women using substances. *J Addict Dis* 2012;31(1):29-44.

(209) Irner TB, Teasdale TW, Nielsen T, Vedal S, Olofsson M. Substance use during pregnancy and postnatal outcomes. *J Addict Dis* 2012;31(1):19-28.

(210) Jacobson SW, Jacobson JL, Sokol RJ, Martier SS, Ager JW. Prenatal alcohol exposure and infant information processing ability. *Child Dev* 1993;64(6):1706-1721.

(211) Andersen SL. Trajectories of brain development: point of vulnerability or window of opportunity? *Neurosci Biobehav Rev* 2003;27(1-2):3-18.

(212) Rice D, Barone S Jr. Critical periods of vulnerability for the developing nervous system: evidence from humans and animal models. *Environ Health Perspect* 2000;108 Suppl 3:511-533.

(213) Bayer SA, Altman J, Russo RJ, Zhang X. Timetables of neurogenesis in the human brain based on experimentally determined patterns in the rat. *Neurotoxicology* 1993;14(1):83-144.

(214) Kodituwakku PW, May PA, Clericuzio CL, Weers D. Emotion-related learning in individuals prenatally exposed to alcohol: an investigation of the relation between set shifting, extinction of responses, and behavior. *Neuropsychologia* 2001;39(7):699-708.

(215) Powell G. Danish studies suggesting low and moderate prenatal alcohol exposure has no adverse effects on children aged 5 years did not use appropriate or effective measures of executive functioning. *BJOG* 2012;119(13):1669-1670.

(216) Goodman R, Ford T, Corbin T, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *Eur Child Adolesc Psychiatry* 2004;13 Suppl 2:II25-31.:II25-II31.

(217) Rimvall MK, Jeppesen P, Niclasen J, Skovgaard AM, et al. The rpedictive validity of the SDQ in preschool children with regards to ADHD diagnoses in school age - a register-based study. *In preparation* 2013.

(218) Parker MO, Brennan CH. Low and moderate alcohol consumption during pregnancy: effects on social behaviour and propensity to develop substance abuse in later life. *BJOG* 2012;119(13):1671-1672.

(219) Astley S, Grant T. Another perspective on 'the effect of different alcohol drinking patterns in early to mid pregnancy on the child's intelligence, attention, and executive function'. *BJOG* 2012;119(13):1672-1675.

(220) Astley SJ. Profile of the first 1,400 patients receiving diagnostic evaluations for fetal alcohol spectrum disorder at the Washington State Fetal Alcohol Syndrome Diagnostic & Prevention Network. *Can J Clin Pharmacol* 2010;17(1):e132-e164.

(221) Rutter M, Pickles A, Murray R, Eaves L. Testing hypotheses on specific environmental causal effects on behavior. *Psychol Bull* 2001;127(3):291-324.

(222) D'Onofrio BM, Van Hulle CA, Waldman ID, Rodgers JL, Rathouz PJ, Lahey BB. Causal inferences regarding prenatal alcohol exposure and childhood externalizing problems. *Arch Gen Psychiatry* 2007;64(11):1296-1304.

Appendix A: Response frequencies for each of the 25 SDQ items

For the pooled 5-7-year-old parent sample (Total N = 56.764)

	Total			**Boys**			**Girls**		
	Not true	Somewhat true	Certainly true	Not true	Somewhat true	Certainly true	Not true	Somewhat true	Certainly true
Hyperactivity									
Restless, overactive..	73.0	20.5	6.5	68.3	23.6	8.0	77.9	17.2	5.0
Constantly fidgeting..	76.2	18.0	5.8	71.4	21.2	7.5	81.3	14.6	4.1
Easily distracted…	62.1	29.5	8.4	56.3	33.2	10.5	68.1	25.7	6.2
Thinks things out …	6.6	59.0	34.4	8.1	61.5	30.4	5.0	56.4	38.5
Sees tasks through…	6.8	43.5	49.8	8.5	48.6	42.9	5.0	38.1	56.9
Emotional									
Often complains of…	81.1	15.5	3.5	84.5	12.8	2.8	77.5	18.3	4.2
Many worries…	68.3	25.0	6.7	68.6	24.5	7.0	68.1	25.5	6.4
Often unhappy…	77.9	18.7	3.4	78.2	18.3	3.5	77.5	19.1	3.4
Nervous or clingy…	57.7	32.5	9.8	58.0	31.4	10.6	57.5	33.6	8.9
Many fears…	79.7	16.9	3.3	79.5	17.0	3.5	79.9	16.9	3.1
Conduct									
Often has temper…	57.1	32.9	9.9	56.7	33.1	10.2	57.6	32.8	9.6
Generally obedient…	2.7	38.6	58.7	3.0	39.8	57.2	2.4	37.3	60.3
Often fights with…	95.6	3.8	0.6	93.4	5.8	0.8	97.9	1.8	0.3

Often lies or cheat	83.9	14.6	1.5	82.6	15.8	1.7	85.3	13.4	1.3
Steals from home…	98.1	1.6	0.3	97.9	1.8	0.3	98.2	1.5	0.3
Peer Problems									
Rather solitary…	81.4	14.5	4.1	78.6	16.5	4.9	84.4	12.4	3.2
Has at least one…	3.6	4.2	92.2	3.7	4.7	91.6	3.6	3.6	92.8
Generally liked…	1.5	8.5	90.0	1.6	9.8	88.5	1.3	7.0	91.6
Picked on or…	89.3	9.1	1.6	88.0	10.2	1.8	90.7	7.9	1.4
Gets on better…	87.1	9.7	3.2	85.7	10.5	3.7	88.6	8.8	2.6
Prosocial									
Considerate of others	0.7	20.4	78.9	1.0	23.5	75.5	0.4	17.1	82.5
Shares readily…	2.2	26.0	71.8	2.6	28.1	69.3	1.8	23.7	74.5
Helpful if someone…	2.1	28.0	69.9	2.9	35.9	61.1	1.2	19.7	79.1
Kind to younger …	1.2	10.1	88.6	1.5	12.6	85.8	0.9	7.5	91.6
Often volunteers…	8.8	47.8	43.4	11.6	52.1	36.2	5.8	43.3	50.9

Appendix B: Factor loadings between the scales for the separate parent and teacher samples

For the modified Models 1 and 2 and non-modified Model 3 as presented in Paper 2

	Parent-rated SDQ				**Teacher-rated SDQ**			
	Younger girls	Younger boys	Older girls	Older boys	Younger girls	Younger boys	Older girls	Older boys
Emotional/ Hyperactivity	0.42	0.43	0.57	0.57	0.57	0.39	0.58	0.64
Emotional/ Conduct	0.55	0.56	0.71	0.71	0.63	0.39	0.63	0.67
Emotional/ Peer problems	0.65	0.67	0.74	0.74	0.76	0.72	0.82	0.79
Emotional/ Prosocial	-0.31	-0.31	-0.34	-0.34	-0.41	-0.38	-0.52	-0.50
Hyperactivity/ Conduct	0.70	0.72	0.76	0.76	0.85	0.80	0.87	0.89
Hyperactivity/ Peer problems	0.46	0.52	0.56	0.56	0.61	0.49	0.61	0.66
Hyperactivity/ Prosocial	-0.42	-0.46	-0.40	-0.40	-0.68	-0.72	-0.70	-0.73
Conduct/ Peer problems	0.59	0.67	0.69	0.69	0.65	0.68	0.75	0.73
Conduct/ Prosocial	-0.75	-0.79	-0.73	-0.73	-0.91	-0.88	-0.86	-0.89
Peer problems/ Prosocial	-0.41	-0.46	-0.48	-0.48	-0.56	-0.57	-0.63	-0.63
Externalising/ Hyperactivity	0.69	0.71	0.75	0.75	0.84	0.81	0.86	0.89
Externalising/ Conduct	1.04	1.04	1.05	1.05	1.03	1.00	1.03	1.02

Internalising/ Emotional	0.77	0.74	0.86	0.86	0.84	0.68	0.85	0.85
Internalising/ Peer problems	0.86	0.91	0.86	0.86	0.92	1.06	0.98	0.94
Externalising/ Internalising	0.72	0.75	0.81	0.81	0.74	0.61	0.74	0.79
Externalising/ Prosocial	-0.70	-0.74	-0.67	-0.67	-0.87	-0.88	-0.84	-0.87
Internalising/ Prosocial	-0.46	-0.50	-0.50	-0.50	-0.59	-0.53	-0.64	-0.66
Total difficulties/ Hyperactivity	0.75	0.69	0.70	0.70	0.90	0.86	0.91	0.92
Total difficulties/ Emotional	0.65	0.66	0.68	0.68	0.75	0.65	0.83	0.80
Total difficulties/ Conduct	0.98	0.96	0.98	0.98	0.99	0.97	0.99	0.99
Total difficulties/ Peer problems	0.80	0.86	0.77	0.77	0.88	0.90	0.96	0.92
Total difficulties/ Prosocial	-0.71	-0.56	-0.57	-0.57	-0.89	-0.89	-0.90	-0.89

Appendix C (1): SDQ cut-off scores for the 5-7-year-old parent ratings

The *"normal"* banding includes app. 80 % of the sample, the *"borderline"* includes app. 10 % of the sample, and the *"clinical"* banding includes app. 10 % of the sample

Parents	**Recommended bandings for boys (N = 28,920)**						**Recommended bandings for girls (N = 27,611)**						**Recommended bandings full (N = 56,531)**					
5-7-year-olds	Normal		Borderline		Clinical		Normal		Borderline		Clinical		Normal		Borderline		Clinical	
Scale	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact
Total dif.	0-9	79.6	10-13	11.7	14-40	8.4	0-8	81.1	9-11	10.6	12-40	8.4	0-9	82.5	10-12	8.9	13-40	8.5
Externalising	0-6	81.1	7-8	9.5	9-20	9.4	0-5	82.8	6-7	9.8	8-20	7.5	0-5	78.1	6-7	11.5	8-20	10.5
Internalising	0-4	82.7	5-6	9.5	7-20	7.8	0-4	84.5	5	5.7	6-20	9.9	0-4	83.6	5	5.6	6-20	10.8
Hyperactivity	0-4	79.7	5-6	12.8	7-10	7.5	0-3	80.6	4-5	13.2	6-10	6.2	0-4	84.0	5	6.8	6-10	9.2
Conduct	0-2	83.9	3	9.2	4-10	6.9	0-2	86.9	3	8.5	4-10	4.7	0-2	85.4	3	8.8	4-10	5.8
Emotional	0-3	86.0	4	6.4	5-10	7.6	0-3	85.4	4	6.7	5-10	7.9	0-3	85.7	4	6.6	5-10	7.7
Peer-probs.	0-1	78.0	2	11.4	3-10	10.6	0-1	82.2	2	10.7	3-10	7.1	0-1	80.1	2	11.1	3-10	8.9
Prosocial	7-10	83.0	6	9.3	0-5	7.8	8-10	81.9	7	9.5	0-6	8.5	8-10	75.0	7	12.1	0-6	12.9
Impact	0	88.8	1	4.1	2-10	7.1	0	92.9	1	2.8	2-10	4.3	0	91.4	1	3.8	2-10	4.8

Appendix C (2): SDQ cut-off scores for the 5-year-old teacher ratings

The *"normal"* banding includes app. 80 % of the sample, the *"borderline"* includes app. 10 % of the sample, and the *"clinical"* banding includes app. 10 % of the sample

Teachers	**Recommended bandings for boys (N = 1,272)**						**Recommended bandings for girls (N = 1,291)**						**Recommended bandings full (N = 2,542)**					
5-year-olds	Normal		Borderline		Clinical		Normal		Borderline		Clinical		Normal		Borderline		Clinical	
Scale	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact
Total difficulties	0-11	81.7	12-15	9.9	16-40	8.4	0-8	82.3	9-11	8.6	12-40	9.1	0-9	80.8	10-13	9.6	14-40	9.7
Externalising	0-7	79.5	8-10	10.0	11-20	10.6	0-4	80.5	5-7	11.5	8-20	8.1	0-6	82.4	7-9	8.5	10-20	9.2
Internalising	0-4	83.8	5-6	7.0	7-20	9.2	0-4	82.2	5-6	9.5	7-20	8.3	0-4	83.0	5-6	8.2	7-20	8.7
Hyperactivity	0-5	81.1	6-7	8.9	8-10	10.1	0-3	81.9	4-5	10.2	6-10	7.9	0-4	80.8	5-6	9.6	7-10	9.7
Conduct	0-2	80.7	3-4	12.3	5-10	6.9	0-1	80.3	2	9.8	3-10	9.8	0-2	85.6	3	5.6	4-10	8.8
Emotional	0-2	81.1	3-4	12.0	5-10	7.0	0-2	77.5	3-4	15.1	5-10	7.4	0-2	79.3	3-4	13.5	5-10	7.2
Peer problems	0	82.0	1	9.3	2-6	8.7	0	89.7	1	5.1	2-6	5.2	0-1	77.2	2-3	14.7	4-10	8.2
Prosocial	5-10	85.8	4	6.1	0-3	8.2	7-10	83.4	6	6.4	0-5	10.2	6-10	82.5	5	8.0	0-4	9.4
Impact	0	82.0	1	9.3	2-10	8.7	0	89.7	1	5.1	2-10	5.2	0	85.9	1	7.2	2-10	2.9

Appendix D (1): SDQ means and SDs for the 5-7-year-old parent sample

Parents	**5-7-year-old boys (N = 28,920)**		**5-7-year-old girls (N = 27,611)**	
	Mean	SD	Mean	SD
Total difficulties	6.42	4.79	5.45	4.17
Externalising	4.01	3.19	3.17	2.74
Internalising	2.41	2.56	2.29	2.36
Hyperactivity	2.73	2.31	2.02	1.98
Conduct	1.28	1.35	1.14	1.22
Emotional	1.59	1.75	1.66	1.76
Peer problems	0.83	1.38	0.64	1.15
Prosocial	8.08	1.66	8.68	1.41
Impact	0.29	1.01	0.15	0.69

Appendix D (2): SDQ means and SDs for the 5-year-old teacher sample

Teachers	**5-year-old boys (N = 1,272)**		**5-year-old girls (N = 1,291)**	
	Mean	SD	Mean	SD
Total difficulties	6,50	5,88	4,69	4,99
Externalising	4,20	4,11	2,45	3,23
Internalising	2,30	2,97	2,23	2,73
Hyperactivity	2,95	2,85	1,70	2,33
Conduct	1,25	1,74	0,76	0,35
Emotional	1,31	1,76	1,41	1,79
Peer problems	0,99	1,71	0,83	1,47
Prosocial	7,19	2,37	8,34	1,90
Impact	0,54	0,78	0,32	0,62

Appendix E: SDQ frequency distributions for 5-7-year-olds

Distributions of scores for the *Total difficulties, Externalising, Internalising, Hyperactivity, Conduct, Emotional, Peer problems, Prosocial scales* and *Impact scores* for the 5-7-year-olds parent ratings and 5-year-old teacher ratings

Total difficulties

Total difficulties	**Parents (5-7-year-olds)**				**Teachers (5-year-olds)**			
	Boys (N = 28,920)		**Girls (N = 27,611)**		**Boys (N = 1,272)**		**Girls (N = 1,291)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	4.2	4.3	5.9	5.9	10.3	10.3	17.9	17.9
1	6.2	10.7	8.6	14.5	8.3	18.6	13.7	31.6
2	8.9	19.8	11.0	25.4	11.3	29.9	10.9	42.5
3	9.8	29.9	12.0	37.4	9.5	39.4	9.8	52.3
4	10.5	40.6	11.3	48.7	7.5	46.9	8.5	60.8
5	10.0	50.9	10.3	59.1	7.5	54.5	7.6	68.4
6	8.8	59.9	8.9	67.9	6.1	60.6	6.0	74.4
7	7.8	67.8	7.3	75.2	5.9	66.5	4.4	78.8
8	6.4	74.4	5.9	81.1	5.0	71.5	3.5	82.3
9	5.1	79.6	4.4	85.5	3.7	75.2	3.7	86.0
10	4.0	83.8	3.4	88.9	4.0	79.2	2.9	88.8
11	3.1	86.9	2.8	91.6	2.4	81.7	2.0	90.9
12	2.6	89.5	1.9	93.5	3.1	84.7	1.3	92.2
13	2.0	91.6	1.5	95.1	2.5	87.3	1.0	93.2
14	1.7	93.3	1.2	96.3	2.1	89.4	0.6	93.8

15	1.3	94.7	0.9	97.1	2.2	91.6	1.2	95.0
16	1.0	95.7	0.7	97.9	0.9	92.5	1.0	96.0
17	0.8	95.5	0.6	98.4	1.0	93.6	1.1	97.1
18	0.7	97.2	0.3	98.7	0.9	94.5	0.6	97.8
19	0.5	97.8	0.3	99.0	1.3	95.8	0.5	98.2
20-40	2.2	100.0	1.0	100.0	4.2	100.0	1.8	100.0

Externalising

Externalising	**Parents (5-7-year-olds)**				**Teachers (5-year-olds)**			
	Boys (N = 28,920)		**Girls (N = 27,611)**		**Boys (N = 1,272)**		**Girls (N = 1,291)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	10.9	10.9	15.2	15.2	20.4	20.4	39.3	39.3
1	12.2	23.1	16.5	31.7	13.1	33.5	14.5	53.8
2	14.2	37.2	16.6	48.3	11.6	45.1	11.1	64.9
3	14.0	51.2	14.7	62.9	9.7	54.8	7.7	72.6
4	12.4	63.6	11.4	74.3	6.8	61.6	7.9	80.5
5	10.0	73.6	8.4	82.8	8.0	69.7	4.4	84.9
6	7.5	81.1	5.8	88.5	5.9	75.6	4.2	89.1
7	5.5	86.6	4.0	92.5	3.9	79.5	2.9	91.9
8	4.0	90.6	2.6	95.0	3.9	83.3	1.9	93.8
9	2.7	93.4	1.8	96.9	3.4	86.7	0.9	94.7
10	2.0	95.3	1.1	98.0	2.7	89.4	1.9	96.7

11	1.5	96.9	0.8	98.7	3.1	92.5	1.0	97.7
12	1.1	98.0	0.5	99.2	2.4	94.9	0.4	98.1
13	0.8	98.8	0.3	99.5	2.0	96.9	0.5	98.6
14	0.5	99.3	0.2	99.8	1.3	98.2	0.8	99.4
15	0.3	99.6	0.1	99.9	0.6	98.7	0.2	99.5
16	0.2	99.8	0.1	99.9	0.6	99.3	0.2	99.8
17	0.1	99.9	0.1	100.0	0.3	99.6	0.0	99.8
18	0.1	100.0	0.0	100.0	0.3	99.9	0.0	99.8
19	0.0	100.0	0.0	100.0	0.0	99.9	0.2	99.9
20	0.0	100.0	0.0	100.0	0.1	100.0	0.1	100.0

Internalising

Internalising	**Parents (5-7-year-olds)**				**Teachers (5-year-olds)**			
	Boys (N = 28,920)		**Girls (N = 27,611)**		**Boys (N = 1,272)**		**Girls (N = 1,291)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	26.1	26.1	25.9	25.9	34.4	34.4	35.6	35.6
1	20.1	46.2	20.5	46.4	18.5	52.9	17.7	53.2
2	16.6	62.8	17.1	63.5	15.5	68.4	13.9	67.1
3	11.8	74.6	12.5	76.0	8.3	76.7	8.8	75.9
4	8.0	82.7	8.5	84.5	7.1	83.8	6.3	82.2
5	5.5	88.2	5.7	90.1	4.6	88.4	6.0	88.2
6	4.0	92.2	3.7	93.8	2.4	90.8	3.5	91.7

7	2.6	94.8	2.3	96.1	2.4	93.2	2.2	93.9
8	1.8	96.6	1.5	97.6	1.7	94.9	1.5	95.4
9	1.2	97.8	1.0	98.6	0.8	95.7	1.7	97.1
10	0.8	98.6	0.6	99.2	0.9	96.6	1.1	98.2
11	0.6	99.2	0.4	99.5	1.3	97.9	0.8	99.0
12	0.3	99.5	0.2	99.7	0.6	98.5	0.5	99.5
13	0.2	99.7	0.2	99.9	0.5	99.0	0.2	99.8
14	0.1	99.9	0.1	99.9	0.5	99.4	0.1	99.8
15	0.1	99.9	0.0	100.0	0.2	99.7	0.1	99.8
16	0.0	100.0	0.0	100.0	0.2	99.8	0.1	99.9
17	0.0	100.0	0.0	100.0	0.1	99.9	0.0	100.0
18	0.0	100.0	0.0	100.0	0.1	100.0	0.0	100.0
19	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0
20	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0

Hyperactivity

Hyperactivity	**Parents (5-7-year-olds)**				**Teachers (5-year-olds)**			
	Boys (N = 28,920)		**Girls (N = 27,611)**		**Boys (N = 1,272)**		**Girls (N = 1,291)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	17.1	17.1	25.1	25.1	24.8	24.8	46.3	46.3
1	18.6	35.7	24.3	49.4	15.7	40.5	16.1	62.4
2	17.8	53.5	17.7	67.1	13.2	53.7	11.0	73.4
3	15.6	69.1	13.5	80.6	11.1	64,8	8.4	81.9
4	10.6	79.7	8.0	88.6	8.9	73.7	5.9	87.8
5	8.3	87.9	5.2	93.8	7.4	81.1	4.3	92.1
6	4.5	92.5	2.7	96.4	4.9	85.9	2.6	94.7
7	2.9	95.3	1.5	98.0	4.0	89.9	1.6	96.3
8	2.0	97.3	1.0	99.0	3.2	93.2	0.8	97.1
9	1.4	98.7	0.6	99.6	4.0	97.2	1.3	98.4
10	1.3	100	0.4	100.0	2.8	100.0	1.6	100.0

Conduct

Conduct	**Parents (5-7-year-olds)**				**Teachers (5-year-olds)**			
	Boys (N = 28,920)		**Girls (N = 27,611)**		**Boys (N = 1,272)**		**Girls (N = 1,291)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	34.6	34.6	37.4	37.4	49.8	49.8	64.5	64.5
1	30.4	64.9	30.7	68.2	19.8	69.7	15.8	80.3
2	19.0	83.9	18.7	86.9	11.1	80.7	9.8	90.2
3	9.2	93.1	8.5	95.3	6.6	87.3	4.8	95.0
4	4.1	97.1	3.2	98.5	5.7	93.1	2.2	97.1
5	1.8	98.9	0.9	99.5	3.6	96.7	1.6	98.8
6	0.6	99.6	0.3	99.8	1.7	98.4	0.6	99.4
7	0.3	99.8	0.1	99.9	0.9	99.4	0.2	99.5
8	0.1	99.9	0.0	100.0	0.5	99.8	0.2	99.8
9	0.0	100.0	0.0	100.0	0.1	99.9	0.2	99.9
10	0.0	100.0	0.0	100.0	0.1	100.0	0.1	100.0

Emotional

Emotional	Parents (5-7-year-olds)				Teachers (5-year-olds)			
	Boys (N = 28,920)		Girls (N = 27,611)		Boys (N = 1,272)		Girls (N = 1,291)	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	35.2	35.2	32.8	32.8	47.2	47.2	44.5	44.5
1	23.9	59.1	24.1	56.9	19.8	67.0	20.4	65.0
2	16.2	75.3	17.3	74.2	14.1	81.1	12.5	77.5
3	10.7	86.0	11.2	85.4	7.1	88.1	8.7	86.2
4	6.4	92.4	6.7	92.1	4.9	93.0	6.4	92.6
5	3.7	96.1	3.9	96.0	2.8	95.8	3.7	96.3
6	2.2	98.3	2.2	98.2	2.3	98.1	1.6	97.9
7	1.0	99.3	1.0	99.2	1.2	99.3	1.7	99.6
8	0.5	99.8	0.5	99.7	0.3	99.6	0.1	99.7
9	0.2	99.9	0.2	99.9	0.3	99.9	0.2	99.8
10	0.1	100.0	0.1	100.0	0.1	100.0	0.2	100.0

Peer-problems

Peer-problems	**Parents (5-7-year-olds)**				**Teachers (5-year-olds)**			
	Boys (N = 28,920)		**Girls (N = 27,611)**		**Boys (N = 1,272)**		**Girls (N = 1,291)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	61.0	61.0	66.9	66.9	61,0	61.0	65.5	65.5
1	17.0	78.0	15.3	82.2	14,4	75.4	13.2	78.8
2	11.4	89.4	10.7	92.9	11,2	86.6	9.0	87.8
3	4.9	94.3	3.6	96.5	4,8	91.4	4.4	92.2
4	2.7	97.0	2.0	98.5	2,8	94.3	3.8	96.0
5	1.4	98.4	0.8	99.3	2,4	96.6	2.2	98.1
6	0.9	99.3	0.4	99.7	1,3	97.9	1.2	99.4
7	0.4	99.7	0.2	99.9	0,6	98.5	0.4	99.8
8	0.2	99.9	0.1	100.0	0,8	99.3	0.2	99.9
9	0.1	100.0	0.0	100.0	0,7	100.0	0.1	100.0
10	0.0	100.0	0.0	100.0	0,0	100,0	0,0	100,0

Prosocial

Prosocial	Parents (5-7-year-olds)				Teachers (5-year-olds)			
	Boys (N = 28,920)		Girls (N = 27,611)		Boys (N = 1,272)		Girls (N = 1,291)	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	0.1	0.1	0.0	0.0	0,9	0.9	0.1	0.1
1	0.1	0.2	0.0	0.1	0,9	1.9	0.2	0.3
2	0.3	0.5	0.1	0.1	1,8	3.7	0.8	1.1
3	0.7	1.2	0.2	0.3	4,5	8.2	1.3	2.4
4	1.9	3.1	0.7	1.1	6,1	14.2	2.3	4.7
5	4.7	7.8	2.2	3.2	10,5	24.8	5.5	10.2
6	9.3	17.0	5.3	8.5	11,0	35.8	6.4	16.6
7	14.6	31.7	9.5	18.1	12,7	48.4	10.1	29.7
8	20.8	52.4	17.8	35.9	15,2	63.6	14.7	41.4
9	25.3	77.7	28.5	64.4	16,4	80.0	20.7	62.1
10	22.3	100.0	35.6	100.0	20,0	100.0	37.9	100.0

Impact scores

Impact score	**Parents (5-7-year-olds)**				**Teachers (5-year-olds)**			
	Boys (N = 28,920)		**Girls (N = 27,611)**		**Boys (N = 1,272)**		**Girls (N = 1,291)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	87.7	87.7	92.9	92.9	82,0	82.0	89.7	89.7
1	5.3	93.0	3.3	96.2	9,3	91.3	5.1	94.8
2	2.8	95.8	1.9	98.1	5,0	96.3	3.1	97.9
3	1.7	97.5	0.9	99.0	2,0	98.3	1.3	99.2
4	1.0	98.5	0.5	99.4	1,2	99.5	0.4	99.6
5	0.6	99.1	0.2	99.6	0,4	99.9	0.2	99.8
6	0.3	99.4	0.2	99.8	0,1	100.0	0.2	100.0
7	0.2	99.6	0.1	99.9	-	-	-	-
8	0.3	99.9	0.1	99.9	-	-	-	-
9	0.1	100.0	0.0	100.0	-	-	-	-
10	0.0	100.0	0.0	100.0	-	-	-	-

Appendix F (1): SDQ cut-off scores for the 10-12-year-old parent ratings

The *"normal"* banding includes app. 80 % of the sample, the *"borderline"* includes app. 10 % of the sample, and the *"clinical"* banding includes app. 10 % of the sample

Parents 10-12-year-olds	**Recommended bandings for boys (N = 3,322)**						**Recommended bandings for girls (N = 3,237)**						**Recommended bandings for boys (N = 6,559)**					
	Normal		Borderline		Clinical		Normal		Borderline		Clinical		Normal		Borderline		Clinical	
	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact
Total	0-10	83.1	11-13	7.1	14-40	9.8	0-8	80.7	9-11	9.2	12-40	10.1	0-9	82.7	10-12	8.2	13-40	9.1
Externalising	0-5	78.6	6-8	13.5	9-20	7.9	0-4	80.8	5-6	10.5	7-20	8.6	0-5	83.5	6-7	8.4	8-20	8.1
Internalising	0-4	80.4	5-7	11.6	8-20	8.0	0-4	79.6	5-7	13.1	8-20	7.3	0-4	80.8	5-6	9.2	7-20	10.0
Hyperactivity	0-4	82.8	5-6	10.4	7-10	6.7	0-3	84.3	4	6.0	5-10	9.7	0-3	79.7	4-5	12.9	6-10	7.4
Conduct	0-2	88.0	3	6.7	4-10	5.3	0-1	78.3	2	13.3	3-10	8.5	0-1	76.8	2	13.6	3-10	9.6
Emotional	0-3	85.0	4	6.1	5-10	8.9	0-3	83.0	4	7.3	5-10	9.7	0-3	84.6	4	6.5	5-10	8.9
Peer problems	0-2	86.2	3	5.3	4-10	8.5	0-2	87.0	3	5.2	4-10	7.9	0-1	76.6	2-3	15.7	4-10	7.7
Prosocial	7-10	83.9	6	8.3	0-5	7.9	8-10	84.6	7	8.0	0-6	7.4	8-10	78.5	7	9.6	0-6	12.0
Impact	0	83.4	1	6.4	2-10	10.2	0	88.0	1	4.2	2-10	7.8	0	86.3	1	5.3	2-10	8.4

Appendix F (2): SDQ cut-off scores for the 10-12-year-old teacher ratings

The *"normal"* banding includes app. 80 % of the sample, the *"borderline"* includes app. 10 % of the sample and the *"clinical"* banding includes app. 10 % of the sample

Teachers 10-12-year-olds	**Recommended bandings for boys (N = 2,790)**						**Recommended bandings for girls (N = 2,805)**						**Recommended bandings for boys (N = 5,595)**					
	Normal		Borderline		Clinical		Normal		Borderline		Clinical		Normal		Borderline		Clinical	
	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact	Raw	Exact
Total	0-11	79.8	12-16	10.2	17-40	10.0	0-8	81.9	9-12	8.8	13-40	9.2	0-9	79.9	10-14	10.7	15-40	9.4
Externalising	0-7	79.6	8-10	10.8	11-20	9.6	0-3	80.4	4-5	9.1	6-20	10.5	0-5	80.2	6-8	10.2	9-20	9.6
Internalising	0-5	83.4	6-7	6.4	8-20	10.2	0-5	82.9	6-7	7.1	8-20	10.0	0-4	78.3	5-7	12.2	8-20	9.5
Hyperactivity	0-5	79.4	6-7	9.7	8-10	10.9	0-2	78.4	3-4	13.1	5-10	8.6	0-4	81.9	5-6	8.9	7-10	9.2
Conduct	0-2	83.6	3	89.7	4-10	10.3	0-1	87.2	2	6.1	3-10	6.7	0-1	81.1	2	8.2	3-10	10.8
Emotional	0-2	80.3	3-4	11.4	5-10	8.3	0-3	84.5	4	6.3	5-10	5.8	0-2	79.1	3-4	12.4	5-10	8.4
Peer problems	0-2	79.4	3-4	11.9	5-10	8.7	0-2	80.7	3-4	11.2	5-10	8.1	0-2	80.3	3-4	11.3	5-10	8.4
Prosocial	5-10	80.5	3-4	12.5	0-2	7.0	6-10	86.4	5	7.4	0-4	6.2	6-10	77.5	5	9.9	0-4	12.6
Impact	0	77.2	1-2	15.4	3-6	7.4	0	85.8	1	5.5	2-6	8.7	0	82.1	1	7.2	2-10	10.7

Appendix G (1): SDQ means and SDs for the 10-12-year-old parent samples

Parents	**10-12-year-old boys (N = 3,322)**		**10-12-year-old girls (N = 3,237)**	
	Mean	SD	Mean	SD
Total difficulties	6.02	5.24	5.27	4.72
Externalising	3.42	3.13	2.58	2.65
Internalising	2.60	2.97	2.69	2.90
Hyperactivity	2.40	2.29	1.72	1.91
Conduct	1.02	1.28	0.86	1.14
Emotional	1.60	1.91	1.76	1.90
Peer problems	1.00	1.61	0.92	1.55
Prosocial	8.19	1.69	8.78	1.38
Impact	0.42	1.18	0.30	1.02

Appendix G (2): SDQ means and SDs for the 10-12-year-old teacher samples

Teachers	**10-12-year-old boys (N = 2,790)**		**10-12-year-old girls (N = 2,805)**	
	Mean	SD	Mean	SD
Total difficulties	6.80	6.42	4.64	5.34
Externalising	4.13	4.21	1.94	2.88
Internalising	2.66	3.37	2.70	3.45
Hyperactivity	3.02	2.92	1.40	2.01
Conduct	1.11	1.72	0.54	1.21
Emotional	1.27	1.92	1.47	2.02
Peer problems	1.39	1.96	1.23	1.91
Prosocial	6.77	2.62	8.07	2.08
Impact	0.48	1.07	0.29	0.82

Appendix H: SDQ frequency distribution for 10-12-year-olds

Distribution of scores for the *Total difficulties, Externalising, Internalising, Hyperactivity, Conduct, Emotional, Peer-problems, Prosocial scores* and *Impact scores* for the 10-12-year-olds parent and teacher ratings

Total difficulties

Total difficulties	**Parents (10-12-year-olds)**				**Teacher (10-12-year-olds)**			
	Boys (N =3,322)		**Girls (N = 3,237)**		**Boys (N = 2,790)**		**Girls (N = 2,805)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	7.4	7.4	9.5	9.5	13.2	13.2	22.5	22.5
1	9.6	17.0	11.0	20.5	8.8	21.9	12.0	34.5
2	11.2	28.2	12.7	33.1	9.7	31.7	12.0	46.6
3	10.9	39.1	10.8	43.9	7.3	39.0	9.6	56.2
4	9.8	48.9	10.0	53.9	7.0	46.0	7.0	63.1
5	8.9	57.8	9.2	63.1	7.2	53.3	5.7	68.8
6	7.0	64.8	7.2	70.3	6.1	59.4	5.0	73.9
7	6.1	71.0	5.8	76.1	5.2	64.6	4.4	78.3
8	4.1	75.0	4.6	80.7	4.9	69.5	3.6	81.9
9	4.1	79.1	4.1	84.8	4.0	73.5	2.9	84.8
10	4.0	83.1	3.1	87.9	3.3	76.8	2.2	87.0
11	3.5	86.6	2.0	89.9	3.0	79.8	1.9	88.9
12	2.0	88.6	1.8	91.8	2.5	82.3	1.8	90.8
13	1.6	90.2	1.5	93.3	2.3	84.6	1.7	92.5
14	1.8	92.0	1.3	94.6	2.0	86.6	1.5	94.0

15	1.4	93.3	1.3	95.9	2.1	88.7	1.0	94.9
16	1.1	94.4	0.6	96.5	1.3	90.0	0.6	95.6
17	1.1	95.5	0.6	97.1	1.8	91.8	0.9	96.5
18	1.1	96.5	0.5	97.6	1.2	93.0	0.7	97.2
19	0.8	97.3	0.6	98.2	1.4	94.4	0.5	97.7
20-40	2.7	100.0	1.8	100.0	5.6	100.0	2.3	100.0

Externalising

Externalising	**Parents (10-12-year-olds)**				**Teachers (10-12-year-olds)**			
	Boys (N = 3,322)		**Girls (N = 3,237)**		**Boys (N = 2,790)**		**Girls (N = 2,805)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	17.9	17.9	24.6	24.6	24.0	24.0	46.8	46.8
1	14.8	32.7	18.1	42.7	10.6	34.6	12.7	59.6
2	14.8	47.4	17.1	59.8	11.6	46.2	13.0	72.6
3	12.8	60.2	12.3	72.1	9.1	55.3	7.8	80.4
4	10.1	70.3	8.7	80.8	8.2	63.5	5.6	86.1
5	8.3	78.6	5.8	86.7	6.4	69.9	3.5	89.5
6	5.9	84.6	4.7	91.4	5.6	75.4	3.0	92.5
7	4.1	88.7	2.7	94.1	4.2	79.6	2.0	94.6
8	3.5	92.1	1.9	96.0	4.6	84.2	1.1	95.7
9	2.7	94.9	1.6	97.6	3.4	87.6	1.1	96.8
10	1.4	96.3	0.9	98.5	2.8	90.4	0.7	97.5

11	1.2	97.5	0.4	98.9	2.0	92.4	0.6	98.0
12	1.0	98.5	0.4	99.4	2.0	94.3	0.6	98.6
13	0.6	99.1	0.3	99.6	1.5	95.9	0.4	99.0
14	0.6	99.6	0.2	99.8	1.4	97.2	0.5	99.5
15	0.2	99.9	0.1	99.9	0.9	98.1	0.2	99.7
16	0.0	99.9	0.0	99.9	0.8	98.9	0.1	99.8
17	0.0	100.0	0.1	100.0	0.5	99.4	0.0	99.8
18	0.1	100.0	0.0	100.0	0.4	99.8	0.1	99.9
19	0.0	100.0	0.0	100.0	0.1	99.9	0.0	99.9
20	0.0	100.0	0.0	100.0	0.1	100.0	0.1	100.0

Internalising

Internalising	**Parent (10-12-year-olds)**				**Teachers (10-12-year-olds)**			
	Boys (N = 3,322)		**Girls (N = 3,237)**		**Boys (N = 2,790)**		**Girls (N = 2,805)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	26.9	26.9	25.3	25.3	34.5	34.5	35.4	35.4
1	20.3	47.1	19.1	44.4	16.6	51.1	15.8	51.2
2	16.0	63.1	15.5	59.9	12.9	63.9	11.9	63.1
3	10.0	73.1	11.1	71.0	8.2	72.2	8.7	71.7
4	7.3	80.4	8.6	79.6	6.1	78.3	5.9	77.6
5	5.7	86.0	5.8	85.4	5.1	83.4	5.3	82.9
6	3.2	89.2	4.0	89.4	3.5	86.9	3.9	86.8

7	2.7	92.0	3.3	92.7	2.9	89.8	3.2	90.0
8	2.1	94.1	2.0	94.7	2.4	92.2	2.3	92.3
9	1.7	95.8	1.1	95.9	1.9	94.1	1.6	93.9
10	1.1	96.9	1.2	97.1	1.8	95.9	1.4	95.3
11	1.1	98.0	1.1	98.2	1.3	97.1	1.2	96.5
12	0.7	98.7	0.7	98.9	0.6	97.8	0.9	97.4
13	0.6	99.3	0.4	99.3	0.8	98.6	0.6	98.0
14	0.3	99.6	0.4	99.8	0.6	99.2	0.4	98.9
15	0.2	99.8	0.2	99.9	0.3	99.5	0.4	99.3
16	0.1	99.9	0.1	100.0	0.1	99.6	0.2	99.6
17	0.1	100.0	0.0	100.0	0.2	99.7	0.1	99.9
18	0.0	100.0	0.0	100.0	0.1	99.9	0.0	100.0
19	0.0	100.0	0.0	100.0	0.1	100.0	0.0	100.0
20	0.0	100.0	0.0	100.0	0.0	100.0	0.0	100.0

Hyperactivity

Hyperactivity	**Parent (10-12-year-olds)**				**Teachers (10-12-year-olds)**			
	Boys (N = 3,322)		**Girls (N = 3,237)**		**Boys (N = 2,790)**		**Girls (N = 2,805)**	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	25.3	25.3	34.8	34.8	26.7	26.7	50.9	50.9
1	17.3	42.6	21.2	56.0	13.0	39.8	14.6	65.5
2	16.3	58.9	16.4	72.5	13.2	52.9	12.8	78.4
3	15.2	74.2	11.8	84.3	10.8	63.7	8.1	86.4
4	8.6	82.8	6.0	90.3	8.0	71.7	5.0	91.4
5	6.5	89.3	4.6	94.8	7.7	79.4	3.1	94.5
6	3.9	93.3	2.5	97.3	5.6	84.9	2.2	96.7
7	2.9	96.2	1.2	98.5	4.1	89.1	1.2	97.9
8	1.8	98.0	0.8	99.3	4.2	93.3	0.9	98.8
9	1.2	99.2	0.6	99.8	3.2	96.4	0.6	99.4
10	0.8	100.0	0.2	100.0	3.6	100.0	0.6	100.0

Conduct

Conduct	Parent (10-12-year-olds)				Teachers (10-12-year-olds)			
	Boys (N = 3322)		Girls (N = 3237)		Boys (N = 2790)		Girls (N = 2805)	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	45.7	45.7	50.0	50.0	54.7	54.7	74.2	74.2
1	27.8	73.5	28.3	78.3	18.7	73.4	12.9	87.2
2	14.5	88.0	13.3	91.5	10.3	83.6	6.1	93.3
3	6.7	94.7	5.0	96.5	6.1	89.7	3.1	96.3
4	3.2	97.9	2.2	98.7	4.5	94.2	1.5	97.9
5	1.3	99.1	0.8	99.5	2.5	96.7	0.8	98.6
6	0.5	99.6	0.3	99.8	1.3	97.9	0.7	99.3
7	0.3	99.9	0.1	99.9	1.1	99.1	0.3	99.6
8	0.1	99.9	0.1	100.0	0.6	99.6	0.2	99.9
9	0.1	99.9	0.0	100.0	0.1	99.8	0.0	99.9
10	0.0	100.0	0.0	100.0	0.2	100.0	0.1	100.0

Emotional

Emotional	Parent (10-12-year-olds)				Teachers (10-12-year-olds)			
	Boys (N = 3,322)		Girls (N = 3,237)		Boys (N = 2,790)		Girls (N = 2,805)	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	38.4	38.4	32.8	32.8	53.6	53.6	47.7	47.7
1	22.6	60.9	22.9	55.7	16.5	70.1	17.8	65.4
2	14.6	75.5	16.0	71.7	10.2	80.3	11.2	76.6
3	9.5	85.0	11.2	83.0	7.2	87.5	7.8	84.5
4	6.1	91.1	7.3	90.3	4.2	91.7	6.3	90.8
5	3.6	94.7	4.6	94.9	3.2	94.9	3.4	94.2
6	2.3	96.9	2.3	97.1	2.4	97.3	2.6	96.8
7	1.4	98.3	1.7	98.8	1.1	98.5	1.3	98.1
8	1.1	99.4	0.6	99.5	0.8	99.3	0.8	98.9
9	0.5	99.9	0.4	99.8	0.4	99.6	0.8	99.7
10	0.1	100.0	0.2	100.0	0.4	100.0	0.3	100.0

Peer-problems

Peer problems	Parent (10-12-year-olds)				Teachers (10-12-year-olds)			
	Boys (N = 3,322)		Girls (N = 3,237)		Boys (N = 2,790)		Girls (N = 2,805)	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	57.3	57.3	61.5	61.5	49.3	49.3	56.1	56.1
1	17.9	75.2	14.4	75.9	17.0	66.3	15.2	71.3
2	11.0	86.2	11.1	87.0	13.0	79.4	9.4	80.7
3	5.3	91.5	5.2	92.1	7.0	86.4	6.4	87.2
4	3.5	95.0	3.6	95.7	4.9	91.3	4.8	91.9
5	2.1	97.1	1.7	97.4	3.4	94.7	3.0	95.0
6	1.3	98.4	1.5	98.9	1.6	96.3	2.3	97.3
7	0.8	99.2	0.5	99.4	1.9	98.2	1.2	98.4
8	0.5	99.8	0.5	99.9	1.0	99.2	1.0	99.4
9	0.2	99.9	0.1	100.0	0.6	99.9	0.5	99.9
10	0.1	100.0	0.0	100.0	0.1	100.0	0.1	100.0

Prosocial scores

Prosocial	Parent (10-12-year-olds)				Teachers (10-12-year-olds)			
	Boys (N = 3,322)		Girls (N = 3,237)		Boys (N = 2,790)		Girls (N = 2,805)	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	0.1	0.1	0.0	0.0	1.9	1.9	0.3	0.3
1	0.2	0.2	0.0	0.0	2.3	4.2	0.5	0.9
2	0.2	0.5	0.1	0.1	2.8	7.0	0.6	1.5
3	0.9	1.4	0.4	0.5	5.0	12.0	1.5	3.0
4	2.3	3.6	0.7	1.2	7.5	19.5	3.2	6.2
5	4.2	7.9	2.2	3.4	12.7	32.2	7.4	13.6
6	8.3	16.2	4.0	7.4	10.7	42.9	9.6	23.2
7	12.0	28.2	8.0	15.4	11.7	54.6	9.8	33.0
8	19.8	48.0	16.5	31.9	12.7	67.3	13.8	46.8
9	27.2	75.1	30.4	62.3	13.8	81.1	17.3	64.1
10	24.9	100.0	37.7	100.0	18.9	100.0	35.9	100.0

Impact scores

Impact score	Parent (10-12-year-olds)				Teachers (10-12-year-olds)			
	Boys (N = 3,322)		Girls (N = 3,237)		Boys (N = 2,790)		Girls (N = 2,805)	
	%	Cumulative	%	Cumulative	%	Cumulative	%	Cumulative
0	83.4	83.4	88.0	88.0	77.2	77.2	85.8	85.8
1	6.4	89.8	4.2	92.2	9.3	86.5	5.5	91.3
2	4.2	94.0	3.2	95.4	6.1	92.6	5.0	96.3
3	2.2	96.2	2.4	97.8	4.1	96.7	2.2	98.5
4	1.7	97.9	0.7	98.5	2.3	99.0	1.0	99.5
5	0.8	98.7	0.6	99.1	0.5	99.6	0.4	99.9
6	0.5	99.2	0.3	99.4	0.4	100.0	0.1	100.0
7	0.4	99.6	0.3	99.7	-	-	-	-
8	0.2	99.9	0.2	99.9	-	-	-	-
9	0.0	99.9	0.1	100.0	-	-	-	-
10	0.1	100.0	0.0	100.0	-	-	-	-

Appendix I: Effect sizes as measured by Cohen's D comparing SDQ scores from the DNBC and SFI birth cohorts

From the 7-year follow-ups. An effect size of Cohen's D of 0.2 to 0.3 is considered small; 0.5 a medium effect size and 0.8 to infinity a large effect

Parents	**Total sample**			**Boys**			**Girls**		
7-year-olds	**DNBC**	**SFI**	**Effect size**	**DNBC**	**SFI**	**Effect size**	**DNBC**	**SFI**	**Effect size**
SDQ scale	Mean (SD)	Mean (SD)	Cohen's D	Mean (SD)	Mean (SD)	Cohen's D	Mean (SD)	Mean (SD)	Cohen's D
Total difficulties	5.87 (4.44)	6.56 (5.07)	0.14	6.35 (4.71)	6.96 (5.30)	0.12	5.37 (4.09)	6.12 (4.77)	0.17
Hyperactivity	2.36 (2.14)	2.59 (2.51)	0.10	2.71 (2.27)	2.90 (2.65)	0.08	2.00 (1.94)	2.25 (2.30)	0.12
Conduct	1.20 (1.26)	1.35 (1.46)	0.11	1.26 (1.33)	1.40 (1.51)	0.10	1.13 (1.19)	1.30 (1.41)	0.13
Emotional	1.59 (1.74)	1.88(1.95)	0.16	1.56 (1.73)	1.84 (1.93)	0.15	1.63 (1.74)	1.94 (2.00)	0.17
Peer Problems	0.72 (1.26)	0.76 (1.30)	0.03	0.82 (1.36)	0.85 (1.41)	0.02	0.63 (1.13)	0.66 (1.17)	0.03
Prosocial	8.35 (1.57)	8.73 (1.53)	0.25	8.05 (1.66)	8.43 (1.67)	0.23	8.66 (1.41)	9.05 (1.29)	0.29

Appendix J:Background characteristics of women defined from early pregnancy intake

[1] GA = gestational age. [2] Gathered at six months post-partum. [3] Fully or partly unplanned pregnancy. [4] Age at birth. [5] Mean (10^{th}/ 90^{th} percentiles). [6] married or cohabiting with the child's biological father six months post-partum. [7] Register-based information on educational level in year 2010. [8] Maternal self-report: has she ever suffered from a psychiatric disorders. [9] Maternal self-report: has she suffered from pre-pregnancy psychiatric disorders in pregnancy? [10] Register-based information on pre-pregnancy contact with the psychiatric system. [11] Definition of binge drinking: intake of five or more alcohol containing units on a single occasion. [12] Cumulated smoking variable: calculated in the same manner as the cumulated alcohol exposure variable from pre-pregnancy and full pregnancy information. [13] In early part of pregnancy. [14] Information from third interview. [15] Paternal smoking in pregnancy: yes/ no; from first interview. [15] BMI: Body Mass Index; normal range: 18.5-24.99). [16] Intake of ≥ 1 liter of Cola per week. [17] Television watching ≥ 2 hours/ day. [18] Tenant, homeless or live with parents. [19] Register-based information. [20] Professional concern that baby was not growing in last part of pregnancy. [21] SGA: small for gestational age. [22] Mean/ days. [23] Percentage of children with an APGAR score <10 after one minute. [24] Child contact with the psychiatric system before the age of seven.

	Average alcohol exposure					**Binge drinking**	
Alcohol group	**Full sample**	**0**	**>0-2**	**>2-4**	**>4**	**No**	**Yes**
N	37,152	20,165	14,935	1680	372	27,820	9332
Sampling characteristics							
GA 1st interview[1]	15	14	15	15	15	15	14
GA 2nd interview[1]	30	29	30	30	29	30	29
Child age 3rd interview [2]	6.3	6.3	6.3	6.3	6.3	6.3	6.3
Unplanned preg. (yes) [3]	22.1 %	21.9 %	21.8 %	24.4 %	30.3 %	20.0 %	28.3 %
Time to preg. (≥ 6 months)	26.4 %	27.4 %	25.1 %	26.1 %	25.8 %	26.9 %	24.9 %
Fertility treatment (yes)	5.7 %	6.1 %	5.5 %	4.9 %	3.5 %	6.8 %	2.7 %
Family characteristics							

Age (M) [4, 5]	30.7	30.2	31.1	32.7	33.7	30.8	30.4
Age (P) [4, 5]	32.5	31.9	32.8	34.5	36.1	32.6	32.1
Married (no) [6]	2.0 %	2.3 %	1.6 %	1.9 %	4.6 %	1.8 %	2.7 %
Education – mandatory (M) [7]	6.8 %	8.3 %	4.9 %	6.3 %	8.3 %	7.1 %	5.8 %
Education – university (M) [7]	15.8 %	14.0 %	17.6 %	20.8 %	13.7 %	15.0 %	17.9 %
Education – mandatory (P) [7]	12.2 %	13.8 %	10.2 %	9.7 %	14.8 %	12.5 %	11.2 %
Education – university (P) [7]	16.4 %	15.2 %	17.8 %	18.1 %	14.2 %	15.9 %	17.8 %
Self-rep. pre-preg. psych. diag. (M) [8]	6.9 %	7.2 %	6.3 %	6.1 %	13.2 %	6.6 %	7.6 %
In-preg. pre-preg. psych. diag. (M) [9]	1.0 %	1.0 %	0.9 %	1.3 %	1.9 %	1.0 %	1.1 %
Pre-preg. psych. diag. (M) [10]	2.5 %	2.9 %	2.0 %	1.9 %	5.4 %	2.5 %	2.5 %
Pre-preg. psych. diag. (P) [10]	1.7 %	1.9 %	1.5 %	1.4 %	2.7 %	1.7 %	1.9 %
In-childhood psych. diag. (M) [10]	3.0 %	3.3 %	2.4 %	2.7 %	4.6 %	2.9 %	3.2 %
In-childhood psych. diag. (P) [10]	2.2 %	2.4 %	2.0 %	1.8 %	2.7 %	2.1 %	2.3 %
Pre-pre. alc. drinks/ week (M) [5]	3.1	2.2	3.7	7.1	11.1	2.6	4.5
Binge drink in preg. (M) [5]	0.5	0.4	0.6	1.1	2.0	0.1	1.8
Binge drink in preg. (yes) (M) [11]	30.9 %	25.4 %	35.4 %	49.4 %	60.0 %	7.7 %	100%
Cum. alc. intake in preg. (M) [5]	33.2	12.3	46.5	124.1	217	28.2	48.1
Cum. smoking in preg. (M) [5, 12]	53	55	47	69	138	48	68
Smoking (yes) [13]	22.8 %	22.6 %	21.9 %	28.6 %	42.3 %	19.8 %	31.5 %
Nicotine substitutes [14]	2.1 %	1.9 %	2.2 %	3.0 %	3.0 %	1.9 %	2.7 %
Partner smoking (yes)[15]	27.8 %	28.4 %	26.3 %	29.8 %	38.2 %	26.8 %	30.7%
Vitamins (no) [14]	15.1 %	14.6 %	15.3 %	18.5 %	21.5 %	15.0 %	15.6 %

Iron supplement (no) [14]	27.8 %	27.9 %	27.3 %	30.6 %	32.9 %	27.9 %	27.7 %
Fish oil supplement (yes) [14]	4.9 %	4.8 %	4.7 %	6.1 %	6.2 %	4.9 %	4.7 %
Fish eating (never) [14]	3.2 %	4.0 %	2.2 %	1.7 %	0.8 %	3.2 %	3.0 %
Fish as warm meal (never)[14]	8.5 %	10.6 %	6.2 %	4.1 %	3.2 %	8.5 %	8.3 %
Pre-pregnant BMI [15]	30.3 %	32.8 %	27.8 %	23.8 %	27.6 %	31.2 %	27.7 %
Cola [14]	15.1 %	16.3 %	13.7 %	13.3 %	15.6 %	15.1 %	15.2 %
Coffee (yes) [14, 16]	41.4 %	34.0 %	48.8 %	64.8 %	73.6 %	60.3 %	44.5 %
Painkillers (yes) [14]	24.1 %	23.1 %	25.0 %	28.0 %	28.3 %	23.8 %	25.1 %
Diabetes in preg. (yes) [14]	1.9 %	2.1 %	1.8 %	1.5 %	1.9 %	2.0 %	1.6 %
Asthma in preg. (yes) [14]	3.2 %	3.5 %	2.9 %	3.3 %	3.8 %	3.3 %	3.1 %
Anaemia in preg. (yes) [14]	3.7 %	3.7 %	3.7 %	3.8 %	5.1 %	3.7 %	3.7 %
Television watching [14, 17]	20.0 %	22.5 %	17.3 %	15.1 %	17.2 %	19.6 %	21.3 %
Tenant [14, 18]	25.7 %	27.9 %	23.5 %	20.7 %	19.4 %	24.2 %	30.2 %
Exercise (yes) [14]	25.7 %	24.3 %	27.5 %	27.3 %	22.6 %	25.0 %	27.8 %
Child characteristics							
Sex (% boys) [19]	51.0 %	50.6 %	51.4 %	52.3 %	52.2 %	51.2 %	50.6 %
Baby not growing [20]	5.3 %	5.3 %	5.1 %	5.2 %	8.3 %	5.3 %	5.0 %
Birth weight [5, 19]	3654	3657	3654	3634	3557	3658	3640
SGA (%) 19, [21]	8.6 %	8.7 %	8.3 %	9.4 %	13.5 %	8.4 %	9.2 %
Gestational age [19, 22]	281.5	281.4	281.6	281.8	281.7	281.4	282.0
APGAR score (% < 10) [19, 23]	7.6 %	7.8 %	7.4 %	7.1 %	6.5 %	7.4 %	8.3 %
Child psych. diagnosis (yes) [24]	1.3 %	1.4 %	1.2 %	1.0 %	0.5 %	1.3 %	1.2 %

Appendix K: Background characteristics of women defined from middle pregnancy intake

[1] GA = gestational age. [2] Gathered at six months post-partum. [3] Fully or partly unplanned pregnancy. [4] Age at birth. [5] Mean (10^{th}/ 90^{th} percentiles). [6] married or cohabiting with the child's biological father six months post-partum. [7] Register-based information on educational level in year 2010. [8] Maternal self-report: has she ever suffered from a psychiatric disorders. [9] Maternal self-report: has she suffered from pre-pregnancy psychiatric disorders in pregnancy? [10] Register-based information on pre-pregnancy contact with the psychiatric system. [11] Definition of binge drinking: intake of five or more alcohol containing units on a single occasion. [12] Cumulated smoking variable: calculated in the same manner as the cumulated alcohol exposure variable from pre-pregnancy and full pregnancy information. [13] In early part of pregnancy. [14] Information from third interview. [15] Paternal smoking in pregnancy: yes/ no; from first interview. [15] BMI: Body Mass Index; normal range: 18.5-24.99). [16] Intake of ≥ 1 liter of Cola per week. [17] Television watching ≥ 2 hours/ day. [18] Tenant, homeless or live with parents. [19] Register-based information. [20] Professional concern that baby was not growing in last part of pregnancy. [21] SGA: small for gestational age. [22] Mean/ days. [23] Percentage of children with an APGAR score <10 after one minute. [24] Child contact with the psychiatric system before the age of seven.

	Average alcohol exposure					**Binge drinking**	
Alcohol group	**Full sample**	**0**	**>0-2**	**>2-4**	**>4**	**No**	**Yes**
N	37,152	17,982	16,271	2,333	566	29,467	7,685
Sampling characteristics							
GA 1st interview[1]	15	15	15	15	15	15	15
GA 2nd interview[1]	30	30	30	29	29	30	30
Child age 3rd interview [2]	6.3	6.3	6.3	6.3	6.3	6.3	6.3
Unplanned preg. (yes) [3]	22.1 %	22.4 %	21.2 %	23.7 %	27.7 %	20.3 %	28.8 %
Time to preg. (≥ 6 months)	26.4 %	27.5 %	25.5 %	24.3 %	28.3 %	26.9 %	24.4 %
Fertility treatment (yes)	5.7 %	6.1 %	5.4 %	5.8 %	4.8 %	6.4 %	3.0 %
Family characteristics							
Age (M) [4, 5]	30.7	30.0	31.1	32.6	34.0	30.8	30.4

Age (P) [4, 5]	32.5	31.7	32.8	34.5	36.3	32.5	32.1
Married (no) [6]	2.0 %	2.5 %	1.4 %	2.0 %	3.2 %	1.8 %	2.8 %
Education – mandatory (M) [7]	6.8 %	9.2 %	4.6 %	4.1 %	7.4 %	6.9 %	6.6 %
Education – university (M) [7]	15.77%	12.00%	18.77%	22.57%	21.06%	15.17%	18.06%
Education – mandatory (P) [7]	12.16%/	14.94%	9.36%	10.38%	12.43%	12.30%	11.62%
Education – university (P) [7]	16.39%	12.82%	19.28%	22.84%	19.64%	16.03%	17.78%
Self-rep. pre-preg.psych.(M)[8]	6.9 %	7.1 %	6.4 %	7.0 %	12.0 %	6.5 %	8.2 %
In-preg. pre-preg. psych.(M)[9]	1.0 %	1.1 %	0.9 %	1.1 %	1.4 %	1.0 %	1.2 %
Pre-preg. psych. diag. (M) [10]	2.5 %	2.9 %	2.0 %	2.3 %	4.6 %	2.4 %	2.8 %
Pre-preg. psych. diag. (P) [10]	1.7 %	1.9 %	1.4 %	2.0 %	3.2 %	1.7 %	2.0 %
In-childhood psych. diag(M)[10]	3.0 %	3.4 %	2. 5 %	2.2 %	5.1 %	2.9 %	3.1 %
In-childhood psych. diag.(P)[10]	2.2 %	2.6 %	1.8 %	2.0 %	3.0 %	2.1 %	2.4 %
Pre-preg. drinks/ week (M)[5]	3.1	1.8	3.7	6.7	9.4	2.7	4.5
Binge drinking in preg. (M) [5]	0.54	0.36	0.62	1.12	1.74	0.19	1.9
Binge drinking in preg. (yes)[11]	30.9 %	23.5 %	35.6 %	48.5 %	56.9 %	12.8 %	100%
Cum. alc. intake in preg. (M)[5]	33.2	7.9	43.0	117.6	206.9	28.8	49.9
Cum. smoking in preg.(M) [5,12]	53	63	41	56	105	49	72
Smoking (yes) [13]	22.8 %	24.3 %	20.3 %	25.1 %	35.4 %	20.4 %	31.8 %
Nicotine substitutes [14]	2.1 %	2.0 %	2.1 %	2.4 %	3.0 %	1.9 %	2.7 %
Partner smoking (yes)[15]	27.8 %	29.5 %	25.5 %	27.7 %	37.1 %	26.8 %	31.4 %
Vitamins (no) [14]	15.1 %	14.9 %	14.8 %	17.2 %	22.2 %	15.0 %	15.7 %
Iron supplement (no) [14]	27.8 %	27.8 %	27.2 %	31.2 %	33.7 %	27.9 %	27.7 %

Fish oil supplement (yes) [14]	4.9 %	4.9 %	4.6 %	5.9 %	6.2 %	4.8 %	5.2 %
Fish eating (never) [14]	3.2 %	4.4 %	2.1 %	1.5 %	1.2 %	3.2 %	3.1 %
Fish as warm meal (never) [14]	8.5 %	11.3 %	6.1 %	4.3 %	3.9 %	8.4 %	8.6 %
Pre-pregnant BMI [15]	30.3 %	34.5 %	26.8 %	23.5 %	24.4 %	30.8 %	28.4 %
Cola [14]	15.1 %	17.5 %	13.0 %	12.5 %	14.5 %	15.0 %	15.8 %
Coffee (yes) [14, 16]	41.4 %	33.9 %	46.4 %	61.3 %	65.5 %	40.5 %	44.7 %
Painkillers (yes) [14]	24.1 %	22.9 %	24.9 %	27.2 %	29.5 %	23.7 %	25.8 %
Diabetes in preg. (yes) [14]	1.9 %	2.2 %	1.6 %	1.7 %	3.2 %	2.0 %	1.8 %
Asthma in preg. (yes) [14]	3.2 %	3.6 %	2.9 %	2.9 %	3.4 %	3.2 %	3.4 %
Anaemia in preg. (yes) [14]	3.7 %	3.6 %	3.6 %	4.7 %	4.6 %	3.7 %	3.8 %
Television watching [14, 17]	20.0 %	23.5 %	17.2 %	14.3 %	15.6 %	19.7 %	21.4 %
Tenant [14, 18]	25.7 %	27.6 %	24.5 %	21.6 %	19.2 %	24.5 %	30.6 %
Exercise (yes) [14]	25.7 %	23.3 %	28.1 %	28.6 %	20.3 %	25.2 %	27.5 %
Child characteristics							
Sex (% boys) [19]	51.0 %	51.0 %	51.1 %	50.5 %	51.2 %	51.0 %	51.0 %
Baby not growing [20]	5.3 %	5.4 %	5.1 %	5.2 %	6.2 %	5.4 %	4.7 %
Birth weight [5, 19]	3654	3652	3660	3638	3606	3657	3641
SGA (%) [19, 21]	8.6 %	9.0 %	7.9 %	8.9 %	13.0 %	8.4 %	9.3 %
Gestational age [19, 22]	281.5	281.3	281.7	281.9	282.3	281.4	282.0
APGAR score (% < 10) [19, 23]	7.6 %	8.0 %	7.4 %	6.6 %	6.0 %	7.5 %	8.2 %
Child psych. diag. (yes) [24]	1.25%	1.49%	1.08%	0.69%	1.06%	1.28%	1.17%

Appendix L: Background characteristics of women defined from late pregnancy intake

[1] GA = gestational age. [2] Gathered at six months post-partum. [3] Fully or partly unplanned pregnancy. [4] Age at birth. [5] Mean (10^{th}/ 90^{th} percentiles). [6] married or cohabiting with the child's biological father six months post-partum. [7] Register-based information on educational level in year 2010. [8] Maternal self-report: has she ever suffered from a psychiatric disorders. [9] Maternal self-report: has she suffered from pre-pregnancy psychiatric disorders in pregnancy? [10] Register-based information on pre-pregnancy contact with the psychiatric system. [11] Definition of binge drinking: intake of five or more alcohol containing units on a single occasion. [12] Cumulated smoking variable: calculated in the same manner as the cumulated alcohol exposure variable from pre-pregnancy and full pregnancy information. [13] In early part of pregnancy. [14] Information from third interview. [15] Paternal smoking in pregnancy: yes/ no; from first interview. [15] BMI: Body Mass Index; normal range: 18.5-24.99). [16] Intake of ≥ 1 liter of Cola per week. [17] Television watching ≥ 2 hours/ day. [18] Tenant, homeless or live with parents. [19] Register-based information. [20] Professional concern that baby was not growing in last part of pregnancy. [21] SGA: small for gestational age. [22] Mean/ days. [23] Percentage of children with an APGAR score <10 after one minute. [24] Child contact with the psychiatric system before the age of seven.

	Average alcohol exposure					**Binge drinking**	
Alcohol group	**Full sample**	**0**	**>0-2**	**>2-4**	**>4**	**No**	**Yes**
N	37,152	19,010	14,559	2739	844	36,799	353
Sampling characteristics							
GA 1st interview [1]	15	15	15	15	15	15	15
GA 2nd interview [1]	30	30	30	30	30	30	29
Child age 3rd interview [2]	6.3	6.3	6.3	6.3	6.3	6.3	6.3
Unplanned preg. (yes) [3]	22.1 %	22.5 %	21.2 %	23.2 %	25.0 %	22.0 %	32.9 %
Time to preg. (≥ 6 months)	26.4 %	27.3 %	25.1 %	26.8 %	26.5 %	26.4 %	27.8 %
Fertility treatment (yes)	5.7 %	6.0 %	5.4 %	5.8 %	4.6 %	5.8 %	3.1 %
Family characteristics							
Age (M) [4, 5]	30.7	30.0	31.1	32.4	33.6	30.7	31.2

Age (P) [4, 5]	32.5	31.8	32.8	34.2	35.8	32.5	33.3
Married (no) [6]	2.0 %	2.5 %	1.3 %	2.0 %	2.8 %	2.0 %	3.1 %
Education – mandatory (M) [7]	6.8 %	8.9 %	4.5 %	4.8 %	6.2 %	6.8 %	11.1 %
Education – university (M) [7]	15.8 %	11.9 %	18.9 %	23.6 %	22.4 %	15.8 %	11.9 %
Education – mandatory (P) [7]	12.2%	14.9 %	9.2 %	9.0 %	11.0 %	12.1 %	17.8 %
Education – university (P) [7]	16.4 %	12.6 %	19.6 %	23.7 %	22.2 %	16.5 %	9.7 %
Self-rep. pre-preg. psych.(M) [8]	6.9 %	7.1 %	6.3 %	7.1 %	10.0 %	6.8 %	12.2 %
In-preg. pre-preg. psych. (M) [9]	1.0 %	1.1 %	0.9 %	1.2 %	1.1 %	1.0 %	2.8 %
Pre-preg. psych. diag. (M) [10]	2.5 %	2.9 %	2.1 %	2.2 %	2.1 %	2.5 %	4.5 %
Pre-preg. psych. diag. (P) [10]	1.7 %	2.0 %	1.3 %	1.9 %	2.4 %	1.7 %	1.7 %
In-childhood psych. diag. (M) [10]	3.0 %	3.5 %	2.4 %	2.5 %	3.6 %	2.9 %	4.3 %
In-childhood psych. diag. (P)[10]	2.17%	2.54%	1.77%	1.68%	2.37%	2.16%	3.12%
Pre-pre. alc. drinks/ week (M) [5]	3.1	1.9	3.7	6.3	8.7	3.0	6.0
Binge drinking in preg. (M) [5]	0.54	0.37	0.62	0.99	1.60	0.52	3.32
Binge drinking in preg. (yes) [11]	30.9 %	23.9 %	35.9 %	45.5 %	53.9 %	30.2 %	100%
Cum. alc. intake in preg. (M) [5]	33.2	9.3	43.1	102.2	177.8	32.5	100.8
Cum. smoking in preg. (M) [5, 12]	53	64	39	49	80	53	130
Smoking (yes) [13]	22.8 %	24.9 %	19.5 %	22.9 %	30.2 %	22.6 %	42.6 %
Nicotine substitutes [14]	2.1 %	2.2 %	1.9 %	2.1 %	2.1 %	2.1 %	4.0 %
Partner smoking (yes) [15]	27.8 %	29.8 %	24.8 %	26.5 %	35.2 %	27.7 %	37.7 %
Vitamins (no) [14]	15.1 %	15.0 %	14.6 %	17.3 %	20.4 %	15.0 %	23.2 %
Iron supplement (no) [14]	27.8 %	28.1 %	27.1 %	29.1 %	31.0 %	27.8 %	28.1 %

Fish oil supplement (yes) [14]	4.9 %	5.0 %	4.6 %	5.0 %	7.0 %	4.8 %	7.1 %
Fish eating (never) [14]	3.2 %	4.3 %	2.2 %	1.1 %	1.4 %	3.2 %	2.0 %
Fish as warm meal (never)[14]	8.5 %	11.3 %	6.0 %	3.4 %	4.2 %	8.5 %	7.4 %
Pre-pregnant BMI [15]	30.3 %	34.8 %	26.5 %	22.4 %	20.9 %	30.3 %	32.5 %
Cola [14]	15.1 %	17.8 %	12.4 %	11.7 %	12.9 %	15.1 %	20.1 %
Coffee (yes) [14, 16]	41.4 %	35.5 %	44.7 %	59.3 %	65.8 %	41.2 %	55.9 %
Painkillers (yes) [14]	24.1 %	22.9 %	24.9 %	27.5 %	28.4 %	24.0 %	36.8 %
Diabetes (yes) [14]	1.9 %	2.1 %	1.7 %	1.6 %	1.9 %	1.9 %	1.4 %
Asthma (yes) [14]	3.2 %	3.6 %	2.8 %	2.9 %	3.1 %	3.2 %	3.1 %
Anaemia (yes) [14]	3.7 %	3.5 %	3.8 %	4.4 %	5.6 %	3.7 %	3.7 %
Television watching [14, 17]	20.0 %	23.3 %	17.2 %	14.7 %	13.7 %	20.0 %	29.0 %
Tenant [14, 18]	25.7 %	27.3 %	24.7 %	22.4 %	20.1 %	25.7 %	25.5 %
Exercise (yes) [14]	25.7 %	22.8 %	28.6 %	29.8 %	27.1 %	25.7 %	19.3 %
Child characteristics							
Sex (% boys) [19]	51.0 %	51.1 %	50.6 %	52.6 %	50.2 %	51.0 %	49.0 %
Baby not growing [20]	5.3 %	5.4 %	5.1 %	5.1 %	6.4 %	5.3 %	4.5 %
Birth weight [5, 19]	3654	3649	3660	3666	3607	3654	3641
SGA (%) [19, 21]	8.6 %	9.0 %	8.0 %	8.5 %	11.3 %	8.6 %	9.4 %
Gestational age [19, 22]	281.5	281.3	281.7	282.2	281.9	281.5	281.7
APGAR score (% < 10) [19, 23]	7.6 %	7.9 %	7.5 %	6.7 %	5.8 %	7.6 %	7.9 %
Child psych. diag. (yes) [24]	1.25%	1.47%	1.09%	0.80%	0.83%	1.26%	1.13%

Contributor statement page

Below follows contributor statement pages of the co-authored papers: Paper 1, Paper 2, Paper 3 and Paper 4. Each contributor statement page contains information on the division of work as well as agreements from the co-authors to include the Papers in this Ph.D. thesis.

Contributor statement page

Article title:

Psychometric Properties of the Danish Strengths and Difficulties Questionnaire: the SDQ Assessed for More than 70,000 Raters in Four Different Cohorts

Janni Niclasen is the first author of the paper

Janni Niclasen conceptualized and designed the study, carried out the initial analyses, discussed the interpretation of the results, drafted the initial manuscript, and approved the final manuscript.

Thomas William Teasdale is the second author of the paper

Dr. Teasdale helped with the analyses, discussed the interpretation of the results, critically reviewed and revised the manuscript, corrected the language, and approved the final manuscript.

Anne-Marie Nybo Andersen is the third author of the paper

Professor Nybo Andersen coordinated and supervised the DNBC data collection, critically reviewed and revised the manuscript and approved the final manuscript.

Anne Mette Skovgaard is the fourth author of the paper

Dr. Skovgaard coordinated and supervised the CCC2000 data collection, helped conceptualizing and designing the study, critically reviewed and revised the manuscript and approved the final manuscript

Hanne Elberling is the fifth author of the paper

Dr. Elberling helped with the coordination and supervision of the CCC2000 data collection, discussed the literature, critically reviewed and revised the manuscript and approved the final manuscript.

Carsten Obel is the sixth and last author of the paper

Dr. Obel helped with the analyses, discussed the interpretation of the results, critically reviewed and revised the manuscript, corrected the language, and approved the final manuscript. He had with Janni Niclasen the final scientific responsibility for the paper.

The authors accept by their signature that the article is made public as part of the PhD thesis.

Janni Niclasen
MSc in Psychology, Ph.D. fellow
Department of Psychology
University of Copenhagen

Thomas William Teasdale
Fil Dr. Dr. Med., associate professor
Department of Psychology
University of Copenhagen

Anne Marie Nybo Andersen
MD, Ph.D., professor
Section of Social Medicine
Department of Public Health

Anne Mette Skovgaard
MD, Fil. Dr., associate professor
Department of Public Health
University of Copenhagen

Hanne Elberling
MD, Ph.D.
Child and adolescent Psychiatric Centre Glostrup
Copenhagen University Hospital

Carsten Obel
MD, Ph.D., associate professor
Institute of General Medical Practice
Department of Public Health

Contributor statement page

Article title:
A Confirmatory Approach to Examining the Factor Structure of the Strengths and Difficulties Questionnaire (SDQ): A Large Scale Cohort Study

Janni Niclasen is the first author of the paper
Janni Niclasen conceptualized and designed the study, carried out the initial analyses, discussed the interpretation of the results, drafted the initial manuscript, and approved the final manuscript.

Anne Mette Skovgaard is the second author of the paper
Dr. Skovgaard coordinated and supervised the CCC2000 data collection, discussed the interpretation of the results, critically reviewed and revised the manuscript and approved the final manuscript.

Anne-Marie Nybo Andersen is the third author of the paper
Professor Nybo Andersen coordinated and supervised the DNBC data collection, critically reviewed and revised the manuscript and approved the final manuscript.

Mikael Julius Sømhovd is the fourth author of the paper
Mr. Sømhovd helped conceptualizing and designing the study, helped with the analyses, discussed the interpretation of the results, critically reviewed and revised the manuscript and approved the final manuscript

Carsten Obel is the fifth author of the paper
Dr. Obel helped with the analyses, discussed the interpretation of the results, critically reviewed and revised the manuscript, corrected the language, and approved the final manuscript. He had with Janni Niclasen the final scientific responsibility for the paper.

The authors accept by their signature that the article is made public as part of the PhD thesis.

Janni Niclasen
MSc in Psychology, Ph.D. fellow
Department of Psychology
University of Copenhagen

Anne Mette Skovgaard
MD, Fil. Dr., associate professor
Department of Public Health
University of Copenhagen

Anne Marie Nybo Andersen
MD, Ph.D., professor
Section of Social Medicine
Department of Public Health

Mikael Julius Sømhovd
MSc in Psychology, Ph.D. fellow
Department of Psychology
University of Copenhagen

Carsten Obel
MD, Ph.D., associate professor
Institute of General Medical Practice
Department of Public Health

Contributor statement page

Article title:

Prenatal exposure to alcohol and gender differences on child mental health at age seven

Janni Niclasen is the first author of the paper

Janni Niclasen conceptualized and designed the study, carried out the initial analyses, discussed the interpretation of the results, drafted the initial manuscript, and approved the final manuscript.

Anne-Marie Nybo Andersen is the second author of the paper

Professor Nybo Andersen coordinated and supervised the DNBC data collection, discussed the interpretation of the results, critically reviewed and revised the manuscript and approved the final manuscript.

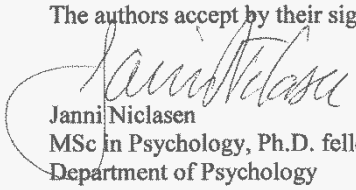
Thomas William Teasdale is the third author of the paper

Dr. Teasdale helped conceptualizing and designing the study, discussed the interpretation of the results, critically reviewed, corrected the language and revised the manuscript and approved the final manuscript.

Katrine Strandberg-Larsen is the fourth author of the paper

Dr. Strandberg-Larsen helped conceptualizing and designing the study, helped with the analyses, discussed the interpretation of the results, critically reviewed and revised the manuscript and approved the final manuscript

The authors accept by their signature that the article is made public as part of the PhD thesis.



Janni Niclasen
MSc in Psychology, Ph.D. fellow
Department of Psychology
University of Copenhagen

Anne-Marie Nybo Andersen
MD, Ph.D., professor
Section of Social Medicine
Department of Public Health

Thomas William Teasdale
Fil Dr. Dr. Med., associate professor
Department of Psychology
University of Copenhagen

Katrine Strandberg-Larsen
Cand. Scient., Ph.D., associate professor
Section of Social Medicine
Department of Public Health

Contributor statement page

Article title:
Is Alcohol Binge Drinking in Early and Late Pregnancy Associations with Behavioural and Emotional Development?

Janni Niclasen is the first author of the paper
Janni Niclasen conceptualized and designed the study, carried out the initial analyses, discussed the interpretation of the results, drafted the initial manuscript, and approved the final manuscript.

Anne-Marie Nybo Andersen is the second author of the paper
Professor Nybo Andersen coordinated and supervised the DNBC data collection, discussed the interpretation of the results, critically reviewed and revised the manuscript and approved the final manuscript.

Katrine Strandberg-Larsen is the third author of the paper
Dr. Strandberg-Larsen discussed the interpretation of the results, critically reviewed and revised the manuscript and approved the final manuscript

Thomas William Teasdale is the fourth author of the paper
Dr. Teasdale helped conceptualizing and designing the study, discussed the interpretation of the results, critically reviewed, corrected the language and revised the manuscript and approved the final manuscript.

The authors accept by their signature that the article is made public as part of the PhD thesis.

Janni Niclasen
MSc in Psychology, Ph.D. fellow
Department of Psychology
University of Copenhagen

Anne-Marie Nybo Andersen
MD, Ph.D., professor
Section of Social Medicine
Department of Public Health

Thomas William Teasdale
Fil Dr. Dr. Med., associate professor
Department of Psychology
University of Copenhagen

Katrine Strandberg-Larsen
Cand. Scient., Ph.D., associate professor
Section of Social Medicine
Department of Public Health

Paper 1

Niclasen J, Teasdale TW, Andersen A-MN, Skovgaard AM, Elberling H, et al. (2012). Psychometric Properties of the Danish Strength and Difficulties Questionnaire: The SDQ Assessed for More than 70,000 Raters in Four Different Cohorts. PLoS ONE 7(2): e32025. doi:10.1371/journal.pone.0032025

OPEN ACCESS Freely available online

PLoS one

Psychometric Properties of the Danish Strength and Difficulties Questionnaire: The SDQ Assessed for More than 70,000 Raters in Four Different Cohorts

Janni Niclasen[1]*, Thomas William Teasdale[1], Anne-Marie Nybo Andersen[2], Anne Mette Skovgaard[3], Hanne Elberling[3], Carsten Obel[4]

1 Department of Psychology, University of Copenhagen, Copenhagen, Denmark, 2 Department of Public Health, University of Copenhagen, Copenhagen, Denmark, 3 Child and Adolescent Psychiatric Centre Glostrup, Copenhagen University Hospital, Copenhagen, Denmark, 4 Department of Public Health, Aarhus University, Aarhus, Denmark

Abstract

Background: The Strength and Difficulties Questionnaire (SDQ) is a brief behavioural five factor instrument developed to assess emotional and behavioural problems in children and adolescents. The aim of the current study was to evaluate the psychometric properties for parent and teacher ratings in the Danish version of SDQ for different age groups of boys and girls.

Methods: The Danish versions of the SDQ were distributed to a total of 71,840 parent and teacher raters of 5-, 7- and 10- to 12-year-old children included in four large scale Danish cohorts. The internal reliability was assessed and exploratory factor analyses were carried out to replicate the originally proposed five factor structure. Mean scores and percentiles were examined in order to differentiate between low, medium and high levels of emotional and behavioural difficulties.

Results: The original five factor structure could be substantially confirmed. The Conduct items however did not solely load on the proposed Conduct scale and the Conduct scale was further contaminated by non-conduct items. Positively worded items tended to load on the Prosocial scale. This was more so the case for teachers than for parents. Parent and teacher means and percentiles were found to be lower compared to British figures but similar to or only slightly lower than those found in the other Nordic countries. The percentiles for girls were generally lower than for boys, markedly so for the teacher hyperactivity ratings.

Conclusions: The study supports the usefulness of the SDQ as a screening tool for boys and girls across age groups and raters in the general Danish population.

Citation: Niclasen J, Teasdale TW, Andersen A-MN, Skovgaard AM, Elberling H, et al. (2012) Psychometric Properties of the Danish Strength and Difficulties Questionnaire: The SDQ Assessed for More than 70,000 Raters in Four Different Cohorts. PLoS ONE 7(2): e32025. doi:10.1371/journal.pone.0032025

Editor: James G. Scott, The University of Queensland, Australia

Received November 28, 2011; **Accepted** January 19, 2012; **Published** February 27, 2012

Copyright: © 2012 Niclasen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Department of Psychology, University of Copenhagen; Ludvig og Sara Elsass Foundation; Aase og Ejnar Danielsens Foundation; Carl J. Becker's Foundation; Lundbeck Foundation; Børne- og Ungdomspsykiatrisk Selskab i Danmark; Dagmar Marshalls Foundation; Fonden til Lægevidenskabens Fremme; Direktør Jakob Madsens Legat. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: Janni.niclasen@psy.ku.dk

Introduction

Mental health problems of children and adolescents occur frequently in the general population with prevalence rates of psychopathology estimated from 7% in rural Brazil and Norway, 10% in Britain and Denmark and up to 15% in Russia and Bangladesh [1–6]. In Denmark one prevalence study indicated that approximately 10% of Danish 8–9-year-olds meet the DSM-IV criteria for a psychiatric diagnosis [7]. However, a considerable discrepancy has been found between prevalence rates and the number of children being treated through childhood and adolescence. This is disturbing as psychopathology developed in childhood shows stability over time and can progress into adult psychiatric disorders. Factors associated with the development of psychopathological disorders include age and gender, socioeconomic markers and family conditions [8]. The strength of these associations may however vary between cultural settings. In order to screen for mental health disorders in children and adolescents there is a need for instruments to assess for behavioural and emotional problems, which have been validated across cultural settings.

The strengths and difficulties questionnaire (SDQ) is a brief instrument developed to screen for child and adolescent psychopathology. It is used worldwide, has been translated into more than 60 languages, and has screening properties comparable with more comprehensive instruments [9]. It consists of 25 items and generates scores within five domains of psychological adjustment: Hyperactivity/Inattention (hereafter Hyperactivity), Peer problems, Conduct problems, Emotional symptoms and Prosocial behaviours. The items are based on key symptoms for

DSM-IV diagnoses and have been grouped into scales using exploratory factor analysis. The five hyperactivity items have for example been selected to assess hyperactivity, inattention and impulsiveness as these constitute the key symptoms for the DSM diagnosis of Attention/Deficit Hyperactivity Disorder (ADHD). The questionnaire is widely used for clinical as well as research purposes [10].

The SDQ appeals to researchers as well as clinicians for several reasons: firstly, because of its brevity, secondly because it covers key aspects of common childhood and adolescence psychopathology, and thirdly because it includes strengths as well as difficulties, which makes it more acceptable for parents, especially in the general population.

The SDQ has been used extensively in European as well as non-European contexts [10,11] since it was developed by Goodman in Britain in the late 1990s as an extension of the early work of Rutter [12]. A recently published review looking into the psychometric properties of the parent and teacher versions of the SDQ included 48 studies from 17 different cultural settings and a total of 131,223 raters [11]. Mean scores and cut-offs have shown some variation across cultural settings indicating some variations in the prevalence of child and adolescence psychopathology. British presented mean scores and cut-offs tend to be higher than northern European mean scores [13], but similar to or lower than the mean scores presented for the southern European countries [14]. Outside European settings, markedly higher than British mean scores have been reported for (non-western) Chinese and Brazilian children [3,15] but similar to American and Australian samples [16,17].

Most studies looking into the factor structure of the SDQ have applied exploratory factor analyses (EFA) and Principal component analysis (PCA). These have by and large found support for Goodman's predicted five factor model [15,18]. Few studies have applied confirmatory factor analysis (CFA) and those that have done so have not found unequivocal support for the five factor model [19–21]. Dickey and Blumberg found support for a three factor structure representing prosocial, internalising and externalising problems in an American sample of 4–17-year-olds and concluded that U.S. parents may construe conduct and peer problems differently from European parents [19]. Along the same lines, a British study concluded that there are advantages to using the broader internalising and externalising subscales for analyses in low-risk epidemiological samples, while one should retain the five subscales when screening for disorders among high-risk children [20]. On the other hand one thorough Norwegian study applying both confirmatory and exploratory factor analyses found none of the alternative models to fit the data better than a slightly modified version of Goodman's five factor model [21].

The discrepancies found in the existing literature for the mean scores and cut-offs, as well as for the factor structure therefore need further investigation. Culture plays a major role in the expression of psychosocial problems and for this reason previous investigations of discrepancies between studies have not been able to identify the extent to which they are expressions of true differences in scores and to what extent they are caused by demographic or cultural variations. In order to rule out any potential cultural and linguistic factors there is therefore a need for a study looking further into these variables within a homogenous cultural and linguistic setting from a large number of raters. Such study would also permit for cross-age, cross-gender and cross-rater comparisons.

Several large scale birth cohorts have been established in Denmark within the last few decades, a number of which have included the SDQ in their follow-up phases. Denmark may therefore, despite its small size, be the country in which the largest number of SDQ ratings has been collected. The aims of the current study were therefore: 1. to evaluate the internal reliability and the five domains of psychological adjustment supposedly evaluated by the SDQ by means of principal component analysis, and 2. to evaluate the mean scores and percentiles across age groups, gender and raters. This is performed for parent and teacher raters, boys and girls and 5-, 7- and 10–12-year-olds separately. It is hypothesised that sound reliability will be established, particularly so for the hyperactivity scale, and that the original proposed five scale factor structure will be confirmed. It is further hypothesised that mean scores and percentiles will be similar to those found in the other Nordic countries but lower compared to other European settings.

Methods

Samples

Included in the present study are data from four general population-based, large-scale birth cohorts, namely the Copenhagen Child Cohort (CCC2000), the Danish National Birth Cohort (DNBC), the Danish National Institute of Social Research's (DNISR) and the Aarhus Birth Cohort (ABC). Data come specifically from the 5-year follow up of the CCC2000, the 7-year follow up of the DNISR and DNBC and the 10–12-year-old of the ABC. Specific participation characteristics of the individual cohorts are shown in Table 1. Details of the methodology of the individual cohorts have been described in more detail elsewhere [22–25]. Drop-out rates were found to vary between cohorts. However, despite different drop-out rates all cohorts had contact information on most participating women (>99%). Thus, most of the non-participating women were non-responders. Compared to the background population the samples were under-represented regarding low socioeconomic resources (education, occupation, income and civil status), parents who were not born in Denmark; younger mothers; low maternal education; parents living separately at the time of birth; and changed family composition in the first five years of life [23,26–28].

The department where the study was carried out did not have an internal review board. However, the collection and analysis of data from the four databases was in each case approved of by regional ethic committees - De Videnskabsetiske Komiteer for Region Hovedstaden for CCC2000, DNBC and DNISR and De Videnskabsetiske Komiteer for Region Midtjylland for ABC. The parents and teachers in each of the four cohorts were in writing made aware that the data was used for research purposes and verbally gave their consent for the data being used for these purposes. The parent consent was required before any approach was made to the child's teacher. The regional ethics committees approved the use of these verbal informed consent procedures for each cohort.

Materials

The SDQ contains 25 questions and an Impact supplement. The 25 questions ask about different positive and negative aspects of the child's behaviour, and can be scored *'not true'*, *'somewhat true'* and *'certainly true'*. Of the 25 questions, 10 are generally thought of as strengths, 14 as difficulties and 1 as a neutral question. The items are divided into five scales (*Hyperactivity, Emotional, Conduct, Peer problem and Prosocial*) of five items each [12]. The first four scales are summed to obtain a total difficulties score whereas the Prosocial scale was included in order to enhance acceptability on part of the rater [12]. The questions have been selected on the basis of contemporary nosological concepts as well as factor analytically derived dimensions [12,18]. An extra Impact supple-

Table 1. Characteristics of the birth cohorts providing SDQ data for the study.

Cohort	Copenhagen Child Cohort	Danish National Birth Cohort	Danish National Institute of Social Research	Aarhus Birth Cohort
Acronym	CCC2000	DNBC	DNISR	ABC
Recruitment period	2000	1996–2002	1995	1990–1992
Study population: Eligible for the included follow-up	5,898	83,315*	5,233	8,244
Parent contribution of SDQ	3,349 (57%)	48,544 (58%)	4,971 (95%)	6,751 (82%)
Teacher contribution of SDQ	2,594 (44%)	-	-	5,631 (68%)
Age at SDQ screening	5	7	7	10–12

*As per October 2009.
doi:10.1371/journal.pone.0032025.t001

ment begins with one screening question asking whether the rater *"overall thinks that the child has difficulties in one or more of the following areas: emotions, concentration, behaviour or being able to get on with other people"*. If the rater answers *"yes"* to this question further items inquire about the severity of these difficulties. The Impact supplement provides an important estimate of the burden of the problems which is an essential part of the diagnostic criteria in the current diagnostic classification systems, ICD-10 and DSM-IV [12,18]. The Danish parent and teacher versions were translated in 2001, implementing standard back-translation procedures and using concepts and terms that were in keeping with time [29].

Statistical analyses

Analyses were carried out using the statistical package SPSS 18 and were conducted on unweighted data. Employed methods include scale reliability analyses, exploratory factor analysis by means of Principal Component Analysis and descriptive statistics. Because of the non-normal distribution of data all statistical group comparisons were carried out by means of the Mann-Whitney's U-test. For the sake of uniformity, responses of five items which were otherwise scored in a positive direction were inverted prior to their inclusion in the different analyses, and the item order was rearranged for visualisation purposes.

Results

Missing data

Goodman suggests that cases be included only when a minimum of three answers are given on any single scale [18]. In the present study the problem of missing values proved to be small and it was for this reason decided to include cases with a total of not more than one missing value. The employed sample sizes were thus 3,349 and 2,594 for parents and teachers of 5-year-olds, 53,515 for parents of 7-year-olds and 6,751 and 5,631 for parents and teachers of 10–12-year-olds comprising a total of 71,840 raters. In all parent samples there was a small overrepresentation of boys whereas there was a small overrepresentation of girls in the two teacher samples (app. 51/49%).

Validation of the scales

Initially, response frequencies for each of the 25 individual items were examined. It appeared that all items for all samples and raters were non-normally distributed with highly positively skewed distributions, especially so for the Conduct and Peer problem items. Particularly skewed were the two conduct items *"fights"* and *"steals"* with only 0.6 and 0.3% of responders agreeing the item to be *"certainly true"* and between 95.6% and 98.1% declaring it *"not true"*.

In order to determine the construct validity of the SDQ inter-item correlations were computed for the 7-year sample. All 20 problem-items as well as the five prosocial items were found to be positively correlated with each other which preliminary indicates that a single latent variable may influence the individual item responses. To further test this hypothesis Cronbach's Alphas were calculated including the option "scale if item is deleted". A higher Alpha appeared from these analyses only for the item *"somatic"* on the Emotional scale indicating that this item may cause some problem for the validity of the scale. However as it was only marginally higher (0.615 and 0.627) it was decided to retain the item for the remaining analyses.

Reliability

Cronbach's Alphas were also calculated for each subscale, the Total difficulties and the Impact score, individually for each subgroup, for parent and teacher raters separately and for boys and girls separately. Notwithstanding the fact that SDQ subscales only comprise five items, the coefficients were generally considered high. Highest estimates were found for the Hyperactivity scale (0.73–0.86) and for the 20 item Total difficulties scale (0.75–0.88) and lowest estimates for the Conduct scale (0.44–0.73). Reliabilities were generally found to be higher for boys than for girls and typically higher for teacher ratings compared to parent ratings for the individual subscales and total difficulties score, but lower so for the impact scores. These somewhat lower reliabilities for the Impact score may be broadly a result of the fact that teacher estimates are calculated on the basis of only three items whereas parent estimates are based on five items.

Inter-rater reliability

The 5- and 10–12-year-old dataset further allowed for an exploration of inter-informant correlations between parents and teacher ratings. For 5-year-olds Pearson's Product moment correlations were found to be: Hyperactivity: 0.42; Emotional: 0.33; Conduct: 0.33; Peer: 0.37; Prosocial: 0.29; Total difficulties: 0.45 and Impact: 0.41. For the 10–12-year-olds the corresponding figures were: Hyperactivity: 0.50; Emotional: 0.37; Conduct: 0.37; Peer: 0.49; Prosocial: 0.30, Total difficulties: 0.53 and Impact: 0.50. Comparing younger and older children it appears from the higher correlations for 10–12-year-olds that parent and teachers consistently rate older children more similar than younger ones.

Factor Structure

Since the internal consistency of the individual subscales and total difficulties scale were considered high, the next step was to determine the dimensionality of the SDQ. Principal component analyses (PCA) with Promax Rotation was carried out. Promax rotation was chosen as this rotation technique is particularly useful for large datasets. It was also chosen as it allows for correlations between factors and it produces both a pattern matrix and a structure matrix both of which are presented below. The values of the structure matrices are presented as they illustrate correlations between items and factors. The values of the pattern matrices are however also presented as they are similar to the easily interpretable values obtained in orthogonal rotations presented by most other researchers.

The analyses were firstly run separately for boys and girls for each of the four samples. The initial PCA analyses showed that the items generally loaded on the same factors between age-groups and gender. For this reason it was decided to pool the data into a large parent sample and a large teacher sample and run the analyses separately for these two groups. The extraction of the PCA were initially based on the number of Eigenvalues greater than 1 which resulted in a five factor solution for parents but a six factor solution for teachers. However, since the sixth factor had an Eigenvalue of 1.008 and only accounted for 4.03% of the variance it was decided to omit this factor from any further analyses and to run the analyses specifying the number of factors to be five.

It appears from Tables 2 and 3 that virtually all 25 items showed the highest loadings on their respective proposed scales. Teacher ratings showed higher loadings on their respective scales than did parent ratings. The values of the pattern matrices for both parents and teachers showed unequivocal high loadings on their proposed scales. The structure matrices on the other hand showed a somewhat more ambiguous picture. Conduct items showed high loadings on the other scales and non-conduct items loaded on the Conduct scale. Positively worded items further loaded on the Prosocial scale. This was more so for teacher raters compared to their parental counterparts.

Table 2. Principal Component Analysis with Promax rotation for parent ratings.

Parents 5–12-year-olds (N = 61,789)					
Principal component	1	2	3	4	5
Initial Eigenvalue:	4.51	2.00	1.65	1.33	1.19
Initial explained variance	18.04	8.00	6.58	5.31	4.76
Extracted Factors	**Hyperactivity**	**Prosocial**	**Emotional**	**Peer**	**Conduct**
Restless	**.75 (.77)**				
Fidgety	**.72 (.77)**				
Distracted	**.77 (.80)**				
Reflects	**.59 (.51)**	−.41			
Attends	**.74 (.78)**				
Somatic			**.41 (.42)**		
Worries			**.68 (.68)**		
Unhappy			**.61 (.55)**		
Clingy			**.68 (.60)**		
Afraid			**.70 (.71)**		
Tantrum					**.54 (.47)**
Obedient	.43	−.41			**.45 (.31)**
Fights				**.41**	**.55 (.47)**
Lies					**.66 (.67)**
Steals					**.53 (.64)**
Loner				**.62 (.62)**	
Friend				**.50 (.56)**	
Popular				**.62 (.58)**	
Bullied				**.53 (.44)**	**.40**
Oldbest				**.68 (.68)**	
Considerate		**.63 (.56)**			**−.43**
Shares		**.57 (.56)**			
Caring		**.67 (.70)**			
Kind		**.50 (.53)**			
Helpout		**.64 (.69)**			

Data from the Structure matrix is presented with data from the Pattern matrix in brackets. Factor loadings between +/−0.4 omitted. The bolded items show the proposed factor loadings.

doi:10.1371/journal.pone.0032025.t002

Table 3. Principal Component Analysis with Promax rotation for teacher ratings.

Teacher 5–12-year-olds (N = 6,829)					
Principal component	1	2	3	4	5
Initial Eigenvalue:	7.22	2.80	1.80	1.36	1.15
Initial explained variance	28.88	11.22	7.19	5.42	4.59
Extracted Factors	**Hyperactivity**	**Prosocial**	**Peer**	**Emotional**	**Conduct**
Restless	**.84 (.89)**				.45
Fidgety	**.80 (.88)**				
Distracted	**.84 (.91)**				
Reflects	.73 (.56)	−.57			.49
Attends	**.80 (.81)**	−.48			
Somatic				.55 (.55)	
Worries				.75 (.75)	
Unhappy			.46	.70 (.63)	
Clingy				.71 (.67)	
Afraid				.74 (.76)	
Tantrum	.42		.44		.67 (.57)
Obedient	.64 (.43)	−.51			.55 (.26)
Fights	.48	−.46			.77 (.67)
Lies	.41				.73 (.74)
Steals					.55 (.70)
Loner			.72 (.75)	.41	
Friend			.76 (.82)		
Popular		−.54	.72 (.59)		.49
Bullied			.82 (.59)		
Oldbest			.75 (.81)		
Considerate	−.51	.79 (.65)			−.54
Shares		.75 (.72)			−.40
Caring		.81 (.86)			
Kind		.73 (.77)			
Helpout		.76 (.88)			

Data from the Structure matrix is presented with data from the Pattern matrix in brackets. Factor loadings between +/−0.4 omitted. The bolded items show the proposed factor loadings.

doi:10.1371/journal.pone.0032025.t003

Mean scores and percentiles

Since the internal consistencies were found to be high and the factor structure could substantially be confirmed for boys and girls, younger and older children and parent and teacher ratings it was decided to examine any potential differences in scores between these groups. Tables 4 and 5 present the means and standard deviations (SD) for each of the five subscales, the Total difficulties and Impact scores for parent and teacher raters respectively. For each sample it appears that girls scored higher than boys on the Emotional and Prosocial subscales whereas boys scored higher on the Externalising (Conduct and Hyperactivity) and Peer scales. Parent and teachers alike rated older children as exhibiting fewer hyperactive and conduct problems and with more prosocial skills compared to younger ones. Teachers furthermore rated older children as also having more peer problems compared to younger ones. The statistical significance of these differences was examined using Mann Whitney-U tests. As could be expected given the very large sample sizes, most comparisons proved to be statistically significant ($P<0.05$). The effect sizes (Cohen's D) were found to be of medium size for the Hyperactivity, Prosocial and Total difficulties for all age groups and raters and also of medium size for teachers. Teachers generally rated girls and boys more dissimilarly than parents.

Following Goodman's recommendations with approximately 80% of children defined as being within a "*normal*" range, 10% in a "*borderline*" range and the highest 10% grouped in an abnormal or "*clinical*" range these percentiles were then calculated for the samples of 5–7- and 10–12-year-olds [12]. The upper percentile for the Total difficulties scores were for boys and girls in the present study found to be between 11 and 14 for parent ratings and between 12 and 18 for teacher ratings. As anticipated on the basis of the mean scores presented above, girls were generally rated as having fewer difficulties than boys, contributing to a broader range of scores for girls in the clinical percentile. This difference was particularly noticeable on the Hyperactivity scale which also contributed to the differences in Total difficulties score. Girls on the other hand had a narrower band of scores in the Prosocial banding indicating higher prosocial ratings. Comparing teacher with parent ratings the differences in scores on the Hyperactivity scale were even more marked, indicating that

Table 4. Mean sum scores and Standard deviations for 5-, 7- and 10–12-year-old parent ratings.

	5-year-olds (N = 3,288)						7-year-olds (N = 53,476)						10–12-year-olds (N = 5,031)					
Parents	Boys		Girls		Gender effects		Boys		Girls		Gender effects		Boys		Girls		Gender effects	
SDQ scale	Mean	SD	Mean	SD	P-values	Cohen's D	mean	SD	mean	SD	P-values	Cohen's D	mean	SD	mean	SD	P-values	Cohen's D
Hyperactivity	2.74	2.27	2.11	1.94	0.000	0.30	2.73	2.31	2.02	1.98	0.000	0.33	2.35	2.25	1.71	1.88	0.000	0.31
Emotional	1.58	1.73	1.59	1.75	0.930	0.01	1.59	1.75	1.66	1.76	0.000	0.04	1.59	1.90	1.72	1.88	0.000	0.07
Conduct	1.32	1.43	1.12	1.27	0.000	0.15	1.28	1.35	1.15	1.21	0.000	0.10	0.98	1.29	0.84	1.12	0.001	0.12
Peer	0.97	1.57	0.79	1.33	0.002	0.12	0.82	1.36	0.63	1.13	0.000	0.15	0.99	1.59	0.89	1.52	0.001	0.06
Prosocial	8.02	1.66	8.57	1.45	0.000	0.35	8.09	1.66	8.69	1.40	0.000	0.39	8.21	1.70	8.80	1.38	0.000	0.38
Total Difficulties	6.60	4.97	5.61	4.26	0.000	0.21	6.41	4.77	5.44	4.16	0.000	0.22	6.85	6.48	4.54	5.25	0.000	0.39
Impact	0.29	1.06	0.14	0.69	0.000	0.17	0.26	0.95	0.13	0.65	0.000	0.16	0.49	1.07	0.27	0.79	0.000	0.24

Gender effects show the 2-tailed p-values (Mann-Whitney U tests) with effect sizes (Cohen's D).
doi:10.1371/journal.pone.0032025.t004

teachers are more likely to rate boys differently than girls differently on this scale (please contact the first author for a table with the full details of the distribution of ranges and percentiles).

The percentiles were also compared to Goodman's British scores. For the Total difficulties scores the British *"clinical"* percentiles were found to be 17 for 5–15-year-old boys for parent as well as teacher raters but 15 and 12 for girls for parent and teacher raters respectively. Applying the parent scores of 17 and 15 for boys and girls respectively only included between 2.9% and 4% of the children in the present cohorts. The scores for teacher of 17 and 12 for boys and girls on the other hand included a larger proportion of the children, namely between 8.3% and 6.3% of the samples thus being more similar to the Danish distribution of scores.

Discussion

This article presents the psychometric properties of the Danish SDQ from a total of 71,265 raters after excluding data on the basis of missing values. To the authors' knowledge this is the first time that data from so many informants from the same cultural setting have been included in the same study. By contrast, a recently published review presented results from 48 studies from across the world with a total of 131,223 raters [11]. This review noted that the methodologies of the included studies varied making it difficult to compare them. Strengths of the present study are the inclusion of studies that apply similar methodologies and are derived from the same cultural setting creating a unique opportunity to investigate the psychometric properties of the SDQ between genders, ages and raters. It appears from the above presented analyses that the psychometric properties of the Danish version of the SDQ are strong, particularly for the teacher version.

The pattern matrices of the EFA replicated Goodman's five factor structure for parents and teachers. It appears from the higher teacher loadings that the questionnaire works a little better for teachers than for parents. Investigating the structure matrices, however, revealed two kinds of scale problems that are worth mentioning: firstly, that Conduct items load on non-conduct scales and conversely non-conduct items load on the Conduct scale and secondly that the positively worded items tend to load on the Prosocial scale. This is more so for teachers than for parents. With regard to the high loading of the Conduct items on the other scales it seems that these items are as much part of a hyperactivity construct as part of a notion of conduct for teachers. This is somewhat in line with a British study [20] applying CFA that concludes that the five subscales may not tap into distinct aspects of child mental health among low-risk, epidemiological samples which is exactly what characterises the four included samples. Instead one should use the broader Internalising and Externalising subscales. In regards to the positively worded items

Table 5. Mean sum scores and Standard deviations for 5- and 10–12-year-old teacher ratings.

	5-year-olds (N = 2,542)						10–12-year-olds (N = 4,264)					
Teachers	Boys		Girls		Gender effects		Boys		Girls		Gender effects	
SDQ scale	Mean	SD	Mean	SD	P-values	Cohen's D	mean	SD	mean	SD	P-values	Cohen's D
Hyperactivity	2.96	2.85	1.68	2.31	0.000	0.50	3.03	2.96	1.39	1.99	0.000	0.66
Emotional	1.31	1.76	1.41	1.79	0.112	0.06	1.28	1.93	1.43	2.00	0.003	0.08
Conduct	1.24	1.74	0.75	1.35	0.000	0.32	1.13	1.73	0.51	1.15	0.000	0.43
Peer	0.99	1.71	0.82	1.46	0.019	0.11	1.41	1.99	1.21	1.92	0.000	0.10
Prosocial	7.19	2.37	8.34	1.90	0.000	0.54	6.75	2.62	8.10	2.09	0.000	0.57
Total Difficulties	6.50	5.87	4.67	4.98	0.000	0.34	6.85	6.48	4.54	5.25	0.000	0.39
Impact	0.33	0.83	0.19	0.66	0.000	0.19	0.49	1.07	0.27	0.79	0.000	0.24

Gender effects show the 2-tailed p-values (Mann-Whitney U tests), with effect sizes (Cohen's D).
doi:10.1371/journal.pone.0032025.t005

this finding is in line with Goodman [18] who also found these items to load on the Prosocial scale. Although the positively worded items are precisely one of the advantages of this questionnaire they also seem to involve some psychometric drawbacks. Thus, although the five dimensions could overall be confirmed by examination of the pattern matrix (indicating no scale problems) the distinctiveness of the factors and some of the items do not seem particularly strong when one investigates the structure matrix that allows for cross-loadings between factors. For the clinicians this means that one should not put too much emphasis on the five individual subscales, much less use the SDQ as a diagnostic tool. These rater differences also illustrates the importance of running rater specific analyses.

The reliability estimates presented above are very similar to those found in other studies [11]. Sound reliability estimates and factor loadings of the hyperactivity scale indicate that the SDQ provides a solid estimate of symptoms of ADHD. The reliability of the Emotional scale has generally been reported as being poorer than what was found in this study, indicating that Danish parent and teachers may be better at reporting Internalising problems compared to other cultural settings. The Conduct subscale was, on the other hand, uniformly found to have the lowest reliability estimates and the lowest factor loadings, indicating a limitation of the usefulness of the scale within a low risk sample.

Lower reliability estimates were found for parents compared to teachers indicating that teachers are more likely than parents to view individual subscale items as measuring the same ability or trait. This may indicate that the subscale items may be viewed as less one-dimensional by parents caused by different tester attitudes. Conversely, teachers may be influenced by some sort of *"halo-effect"* which in the literature is referred to as the impact of one class of behaviour on the perception of another [30]. This means that children exhibiting problem behaviours in one area are more likely to be rated as problematic in other areas as well. Support for this hypothesis also comes from the teacher factor loadings where several items show high loadings on more than one subscale. Halo-effects have in the literature been found to show a different pattern for boys and girls and these tendencies could also contribute to the higher reliability estimates for boys than for girls [30].

The means and percentiles presented above are in line with those reported for other Scandinavian studies and somewhat lower on the Hyperactivity, Peer and Total difficulties scales compared to those found in other European and non-European studies [10,11]. The 90th percentile for the Total difficulties scores were for boys and girls in the present study found to be between 11 and 14 for parent ratings and between 12 and 18 for teacher ratings. These parent ratings are somewhat lower than the British recommendation of 17 [18] and Swedish of 14 [31] indicating that children of all the included age groups are rated as exhibiting fewer emotional and behavioural problems compared to other samples. Different explanations for the above described differences can be given. Firstly, they may indicate that Danish parents and teachers rate children more positively than do British parents and teachers. When the upper 10% British percentiles for boys and girls were applied for parent and teacher raters it appeared that the teacher ratings were more similar across cultures than the parent ratings indicating that this is only so for the parents. Secondly, it may be that the included samples are more selective and therefore less representative of the general population compared to the samples included in other studies. The present study is characterised by four large scale cohorts with attrition rates between 5 and 56% making the samples more or less non-representative of the general population biasing the included children toward a psychiatrically low-risk sample. This was particularly true for the large DNBC cohort. Since data were included in the analyses without compensatory weightings for underrepresented groups this may have introduced a potential source of bias. Thirdly, it may reflect actual behavioural and emotional differences in the Nordic countries characterised by better social security, low poverty, high living standards and less economic and social inequality. Meltzer et al. [8] found that children with mental disorder were more likely to live in lower income households, with a lone parent and in social sector housing. Denmark is characterised by a relatively homogenous population with a high level of social security which may cause fewer behavioural and emotional problems in the general populations.

Looking into potential gender differences boys were found to score higher than girls on the Hyperactivity, Conduct and Peer subscales and Total difficulties and Impact scores. Girls on the other hand were rated higher on the Emotional and Prosocial scales. Few other studies have reported potential significant differences between boys and girls [32]. The present study found medium to large effect sizes between boys and girls on the Hyperactivity, Conduct and Prosocial scales and Total difficulties scores. The present large-scale study has thus shown the importance of running the analyses separately for boys and girls. Failure to do so may potentially mask large differences between the sexes.

Younger children were in the present study found to score higher than older ones on the two Externalising subscales (Hyperactivity and Conduct scales). This is in similar vein to a German study [33] reporting younger children exhibiting more hyperactivity compared to older ones and a Dutch study [34] reporting a decline in parent ratings of total difficulties, emotional and hyperactivity scores with age. Interestingly, this same study reported increased total difficulties, conduct and emotional scores for teacher ratings as compared to parental ratings. Again, these results show the importance of running separate analyses for younger and older children.

Some limitations of the present study should be noted. The questionnaires from all the cohorts were mainly completed by mothers rather than fathers and this may have had an impact on the distribution of scores. Generally, other studies do not report on the gender distribution of the rater and this may cause some of the variability of scores across studies. Additionally, future studies would benefit from including information on socioeconomic risk factors. One study did find strong effects of social class on the Hyperactivity scale and somewhat less on the Peer scale [33] so controlling for a number of risk factors as for example second order factors in confirmatory factor analyses will further improve the findings of future studies. Future studies should further investigate different factor models using a confirmatory factor analytic approach. Finally, the SDQ is a widely used instrument in Danish epidemiologic studies and future work could advantageously examine the predictive validity of the SDQ as this is of prime importance in order to know how well the SDQ predicts future child, adolescent and adult psychiatric illness.

In conclusion, despite the above mentioned limitations this study demonstrates that SDQ is a well-functioning questionnaire with sound psychometric properties. The internal consistency is high, the factor structure could largely be confirmed and the means and percentiles make theoretical sense.

Acknowledgments

Many thanks to Professor Robert Goodman who constructively commented on an earlier draft of the manuscript.

Author Contributions

Conceived and designed the experiments: JN TWT AMNA AMS HE CO. Performed the experiments: JN TWT AMNA AMS HE CO. Analyzed the data: JN TWT CO. Contributed reagents/materials/analysis tools: JN TWT AMNA AMS HE CO. Wrote the paper: JN TWT AMNA AMS HE CO. Literature search: JN. Study design: JN CO TWT.

References

1. Ford T, Goodman R, Meltzer H (2003) The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. J Am Acad Child Adolesc Psychiatry 42(10): 1203–11.
2. Goodman R, Renfrew D, Mullick M (2000) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. Eur Child Adolesc Psychiatry 9(2): 129–34.
3. Goodman R, Neves dos SD, Robatto Nunes AP, Pereira de MD, Fleitlich-Bilyk B, et al. (2005) The Ilha de Mare study: a survey of child mental health problems in a predominantly African-Brazilian rural community. Soc Psychiatry Psychiatr Epidemiol 40(1): 11–7.
4. Goodman R, Slobodskaya H, Knyazev G (2005) Russian child mental health–a cross-sectional study of prevalence and risk factors. Eur Child Adolesc Psychiatry 14(1): 28–33.
5. Heiervang E, Stormark KM, Lundervold AJ, Heimann M, Goodman R, et al. (2007) Psychiatric disorders in Norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. J Am Acad Child Adolesc Psychiatry 46(4): 438–47.
6. Mullick MS, Goodman R (2005) The prevalence of psychiatric disorders among 5-10 year olds in rural, urban and slum areas in Bangladesh: an exploratory study. Soc Psychiatry Psychiatr Epidemiol 40(8): 663–71.
7. Bilenberg N, Petersen DJ, Hoerder K, Gillberg C (2005) The prevalence of child-psychiatric disorders among 8–9-year-old children in Danish mainstream schools. Acta Psychiatr Scand 111(1): 59–67.
8. Meltzer H, Gatward R, Goodman R, Ford T (2003) Mental health of children and adolescents in Great Britain. Int Rev Psychiatry 15(1–2): 185–7.
9. Goodman R, Scott S (1999) Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? J Abnorm Child Psychol 27(1): 17–24.
10. Achenbach TM, Becker A, Dopfner M, Heiervang E, Roessner V, et al. (2008) Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. J Child Psychol Psychiatry 49(3): 251–75.
11. Stone LL, Otten R, Engels RC, Vermulst AA, Janssens JM (2010) Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-Olds: A Review. Clin Child Fam Psychol Rev 13(3): 254–274.
12. Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 38(5): 581–6.
13. Obel C, Heiervang E, Rodriguez A, Heyerdahl S, Smedje H, et al. (2004) The Strengths and Difficulties Questionnaire in the Nordic countries. Eur Child Adolesc Psychiatry 13 Suppl 2: II32-9.:II32-II39.
14. Marzocchi GM, Capron C, Di PM, Duran TE, Duyme M, et al. (2004) The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. Eur Child Adolesc Psychiatry 13 Suppl 2: II40-6.:II40-II46.
15. Du Y, Kou J, Coghill D (2008) The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. Child Adolesc Psychiatry Ment Health 29;2(1): 8.
16. Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz DS (2005) The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. J Am Acad Child Adolesc Psychiatry 44(6): 557–64.
17. Mellor D (2004) Furthering the use of the strengths and difficulties questionnaire: reliability with younger child respondents. Psychol Assess 16(4): 396–401.
18. Goodman R (2001) Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry 40(11): 1337–45.
19. Dickey WC, Blumberg SJ (2004) Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. J Am Acad Child Adolesc Psychiatry 43(9): 1159–67.
20. Goodman A, Lamping DL, Ploubidis GB (2010) When to Use Broader Internalising and Externalising Subscales Instead of the Hypothesised Five Subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British Parents, Teachers and Children. J Abnorm Child Psychol 38(8): 1179–91.
21. Sanne B, Torsheim T, Heiervang E, Stormark KM (2009) The Strengths and Difficulties Questionnaire in the Bergen Child Study: a conceptually and methodically motivated structural analysis. Psychol Assess 21(3): 352–64.
22. Aarhus Birth Cohort website. Available: www.aarhus-born.dk. Accessed 2008 Apr 21.
23. Christensen E (2004) 7 års børneliv. Velfærd, sundhed og trivsel hos børn født i 1995. København: Socialforskningsinstituttet. Report No.: 04:13.
24. Olsen J, Melbye M, Olsen SF, Sorensen TI, Aaby P, et al. (2001) The Danish National Birth Cohort–its background, structure and aim. Scand J Public Health 29(4): 300–7.
25. Skovgaard AM, Olsen EM, Houmann T, Christiansen E, Samberg V, et al. (2005) The Copenhagen County child cohort: design of a longitudinal study of child mental health. Scand J Public Health 33(3): 197–202.
26. Elberling H, Linneberg A, Olsen EM, Goodman R, Skovgaard AM (2010) The prevalence of SDQ-measured mental health problems at age 5–7 years and identification of predictors from birth to preschool age in a Danish birth cohort: The Copenhagen Child Cohort 2000. Eur Child Adolesc Psychiatry 19(9): 725–735.
27. Jacobsen TN, Nohr EA, Frydenberg M (2010) Selection by socioeconomic factors into the Danish National Birth Cohort. Eur J Epidemiol 25(5): 349–355.
28. Nohr EA, Frydenberg M, Henriksen TB, Olsen J (2006) Does low participation in cohort studies induce bias? Epidemiology 17(4): 413–8.
29. Obel C, Dalsgaard S, Stax HP, Bilenberg N (2003) Spørgeskema om barnets styrker og vanskeligheder (SDQ-Dan). et nyt instrument til screening for psykopatologi i alderen 4–16 år. Danish Medical Bullentin 27;165/5: 462–5.
30. Jackson DA, King AR (2004) Gender differences in the effects of oppositional behavior on teacher ratings of ADHD symptoms. J Abnorm Child Psychol 32(2): 215–24.
31. Smedje H, Broman JE, Hetta J, von Knorring AL (1999) Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". Eur Child Adolesc Psychiatry 8(2): 63–70.
32. Shojaei T, Wazana A, Pitrou I, Kovess V (2009) The strengths and difficulties questionnaire: validation study in French school-aged children and cross-cultural comparisons. Soc Psychiatry Psychiatr Epidemiol 44(9): 740–7.
33. Woerner W, Becker A, Rothenberger A (2004) Normative data and scale properties of the German parent SDQ. Eur Child Adolesc Psychiatry 13 Suppl 2: II3-10.:II3-10.
34. van Widenfelt BM, Goedhart AW, Treffers PD, Goodman R (2003) Dutch version of the Strengths and Difficulties Questionnaire (SDQ). Eur Child Adolesc Psychiatry 12(6): 281–9.

Paper 2

Niclasen J, Skovgaard AM, Andersen A-MN, Sømhovd, MJ, Obel, C. (2012). A Confirmatory Approach to Examining the Factor Structure of the Strengths and Difficulties Questionnaire (SDQ): A Large Scale Cohort Study. Journal of Abnormal Child Psychology, 2013, Apr, 41(3), 355-365. DOI 10.1007/s10802-012-9683-y

J Abnorm Child Psychol
DOI 10.1007/s10802-012-9683-y

A Confirmatory Approach to Examining the Factor Structure of the Strengths and Difficulties Questionnaire (SDQ): A Large Scale Cohort Study

Janni Niclasen · Anne Mette Skovgaard · Anne-Marie Nybo Andersen · Mikael Julius Sømhovd · Carsten Obel

© Springer Science+Business Media New York 2012

Abstract The aim of this study was to examine the factor structure of the Strengths and Difficulties Questionnaire (SDQ) using a Structural Confirmatory Factor Analytic approach. The Danish translation of the SDQ was distributed to 71,840 parents and teachers of 5–7 and 10–12-year-old boys and girls from four large scale cohorts. Three theoretical models were examined: 1. a model with five first order factors (i.e., hyperactivity/inattention, conduct, emotional, peer problems and prosocial), 2. a model adding two internalising and externalising second order factors to model 1, and 3. a model adding a total difficulties second order factor to model 1. Model fits were evaluated, multi-group analyses were carried out and average variance extracted (AVE) and composite reliability (CR) estimates were examined. In this general population sample, low risk sample models 1 and 2 showed similar good overall fits. Best model fits were found when two positively worded items were allowed to cross load with the prosocial scale, and cross loadings were allowed for among three sets of indicators. The analyses also revealed that model fits were slightly better for teachers than for parents and better for older children than for younger children. No convincing differences were found between boys and girls. Factor loadings were acceptable for all groups, especially for older children rated by teachers. Some emotional, peer, conduct and prosocial subscale problems were revealed for younger children rated by parents. The analyses revealed more internal consistency for older children rated by teachers than for younger children rated by parents. It is recommended that model 1 comprising five first order factors, or alternatively model 2 with additionally two internalising/externalising second order factors, should be used when employing the SDQ in low risk epidemiological samples.

Keywords Strengths and difficulties questionnaire · SDQ · Psychometric properties · Factor structure · Confirmatory factor analysis · CFA · CR reliability · AVE reliability · Psychopathology · Mental health · Children · Adolescents · Cohort · Questionnaire

J. Niclasen (✉) · M. J. Sømhovd
Department of Psychology, University of Copenhagen,
Øster Farimagsgade 2A,
1353 Copenhagen K, Denmark
e-mail: janni.niclasen@psy.ku.dk

A. M. Skovgaard
Child and Adolescent Psychiatric Centre Glostrup,
Copenhagen University Hospital,
Copenhagen, Denmark

A.-M. N. Andersen
Department of Public Health, University of Copenhagen,
Copenhagen, Denmark

C. Obel
Department of Public Health, University of Aarhus,
Bartholins Allé 2, building 1260, room 126,
8000 Aarhus C, Denmark

Published online: 25 September 2012

The Strengths and Difficulties Questionnaire (SDQ) was developed by Goodman in the mid-1990's as a screening instrument aimed to cover the most prevalent areas of psychopathology in children and adolescents and designed to correspond to the diagnostic categories recognised by the two major diagnostic classification systems, i.e., the International Classification of Diagnosis (ICD-10) (World Health Organisation 1993) and the Diagnostic and Statistical Manual (DSM-IV) (American Psychiatric Association 1994) (Goodman 1994). The 25 SDQ items ask about five distinct domains of psychological adjustment among children and adolescents namely: hyperactivity/inattention, emotional symptoms, conduct problems, peer problems and prosocial behaviours. Apart from the five prosocial items, five problem items are also positively worded in order to enhance acceptability of the questionnaire in the

general population where the majority of children experience relatively few psychopathological difficulties (Goodman 1997; Goodman and Scott 1999).

The factor structure of the 25 SDQ items has been extensively assessed in different cultural settings by means of exploratory factor analysis (EFA) and most studies have been able to confirm the five factor structure (Goodman 2001; Koskelainen et al. 2000; Niclasen et al. 2012). However, as the development of the SDQ was theory driven and since it is assumed that the 25 items reflect five underlying latent dimensions, it seems more appropriate to validate the five scales by means of confirmatory factor analysis (CFA). CFA constitutes the measurement part of structural equation modeling (SEM). It is a technique that analyses measurement models in which both the number of factors and their corresponding indicators are explicitly specified a priori. Relatively few studies have employed structural confirmatory methods in relation to the SDQ and their results vary (Sanne et al. 2009; Van et al. 2008). Thus, some studies have found support for a five-factor model (Palmieri and Smith 2007; Sanne et al. 2009; Van et al. 2008) and others for a three-factor solution (Dickey and Blumberg 2004; Goodman et al. 2010). A study by Goodman et al. (2010) found a three-factor model (internalising/externalising/prosocial) to have a better fit in a low risk epidemiological sample of 5–16-year-olds, but that a five factor model was superior in high risk samples.

While one central issue is concerned with whether SDQ items are truly valid indicators of the proposed five behavioural domains or whether an even simpler structure would be superior, another key issue concerns the impact of the positively worded items. The inclusion of these items was originally intended to increase the acceptability of the SDQ to respondents, making it particularly suitable for use in non-clinical, epidemiological studies. The disadvantage however is, as several studies have pointed out, that positively worded items can confound the factor structure (Goodman 2001; Palmieri and Smith 2007). One study which included proxy data from custodial grandmothers found that a model which contained a positive construct method factor fitted the data better than the three- and five-factor models (Palmieri and Smith 2007). Similarly, a Norwegian study using self-rating data also found a significant improvement of the model fit by introducing a positive construct factor (Van et al. 2008). On the other hand, Sanne et al. (2009) did not find support for a positive construct factor for parent and teacher proxy data.

Thus, the advantages of the structural confirmatory methods are that they provide a comprehensive means for assessing and modifying theoretical models and therefore have a great potential for furthering theory development. The aims of the present paper are three fold. First, to examine how well three overall theoretical models fit data: Model 1. a five factor model (*hyperactivity/inattention, emotional, conduct, peer problems* and *prosocial*); Model 2. a five factor model with 2 s order factors (*internalising/externalising*); and Model 3. a five factor model with one latent *total difficulties* factor (Fig. 1). The three theoretical models are included as Goodman found the internalising/externalising model to have better overall fit as compared to the five-factor model in a low risk sample but did not test whether these two models were superior to the original proposed model with a total difficulties second order factor (Goodman et al. 2010). The models are here examined separately for parent ratings and teacher ratings, separately for both 5–7- and 10–12-year-old children and separately for boys and girls. Secondly, after examining the overall model fits, multi-group analyses are carried out in order to test for the presence of multi-group invariance, and to investigate in what ways the groups differ. Thirdly, two measures of reliability, average variance extracted (AVE) and composite reliability (CR), are examined.

Materials and Method

Samples

Data from the four population based, large scale birth cohorts, the Copenhagen Child Cohort (CCC2000), the Danish National Birth Cohort (DNBC), the Danish National Institute of Social Research's (DNISR) and the Aarhus Birth Cohort (ABC) were included in the present study (Table 1). Teacher ratings were available for the ABC and CCC2000 cohorts. The parent samples all had a small overrepresentation of boys whereas the opposite was true for the teacher samples and in all parent samples the questionnaires were mainly filled in by the mothers. As no differences in any analyses were found between the 5- and 7-year samples these were pooled for all analyses presented below and are denoted as younger children. In this way, the parent sample included a total of 63,615 ratings whereas the teacher samples added up to a total of 8,225 ratings.

Loss to follow up varied between the cohorts and various reasons may be responsible for these different response rates (Table 1). One explanation for the relatively low response rate of the DNBC could, for example, be that a large number of general practitioners refused to inform the pregnant women of the study. Similar for all samples, however, was that compared to the background population the samples were underrepresented regarding low socioeconomic resources (education, occupation, income and civil status), parents who were not born in Denmark; younger mothers; parents living separately at the time of birth; and changed family composition in the first 5 years of life (Aarhus Birth Cohort

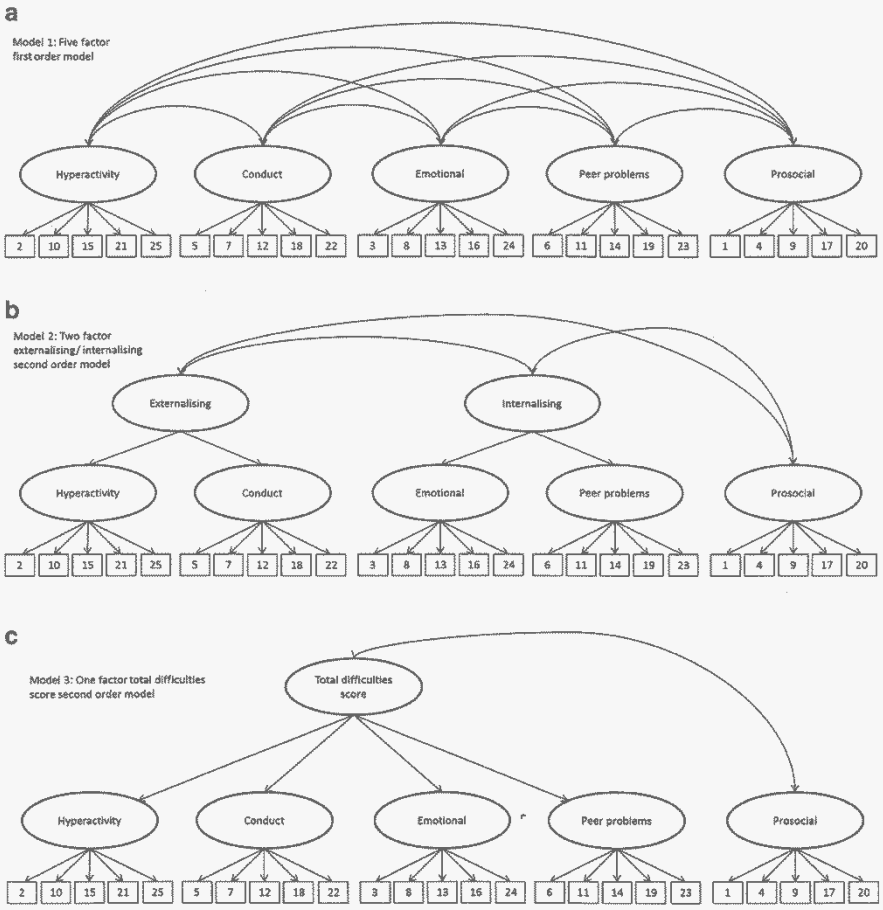


Fig. 1 Three theoretical models tested in CFA for each of the eight subgroups

2008; Christensen 2004; Elberling et al. 2010; Jacobsen et al. 2010; Nohr et al. 2006). The individual cohorts have been described in more detail elsewhere (Aarhus Birth Cohort 2008; Christensen 2004; Elberling et al. 2010;

Table 1 Characteristics of the birth cohorts providing SDQ data for the study

Cohort	Copenhagen Child Cohort	Danish National Birth Cohort	Danish National Institute of Social Research	Aarhus Birth Cohort
Acronym	CCC2000	DNBC	DNISR	ABC
Recruitment period	2000	1996–2002	1995	1990–1992
Study population: Eligible for the included follow-up	5,898	83,315[a]	5,233	8,244
Parent contribution of SDQ	3,349 (57 %)	48,544 (58 %)	4,971 (95 %)	6,751 (82 %)
Teacher contribution of SDQ	2,594 (44 %)	N/A[b]	N/A[b]	5,631 (68 %)
Age at SDQ screening	5	7	7	10–12

[a] As per October 2009; [b] *N/A* Not applicable

Olsen et al. 2001). Ethical approval was obtained for all of the studies.

Materials

The SDQ contains 25 questions asking about different positive and negative aspects of the child's behaviour. Responses are made on a three point Likert scale; *'not true','somewhat true'* and *'certainly true'*. Following the scoring recommendations, the items are divided into five subscales (*hyperactivity scale, emotional symptoms scale, conduct problem scale, peer problem scale and prosocial scale*) each comprising five items. The sum score of the first four subscales yields a *total difficulties score*. Parallel versions of the SDQ have been developed for parents, teachers and young persons (Goodman 1997; Goodman and Scott 1999).

Statistical Analyses

The method of Confirmatory Factor Analysis (CFA) was chosen as the appropriate means to test the three hypothesised models as it takes measurement error into account. All analyses were performed using the statistical package MPlus version 6.12. As the 25 items were rated on a non-redundant 3-point Likert scale and all items had skewed or indeed very skewed distributions, the data were treated categorically.

Previous research has found the weighted least square (WLS) method to be the superior estimator for CFA modelling of categorical data of exceptionally large samples sizes (Jöreskog and Sörbom 1996) and this estimator was applied for the two samples of younger boys and girls rated by parents (N=28,920 and 27,611 respectively). The weighted least square means and variance adjusted (WLSMV) on the other hand has been found to be superior with small to medium sample sizes (Brown 2006) and was initially applied for all analyses for the remaining six samples that varied in size between 1,272 and 3,322. The WLS estimator proved superior to the WLSMV within all samples and was therefore applied for all analyses for all samples throughout the study.

Model fits were evaluated by means of chi square test of model fit where 0 indicates a perfect fit, the Steiger-Lind root mean square error of approximation (RMSEA) where an RMSEA <0.08 indicates an acceptable model fit and <0.05 a good model fit, and Bentler comparative fit index (CFI) and Tucker-Lewis fit index (TFI), where CFI and TLI >0.90 signifies acceptable fits and >0.95 signifies good fits respectively (Schreiber et al. 2006). When certain parts of the model did not show acceptable fits, cross-loadings between specific indicators were allowed for on the basis of residual correlations and between factors and indicators based on modification indices. These modifications were only allowed for if they were considered to be theoretically meaningful.

Results

Missing Data

Kline suggests that less than 5 % of data missing on a single variable should be of little concern (Kline 2011). In the present study missing values were considered as missing at random (MAR); they constituted less than 0.05 % of all data and resulted in listwise deletion of cases. A further eleven cases were deleted due to lack of information on gender. The 71,840 cases were on these grounds reduced to 71,248.

Overall Model Fits: Factor Structure of the SDQ

Three different models were examined in the present study (Fig. 1). Model 1 was identical to Goodman's original factor structure with five hypothesised first-order factors (*hyperactivity/inattention, emotional, peer problems, conduct and prosocial).* Model 2 added 2 second-order *internalising/externalising* factors to Model 1 and Model 3 added 1 second-order *total difficulties* factor to Model 1. All models were tested separately as a function of informants (parent and teachers), ages (younger and older) and gender (boys and girls), yielding a total of eight subgroups.

Initially the five separate scales (*hyperactivity, conduct, emotional, peer problems* and *prosocial*) were examined as five individual models with one factor and five indicators each in order to specify five separate well working models. This procedure was carried out for each of the eight subgroups separately. These were then aggregated to a full Model 1 for each sample. Having identified eight best working, theoretically justified models, a number of cross-loadings that improved the models for all of the eight subsamples were identified. This was done in order to identify one overall well working model for all subsamples. The following three cross-loadings between indicators were identified as yielding improved model fits across all samples: item 22 ("*steals from home school or elsewhere*") and item 18 ("*often lies or cheat*"); item 10 ("*constantly fidgeting or squirming*") and item 2 ("*restless, overactive, cannot stay still for long*"); and item 20 ("*often volunteers to help others (parents, teachers, other children)*") with item 9 ("*helpful if someone is hurt, upset or feeling ill*"). These cross-loadings were not only permitted as they significantly improved model fits but also because they were considered theoretically meaningful. Items 22 and 18 are both concerned with delinquent behaviour, items 10 and 2 with

problems of keeping calm and sitting still and items 20 and 9 are both associated with helpful behaviour. Further, cross-loadings between the two positively worded items 21 ("*thinks things out before acting*") and item 14 ("*generally liked by other children*") were allowed to cross-load with the prosocial factor as this improved fit statistics significantly and was considered an appropriate means to capture response bias. Running Model 3 with these modifications resulted in non-convergent models with the implication that factor loadings could not be computed. This could indicate misspecifications in the model, or it could indicate that the model was overpowered because of the large sample sizes. As Model 3 was consistently found to have the poorest fits, these problems were not pursued further within the scope of this article. Thus for Model 3 only the raw, unadjusted model fits are presented.

Tables 2 and 3 present the initial measured, unadjusted model fits (in parentheses) along with the fits for the slightly modified models. The RMSEA values were considered good for all samples whereas the CFI and TLI were considered good for the teacher samples and acceptable for the parent samples. Inspection of the RMSEA, CFI and TLI revealed that Model 3 consistently had the poorest fits and further that the fits of Model 1 were generally somewhat better than the fits for Model 2. However, considering that Model 2 was the more parsimonious of the two models and considering that the differences of the fit statistics actually were minor, the fits for Model 1 and Model 2 were considered equally good. The two models could be compared statistically by means of the chi square difference test. However, since all the samples were large or extremely large, all yielded very large chi square values (Tables 2 and 3) and all chi square difference tests would in return be expected to prove highly significant. When such chi square difference tests were carried out they were indeed highly significant. This is because the data sets are so large and therefore overpowered, which means that even minor and trivial differences between the models will be found to be statistically highly significant. Because of this, the results of these analyses are not reported here. Another possible way to investigate whether there are true and meaningful differences between the models is by randomly selecting a number of smaller samples (e.g., N=250 or 500) drawn from the full cohort. If the differences remain in these smaller samples they can be considered nontrivial and important. This approach was carried out with N=250, 500 and 1000. However, most analyses resulted in non-identified models and results of these analyses are therefore not presented here.

Multi-group Analyses

In order to test for multi-group invariance the chi square contributions from each sample were used to carry out multi-group analyses for the modified Model 1 between parent and teacher raters and between boys and girls (Tables 4 and 5). As no information was available for the different age groups within the same samples, these analyses were not carried out. From the chi square values it appeared that the data fit Model 1 more convincingly for teachers than for parents. Possible reasons for this are described in more detail immediately below. It seems that higher factor loading and more explained total variance for individual items can explain at least part of the lower (and thus better) chi square values for teachers than for parents.

Table 2 SDQ parent Chi Square model fits, RMSEA, CFI and TLI for younger and older children and boys and girls separately. Fits with modifications (items 22–18, 10–2, 20–9 and the prosocial factor with positively worded items 21 and 14) are presented as are fits without modifications (in brackets)

Parent SDQ	Model	Chi Square	DF	RMSEA	CFI	TLI
Younger girls (N=27,611)	Model 1	7159 (10002)	260 (265)	0.031 (0.036)	0.893 (0.849)	0.877 (0.829)
	Model 2	7385 (10056)	263 (268)	0.031 (0.036)	0.890 (0.848)	0.874 (0.830)
	Model 3	(10688)	(270)	(0.037)	(0.839)	(0.821)
Younger Boys (N=28,920)	Model 1	8790 (12782)	260 (265)	0.034 (0.040)	0.906 (0.863)	0.892 (0.844)
	Model 2	9089 (12879)	263 (268)	0.034 (0.040)	0.903 (0.861)	0.889 (0.845)
	Model 3	(13642)	(270)	(0.041)	(0.853)	(0.837)
Older girls (N=3,237)	Model 1	1253 (1700)	260 (265)	0.034 (0.041)	0.934 (0.905)	0.924 (0.892)
	Model 2	1341 (1736)	263 (268)	0.036 (0.041)	0.929 (0.903)	0.919 (0.891)
	Model 3	(2123)	(270)	(0.046)	(0.911)	(0.901)
Older boys (N=3,322)	Model 1	1501 (2150)	260 (265)	0.038 (0.046)	0.938 (0.906)	0.928 (0.893)
	Model 2	1570 (2169)	263 (268)	0.039 (0.046)	0.935 (0.905)	0.925 (0.893)
	Model 3	(2265)	(270)	(0.047)	(0.900)	(0.889)

Table 3 SDQ teacher Chi Square model fits, RMSEA, CFI and TLI for younger and older children and boys and girls separately. Fits with modifications (items 22–18, 10–2, 20–9 and the prosocial factor with positively worded items 21 and 14) are presented as are fits without modifications (in brackets)

Teacher SDQ	Model	Chi Square	DF	RMSEA	CFI	TLI
Younger girls (N=1,291)	Model 1	1043 (1308)	260 (265)	0.048 (0.055)	0.955 (0.940)	0.948 (0.932)
	Model 2	1097 (1349)	263 (268)	0.050 (0.056)	0.952 (0.937)	0.945 (0.930)
	Model 3	(1458)	(270)	(0.058)	(0.931)	(0.924)
Younger boys (N=1,272)	Model 1	1100 (1502)	260 (265)	0.050 (0.061)	0.961 (0.943)	0.955 (0.935)
	Model 2	1132 (1542)	263 (268)	0.051 (0.061)	0.960 (0.941)	0.954 (0.934)
	Model 3	(1673)	265 (270)	(0.064)	(0.935)	(0.928)
Older girls (N=2,805)	Model 1	1491 (1903)	260 (265)	0.041 (0.047)	0.967 (0.957)	0.962 (0.951)
	Model 2	1513 (1935)	263 (268)	0.041 (0.047)	0.967 (0.956)	0.962 (0.951)
	Model 3	(2165)	(270)	(0.050)	(0.950)	(0.944)
Older boys (N=2,790)	Model 1	1903 (2515)	260 (265)	0.048 (0.055)	0.973 (0.963)	0.969 (0.958)
	Model 2	1953 (2535)	263 (268)	0.048 (0.055)	0.972 (0.963)	0.968 (0.958)
	Model 3	(2663)	(270)	(0.056)	(0.961)	(0.956)

Standardised Factor Loadings

One possible explanation for the differences in the multi-group analyses could be the observed differences in the standardised factor loadings; i.e., it is expected that the items (e.g., the five hyperactivity items) of an underlying factor (e.g., the hyperactivity scale) should show relatively high standardised loadings on that particular factor, but low loadings on other factors. Overall, higher loadings were found for the teacher samples compared to the parent samples (Table 6). Highest loadings were found for older children rated by teachers whereas lowest loadings were observed for younger children rated by their parents. No noteworthy differences were found between boys and girls. For all subsamples, the best parameter estimates were established for the hyperactivity scale indicating this to be psychometrically most satisfactory scale. Virtually all items on all scales were considered high for the teacher ratings and were all considered good. However, low standardised loadings were consistently found in most samples for the emotional item 3 ("*often complains of headaches, stomach-aches or sickness*"). It should be noted that the relatively low loadings of items 14 *("generally liked by other children")* and 21 *("thinks things out before acting")* are caused by their cross-loadings with the prosocial factor.

Explained Total Variances for the Observed Variables

Another plausible explanation for the differences reported in the multi-group analyses above are differences in the values of R^2 (Table 7). R^2 refers to the magnitude of proportion of variance for each observed variable that is accounted for by its related latent factor. Values of R^2 are computed by subtracting the square of the residual from 1. The values of R^2 should preferably be >0.50 indicating that at least 50 % of the total variance of that indicator has been explained by the model, with the remaining unexplained parts of the variance being attributable to other, residual factors. Values of R^2< 0.50 are considered critically low since more than 50 % of the variance is then explained by factors other than the test item itself. The values of R^2 were consistently found to be much higher for teacher ratings than for parent ratings and also markedly higher for older children than for younger children. For older children with teacher raters, all R^2values explained more than 50 % of the total variance indicating that all items work well. By contrast, for younger children rated by their parents as many as 16 and 14 out of the 25 items (for girls and boys respectively) explained <0.50 of the total variance indicating severe problems with several test items for this age groups with parent raters. These

Table 4 chi square multi-group comparisons between parents and teachers. Chi Square contributions from each subsample is presented

	Parents	Teachers
Younger girls	1896 (N=1630)	1699 (N=1291)
Younger boys	1964 (N=1694)	1584 (N=1272)
Older girls	2263 (N=3237)	2153 (N=2805)
Older boys	2898 (N=3322)	2723 (N=2790)

Table 5 Multi-group comparisons between boys and girls. Chi Square contributions from each subsample is presented

	Boys	Girls
Parents younger children	9685 (N=28920)	8398 (N=27611)
Parents older children	1671 (N=3322)	1471 (N=3237)
Teachers younger children	1305 (N=1272)	1349 (N=1291)
Teachers older children	2120 (N=2790)	1772 (N=2805)

Table 6 Factor loadings for the separate parent and teacher samples for each of the indicators of the five latent variables (for the modified Model 1 that allows unique variance to correlate between factors and indicators)

	Items	Parent SDQ				Teacher SDQ			
		Younger girls	Younger boys	Older girls	Older boys	Younger girls	Younger boys	Older girls	Older boys
Hyperactivity/ Inattention	2	0.73	0.77	0.74	0.75	0.94	0.90	0.92	0.90
	10	0.69	0.72	0.71	0.66	0.93	0.88	0.85	0.86
	15	0.87	0.87	0.93	0.93	0.94	0.95	0.89	0.95
	21	0.49	0.51	0.53	0.69	0.64	0.53	0.61	0.60
	25	0.81	0.82	0.85	0.87	0.88	0.90	0.94	0.94
Emotional Problems	3	0.40	0.37	0.48	0.52	0.61	0.43	0.74	0.79
	8	0.61	0.62	0.70	0.74	0.81	0.82	0.82	0.80
	13	0.76	0.75	0.83	0.84	0.88	0.78	0.97	0.93
	16	0.66	0.67	0.75	0.82	0.80	0.84	0.87	0.86
	24	0.73	0.74	0.73	0.76	0.75	0.89	0.79	0.88
Conduct problems	5	0.57	0.62	0.67	0.71	0.87	0.87	0.90	0.87
	7	0.63	0.63	0.68	0.61	0.80	0.85	0.87	0.88
	12	0.77	0.82	0.85	0.81	0.95	0.91	0.96	0.93
	18	0.62	0.58	0.73	0.65	0.84	0.73	0.94	0.87
	22	0.48	0.46	0.56	0.53	0.87	0.66	0.87	0.84
Peer problems	6	0.60	0.69	0.65	0.73	0.84	0.89	0.86	0.85
	11	0.47	0.54	0.63	0.61	0.70	0.82	0.90	0.92
	14	0.58	0.61	0.68	0.74	0.62	0.55	0.65	0.58
	19	0.71	0.74	0.88	0.84	0.68	0.81	0.89	0.83
	23	0.67	0.75	0.81	0.85	0.88	0.88	0.84	0.86
Prosocial	1	0.84	0.87	0.83	0.91	0.92	0.95	0.96	0.96
	4	0.62	0.60	0.70	0.68	0.82	0.81	0.86	0.87
	9	0.58	0.59	0.63	0.66	0.78	0.82	0.81	0.87
	17	0.53	0.56	0.59	0.56	0.84	0.79	0.75	0.80
	20	0.46	0.45	0.54	0.52	0.53	0.65	0.71	0.73

marked differences of R^2 values between parent and teacher ratings can explain some of the differences in the multigroup analyses found above. Virtually no differences were observed in R^2 values between boys and girls. The value of $R^2$for item 3 *("often complaints of headaches...")* was the lowest for virtually all subsamples. For parent raters and younger children the item was consistently and critically low (e.g., for parents rating younger boys: 0.367^2=13.5 % of the total variance, leaving 86.5 % unexplained). Neither allowing the item to load on to other items, or factors nor removing the item altogether, increased either the general fits of the models, or the total variance explained by that item.

Reliability Measures

In order to evaluate the internal consistency of the individual scales, i.e., to what degree the scores are free from random measurement error, composite reliability (CR) and average variance extracted (AVE) were calculated. Although Cronbach's Alpha is the most commonly used measure of reliability in the literature, it is not reported here as it is a conservative measure of reliability which assumes that all items contribute equally to the reliability, i.e., it estimates how the full scale works rather than taking account of the variance and measurement error of the individual items. The AVE and CR on the other hand are reported here as they take complexity into account and do not assume that all items add equally to the reliability of the factor in question. CR is specifically concerned with the composite of the items taking into account the standardised loadings and the measurement errors of each of them. If CR >0.70 then satisfactory scale reliability is typically considered to have been established. AVE on the other hand is a measure that indicated how much variance is, on average, explained. If an item is overall poor for its scale it will result in a low AVE (<0.50) (Fornell and Larcker 1981). It appears from Table 8 that all CR's were above 0.7 indicating good scale reliability for all scales for all subsamples. It should be noted, however, that the lowest values of CR were found for younger children with parent raters and highest values were found for older

Table 7 values of R^2 for the separate parent and teacher samples for each of the indicators of the five latent variables (for the modified Model 1 that allows unique variance to correlate between factors and indicators)

	Items	Parent SDQ				Teacher SDQ			
		Younger girls	Younger boys	Older girls	Older boys	Younger girls	Younger boys	Older girls	Older boys
Hyperactivity/ Inattention	2	0.53	0.59	0.54	0.56	0.89	0.80	0.85	0.81
	10	0.47	0.53	0.50	0.43	0.86	0.77	0.72	0.74
	15	0.76	0.76	0.86	0.87	0.88	0.90	0.80	0.90
	21	0.44	0.49	0.52	0.56	0.75	0.68	0.74	0.76
	25	0.65	0.67	0.73	0.76	0.77	0.80	0.89	0.88
Emotional problems	3	0.16	0.14	0.23	0.27	0.37	0.18	0.55	0.62
	8	0.37	0.39	0.49	0.55	0.66	0.66	0.68	0.64
	13	0.58	0.56	0.69	0.70	0.77	0.61	0.94	0.86
	16	0.43	0.46	0.56	0.67	0.64	0.71	0.75	0.74
	24	0.53	0.55	0.54	0.58	0.57	0.79	0.62	0.77
Conduct problems	5	0.32	0.39	0.45	0.51	0.75	0.76	0.81	0.76
	7	0.40	0.40	0.46	0.37	0.63	0.72	0.76	0.78
	12	0.59	0.67	0.73	0.66	0.90	0.83	0.92	0.87
	18	0.39	0.33	0.53	0.42	0.71	0.53	0.89	0.75
	22	0.23	0.21	0.32	0.28	0.75	0.43	0.76	0.70
Peer problems	6	0.36	0.47	0.42	0.53	0.71	0.79	0.73	0.72
	11	0.22	0.29	0.40	0.37	0.50	0.68	0.80	0.85
	14	0.54	0.64	0.73	0.76	0.86	0.87	0.91	0.90
	19	0.51	0.54	0.78	0.71	0.47	0.65	0.79	0.69
	23	0.45	0.56	0.66	0.72	0.78	0.76	0.71	0.74
Prosocial	1	0.71	0.76	0.69	0.83	0.85	0.91	0.93	0.91
	4	0.38	0.36	0.49	0.46	0.68	0.66	0.73	0.76
	9	0.33	0.35	0.40	0.43	0.62	0.67	0.65	0.75
	17	0.28	0.31	0.35	0.31	0.70	0.63	0.56	0.64
	20	0.21	0.20	0.30	0.27	0.28	0.43	0.51	0.54

Table 8 Composite Reliability (CR) and Average Variance Extracted (AVE) for the separate parent and teacher subsamples

	Reliability	SDQ parents				SDQ teachers			
		Younger girls	Younger boys	Older girls	Older boys	Younger girls	Younger boys	Older girls	Older boys
Hyperactivity/ inattention	CR	0.86	0.82	0.88	0.89	0.96	0.94	0.95	0.95
	AVE	0.55	0.48	0.61	0.62	0.82	0.77	0.78	0.80
Emotional problems	CR	0.77	0.77	0.83	0.86	0.88	0.87	0.92	0.93
	AVE	0.41	0.42	0.50	0.55	0.60	0.59	0.71	0.72
Conduct problems	CR	0.75	0.76	0.83	0.80	0.94	0.90	0.96	0.94
	AVE	0.39	0.40	0.50	0.45	0.75	0.65	0.82	0.77
Peer problems	CR	0.76	0.82	0.87	0.88	0.89	0.93	0.94	0.94
	AVE	0.39	0.47	0.58	0.60	0.63	0.72	0.77	0.75
Prosocial	CR	0.75	0.76	0.79	0.80	0.89	0.90	0.91	0.93
	AVE	0.38	0.40	0.44	0.46	0.62	0.66	0.68	0.72
Externalising	CR	0.87	0.88	0.91	0.90	0.94	0.90	0.95	0.95
	AVE	0.78	0.79	0.83	0.82	0.89	0.82	0.90	0.91
Internalising	CR	0.79	0.82	0.85	0.88	0.88	0.88	0.91	0.89
	AVE	0.66	0.69	0.73	0.78	0.78	0.79	0.84	0.81
Total	CR	0.88	0.87	0.87	0.92	0.94	0.91	0.96	0.95
	AVE	0.64	0.64	0.63	0.76	0.78	0.73	0.85	0.83

children with teacher raters, indicating that the individual scales work better in the latter situation. No substantial differences were found between boys and girls. From the sizes of AVE it appears that all factors work well for older children rated by teachers and also that no items from the hyperactivity/inattention subscale are problematic for any of the subsamples. Single items on the emotional, conduct, peer and prosocial scales, on the other hand, do create problems for these scales for younger children rated by parents, resulting in poor values of AVE. This is, however, not surprising since 14 items and 16 items out of 25 explained <0.50 of the total variance for these samples of boys and girls respectively.

Discussion

The aim of the present study was to examine how well 71,248 SDQ ratings, divided into eight subgroups, fit three theoretically based models by means of confirmatory factor analysis. It was concluded that Model 1including five latent first order factors and Model 2 including a further two *internalising/externalising* second order factors both have good fits and work equally well. Model 3, which included one *total difficulties* second order factor was throughout all samples found to be less satisfactory than Model 1 and Model 2. Also, data from teachers seem to fit the models better than data from parents, and from older children better than for younger ones. No differences were found between boys and girls. Although Model 1 and Model 2 overall are well working several of the findings call for closer inspection.

Firstly, it appears that both Model 1 and Model 2 show good overall fit statistics in the present study of low risk epidemiological samples. This finding is somewhat in contrast to Goodman et al. (2010) who concluded that the broader internalising and externalising SDQ subscales of Model 2 are superior in low risk epidemiological samples. Contrary to the present study, however, Goodman et al. did not subdivide their sample on the basis of gender or age. Considering then that the present study did so and did find rather large intergroup differences on the basis of age and rater (but not on the basis of gender), this may partly explain these somewhat contradictory findings. The findings of the present study suggest that different models can be advantageously examined by subdividing the sample on the basis of age and rater (but not necessarily on gender) leading to better and more accurate model fits. The sample used in the study by Goodman et al. included 5–16-year-old children and this large age span may have masked potential differences between subgroups. Another potential explanation for this discrepancy in results is that the differences are genuine and are caused by cultural differences. Compared to Goodman's British cultural setting, Denmark is probably more homogenous in terms of access to the education and health care systems—services that are all tax-financed and free of charge for the citizens. Examining the influence of such cultural and societal differences on the factor structure of the SDQ remains to be carried out, but it certainly would be both an interesting and highly relevant study for future research considered the global and widespread use of the SDQ.

Secondly, all models significantly benefitted from minor model modifications, i.e., allowing the two positively worded items 14 and 21 to cross-load with the prosocial factor and allowing cross-loadings between items 22–18, 10–2 and 20–9. These modifications represent systematic, rather than random, measurement errors in item responses and they may derive from characteristics which are specific either to the test items or to the respondents (Byrne 2011). In other words, reversed items 14 and 21 not only relate to their respective factors (*peer problems* and *hyperactivity*) but they also reflect response bias and some underlying prosocial behaviour. It is recommended that that the above model modifications should be applied for future research purposes. This is, however, not feasible in clinical settings and there it is instead recommended that sum scores be retained, as also originally recommended by Goodman.

Thirdly, one of the major advantages of structural equation modeling is that it provides a comprehensive means for assessing and modifying theoretical models. The findings presented in the present paper suggest a cautious future use of positively worded, reversed items in questionnaires of this type, as this may contaminate the factor structure of the questionnaire. The present study tested for, but did not find, support for a positive construct factor (the results were not presented here). However, there are still many and very good reasons to include positively worded items in a questionnaire of this type. Firstly, as noted by Goodman, because it enhances acceptability of the questionnaire on the part of the rater, especially so in the general population (Goodman 1994). Secondly, because it expands the description of the mental health functioning of the child by including non-pathological traits. By adding assessment of mental health strengths, the questionnaire informs about possible protective or resilience factors, which might be of particular importance in the investigation of developmental psychopathology.

Fourthly, item 3 *"often complains of headaches, stomach-aches or sickness"* repeatedly showed poor factor loadings and explained critically little of the total variance throughout most analyses. Neither removing the item, nor allowing the item to cross-load with other items or scales improved the model. However, it was retained in the model as it did contribute significantly to the overall model fits. There may be several reasons for this item fitting the SDQ so poorly: 1. from a closer inspection of the wording of the item, it appears that it is actually the only one of the 25 items that relies on some sort of self-report on the part of the child. The remaining 24 items solely rely on evaluation on part of

the rater. 2. The item appears to be *state* dependent reflecting the state of the child at a particular moment in time, whereas the remaining 24 items appear to reflect traits i.e., relatively time-stable individual characteristics. In other words, the item may represent an unspecific marker of impact, probably expressed by age appropriate somatic symptoms, rather than as a direct psychopathological trait or symptom. 3. very little of item 3's total variance is explained. In other words, when children complain of headaches, one cannot be certain that they actually have a headache. Instead, it may indicate that they experience other sorts of unspecified problems.

Finally, the questionnaire was found to be superior for teacher compared to parent raters and for older children compared to younger ones. These differences were found between the different subsamples on all levels of analyses, namely on an overall model level, a factor level and an item level and they point to the importance of running at least age and rater specific analyses in future work with the SDQ.

Limitations and Future Work

A limitation of the present study is the lack of access to a high risk sample. It is not known whether one model would prove superior to the other within such a setting, as was concluded in the study by Goodman et al. (2010). Future studies should replicate the analyses of the present study using high risk, clinical samples, in order to investigate whether the present findings hold true across such groups. The participation rates in some of the published samples are rather low and this could potentially have had an effect on the results. The substantial size of the sample has allowed for very specific comparisons of item functioning across the different samples of ages, raters and gender. Such analyses were beyond the scope of the present article but will be a highly relevant focus for future studies.

Acknowledgments I would like to thank associate professor Thomas William Teasdale, PhD DM Sci. for commenting on an earlier draft of this manuscript and Jan Ivanouw, M.Sc. PhD for psychometrics advice.

Funding The study was financially supported by Institute of Psychology, University of Copenhagen; the Lundbeck Foundation; The Carl J Beckers Foundation; Børne- og Ungdomspsykiatrisk Selskab i Danmark; Direktør Jacob Madsen og Hustru Olga Madsens Foundation; The A.P. Møller Foundation for the Advancement of Medical Science; The Dagmar Marshall Foundation; Aase og Ejnar Danielsens Foundation; The Ludvig and Sara Elsass Foundation.

References

Aarhus Birth Cohort. (2008). www.aarhus-born.dk.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*. Washington DC: DSM-IV.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Byrne, B. (2011). *Structural equation modeling with MPlus*. Routledge.

Christensen, E. (2004). *7 års børneliv. Velfærd, sundhed og trivsel hos børn født i 1995 (Rep. No. 04:13)*. København: Socialforskningsinstituttet.

Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 1159–1167.

Elberling, H., Linneberg, A., Olsen, E. M., Goodman, R., & Skovgaard, A. M. (2010). The prevalence of SDQ-measured mental health problems at age 5–7 years and identification of predictors from birth to preschool age in a Danish birth cohort: The Copenhagen Child Cohort 2000. *European Child & Adolescent Psychiatry, 19*, 725–735.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39–50.

Goodman, R. (1994). A modified version of the Rutter parent questionnaire including extra items on children's strengths: a research note. *Journal of Child Psychology and Psychiatry, 35*, 1483–1494.

Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1337–1345.

Goodman, R., & Scott, S. (1999). Comparing the strengths and difficulties questionnaire and the child behavior checklist: is small beautiful? *Journal of Abnormal Child Psychology, 27*, 17–24.

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the strengths and difficulties questionnaire (SDQ): data from british parents, teachers and children. *Journal of Abnormal Child Psychology, 38*, 1179–1191.

Jacobsen, T. N., Nohr, E. A., & Frydenberg, M. (2010). Selection by socioeconomic factors into the Danish National Birth Cohort. *European Journal of Epidemiology, 25*, 349–355.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference gouide*. Chicago: Scientific Software International.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. (Third ed.) Guilford.

Koskelainen, M., Sourander, A., & Kaljonen, A. (2000). The strengths and difficulties questionnaire among Finnish school-aged children and adolescents. *European Child & Adolescent Psychiatry, 9*, 277–284.

Niclasen, J., Teasdale, T. W., Andersen, A. M., Skovgaard, A. M., Elberling, H., & Obel, C. (2012). Psychometric properties of the Danish strength and difficulties questionnaire: The SDQ assessed for more than 70,000 raters in four different cohorts. *PLoS One, 7*, e32025.

Nohr, E. A., Frydenberg, M., Henriksen, T. B., & Olsen, J. (2006). Does low participation in cohort studies induce bias? *Epidemiology, 17*, 413–418.

Olsen, J., Melbye, M., Olsen, S. F., Sorensen, T. I., Aaby, P., Andersen, A. M., et al. (2001). The Danish national birth cohort—its background, structure and aim. *Scandinavian Journal of Public Health, 29*, 300–307.

Palmieri, P. A., & Smith, G. C. (2007). Examining the structural validity of the strengths and difficulties questionnaire (SDQ) in a U.S. sample of custodial grandmothers. *Psychological Assessment, 19*, 189–198.

Sanne, B., Torsheim, T., Heiervang, E., & Stormark, K. M. (2009). The strengths and difficulties questionnaire in the Bergen child study:

a conceptually and methodically motivated structural analysis. *Psychological Assessment, 21*, 352–364.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Educational Research, 99*, 323–337.

Van, R. B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor strengths and difficulties questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry, 49*, 1304–1312.

World Health Organisation. (1993). The ICD-10 classification of mental and behavioural disorders diagnostic criteria for research.

Paper 3

Niclasen, J., Andersen, AMN., Teasdale, TW., Strandberg-Larsen, K. Prenatal exposure to alcohol and gender differences on child mental health at age seven (submitted for publication)

Prenatal exposure to alcohol and gender differences on child mental health at age seven

Niclasen, J., Andersen, AMN., Teasdale, TW., Strandberg-Larsen, K.

Authors:

Janni Niclasen. M.Sc., Ph.D. fellow, Institute of Psychology, University of Copenhagen, janni.niclasen@psy.ku.dk

Anne Marie Nybo Andersen, MD, Ph.D., professor, Department of Public Health, University of Copenhagen, amny@sund.ku.dk

Thomas William Teasdale, M.Sc., Fil. Dr., Dr. Med., associate professor, Institute of Psychology, University of Copenhagen, tom.teasdale@psy.ku.dk

Katrine Strandberg-Larsen, M.Sc., Ph.D., associate professor, Department of Public Health, University of Copenhagen, ksla@sund.ku.dk

Corresponding author:
Janni Niclasen
Øster Farimagsgade 2A
1353 Copenhagen K
+45 3532 4891
Janni.niclasen@psy.ku.dk

Abbreviations: CI – Confidence Intervals; CNS – Central Nervous System; FAS – Fetal Alcohol Syndrome; OR – Odds Ratio; SDQ – Strengths and Difficulties Questionnaire;

Keywords: prenatal alcohol exposure; SDQ; neurobehavioural development; emotional; behavioural; follow-up; binge drinking; average alcohol intake; cumulated alcohol, exposure; Danish National Birth Cohort; DNBC; cohort; Strengths and Difficulties Questionnaire

Short title: prenatal alcohol exposure and behavioural development

What's known on this subject:

Prenatal exposure to high levels of alcohol is known to be associated with childhood mental health problems. Evidence of mental health problems caused by small to moderate levels of alcohol has, however been less conclusive.

What this study adds:

Binge drinking is weakly associated with behavioural and emotional development at age seven. Large differences in background characteristics were observed between the groups defined by cumulated alcohol exposure, i.e. low-moderate doses of alcohol, leaving these interpretations of findings uncertain

ABSTRACT

Background: It remains uncertain whether exposure to lower doses of alcohol is damaging to the developing foetus. The present study aimed to investigate associations for boys and girls between prenatal exposure to binge drinking and lower doses of alcohol in pregnancy and parent reported behavioural and emotional development at age seven.

Methods: This study used data from the Danish National Birth Cohort. Associations between cumulated alcohol exposure and binge drinking from full pregnancy and parent scores on the Strengths and Difficulties Questionnaire (SDQ) measured at age seven were investigated. The SDQ was used as continuous externalising/internalising scores and as above/below cut-off for the specific scales of hyperactivity/inattention, conduct, emotional and peer-problems. Inclusion criteria were information on alcohol exposure from three interviews, SDQ scores at age seven and being born full-term (N=37,152).

Results: Controlling for relevant confounders, small positive associations were observed between binge drinking and internalising (relative change in mean: 1.04-1.06), externalising scores (relative change in mean: 1.01-1.07), and conduct scores (OR 1.12-1.23) for boys. No associations were observed with lower doses of alcohol.

Conclusions: Exposure to binge drinking is weakly associated with impaired behavioural and emotional development measured at age seven. Large differences in background characteristics were observed between the groups defined by cumulated alcohol exposure, leaving the interpretations of findings uncertain.

INTRODUCTION

Today it is recognised that prenatal exposure to large amounts of alcohol can have long-term adverse neurobehavioural consequences for the child. At the extreme end, foetal alcohol syndrome (FAS), caused by exposure to excessive amount of alcohol in pregnancy, is characterised by growth retardation, facial abnormalities and dysfunctions of the central nervous system (CNS) [1]. Evidence of CNS impairments caused by small to moderate levels of alcohol in humans, i.e. <1 unit/day, has, however been less conclusive [2,3]. Some studies have suggested subtle neuropsychological deficits later in life [4,5], others have not [6,7]. Animal studies have been somewhat more conclusive and have largely found negative associations with learning, memory, and social behaviour later in life [8,9]. Most human studies concerned with alcohol intake in pregnancy distinguish between average alcohol intake (i.e. low/moderate/high doses as described above) and binge drinking (most often defined as an intake of minimum five units of alcohol on a single occasion) and the latter is generally considered to be more devastating for the developing CNS. Results from human binge-drinking studies have found negative effects on neurodevelopment including specific psychiatric disorders [10], hyperactivity and inattention [11], and IQ and delinquent behaviour [12], whereas others have not [6,7]. Animal studies concerned with binge-like exposures have largely focused on brain development, and have found associations with Purkinje cell loss [13], vulnerability of developing white matter [14] and neuronal reduction in the frontal cortex [15].

Pre- and post-natal brain development in males and females has been observed in animal studies to follow a somewhat different trajectory, despite no obvious anatomical differences [16,17]. In humans, gender differences are found regarding psychopathological prevalence rates [17], different ages of onset [17], and differences in responses to psychotropic medication [17]. Despite these differences few human studies have investigated gender-specific effects of prenatal exposure to alcohol. One study concluded that exposure to <1 unit/week in early pregnancy was associated with later mental health problems in girls, but not in boys [5] However, another study inferred that girls as well as boys born to mothers who drank up to 1–2 drinks per week during pregnancy were not at increased risk of clinically relevant behavioural difficulties at age five compared with non-exposed children [18]. The aim of the present study is to investigate association between exposure to total amount of alcohol and binge drinking in full pregnancy, irrespective of the timing of the exposure, and parent rated child behavioural and emotional development at age seven. These associations were investigated separately for boys and girls.

METHODS

Sample

Data are derived from the population-based Danish National Birth Cohort (DNBC) that comprises information on 100,418 pregnancies. The intention of the DNBC was to investigate potential associations between diverse exposures early in life and the health and development of the children from a life-course perspective [19]. Between 1996 and 2002 pregnant women were enrolled in the cohort nationwide at their first antenatal visit. The women were approached twice in pregnancy at approximately weeks 16 and 30 and again at six months postpartum. When the offspring reached the age of seven a questionnaire regarding the child's health and development was sent to the mother. The collection and analyses of data was approved by the regional ethical committee, den videnskabsetiske komite for region hovedstaden.

Restriction of sample

The sample for the present study was restricted to women with full information on key alcohol (average alcohol intake and binge drinking) and Strengths and Difficulties Questionnaire (SDQ) variables (hyperactivity, conduct, emotional and peer-problems) and to live-born, term singletons, i.e. gestational age of ≥ 37 completed weeks. This left a total of 37,152 mother-child dyads in the study.

Prenatal alcohol exposure

Alcohol exposure was assessed at three points in time from maternal self-reports: approximately in week 16 concerning pre- and early pregnancy intake, approximately in week 30 regarding intake in the middle part of pregnancy, and six months post-partum concerning alcohol intake in the last part of pregnancy. The women answered separate questions regarding their weekly average intake of alcohol (beer, wine and spirits) and binge episodes, defined as an intake of five or more units of alcohol on a single occasion. Because the focus was to investigate possible associations on behavioural and emotional development with total exposure to alcohol throughout pregnancy a single value for the cumulated intake of alcoholic drinks across the entire pregnancy was summed (Figure 1). This was done by multiplying the reported intake from each interview

with the number of weeks between each interview. Because the focus was to evaluate potential associations with very low exposures (i.e. down to >0-5 units of alcohol throughout pregnancy) it was decide not to divide the summed total by actual number of weeks of pregnancy. The following categories were adopted: 0, >0-5, >5-15, >15-45, >45-90 and >90. Because the >15-45 group was the largest it was chosen as the reference group. When a woman reported occasions of binge drinking she was asked about the number of such episodes. The women were grouped as follows: 0, 1, 2-3, 4+ binge episodes during pregnancy.

Outcome measure: parent-rated SDQ

The parent version of the SDQ contains 25 items concerned with five domains of psychological adjustment: hyperactivity/inattention (hereafter hyperactivity), conduct, emotional, peer-problems and prosocial behaviours. Each item is scored on a three point Likert scale: 'not true', 'somewhat true' and 'certainly true' yielding scores between 0-2 for each question [20-22]. Because the aim was to identify problem behaviours the prosocial scale was not used. The problem scales were used both as four separate scale models (i.e. hyperactivity, conduct, emotional and peer-problems) and as a broader model of externalising and internalising scales (combining the hyperactivity and conduct scales and the emotional and peer-problem scales, respectively). The four scale model used to identify children above 10% clinical cut-off whereas the two scale model was used to investigate mean differences between exposure groups. Both models have in the literature been found to have equally good model fits as tested by confirmatory factor analysis [23]. The following nationally-developed, partially gender-specific cut-offs were adapted: hyperactivity (≥ 7 for boys and ≥ 6 for girls), emotional (≥ 5 for boys and girls), peer-problems (≥ 3 for boys and girls) and conduct problems scores (≥ 4 for boys and girls) [24].

Confounding factors

The following covariates were statistically controlled for: maternal cumulated smoking in pregnancy (0, >0-100, >100- 300, >300 cigarettes); paternal smoking (yes/ no), maternal and paternal education (9 years or less, 10–12 years, 13 years or more); maternal and paternal past history of psychiatric diagnosis (yes/ no); and maternal well-being in pregnancy (good/ somewhat good/ severe problems). The maternal smoking variable was cumulated in the same way as the cumulated alcohol exposure variable (Figure 1). Information

on past psychiatric history and education came from the Danish registers, and the remaining variables from the structured interviews.

Statistical analyses

All analyses were carried out using SAS version 9.2. The overall aim was to investigate possible associations between prenatal exposure to alcohol and parent-rated SDQ scores at age seven. The first aim was to thoroughly describe the background characteristics of the mothers in relation to their alcohol intake during pregnancy. Secondly, multivariate linear regressions were used to model prenatal exposure to low/moderate doses of alcohol and binge drinking and associations with continuous externalising and internalising SDQ scores. Because the distributions of SDQ scores have been found to be positively skewed [25] these scores were log-transformed and the outcomes thus reflect a relative change in mean. The four dichotomised problem scales (hyperactivity, conduct, emotional and peer-problems) were assessed using logistic regression models with appropriate cut-offs identifying the 10% of the sample with the highest problems scores [24]. It was a-priori decided to carry out all analyses separately for boys and girls.

To test the robustness of the results when making minor changes to the analytical strategy, we conducted the following analyses: 1. All analyses were re-run using early pregnancy exposure only. 2. Combined cumulated alcohol and binge exposure categories were constructed and all analyses were re-run with this compound exposure variable. 3. All analyses were re-run excluding the all-time abstaining women from the analyses. 4. All analyses were re-run including children born before 37 full gestational weeks. 5. All analyses were re-run excluding siblings from the analyses.

RESULTS

Background characteristics

Cumulated alcohol intake: Abstainers (0 alcoholic drinks in pregnancy) and high intakers (>90 alcoholic drinks in pregnancy) distinguish themselves from the women with a low-to-moderate intake on most characteristics

(Table 1). The abstainers were younger, had high frequencies on psychiatric variables, more likely to drink cola, watch television, smoke cigarettes, and have a pre-pregnancy BMI outside the normal range and they were the least educated. 10,1% of the otherwise abstainers reported at least one binge episode in pregnancy. Because of the large sample size highly statistically significant differences (P >0.0001) were observed for virtually all variables and are thus not reported here.

The women with the highest alcohol intake (>90 drinks) were the oldest, the most well educated, most likely to drink coffee and do exercise. They had high smoking frequencies, but were the least likely to watch television, drink cola and have a pre-pregnancy BMI outside the normal range. 52.7% reported binge drinking in pregnancy. The frequencies for the low-to-moderate exposure groups were generally in-between these two extreme groups and appeared rather similar on most characteristics.

Binge drinking: the women in the four binge drinking groups did not differ as markedly as the cumulated alcohol exposure groups. However, the 4+ binge group did stand: they were more likely to have been in contact with the psychiatric system, to smoke, drink coffee and alcohol, but less likely to have a pre-pregnancy BMI outside the normal range. All four binge-groups had similar educational levels. Statistical differences were observed between the binge exposure groups, however not on as many variables and not as highly significant as the cumulated exposure groups.

Prenatal alcohol exposure and continuous scores

From the adjusted model, no associations were observed between low/moderate doses of alcohol in pregnancy and the parent-rated SDQ externalising and internalising scores at age seven (Table 2). However, an apparent protective association was found for the high exposure group (>90). Small, but statistically significant, elevated risks were found between binge drinking and internalising (relative change in mean: 1.03-1.07) and externalising scores (relative change in mean: 1.01-1.07) for boys. No associations were observed for girls.

Prenatal alcohol exposure and above cut-off SDQ scores

Cumulated alcohol exposure: From the adjusted model no overall association with any of the four problem scales appear (hyperactivity, conduct, emotional and peer-problems) (Table 3). However, few significant associations in the opposite of expected direction were present for the abstaining (0) group for boys: peer-problems (OR: 1.21 (CI 1.03-1.43)); and for girls: for conduct (OR: 1.17 (CI 1.00-1.37)), and peer-problems (OR: 1.40 (CI 1.15-1.71)), and for the high exposure group (>90) for boys for hyperactivity (OR: 0.79 (CI 0.65-0.96)) and emotional (OR: 0.71 (CI 0.59-0.85)), and for girls emotional (OR: 0.82 CI (0.69-0.98)).

Binge drinking: From the adjusted model in Table 3 it appears that binge drinking overall was associated with above clinical cut-off conduct scores (OR: 1.12-1.23) but only for boys. No dose-response effects were present, i.e. no increased OR was found with increased exposure to binge episodes. No associations were observed for girls.

Sensitivity analyses

Alcohol consumption is in the literature most often defined from early pregnancy intake only. In order to make the analyses from the present study comparable to the remaining literature the analyses were replicated with average alcohol intake in the early part of pregnancy as the exposure variable. The total sample added up to 37,152 pregnancies and was defined as follows: 0, >0-2, >2-4 and >4 units/ week. The results for the average alcohol intake revealed significant associations in the opposite of the expected direction, i.e. the higher intake the lower score for boys with internalising (>0-2: 0.95 (CI 0.92-0.97; >2-4: 0.90 (CI 0.85-0.95) and >4: 0.85 (CI 0.75-0.95) relative changes in mean), and emotional scores (>0-2: 0.83 (CI 0.76-0.91; >2-4: 0.71 (CI 0.56-0.88) and >4: 0.56 (CI 0.33-0.99) relative changes in mean). The binge drinking variable was defined as 0, 1, 2-3 and 4 episodes in the early part of pregnancy and these sub-analyses revealed (mostly) significant associations for boys with externalising (1: 1.04 (CI 1.01-1.07); 2-3: 1.05 (CI 1.00-1.09; 4+: 0.96 CI 0.88-1.06) relative change in mean) and internalising scores (1: 1.04 (CI 1.01-1.08); 2-3: 1.06 (CI 1.01-1.11; 4+: 1.02 CI 0.93-1.13) relative change in mean). No associations were observed for girls.

The cumulated alcohol estimates and the number of binge episodes were combined appropriately into six exposure groups (N = 37,152) and the analyses were re-run with this combined measure as the exposure variable. The analyses of the combined alcohol measure revealed no associations with the SDQ internalising

and externalising scores. For the four problem scales the high exposure group almost consistently had the most extreme estimates, but in both the expected and unexpected directions.

The analyses were further re-run where all-time abstainers were excluded from the analyses and where exposure was defined from first early pregnancy intake only (N = 32,733). No important differences were found between these analyses and the analyses excluding pre-pregnancy abstaining women. Likewise, the analyses were re-run including non-term born babies (N = 38,421), and re-run excluding siblings from the analyses (N = 35,635). These results were all virtually identical to the main analyses presented above.

DISCUSSION

After controlling for a wide range of confounding factors, the analyses revealed significant, positive associations between exposure to binge drinking in pregnancy and internalising, externalising, and conduct scores at age seven for boys, but not for girls. The findings are somewhat contradictory to another study that inferred that prenatal exposure to alcohol is more damaging to girls [5]. However, these authors actually hypothesised that any associations would be more readily detectable in boys, and further concluded that their finding might be chance [5]. Another study investigating associations between exposure to smoking in pregnancy and conduct disorder in childhood did find associations for boys only [26]. Thus, it may be that the brain development trajectory for boys is somewhat more vulnerable to prenatal exposures than females [16,17].

No associations were observed between lower doses of alcohol and any of the outcomes. On the contrary, the main analyses revealed poorest mental health outcomes for children of abstainers, but most advanced outcomes for children of the high intakers. Very large differences on background characteristics were observed between the groups defined on the basis of cumulated alcohol exposure. The high intakers were older, and much more well educated than the abstainers who, were the least educated, the ones with the highest frequencies of mental disorders and poorest lifestyles habits. These characteristics may well be mentally protective for the high exposed children but disadvantageous for the unexposed children [27,28]. The expectedly large positive impact of the home environments of the well-educated may masks the potential small negative effects of being exposed to low doses of alcohol. This has similarly been inferred in a study by Kelly where the odds of behavioural problems in children of never drinkers were similar to those of children

exposed to high levels [18]. For the binge groups the distribution of covariates were less variable and the results will, all other things being equal, be less confounded.

The sensitivity analyses defining alcohol exposure from early pregnancy, rather than full pregnancy, revealed some associations that were sometimes found to be in the same direction as the main analyses and sometimes in the opposite direction. Two explanations could account for these differences. The first is, that the differences are real because early pregnancy exposure and full pregnancy exposure are associated with different observable behaviours at age seven. Another possible explanation is that the results are due to unmeasured and residual confounding. Large intergroup differences in characteristics were observed for the groups defined on the basis of full pregnancy exposure on key covariates, e.g. age, education, psychiatric difficulties and lifestyle factors. These different patterns in covariates may explain the different results observed in the main analyses and the sensitivity analyses. These characteristics, along with other unmeasured confounding variables such as IQ, attachment style and personality could be mentally protective for the high exposed children, but disadvantageous for the unexposed children. It is today well known that the quality of the mother-child relationship has lasting consequences for a wide range of developmental cognitive and mental health outcomes [29]. Infants who develop a secure attachment style has a better emotional regulation, higher self-esteem, and more develop coping skills, that in turn makes them better able to handle stressful or challenging situations and lowers the risk for poorer mental health outcomes later in life. On the other hand, insecurely attached children are at greater risk for poor mental health outcomes [29,30]. Thus, different home environments create different conditions for the developing brain. Because the home environment has such great impact on brain development, a potential damaging effect of being exposed to a small amount of alcohol is masked. Further, because the distribution of covariates in the binge exposure groups were less variable compared to the cumulated alcohol exposure groups less confounding will mask potential associations in the binge exposure groups.

Strengths and limitations

The tremendous size of the sample and the use of well-established outcome measure are obvious advantages of the present study. Also, exposure from full pregnancy was used, rather than just early

pregnancy exposure. The construction of the cumulated alcohol measure can, on the other hand, also be considered a limitation as it may include some misclassification. However, the strong linear trends for many of the covariates validate the method. Finally, the timing of the exposure has been ignored in the present study. Other studies have demonstrated that ignoring this factor can mask potential associations low and moderate prenatal alcohol exposure and fetal effects [31].

Acknowledgements: The study was supported financially by the Department of Psychology, University of Copenhagen; Ludvig og Sara Elsass Foundation; Aase og Ejnar Danielsens Foundation; Carl J. Becker's Foundation; the Lundbeck Foundation; Børne- og Ungdomspsykiatrisk Selskab i Danmark; Dagmar Marshalls Foundation; The A.P. Møller Foundation for the Advancement of Medical Science; Direktør Jakob Madsens Legat.

Reference List

(1) Calhoun F, Warren K. Fetal alcohol syndrome: historical perspectives. Neurosci Biobehav Rev 2007;31(2):168-171.

(2) Gray R, Henderson J. Rewiev of the Fetal Effects of Prenatal Alcohol Exposure. National Perinatal Epidemiology Unit, University of Oxford; 2007 May.

(3) Henderson J, Gray R, Brocklehurst P. Systematic review of effects of low-moderate prenatal alcohol exposure on pregnancy outcome. BJOG 2007;114(3):243-252.

(4) Kapil Sayal, Jon Heron, Jean Golding, Alan Emond. Prenatal Alcohol Exposure and Gender Differences in Childhood Mental Health Problems: A Longitudinal Population-Based Study . PEDIATRICS 2007;119:426-434.

(5) Sayal K, Heron J, Golding J, Emond A. Prenatal alcohol exposure and gender differences in childhood mental health problems: a longitudinal population-based study. PEDIATRICS 2007;119(2):e426-e434.

(6) Skogerbo A, Kesmodel US, Wimberley T et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on executive function in 5-year-old children. BJOG 2012;119(10):1201-1210.

(7) Underbjerg M, Kesmodel US, Landro NI et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on selective and sustained attention in 5-year-old children. BJOG 2012;119(10):1211-1221.

(8) Schneider ML, Moore CF, Adkins MM. The Effects of Prenatal Alcohol Exposure on Behavior: Rodent and Primate Studies. Neuropsychol Rev 2011.

(9) Valenzuela CF, Morton RA, Diaz MR, Topper L. Does moderate drinking harm the fetal brain? Insights from animal models. Trends Neurosci 2012;35(5):284-292.

(10) Barr HM, Bookstein FL, O'Malley KD, Connor PD, Huggins JE, Streissguth AP. Binge drinking during pregnancy as a predictor of psychiatric disorders on the Structured Clinical Interview for DSM-IV in young adult offspring. Am J Psychiatry 2006;163(6):1061-1065.

(11) Sayal K, Heron J, Golding J et al. Binge pattern of alcohol consumption during pregnancy and childhood mental health outcomes: longitudinal population-based study. PEDIATRICS 2009;123(2):e289-e296.

(12) Bailey BN, aney-Black V, Covington CY et al. Prenatal exposure to binge drinking and cognitive and behavioral outcomes at age 7 years. Am J Obstet Gynecol 2004;191(3):1037-1043.

(13) Ikonomidrus NM, Napper RM. Acute and long-term Purkinje cell loss following a single ethanol binge during the early third trimester equivalent in the rat. Alcohol Clin Exp Res 2012;36(8):1365-1373.

(14) Watari H, Born DE, Gleason CA. Effects of first trimester binge alcohol exposure on developing white matter in fetal sheep. Pediatr Res 2006;59(4 Pt 1):560-564.

(15) Burke MW, Palmour RM, Ervin FR, Ptito M. Neuronal reduction in frontal cortex of primates after prenatal alcohol exposure. Neuroreport 2009;20(1):13-17.

(16) Goodlett CR, Johnson TB. Temporal windows of vulnerability within the third trimester equivalent: why ''knowing when'' matters. In: Hannigan JH, Goodlett CR, Spear LP, Spear NE, editors. ALcohol and alcoholism: effects on brain and development. Mahwah, NJ: L. Earlbaum Associates; 1999. 59-91.

(17) Andersen SL. Trajectories of brain development: point of vulnerability or window of opportunity? Neurosci Biobehav Rev 2003;27(1-2):3-18.

(18) Kelly YJ, Sacker A, Gray R et al. Light drinking during pregnancy: still no increased risk for socioemotional difficulties or cognitive deficits at 5 years of age? J Epidemiol Community Health 2010.

(19) Olsen J, Melbye M, Olsen SF et al. The Danish National Birth Cohort--its background, structure and aim. Scand J Public Health 2001;29(4):300-307.

(20) Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 1997;38(5):581-586.

(21) Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? J Abnorm Child Psychol 1999;27(1):17-24.

(22) Goodman R. Psychometric properties of the strengths and difficulties questionnaire. Journal of the American Academy of Child & Adolescent Psychiatry 2001;40(11):1337-1345.

(23) Niclasen J, Skovgaard AM, Andersen AM, Somhovd MJ, Obel C. A Confirmatory Approach to Examining the Factor Structure of the Strengths and Difficulties Questionnaire (SDQ): A Large Scale Cohort Study. J Abnorm Child Psychol 2012.

(24) Niclasen J. SDQ Danish cut off scores: 5-7-year-olds. 18-12-2012. www.sdqinfo.com.

(25) Niclasen J, Teasdale TW, Andersen AM, Skovgaard AM, Elberling H, Obel C. Psychometric properties of the Danish strength and difficulties questionnaire: The SDQ assessed for more than 70,000 raters in four different cohorts. PLoS One 2012;7(2):e32025.

(26) Fergusson DM, Woodward LJ, Horwood LJ. Maternal smoking during pregnancy and psychiatric adjustment in late adolescence. Arch Gen Psychiatry 1998;55(8):721-727.

(27) Levitt P. Structural and functional maturation of the developing primate brain. J Pediatr 2003;143(4 Suppl):S35-S45.

(28) Sroufe LA. Attachment and development: a prospective, longitudinal study from birth to adulthood. Attach Hum Dev 2005;7(4):349-367.

(29) Bowlby J. Research into the Origins of Delinquent Behaviour. Br Med J 1950;1(4653):570-573.

(30) Ainsworth MD, Bell SM. Attachment, exploration, and separation: illustrated by the behavior of one-year-olds in a strange situation. Child Dev 1970;41(1):49-67.

(31) O'Leary CM, Bower C, Zubrick SR, Geelhoed E, Kurinczuk JJ, Nassar N. A new method of prenatal alcohol classification accounting for dose, pattern and timing of exposure: improving our ability to examine fetal effects from low to moderate alcohol. J Epidemiol Community Health 2010;64(11):956-962.

Figure 1: timeline showing the period of which information from each interview was used.

		Cumulated alcohol exposure (total no. of alcohol containing units in entire pregnancy)						Binge drinking (no. of binge episodes in entire pregnancy)			
Alcohol group	**Full sample**	**0**	**>0-5**	**>5 - 15**	**>15-45**	**>45 -90**	**>90**	**0**	**1**	**2-3**	**4+**
N	37,152	3910	6739	7156	9929	6091	3327	25,692	6833	3779	848
Age (M) [1,2]	30.7	29.7	29.9	30.4	30.8	31.4	32.7	30.8	30.3	30.4	31.1
Age (P) [1,3]	32.5	31.6	31.6	32.1	32.4	33.2	34.7	32.6	32.0	32.1	33.0
Unplanned preg.[4]	22.1%	24.3%	19.7%	21.4%	21.3%	23.9%	24.9%	19.4%	25.4%	31.1%	36.2%
Time to preg.[5]	26.4%	28.1%	27.9%	26.9%	25.1%	25.2%	26.4%	27.0%	25.9%	23.7%	24.9%
Fertility treatment [6]	5.7%	5.6%	6.5%	6.2%	5.3%	5.6%	5.0%	7.0%	3.3%	2.4%	1.9%
Married (no) [7]	2.0%	3.3%	1.9%	2.1%	1.6%	1.7%	2.2%	1.7%	1.9%	3.0%	6.4%
Education											
Mandatory (M) [2,8]	6.8 %	14.6%	8.4%	6.5%	4.7%	4.7%	5.4%	7.0%	6.1%	6.5%	8.1%
University (M) [2,8]	15.8%	6.8%	11.6%	14.1%	17.6%	21.1%	22.9%	15.1%	16.5%	18.8%	16.9%

Mandatory (P) [3,8]	12.2%	20.9%	14.2%	12.2%	10.0%	9.2%	10.0%	12.4%	11.3%	11.9%	14.0%
University (P) [3,8]	16.4%	7.8%	12.7	15.3%	18.1%	21.0%	22.5%	16.0%	17.2%	17.5%	17.2%
Contact with psychiatric system											
Pre-preg. (M) [2,9]	2.5%	4.0%	2.9%	2.1%	2.1%	2.3%	2.4%	2.4%	2.2%	3.1%	4.0%
Pre-preg. (P) [3,9]	1.7%	2.5%	1.9%	1.6%	1.3%	1.6%	1.9%	1.6%	1.9%	1.9%	2.4%
Maternal lifestyle factors in pregnancy											
Binge drinking [2,10]	30.9%	10.1%	19.3%	28.1%	34.3%	42.7%	52.7%	0%	100%	100%	100%
Cumulated alc. intake [2,11]	33.2	0	2.7	9.8	28.5	64.6	140.8	27	38	55	93
Cumulated smoking [2,12]	53 (22.8%)	85 (28.1%)	54 (21.5%)	51 (21.2%)	43 (20.3%)	46 (23.3%)	67 (28.8%)	47 (18.9%)	59 (27.8%)	74 (34.7%)	122 (47.5%)
Partner smoking [14]	27.8%	34.8%	27.5%	27.1%	25.5%	26.6%	30.2%	26.2%	29.7%	32.0%	39.2%
Coffee [13]	41.4%	31.9%	30.6%	26.9%	43.0%	51.3%	63.7%	40.0%	42.3%	45.0%	56.5%

Vitamins [14]	84.9%	83.1%	85.9%	86.5%	85.1%	84.3%	81.9%	85.2%	84.7%	84.6%	80.2%
Fish oil [14]	4.9%	4.7%	4.9%	5.0%	4.6%	4.8%	5.4%	4.8%	4.8%	5.1%	6.0%
Fish eating (never) [14]	3.2%	6.5%	4.1%	3.1%	2.5%	1.9%	1.5%	3.2%	3.0%	3.0%	2.6%
BMI [15]	30.3%	41.1%	35.2%	31.4%	27.9%	24.7%	22.6%	31.2%	28.7%	28.2%	25.1%
Cola [16]	15.1%	23.2%	16.2%	15.0%	13.4%	12.7%	13.4%	15.0%	15.3%	15.7%	16.3%, NS
TV [17]	20.0%	27.7%	23.4%	20.4%	18.3%	16.6%	14.8%	19.3%	20.8%	22.6%	25.2%
Tenant [18]	25.7%	28.6%	26.6%	26.6%	25.0%	25.2%	22.1%	23.8%	28.5%	31.9%	34.4%
Exercise [19]	25.7%	18.6%	22.7%	25.3%	27.6%	29.5%	28.1%	24.9%	26.5%	29.2%	26.4%
Child characteristics											
Sex (boys)	51.0%	51.4%	50.6%	51.7%	50.9%	50.7%	50.9%	51.2%	50.8%	50.1%	50.7%
Birth weight[20]	3654	3636	3660	3656	3655	3658	3643	3659	3651	3632	3596
SGA (%) [21]	8.6%	9.2%	8.7%	8.5%	8.2%	8.0%	9.8%	8.3%	9.0%	9.3%	12.2%
G.A. [22]	281.5	280.6	281.3	281.3	281.6	282.0	282.6	281.3	281.9	282.0	282.6
Psych. diagnosis [23]	1.3%	1.6%	1.3%	1.4%	0.8%	1.1%	1.2%	1.3%	1.0%	1.4%	1.5%

Table 1: background characteristics across levels of cumulated exposure to alcohol and binge drinking episodes. [1] Age at birth . [2] M = Maternal. [3] P = Paternal. [4] Fully or partly unplanned pregnancy. [5] Time to pregnancy (≥ 6 months). [6] Fertility treatment – yes. [7] married or cohabiting with the child's biological father six months post-partum. [8] Register-based information on educational level in year 2010. [9] Register-based information on contact with the psychiatric system. [10] Binge drinking – yes. [11] Maternal cumulated alcohol intake In pregnancy. [12] Cumulated smoking in pregnancy (smoking in pregnancy – yes). [13] Maternal intake of coffee in pregnancy – yes. [14] Maternal intake in last part of pregnancy. [15] Pre-pregnant body mass index (BMI) outside normal range. BMI normal range: 18.5-24.99. [16] Intake of ≥ 1 liter of Cola per week. [17] Television watching ≥ 2 hours/ day in last part of pregnancy. [18] Tenant, homeless or live with parents. [19] Maternal exercise in last part of pregnancy. [20] Birth weight in grams. [21] SGA: small for gestational age. [22] G.A.: Gestational age. [23] Child contact with the psychiatric system before the age of seven

	Boys		Girls	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Exposure categories	Univariable Relative change in mean	Multivariable Relative change in mean (95% CI)	Univariable Relative change in mean	Multivariable Relative change in mean (95% CI)
SDQ externalizing scores at age seven and *cumulated alcohol exposure* in pregnancy				
0	1.10	1.02 (0.98-1.06)	1.06	0.99 (0.95-1.03)
>0–5	1.01	0.98 (0.95-1.01)	1.00	0.97 (0.94-1.01)
>5–15	1.00	0.98 (0.95-1.01)	0.99	0.97 (0.94-1.00)
>15–45	1	1	1	1
>45–90	0.98	0.98 (0.95-1.02)	0.97	0.98 (0.94-1.01)
>90+	0.95	0.95 (0.91-0.99)	1.00	0.99 (0.95-1.04)
SDQ internalizing scores at age seven and *cumulated alcohol exposure* in pregnancy				
0	1.08	1.03 (0.99-1.08)	1.09	1.03 (0.98-1.07)
>0–5	1.04	1.02 (0.98-1.06)	1.01	0.99 (0.96-1.03)
>5–15	1.03	1.02 (0.99-1.06)	1.02	1.01 (0.97-1.04)
>15–45	1	1	1	1
>45–90	0.99	0.99 (0.96-1.03)	0.96	0.97 (0.94-1.01)
>90+	0.93	0.92 (0.88-0.97)	0.97	0.97 (0.92-1.01)
SDQ externalizing scores at age seven and *binge drinking* exposure in pregnancy				
0	1	1	1	1
1	1.04	1.04 (1.01-1.07)	1.02	1.02 (0.99-1.05)
2-3	1.08	1.07 (1.04-1.11)	1.03	1.03 (0.99-1.07)
4+	1.05	1.01 (0.94-1.09)	1.10	1.07 (0.99-1.15)
SDQ internalizing scores at age seven and *binge drinking* exposure in pregnancy				
0	1	1	1	1

1	1.03	1.04 (1.00-1.07)	1.01	1.01 (0.98-1.04)
2-3	1.04	1.04 (1.00-1.08)	1.00	1.01 (0.97-1.05)
4+	1.08	1.06 (0.98-1.15)	1.01	0.97 (0.90-1.05)

Table 2: Relative change in mean between cumulated alcohol exposure and binge drinking in pregnancy, and continuous SDQ externalising and internalising scores at age seven. Adjusted model: adjusted for the following confounders: parental smoking, parental education, parental pre-pregnancy psychiatric diagnoses, and maternal psychological well-being in pregnancy

	Boys		Girls	
	Unadjusted	Adjusted	Unadjusted	Adjusted
Exposure categories	Univariable OR	Multivariable OR (95% CI)	Univariable OR	Multivariable OR (95% CI)
SDQ hyperactivity/ inattention scores at age seven and *cumulated alcohol exposure* in pregnancy				
0	1.42	1.13 (0.96-1.32)	1.39	1.12 (0.95-1.32)
>0–5	1.10	1.02 (0.88-1.17)	1.05	0.97 (0.84-1.12)
>5–15	0.98	0.94 (0.81-1.08)	0.93	0.89 (0.76-1.03)
>15–45	1	1	1	1
>45–90	1.02	1.05 (0.90-1.21)	0.83	0.84 (0.71-0.98)
>90+	0.80	0.79 (0.65-0.96)	0.99	0.94 (0.78-1.14)
SDQ conduct scores at age seven and *cumulated alcohol exposure* in pregnancy				
0	1.34	1.08 (0.94-1.24)	1.47	1.17 (1.00-1.37)
>0–5	1.14	1.05 (0.93-1.18)	1.02	0.93 (0.81-1.07)
>5–15	0.96	0.89 (0.79-1.01)	0.92	0.87 (0.76-1.00)
>15–45	1	1	1	1
>45–90	1.03	1.06 (0.94-1.20)	1.02	1.05 (0.91-1.20)
>90+	0.88	0.86 (0.73-1.02)	0.89	0.89 (0.74-1.06)
SDQ emotional scores at age seven and *cumulated alcohol exposure* in pregnancy				
0	1.24	1.08 (0.93-1.25)	1.10	0.96 (0.82-1.12)
>0–5	1.15	1.08 (0.95-1.23)	1.04	0.99 (0.87-1.13)
>5–15	1.12	1.10 (0.97-1.25)	0.98	0.96 (0.84-1.09)
>15–45	1	1	1	1
>45–90	0.94	0.95 (0.83-1.09)	0.84	0.86 (0.75-0.99)
>90+	0.73	0.71 (0.59-0.85)	0.82	0.82 (0.69-0.98)
SDQ peer problems scores at age seven and *cumulated alcohol exposure* in pregnancy				
0	1.45	1.21 (1.03-1.43)	1.75	1.40 (1.15-1.71)

>0–5	1.11	1.05 (0.90-1.21)	1.11	1.01 (0.84-1.21)
>5–15	1.04	1.00 (0.86-1.15)	1.11	1.05 (0.87-1.26)
>15–45	1	1	1	1
>45–90	0.89	0.91 (0.78-1.06)	1.02	1.04 (0.85-1.26)
>90+	0.93	0.92 (0.76-1.12)	0.89	0.83 (0.64-1.07)
SDQ hyperactivity/ inattention scores at age seven and *binge drinking* exposure in pregnancy				
0	1	1	1	1
1	1.15	1.16 (1.02-1.30)	1.11	1.11 (0.98-1.25)
2-3	1.16	1.16 (0.99-1.35)	1.21	1.14 (0.97-1.33)
4+	1.18	1.06 (0.77-1.43)	1.32	1.14 (0.83-1.54)
SDQ conduct scores at age seven and *binge drinking* exposure in pregnancy				
0	1	1	1	1
1	1.11	1.14 (1.03-1.27)	0.94	0.93 (0.82-1.05)
2-3	1.22	1.23 (1.08-1.40)	0.95	0.93 (0.80-1.09)
4+	1.20	1.12 (0.86-1.45)	1.02	0.89 (0.65-1.21)
SDQ emotional scores at age seven and *binge drinking* exposure in pregnancy				
0	1	1	1	1
1	1.18	1.17 (1.05-1.31)	1.12	1.11 (0.99-1.24)
2-3	1.11	1.10 (0.96-1.27)	1.08	1.13 (0.98-1.30)
4+	1.33	1.20 (0.90-1.56)	1.03	0.90 (0.66-1.21)
SDQ peer problem scores at age seven and *binge drinking* exposure in pregnancy				
0	1	1	1	1
1	1.03	1.07 (0.94-1.21)	1.11	1.14 (0.98-1.32)
2-3	1.12	1.15 (0.98-1.35)	1.07	0.98 (0.80-1.20)
4+	1.15	1.11 (0.80-1.50)	1.13	0.93 (0.60-1.38)

Table 3: Cumulated alcohol exposure and binge drinking in pregnancy and SDQ hyperactivity/inattention, conduct, emotional and peer problem scores above clinical cut-off at age seven. Adjusted model: adjusted for the following confounders: parental smoking, parental education, parental past history of contact with psychiatric system and maternal well-being in pregnancy

Paper 4

Niclasen, J, Teasdale, T., Strandberg-Larsen, K., Nybo Andersen, AM. Is Alcohol Binge Drinking in Early and Late Pregnancy associated with Behavioural and Emotional Development at Age Seven? (submitted for publication)

Is Alcohol Binge Drinking in Early and Late Pregnancy Associations with Behavioural and Emotional Development at age seven?

Niclasen, J., Andersen, AMN., Strandberg-Larsen, K., Teasdale, TW.,

Authors:

Janni Niclasen. M.Sc., Ph.D. fellow, janni.niclasen@psy.ku.dk

Anne Marie Nybo Andersen, Ph.D., professor amny@sund.ku.dk

Katrine Strandberg-Larsen, Ph.D., associate professor, ksla@sund.ku.dk

Thomas William Teasdale, Dr. Med., associate professor, tom.teasdale@psy.ku.dk

Corresponding author:

Janni Niclasen

Øster Farimagsgade 2A

1353 Copenhagen K

+45 3532 4891

Janni.niclasen@psy.ku.dk

Keywords: prenatal alcohol exposure; SDQ; binge drinking; Danish National Birth Cohort; timing; Strengths and Difficulties Questionnaire

Abstract

Introduction: the purpose was to investigate associations of maternal binge drinking in early and late pregnancy with child behavioural and emotional development at age seven. It was hypothesised that late exposure was associated with more negative outcomes than early exposure. Differences were expected on the continuous outcome measures, but not on above cut-off scale scores.

Methods: Data were derived from the Danish National Birth Cohort (DNBC). The three exposure groups, were defined according to binge drinking from three interviews regarding binge episodes in early, middle and late pregnancy. The 'no binge' group included women with no binge episodes in any of the interviews, the 'early bingers' reported episodes in the first interview, and the 'late bingers' in the last part of pregnancy only. The outcome measure was the Strengths and Difficulties Questionnaire (SDQ) used as continuous externalising/internalising scores and above cut-off hyperactivity/inattention, conduct, emotional and peer-problems scores. Women with full information on binge drinking from the three interviews, full-scale SDQ information at age seven and being term-born were included in the study (N = 37,315).

Results: after adjustment for maternal education, psychiatric diagnoses, age and smoking, children exposed to binge drinking in late pregnancy compared with unexposed children had significantly higher mean externalising scores at age seven (relative change in mean 1.21 (1.04-1.42)). No associations were observed for any of the above cut-off outcomes.

Conclusion: exposure to binge drinking in late pregnancy is associated with elevated externalising scores, but not with increased risk of above cut-off scale scores.

Introduction

It is widely recognised today that prenatal exposure to alcohol, particularly at high levels is negatively associated with neurobehavioural development[1]. It is also recognised that exposure to larger amounts of alcohol over a short period, say one day, is more teratogenic for the developing central nervous system (CNS) than is exposure to a comparable amount of alcohol spread over several days[2]. What is less known is the effects of timing, i.e. whether the developing CNS is particularly vulnerable to alcohol exposure in early and/ or late pregnancy. Few human studies have investigated the importance of the timing of alcohol exposure during pregnancy on mental health outcomes[3, 4] and no human studies have to our knowledge investigated whether exposure to binge drinking, defined as sporadic high intake, in early and/ or late pregnancy is associated with different mental health outcomes in childhood. This lack of evidence may be problematic as prenatal exposure to alcohol potentially interferes with all stages of brain development and different brain regions have unique courses of ontogeny[3-7]. Broadly speaking, first trimester is mostly concerned with cell proliferation and migration of cells[5], second trimester with neuronal and synapse formation, axonal and dendritic outgrowth and programmed cell death[6], whereas in the third trimester the brain is growing larger with synapse formation and myelination taking place[6].

Whereas there is a general lack of human studies investigating the effects of timing on prenatal exposure to alcohol binge drinking, some animal studies have investigated such possible associations. One study[8] that examined the effects of gestational timing of alcohol exposure on neurobehavioural development in rhesus monkeys found that exposure to alcohol during early gestation significantly decreased scores on infant behavioural tests, whereas mid- to late-gestation exposure resulted in reduced motor maturity but did not affect overall neurobehavioural performance. According to the authors, this indicates that early-gestation alcohol exposure is as deleterious to neonatal neurobehaviour as late gestation or continuous exposure and that neurobehaviour is a more sensitive marker of early-gestation moderate alcohol exposure than growth parameters. Another study looking at rats[9] suggested that exposure during the equivalent of all three trimesters and third trimester equivalent significantly reduced Purkinje cell number compared with first and second trimester equivalent exposure, and that third trimester equivalent resulted in a decrement in the number of olfactory bulb mitral cell number compared to first and second trimester equivalent exposure. Another rat study[10] investigated the effects of age, sex, and timing of prenatal exposure to ethanol on

social behaviour and inferred that early pregnancy exposure resulted in mild changes of social behaviour in young adolescents, whereas mid-gestation exposure resulted in pronounced behavioural deficits throughout ontogeny, with deficits being most robust in male off-spring. Males exposed to ethanol in mid-gestation showed decreases in social investigation, contact behaviour, and play fighting, whereas a decrease in social motivation was evident in adolescence regardless of sex. It was concluded that exposure to ethanol alters social behaviour, and that the timing of the exposure defines the behavioural outcome.

The present study aimed to investigate possible associations of timing on prenatal exposure to binge drinking and behavioural and emotional development at age seven. Most previous studies based on human samples have defined prenatal exposure to alcohol solely from first trimester exposure. The aim of the present study was to examine whether exposure to binge drinking in either early or late pregnancy was differently associated with behavioural and emotional development at age seven.

Methods

Sample

Data came from the population-based birth cohort, the Danish National Birth Cohort (DNBC) that comprises information on 100,418 pregnancies. The intention of the DNBC was to look at the association between exposures early in life and the health and development of the children from a longitudinal, life-course perspective[11]. Between 1996 and 2002 pregnant women were enrolled nationwide at their first antenatal visit. The women were approached twice in pregnancy at approximately weeks 16 and 30 and again six month post-partum. When the child reached seven years of age a questionnaire regarding the child's health and development was sent to the mother. All questionnaires are available in English at www.dnbc.dk.

Restriction of sample

The sample was restricted to mothers with full information on the binge drinking variables obtained in the first three interviews, information on the Strengths and Difficulties Questionnaire (SDQ) at age seven and for

whom the child was a singleton with a gestational age of ≥ 37 completed weeks. These restrictions resulted in a total of 37,315 mother-child dyads. The mothers were informed that the data were being collected for research purposes and gave oral consent to this. They were informed that statistical results would be reported in such a way that individuals could not be identified. Ethical approval was obtained for the study.

Of the total of 37,315 mothers, 25,781 women reported no binge drinking episodes in pregnancy and these constituted a 'no binge' control group. The early exposure group consisted of women reporting binge drinking in the first interview, but no binge episodes in the second and third interview (N=3,654). The late exposure group consisted of women reporting binge episodes in the third interview, but no binge episodes in the first and second interviews (N=94). The remaining exposure groups were excluded from all analyses and included: reported binge episode(s) in the second interview only (N=1,990), binge episodes according to both the first and third interviews (N=48), binge episodes according to all three interviews (N=155), binge episodes in first and second interviews (N=5,535) and binge episodes in the second and third interviews (N=58).

Exposure: alcohol binge drinking

Binge drinking was defined as an intake of five or more alcohol containing units on a single occasion. It was assessed separately on the basis of information from three interviews conducted approximately in week 16 regarding early pregnancy intake, approximately in week 30 regarding middle pregnancy intake and six months post-partum concerning intake in the last part of pregnancy. If a woman reported any binge episodes in the first interview and none later, she was considered an 'early' pregnancy binger, if she reported any such episodes in the second interview she was considered a middle pregnancy (and therefore excluded) binge drinker and if she reported occasions of binge drinking in the third interview with none earlier she was considered a late pregnancy binge drinker. All women reporting any occasions of binge drinking were further asked about the total number and pregnancy week of each binge episode. The present study is concerned with women who reported binge drinking exclusively early or exclusively late in pregnancy as defined here. A control group was included comprising women who reported never binge drinking during pregnancy.

Outcome: parent-rated SDQ

The SDQ contains 25 items concerned with five domains of psychological adjustment: hyperactivity/inattention (hereafter hyperactivity), conduct, emotional, peer problems and prosocial behaviours. Each item is scored on a three point Likert scale; 'not true', 'somewhat true' and 'certainly true' yielding scores between 0-2 for each question [12-14]. Because the aim was to identify problem behaviours the prosocial scale was not used in the present study. The four problem scales were used both as four separate scale models (i.e. hyperactivity, conduct, emotional and peer problems) and in a broader model of externalising (combining the hyperactivity and conduct scales) and internalising scales (combining the emotional and peer problem scales) as these two models have been found to have equally good statistical fits [15]. The externalising and internalising scales were used as continuous outcome variables and the four problem scales as above clinical cut-offs (approximately the highest scoring 20%). The following cut-offs were adapted: hyperactivity (≥ 5 for boys and ≥ 4 for girls), emotional (≥ 4 for boys and girls), peer problems (≥ 2 for boys and girls) and conduct problems scores (≥ 3 for boys and girls) [16].

Confounders

The following confounding factors were controlled for: maternal education (9 years or less, 10–12 years, 13 years or more); maternal psychiatric diagnosis up to the age of seven of the child (yes/ no), maternal age and maternal smoking in pregnancy (yes/ no). Information on past psychiatric history, education and age came from registries, the smoking variable from self-reports. Because only women who have been in contact with the psychiatric services are recorded in the Danish psychiatric central registry and because it is the vulnerability of the women that is considered a confounder all women that had been in contact with the psychiatric system up to the age of seven of their child were included as cases in the psychiatric diagnoses variable.

Statistical analyses

All analyses were carried out in SAS version 9.2. Overall, the aims were to estimate the associations between prenatal exposure to maternal binge drinking in either the early or late part of pregnancy and child SDQ scores at age seven. Specifically the aims were to investigate 1. associations between exposure to binge drinking in early and late pregnancy and continuous internalising/ externalising SDQ scores at age seven, 2.

Associations between binge drinking in early and late pregnancy and above cut-off hyperactivity, conduct, emotional and peer problems scores at age seven. Multivariable linear regressions were used to model binge drinking and associations with continuous externalising/ internalising SDQ scores. Because the distribution of SDQ scores have been found to be positively skewed[17] the internalising/ externalising scale scores were log transformed before used as continuous variables in the multivariable linear regressions and the outcomes therefore reflect relative changes in mean. The four problem scales (hyperactivity, conduct problems, peer problems and emotional symptoms) were used in logistic regression models with appropriate cut-offs identifying the 20% of the sample with the highest problems scores. The following partially gender-specific cut-offs were adapted: hyperactivity (≥ 5 for boys and ≥ 4 for girls), emotional (≥ 4 for boys and girls), peer problems (≥ 2 for boys and girls) and conduct problems scores (≥ 3 for boys and girls)[16].

In the main analyses all women, regardless of number of binge episodes, were included. In order to test the robustness of the results the analyses were rerun including women with one or two binge episodes only, i.e. excluding women with three or more episodes. Further, all analyses were rerun including only women who were interviewed before 16 completed gestational weeks in the early exposure group, in order to increase the reliability of the early binge drinking reports.

Results

Table 1 shows the background characteristics of the three exposure groups. It appears that both binge exposure groups had higher percentages of women being a tenant (29.3% and 29.8%), but probably because of the small size of the late binge group only the early binge group differed significantly from the no binge group. Both exposure groups differed significantly from the no binge group on the maternal education variable, however in opposite directions. Only 38.3% of the late binge drinkers had 13 years or more of education compared to 52.3% of the non-bingers. It is also apparent from Table 1 that 10.6% of the mothers in the late binge group have had contact with the psychiatric system up to the age of seven of the child. This is twice as many as in the other groups and significantly different from the no bingers. Interestingly, the opposite is true for the fathers of the late binge group where only 1.1% have had contact with the psychiatric system before the age of seven of the child. As was expected, significantly more early and late binge drinking

mothers smoked compare to the non-binge group and the two binge exposure groups also had a lower proportion of planned pregnancies.

Table 2 and Table 3 present the adjusted relative changes in means for the continuous externalising and internalising scores and number of cases as well as odds ratios (OR) for the dichotomised measures on the hyperactivity, conduct, peer problem and emotional scales, respectively. The models were initially controlled for maternal education, psychiatric diagnoses, age, cumulated alcohol intake in pregnancy and smoking in pregnancy. However, as the cumulated alcohol variable did not contribute significantly to the model it was removed and thus not controlled for in any of the analyses. Table 2 shows that externalising scores are significantly associated with binge drinking in early (relative change in mean 1.02 (CI 1.00-1.05) and late pregnancy (relative change in mean 1.21 (CI 1.04-1.42). Further when the early and late exposure groups are tested statistically the difference is significant (results not shown here).For the above cut-off clinical scales no statistically significant associations were present and the point estimates did not indicate any strong association (Table 3).

In the main analyses the early and late binge groups were compared to the no binge group irrespective of the number of binge episodes of each woman. In order to look at possible effects of the number of binge episodes the analyses were replicated only including women with either one or two binge episodes. This left at total of 3,309 and 91 women in the early and late exposure groups, respectively. The results revealed virtually identical estimates for all outcomes. However, the OR estimates for the hyperactivity scores were a little lower compared to the main results. Further, the sensitivity analyses including only women interviewed before 16 completed weeks also revealed virtually identical results (data not shown).

Discussion

No significant differences were observed in the main analyses looking at associations of binge drinking on any of the above clinical cut-off scale scores. However, since previous studies looking at prenatal exposure to binge drinking and associations with behavioural and emotional outcomes generally have only found weak associations, we did not a priori expect such associations[18, 19]. From this perspective it was not expected that exposure to one or two binge episodes would lead to substantial childhood behavioural and emotional

problems, i.e. significant increase in the number of children with above clinical cut-off scores. However, it was expected that it would lead to an increase in mean score, i.e. lead to subtle behavioural differences.

The analyses did reveal significant, elevated externalising estimates for children of late bingers, estimates that were stable even after excluding women with more than two binge episodes. In other words, being exposed to just one or two binge drinking episodes late in pregnancy was associated with an increased risk of subtle behavioural differences at age seven. For the early bingers there was, a significant albeit very weak increase in externalising scores at age seven. One systematic review looking at human binge studies concluded that there was a possible, but generally quite small effect of exposure to binge drinking on neurodevelopmental outcomes[19]. The studies reported in the review all investigated exposure to early pregnancy binge drinking with results that are in line with the findings from the present study, where very weak associations were observed with early pregnancy exposure. Another study including women with binge episodes throughout pregnancy found similar weak associations with binge drinking, and only for boys[18]. However, as is apparent from the present study far fewer women report binge episodes in the last part of pregnancy compared to early pregnancy masking potentially stronger effects of late exposure. The much lower group size here results in less reliable estimates and lower statistical power. To our knowledge no other humans studies have to date investigated the specific associations of being exposed to binge drinking exclusively in late pregnancy, and the results indicate that late exposure is actually associated with significantly more externalising problems at age seven compared to the group exposed in early pregnancy.

Previous studies have typically used data on maternal binge drinking limited to the early period of pregnancy and have most often found no or at most modest associations with subsequent development in the child[20, 21]. The present findings suggest that it actually is important also to examine binge drinking late in pregnancy. An obvious limitation of the present study is that the late binge group consist of a highly selected group of women and the generalizability of the results to other women may be limited. Also, parent ratings were used as the outcome measure in the study, despite teacher rating may be better at assessing the behaviour of the children. Researches are thus encouraged to investigate whether the findings from our study can be replicated across other samples.

Conflict of interest: there are no declared conflict of interests for any of the authors.

Acknowledgements: The study was supported financially by the Department of Psychology, University of Copenhagen; Ludvig og Sara Elsass Foundation; Aase og Ejnar Danielsens Foundation; Carl J. Becker's Foundation; the Lundbeck Foundation; Børne- og Ungdomspsykiatrisk Selskab i Danmark; Dagmar Marshalls Foundation; The A.P. Møller Foundation for the Advancement of Medical Science; Direktør Jakob Madsens Legat.

Reference List

[1] Irner TB. Substance exposure in utero and developmental consequences in adolescence: A systematic review. Child Neuropsychol 2011 Nov 25.

[2] Gray R, Henderson J. Rewiev of the Fetal Effects of Prenatal Alcohol Exposure. National Perinatal Epidemiology Unit, University of Oxford; 2007 May.

[3] O'Leary CM, Zubrick SR, Taylor CL, Dixon G, Bower C. Prenatal alcohol exposure and language delay in 2-year-old children: the importance of dose and timing on risk. PEDIATRICS 2009 Feb;123(2):547-54.

[4] Willford J, Leech S, Day N. Moderate prenatal alcohol exposure and cognitive status of children at age 10. Alcohol Clin Exp Res 2006 Jun;30(6):1051-9.

[5] Andersen SL. Trajectories of brain development: point of vulnerability or window of opportunity? Neurosci Biobehav Rev 2003 Jan;27(1-2):3-18.

[6] Rice D, Barone S Jr. Critical periods of vulnerability for the developing nervous system: evidence from humans and animal models. Environ Health Perspect 2000 Jun;108 Suppl 3:511-33.

[7] Schlotz W, Phillips DI. Fetal origins of mental health: evidence and mechanisms. Brain Behav Immun 2009 Oct;23(7):905-16.

[8] Schneider ML, Moore CF, Becker EF. Timing of moderate alcohol exposure during pregnancy and neonatal outcome in rhesus monkeys (Macaca mulatta). Alcohol Clin Exp Res 2001 Aug;25(8):1238-45.

[9] Maier SE, Miller JA, Blackwell JM, West JR. Fetal alcohol exposure and temporal vulnerability: regional differences in cell loss as a function of the timing of binge-like alcohol exposure during brain development. Alcohol Clin Exp Res 1999 Apr;23(4):726-34.

[10] Mooney SM, Varlinskaya EI. Acute prenatal exposure to ethanol and social behavior: effects of age, sex, and timing of exposure. Behav Brain Res 2011 Jan 1;216(1):358-64.

[11] Olsen J, Melbye M, Olsen SF, Sorensen TI, Aaby P, Andersen AM, et al. The Danish National Birth Cohort--its background, structure and aim. Scand J Public Health 2001 Dec;29(4):300-7.

[12] Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 1997 Jul;38(5):581-6.

[13] Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? J Abnorm Child Psychol 1999 Feb;27(1):17-24.

[14] Goodman R. Psychometric properties of the strengths and difficulties questionnaire. Journal of the American Academy of Child & Adolescent Psychiatry 2001 Nov;40(11):1337-45.

[15] Niclasen J, Skovgaard AM, Andersen AM, Somhovd MJ, Obel C. A Confirmatory Approach to Examining the Factor Structure of the Strengths and Difficulties Questionnaire (SDQ): A Large Scale Cohort Study. J Abnorm Child Psychol 2012 Sep 25.

[16] Niclasen J. SDQ Danish cut off scores: 5-7-year-olds. 2012.

[17] Niclasen J, Teasdale TW, Andersen AM, Skovgaard AM, Elberling H, Obel C. Psychometric properties of the Danish strength and difficulties questionnaire: The SDQ assessed for more than 70,000 raters in four different cohorts. PLoS One 2012;7(2):e32025.

[18] Niclasen J, Andersen AM, Teasdale TW, Strandberg-Larsen K. Binge drinking and cumulated alcohol exposure in pregnancy: behavioural and emotional development at age seven. 2013.

[19] Henderson J, Kesmodel U, Gray R. Systematic review of the fetal effects of prenatal binge drinking. J Epidemiol Community Health 2007;61:1069-73.

[20] Kesmodel US, Eriksen HL, Underbjerg M, Kilburn TR, Stovring H, Wimberley T, et al. The effect of alcohol binge drinking in early pregnancy on general intelligence in children. BJOG 2012 Sep;119(10):1222-31.

[21] Sayal K, Heron J, Golding J, Alati R, Smith GD, Gray R, et al. Binge pattern of alcohol consumption during pregnancy and childhood mental health outcomes: longitudinal population-based study. PEDIATRICS 2009 Feb;123(2):e289-e296.

	No binge	Early binge	Late binge
N	25781	3654	94
	Means (SD)	**Means (SD)**	**Means (SD)**
Maternal age	30.8 (4.2)	30.4 (4.2)****	31.0 (4.5)
Paternal age	32.6 (5.1)	32.1 (5.2)****	32.6 (6.0)
Cumulated alcohol intake in pregnancy[3]	26.7 (38.4)	42.0 (51.3)****	64.1 (63.7)****
	Percentages	**Percentages**	**Percentages**
Tenant (yes)[1]	23.8%	29.3%****	29.8%
Maternal education			
0-9 years of education	7.0%	5.9%	7.5%
10-12 years of education	40.7%	39.8%	54.3%
>13 of education	52.3%	54.2%*	38.3%*
Paternal education			
0-9 years of education	12.4%	11.5%	18.1%
10-12 years of education	47.2%	47.4%	43.6%
>13 years of education	40.4%	41.1%	38.3%
Maternal psychiatric diagnosis[2]	4.8%	5.2%	10.6%**
Paternal psychiatric diagnosis[2]	3.4%	3.7%	1.1%
Maternal smoking	18.91%	30.62%****	32.26%***
Paternal smoking	26.2%	30.5%****	27.7%
BMI[4]	68.81%	72.15%****	70.97%
Planned pregnancy (yes)	80.59%	73.67%****	67.02%***

Table 1: background characteristics of the three exposure groups. Two tailed ANOVA and Chi Square probability testing comparing the 'no binge' group with the 'early' and 'late' binge groups, separately. *: sig. at 0.05; **: sig. at 0.01; ***: sig. at 0.001; ****: sig. at 0.0001. [1]Tenant, homeless or living with parents; [2]appears in the Psychiatric Central Registry before the age of seven of the child; [3]intake of total number of drinks in the entire pregnancy, excluding binge episodes; [4]Pre-pregnancy body Mass Index within the normal range of 18.5-<25

	No binge (N=25,781)	Early binge (N=3,654)	Late binge (N=94)
	Relative changes in means	***Relative changes in means***	***Relative changes in means***
Externalising	Ref.	**1.02 (1.00-1.05)**	**1.21 (1.04-1.42)**
Internalising	Ref.	1.00 (0.98-1.03)	1.05 (0.89-1.24)

Table 2: Adjusted relative changes in means for the log-transformed externalising/ internalising scores. The analyses are adjusted for the following confounders: maternal education, psychiatric diagnoses, age and smoking

	No binge (N=25,781)		Early binge (N=3,654)		Late binge (N=94)	
	Cases (%)		***Cases (%)***	***OR***	***Cases (%)***	***OR***
Hyperactivity	4748 (18.42%)	Ref.	738 (20.20%)	0.93 (0.85-1.02)	19 (20.21%)	0.98 (0.60-1.68)
Conduct	3596 (13.95%)	Ref.	521 (14.26%)	1.01 (0.91-1.11)	17 (18.09%)	0.81 (0.49-1.43)
Emotional	3401 (13.19%)	Ref.	530 (14.50%)	0.93 (0.84-1.03)	15 (15.96%)	0.86 (0.51-1.57)
Peer problems	5019 (19.47%)	Ref.	690 (18.88%)	1.04 (0.96-1.14)	17 (18.09%)	1.17 (0.70-2.05)

Table 3: Adjusted ORs for the above cut-off hyperactivity, conduct, peer problems and emotional scores. The analyses are adjusted for the following confounders: maternal education, psychiatric diagnoses, age and smoking

Paper 5

Niclasen, J. Drinking or not drinking in pregnancy: the multiplicity of confounding influences (accepted for publication in *"Alcohol and Alcoholism"*).

Drinking or not drinking in pregnancy: the multiplicity of confounding influences

Niclasen, J.

Janni Niclasen

Øster Farimagsgade 2A

1353 Copenhagen K

Telephone +45 40817776

Janni.niclasen@psy.ku.dk

Running title: Drinking or not drinking in pregnancy: the role of confounding

Keywords: prenatal alcohol exposure, prenatal alcohol intake, background characteristics, psychological outcomes, cohort, methodology, lifestyle, confounders, confounding, reproductive, DNBC, prenatal, alcohol, lifestyle, exploratory factor analysis, EFA, principal component analysis, factor scores

This work was supported by Department of Psychology, University of Copenhagen; Ludvig og Sara Elsass Foundation; Aase og Ejnar Danielsens Foundation; Carl J. Becker's Foundation; the Lundbeck Foundation; Børne- og Ungdomspsykiatrisk Selskab i Danmark; Dagmar Marshalls Foundation; The A.P. Møller Foundation for the Advancement of Medical Science; and Direktør Jakob Madsens Legat.

Abstract

Background: Studies investigating associations between prenatal exposure to low-moderate doses of alcohol and mental health development in childhood are inconsistent. The aim of the present study was to compare women who drink and who do not drink alcohol in pregnancy on a number of potential confounding variables, and to investigate whether any latent variables could be identified among these.

Methods: Data were obtained from the Danish National Birth Cohort. Exposure: cumulated alcohol intake in full pregnancy (N=63,464). The women were subdivided into intake groups 0, >0-10, >10-30, >30-90 and >90 units of alcohol in pregnancy. Hereafter, the abstainers were subdivided into an all-time and a pregnancy-abstaining group, and the high intakers (>90) into a high (>90-180) and a very-high (>180) intake group. Outcome: self-reported and register-based information on socio-demographic and lifestyle factors, and latent variables from an exploratory factor analysis.

Results: Significant differences were observed between the intake groups on virtually all parameters. Significant differences were observed between the abstaining and the high-intake groups. The exploratory factor analyses identified a number of latent variables between the potential confounding variables.

Conclusions: Differences on confounding factors may in part explain the lack of consistency in the literature. It is cautiously concluded that the failure to control for these factors introduce residual confounding into the analyses, and thus masks the potential (small) effect of being exposed to low doses of alcohol in pregnancy. It is recommended that future studies control for factor scores rather than for the observed variables as is practice today.

Introduction

Forty years ago the Lancet published Jones and Smith's now legendary article on 'Recognition of the Fetal Alcohol Syndrome in early infancy', describing the first eight identified clinical cases of fetal alcohol syndrome (FAS) (Jones and Smith 1973). The focus in the early years was on identifying effects of large doses of alcohol on mental health outcomes. Since then, much research has been carried out investigating associations between exposure to more moderate doses of alcohol, typically <1 unit/week, and binge drinking on the one hand and mental health outcomes in childhood on the other (Gray and Henderson 2007;

Henderson et al. 2007; Sayal et al. 2007a; Sayal et al. 2009; Sayal et al. 2013). Thus, the focus today is largely on establishing whether there is a 'safe' lower threshold below which drinking alcohol in pregnancy is not associated with any damages to the developing fetus. The findings from these observational studies are somewhat contradictory. Some studies have concluded that prenatal exposure to lower doses of alcohol is indeed negatively associated with mental health development in childhood (Fried and Watkinson 1988; Streissguth et al. 1990; Olson et al. 1997; Larroque and Kaminski 1998; NIAAA: National Institute of Alcohol Abuse and Alcoholism 2000; Sood et al. 2001; Sayal et al. 2007; Testa et al. 2007), whereas others have not found such associations (O'Leary et al. 2009; Rodriguez et al. 2009; Skogerbo et al. 2012; Underbjerg et al. 2012; Sayal et al. 2013). Many studies even report on a J-shaped association between alcohol exposure and mental health outcomes in childhood, such that exposure to low doses of alcohol has an apparently protective effect on the foetus (Kelly et al. 2009; Robinson et al. 2010; Kelly et al. 2012). Observational studies today generally apply multivariate research designs and control for what are considered the most relevant confounding factors. However, other confounding factors that may be relevant for mental health development are not controlled for, and the examination of potential mediating factors is virtually absent.

The primary aim of the present study was to thoroughly describe women who drink and women who do not drink alcohol in pregnancy on a large number of background characteristics these being, potential confounding factors, including socio-demographic and lifestyle factors. All-time abstainers and pregnancy-abstainers were then compared separately as were high intakers and very-high intakers. It was hypothesized that if large variations were observed between exposure groups, and such variables were not controlled for in the statistical analyses this would introduce residual confounding. This residual confounding could in turn mask potential effects of exposure to lower doses of alcohol, and thus explain the lack of consistency in studies hitherto concerned with prenatal exposure to low doses of alcohol and mental health development in childhood. The secondary aim was to carry out exploratory factor analyses in order to identify possible subtle latent variables underlying the reported background characteristics.

Methods

Sample

Data were derived from the population-based, large-scale birth cohort, the Danish National Birth Cohort (DNBC), that comprises information on 100,418 pregnancies (Olsen et al. 2001). The intention of the DNBC was to examine associations between diverse forms of exposures early in pre- and postnatal life and the health and development of the children from a life-course perspective. Between 1996 and 2002 pregnant women were enrolled nationwide at their first antenatal visit. The women were approached twice in pregnancy at approximately weeks 16 and 30 regarding their lifestyle in the early and middle part of pregnancy and then again at age six months of their child regarding their lifestyle in the last part of pregnancy. The sample for the present study was restricted to women with full information on key alcohol variables (variables with information on weekly average alcohol intake of wine, beer and spirits from in all three interviews) leaving a total of 63,464 women in the study.

Prenatal alcohol intake

In the three interviews the women answered separate questions regarding their weekly average intake of beer, wine and spirits. In order to compute a single estimate for the cumulated intake of alcoholic drinks in the entire pregnancy the reported intake from each interview was multiplied by the number of weeks between each interview (See Figure 1). Because the focus was to investigate the background characteristics of women with different estimates of cumulated alcohol intake in pregnancy, information from the three interviews was cumulated in order to obtain a total intake sum-scores. The following categories of cumulated alcohol intake were adopted: 0, >0-10, >10-30, >30-90, >90. These categories were selected in order to be able distinguish between intake groups with what can be considered very low intakes of alcohol, and further to include a fairly large number of women in each of the intake groups.

Outcome measure: background characteristics

The outcomes included both self-reported and register-based information on background characteristics including socio-demographic and lifestyle factors. Variables were included if they were a priori identified as possible confounders of the association between prenatal exposure to low-moderate doses of alcohol and mental health development. The socio-demographic factors included information on: maternal and paternal age, marriage, owning one's own house, maternal and paternal education, and a number of variables on self-

reported and register-based psychological problems and psychiatric diagnoses. The lifestyle factors included prior-to-pregnancy and in-pregnancy information on average alcohol consumption and binge drinking, pre-pregnancy BMI, in-pregnancy information on maternal and paternal smoking, in-pregnancy information on maternal intake of coffee, cola, vitamins, iron, fish oil, fish, analgesics and sleep medication, in-pregnancy maternal habits of TV watching and exercise, as well as in-pregnancy occurrences of maternal diabetes, asthma and anaemia.

Statistical analyses

The analyses were carried out using SAS version 9.2 and SPSS version 19. By means of SAS, the data were analysed using ANOVA (for continuous variables) and Chi Square (for categorical variables) tests to test for differences between intake groups and LR tests for linear trends. In order to identify possible latent factors between the observed variables (i.e. the background characteristics), exploratory factor analyses (EFA) with Varimax rotation were carried out using SPSS.

Results

Of the 63,464 women included in the sample a total of 56,258 (88.7%) women reported drinking average doses of alcohol on at least one occasion in the recognized or unrecognized part of pregnancy (Table 1). The majority of women reported drinking prior to pregnancy (87.2%) whereas similar percentages of women reported drinking average quantities of alcohol in early (44.8%), middle (49.5%) and late (46.9%) pregnancy.

Background characteristics of women who drink and do not drink alcohol in pregnancy

ANOVA and Chi Square tests between the intake groups revealed significant inter-group differences on most variables (Table 1). Socio-demographic factors: significant differences between the groups were observed for all variables. Significant linear trends were further observed for most variables. Maternal and paternal ages were positively associated with alcohol intake, as was university education. Housing (owning your own house), on the other hand, was negatively associated with cumulated alcohol intake. For the remaining

variables a J-shape or reversed J-shape curve appeared, i.e. a linear trend was observed for four intake groups, whereas the fifth group (the abstaining or the high-intake group) had lower or higher estimates forming the curve on the J-shaped association between intake groups. This was observed for the married, mandatory education, and all of the psychological problems/ psychiatric diagnoses variables. Compared to the high intake group (>90 alcohol containing units, hereafter units), three times as many abstainers (0 units) had mandatory education only. The opposite was true for university education. For the psychological problems/ psychiatric diagnoses variables the highest percentages were observed for the abstainers (0 units) and the high intakers (>90 units), and this was true for the self-reported as well as for the register-based variables. Lifestyle factors: apart from the fish oil variable significant differences were observed between the intake groups for all variables and linear trends were also observed for virtually all variables. The cumulated alcohol intake variable was found to be linearly associated (positively or negatively) with coffee, all alcohol variables, fish eating, BMI, TV, diabetes and asthma. For the remaining variables, namely smoking variables, vitamin, iron, analgesics and anaemia variables a J-shape trend appeared, i.e. the low and high intake groups had similar highest or lowest estimates whereas the three remaining mid-intake groups revealed linear trends. For example, the abstaining (0 units) and high intake (>90 units) groups had mean cumulated cigarettes estimates of 99 and 78 (total number of cigarettes in full pregnancy), respectively, whereas the means for the remaining groups were 62, 54 and 50 cigarettes in full pregnancy.

To further investigate the curve in the shape function observed for many of the variables, the abstaining groups was subdivided into an all-time abstaining group and a group of women ceasing to drink alcohol once recognising their pregnancy. Correspondingly, the high intake group (>90 units) was subdivided into a high (>90-180 units) and a very-high (>180 units) intake group.

Differences in background characteristics between all-time abstainers and pregnancy-only abstainers

The abstaining (0 units) group was subdivided into a group of all-time abstainers (N=7,206) and a group of women drinking prior to pregnancy but ceasing once recognising their pregnancy (N=16,563) (Table 2). The groups differed significantly on a large number of variables, and the means and percentages for the two groups were often lying on opposite sides of the total sample mean. Socio-demographic factors: significant differences were observed for all the education variables. The percentages for the pregnancy-abstainers were generally very close to the total sample. The all-time abstainers on the other hand, included more than

twice as many women with mandatory education only, and correspondingly less than half as many women with a university degree. Furthermore, significantly more women had psychological problems/ psychiatric diagnoses in the all-time abstaining group compared to the pregnancy-abstaining group. Lifestyle factors: The means and percentages for the pregnancy-abstainers on the lifestyle factors resembled those for the total sample, whereas the estimates for the all-time abstaining group were significantly different from the pregnancy-abstaining group. This was, for example, true for the smoking variables: significant differences were observed between the two intake groups, the all-time abstainers showing high smoking estimates and the pregnancy-abstainers estimates similar to those observed for the total sample. For the remaining lifestyle factor variables two overall patterns emerged: one where the estimates for the pregnancy-abstaining group resembled those for the total sample and one where the estimates for the all-time abstainers were more adverse. This was so for the fish-eating variable, BMI, Cola, TV, exercise, diabetes and asthma variables. The other pattern that appeared was one where the estimates for the all-time abstainers were similar to the total sample, but with the pregnancy-abstainers having more beneficial estimates. This was true for the vitamin, analgesics and anaemia variables.

Differences in background characteristics between high (>90-180) and very-high (>180) intakers

Subdividing the high intake (>90 units) group into a high (>90-180 units) (N = 4,605) and a very-high (>180 units) (N = 871) intake group revealed two groups that were statistically significant on some variables (Table 3) but fewer than appeared between the two abstaining groups. Socio-demographic factors: The very-high intake group (>180 units) included significantly older fathers and twice as many unmarried women. No significant differences were observed for maternal age and housing. Significantly more women had mandatory education only in the very-high intake (>180 units) group (9.1% vs. 6.0%), whereas significantly more women had a university degree in the high intake (>90-180 units) group (21.7% vs. 18.2%). Similar patterns were observed for fathers. The percentages for the psychological problems/ psychiatric diagnoses variables for the high-intake group (>90-180 units) resembled those of the total sample. The percentages for the very-high intake group (>180 units) on the other hand were almost twice as large compared to the high-intake group (>90-180 units) for both self-reported and register-based variables. The differences were mostly statistically significant. Regarding alcohol intake the groups differed on prior-to-pregnancy and in-pregnancy alcohol intake. The very-high intake group (>180 units) reported on average twice as many episodes of binge

drinking (2.3 versus 1.2 binge episodes in pregnancy), more than twice as high cumulated alcohol intake in pregnancy (263 versus 121) and significantly higher weekly average alcohol intake prior to pregnancy (11.4 versus 7.7 units of alcohol per week). Lifestyle factors: the very-high intake group (>180 units) reported twice as high cumulated smoking frequencies (138 versus 66 cigarettes in pregnancy), more smoking partner (42.5% versus 29.8%) vs. more women reported drinking coffee (70.6% versus 64.0%) and ≥1 litre of cola/ week (17.7% versus 14.9%), more women reporting watching ≥2½ hours of TV/ day (19.1% versus 15.1%), but fewer women reported doing exercise in the last part of pregnancy (28.1% vs. 22.7%). Significantly more women in the very-high intake group (>180 units) reported eating vitamins (84.1% versus 77.1%). All differences were statistically significant.

Factor analysis

It was hypothesized that the large number of potential confounding factors presented above could possible inter-correlate to a lesser or greater degree. For this reason it was decided to carry out exploratory factor analyses (EFA) to investigate the magnitude of inter-correlations between the observed variables and potentially to identify subtle latent variables underlying these. Specifically, principal component analysis (PCA) with Varimax rotation was employed. The results revealed that a total of 11 components showed an initial Eigenvalue of >1. Because of this large number, a six factor solution was decided on following examination of the scree plot. Table 4 shows that the first 'Average alcohol consumption' factor explained 11.5% of the total variance, and included all of the non-binge alcohol variables. Likewise, the second 'Stimulants' factor explained 7.4% of the total variance and included all the smoking as well as the coffee variable. The third 'Parental age' factor explained 4.9% of the total variance, and included the age and housing variables. The fourth factor 'Maternal mental health' factor explained 4.4% of the total variance and included the four maternal psychiatry variables, but none of the paternal psychiatry variables. The fifth 'binge drinking' factor explained 3.7% of the total variance and included the binge drinking variables, whereas the sixth 'educational-related lifestyle' factor explained 3.4% of the total variance and included both the maternal and paternal educational variables as well as the BMI, exercise and TV variables. The variables not presented in Table 4 correlated <0.30 with the six extracted factors. However, the variables were included in the factor analyses. Hereafter, another series of PCAs was carried out (shown in brackets) specifying the number of factors to one. Each of these PCAs only included the variables identified as having

large >0.30 factor loading in the initial PCA. These figures thus represent factor loadings not 'contaminated' by the variables from the other factors.

Discussion

The main results revealed significant differences between intake groups on most of the background characteristics. Compared to the abstainers, the alcohol intakers were older, more likely to have a university degree, eat fish and to have a pre-pregnancy BMI within the normal range. Further, they watch TV and drank cola less. The abstainers on the other hand were younger, more likely to have mandatory education only, drink cola, watch TV, smoke cigarettes, live alone and have psychiatric problems. They were less likely to do exercise and eat fish. Linear trends were observed between all of the intake groups despite the fact that the low alcohol exposure group in the present study included women who reported drinking as little as less than ten units of alcohol throughout the entire pregnancy, i.e. less than a quarter of a drink per week. These results are important as they shed light on the fundamental differences that exist between women who drink and women who do not drink alcohol in pregnancy.

The results should be of interest to researchers doing observational studies on prenatal exposure to low-moderate doses of alcohol and child mental health outcomes. Currently, findings from such observational studies are ambiguous. Sometimes they report a negative association between prenatal exposure to low doses of alcohol and mental health development in childhood (Olson et al. 1997; Sood et al. 2001; Sayal et al. 2007a;), but sometimes they do not (O'Leary et al. 2009; Rodriguez et al. 2009; Skogerbo et al. 2012; Underbjerg et al. 2012; Sayal et al. 2013). Very often they even report a J-shaped association (Kelly et al. 2009; Robinson et al. 2010; Kelly et al. 2012), where exposure to lower doses of alcohol apparently act as a protective factor for the development of mental health problems in childhood. Part of this ambiguity in results may be explained if similar variations in background characteristics exist in the samples applied in the existing observational literature. Unfortunately, it is not standard for observational studies to report on such background variables. Most studies only control for what are considered the most important confounding factors such as maternal age, smoking and education – factors that influence the causal pathway between prenatal alcohol exposure and child development. However, these differences may explain (at least some) of the inconsistency in the observational epidemiological literature. If the effect of low doses of alcohol on

mental health in childhood is expected to be low, the total sum of the residual confounding of the socio-demographic and lifestyle factors that are not controlled will add up to a larger apparent 'protective' factor compared to the small 'negative' effect of being exposed to low doses of alcohol prenatally.

Socio-demographic and lifestyle factors, similar to the ones presented above, are part of many birth cohort studies. Information on other strong confounders, such as personality, IQ and social support is on the other hand rarely included, and therefore not possible to control for. When such strong confounders are not controlled for residual confounding remains that bias the results. For example, parental psychological problems constitute a very strong risk factor for the development of child mental health problems (Downey and Coyne 1990; Todorow et al. 2010). Todorow et al (2010) argue that the differences in mental health between abstainer and light drinkers can directly contribute to the improved mental health outcomes in the latter group, because they have the ability to provide more effective parenting and model more adaptive behavioural and emotional regulation. Significantly higher rates of psychological problems were observed for the abstaining group compared to the low intake group and a similar tendency was observed for all-time abstaining vs. pregnancy-abstaining intake groups.

Similarly, few studies control for mediating factors; i.e. factors that causes variation in the outcome variable (child mental health) and are themselves caused to vary by the exposure variable (prenatal alcohol exposure) (Porta 2008). Strong mediators include attachment, family functioning, parent-infant interaction and child IQ. If such factors are not controlled for these will induce bias to the reported results. For example, the close association between attachment style and mental health development in children is well known. Higher rates of secure attachment styles are observed among well-educated groups compared to low educational groups. In the 1950's Bowlby was the first to demonstrate the lasting consequences that the quality of the mother-child relationship has for a wide range of developmental cognitive and mental health outcomes (Bowlby 1950). He concluded that infants who develop a secure attachment style are those with a history of sensitive and responsive maternal care (Bowlby 1950). Furthermore, this attachment style is associated with better emotional regulation, higher self-esteem, and more developed coping skills in the child. In turn, these factors make the children better able to handle stressful or challenging situations and it lowers the risk for poorer mental health outcomes later in life. On the other hand, children with an insecure attachment are at greater risk for poor mental health outcomes (Ainsworth and Bell 1970; Sroufe 2005).

Where one problem with the observational literature is that only a limited number of variables are controlled for, another reason for the ambiguity in the literature could be attributed to the division of the intake groups. The abstaining group is most often not subdivided on the basis of pre-pregnancy abstinence (Sayal et al. 2007; Rodriguez et al. 2009; Kelly et al. 2009; Sayal et al. 2013). However one thorough study did report that the socio-economic profile of mothers in the pregnancy-abstaining group was more advantaged than the all-time abstaining group (Kelly et al. 2012). The two abstaining groups also differed significantly in the present study, with observed means and percentages on each side of the total sample mean. It is on this basis recommended that future studies also divide pregnancy abstainers according to pre-pregnancy abstinence.

Differences for the high intake groups were observed on some, but not on as many variables. However, this could in part be explained by the relatively small size of the very-high exposure group. The high and the very-high intake groups did differ significantly on most other variables, indicating that careful consideration should also be given to the subdivision of the high intake groups in observational studies. Another point needs to be made regarding the very-high intakers. This group reported twice as many binge episodes as the high intakers and "binge drinking" was identified as the fifth factor in the EFA explaining 3.7% of the total variance. It is on this basis suggested that future observational epidemiological studies should control for number of binge episodes.

Taken together, the large differences within the extreme groups (both the abstainers and high intakers) may explain the J-shaped function of alcohol intake. Together with the large variabilities in background characteristics described above, and the potential large differences between groups on other important confounding and mediating factors it is suggested that the inconsistency in the existing literature, at least in part, can be explained by these factors. In other words, the lack of consistency may reflect spurious associations between unmeasured and residual confounding (and mediating) factors in the existing literature.

The factor analyses identified a number of underlying, latent factors - factors that can be directly applied in epidemiological studies looking at associations between prenatal exposure to alcohol and mental health outcomes in childhood. It is recommended that future observational studies set out by conducting an EFA after which the extracted factors are controlled for, instead of controlling for the observed background variables individually. The factor scores for each factor, for each person in the sample, will, all other things

being equal, be more precise because they are calculated on the basis of information from all the variables constituting that particular factor. This way the method will allow for more factors to be controlled for, without the loss of statistical power.

Limitations

Firstly, It was decided to describe the background characteristics on the basis of full pregnancy intake and investigate differences between the groups with a very low intake. This was decided on despite the fact that most studies investigate the effects of alcohol intake in the early part of pregnancy. The tendencies between intake groups described above might have been different had they been defined on the basis of first trimester intake only. However, since the aim of the present study was not to specifically show differences between these particular groups, but instead to report potential tendencies between intake groups, this point is of less relevance. Future work however, could replicate the present study with the groups defined differently, and from other cohorts to investigate to what degree the tendencies reported in present study are replicable. Secondly, a similar study could be carried out describing differences between intake groups on mediating factors. Thirdly, the present study did not report on uses of psychotropic medication, e.g. anti-depressive and benzodiazepines or the use of illicit drugs. It could be relevant to include these variables in future descriptive and observational studies.

References

Ainsworth MD & Bell SM 1970. Attachment, exploration, and separation: illustrated by the behavior of one-year-olds in a strange situation. Child Dev 41:49-67 available from: PM:5490680

Bowlby J 1950. Research into the origins of delinquent behaviour. Br Med J 1: 570-573 available from: PM:20787782

Downey G & Coyne JC 1990. Children of depressed parents: an integrative review. Psychol Bull 108: 50-76 available from: PM:2200073

Fried PA & Watkinson B 1988. 12- and 24-month neurobehavioural follow-up of children prenatally exposed to marihuana, cigarettes and alcohol. Neurotoxicol Teratol 10: 305-313

Gray R & Henderson J 2007, Review of the fetal effects of prenatal alcohol exposure, National perinatal epidemiology unit, University of Oxford.

Henderson J, Gray R & Brocklehurst P 2007. Systematic review of effects of low-moderate prenatal alcohol exposure on pregnancy outcome. BJOG 114: 243-252

Jones KL & Smith DW 1973. Recognition of the fetal alcohol syndrome in early infancy. Lancet 302: 999-1001 available from: PM:4127281

Kelly Y, Sacker A, Gray R, et al. 2009. Light drinking in pregnancy, a risk for behavioural problems and cognitive deficits at 3 years of age? Int J Epidemiol 38: 129-140 available from: PM:18974425

Kelly, YJ, Sacker A, Gray R. et al. 2012. Light drinking during pregnancy: still no increased risk for socioemotional difficulties or cognitive deficits at 5 years of age? J Epidemiol Community Health 66: 41-48 available from: PM:20924051

Larroque B. & Kaminski M. 1998. Prenatal alcohol exposure and development at preschool age: main results of a French study. Alcohol Clin Exp Res 22: 295-303 available from: PM:9581632

NIAAA: National Institute of Alcohol Abuse and Alcoholism 2000, 10th special report to the US congress on alcohol and health. Highlights from current research, MD: Department of health and human services, Bethesda, 00-1583.

O'Leary CM, Zubrick SR, Taylor CL, et al. 2009. Prenatal alcohol exposure and language delay in 2-year-old children: the importance of dose and timing on risk. Pediatrics 123: 547-554 available from: PM:19171621

Olsen J, Melbye M, Olsen SF, et al. 2001. The Danish National Birth Cohort--its background, structure and aim. Scand J Public Health 29: 300-307

Olson HC, Streissguth AP, Sampson PD, et al. 1997. Association of prenatal alcohol exposure with behavioral and learning problems in early adolescence. J Am Acad Child Adolesc Psychiatry, 36: 1187-1194

Porta, M 2008. A dictionary of epidemiology Oxford University Press.

Robinson M, Oddy WH, McLean NJ, et al. 2010. Low-moderate prenatal alcohol exposure and risk to child behavioural development: a prospective cohort study. BJOG 117: 1139-1150 available from: PM:20528867

Rodriguez A, Olsen J, Kotimaa AJ, et al. 2009. Is prenatal alcohol exposure related to inattention and hyperactivity symptoms in children? Disentangling the effects of social adversity. J Child Psychol Psychiatry 50: 1073-1083 available from: PM:19298478

Sayal K 2007. Alcohol consumption in pregnancy as a risk factor for later mental health problems. Evid Based Ment Health 10: 98-100 available from: PM:17962644

Sayal K, Draper ES, Fraser R, et al. 2013. Light drinking in pregnancy and mid-childhood mental health and learning outcomes. Arch Dis Child 98: 107-111 available from: PM:23322857

Sayal K, Heron J, Golding J, et al. 2007a. Prenatal alcohol exposure and gender differences in childhood mental health problems: a longitudinal population-based study. Pediatrics 119: e426-e434 available from: PM:17272604

Sayal K, Heron J, Golding J, et al. 2009. Binge pattern of alcohol consumption during pregnancy and childhood mental health outcomes: longitudinal population-based study. Pediatrics 123: e289-e296 available from: PM:19171582

Skogerbo A, Kesmodel US, Wimberley T, et al. 2012. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on executive function in 5-year-old children. BJOG 119: 1201-1210 available from: PM:22712874

Sood B, Delaney-Black V, Covington C, et al 2001. Prenatal alcohol exposure and childhood behavior at age 6 to 7 years: I. dose-response effect. Pediatrics 108: E34

Sroufe LA 2005. Attachment and development: a prospective, longitudinal study from birth to adulthood. Attach Hum Dev 7: 349-367 available from: PM:16332580

Streissguth AP, Barr HM, & Sampson PD 1990. Moderate prenatal alcohol exposure: Effects on child IQ and learning problems at age 7½ years. Alcoholism: Clinical and Experimental Research, 14: 662-669

Testa M, Quigley BM, & Eiden RD 2007. The effects of prenatal alcohol exposure on infant mental development: a meta-analytical review. Alcohol Alcohol 38: 295-304

Todorow M, Moore TE & Koren G 2010. Investigating the effects of low to moderate levels of prenatal alcohol exposure on child behaviour: a critical review. J Popul Ther Clin Pharmacol 17: e323-e330 available from: PM:20962358

Underbjerg M, Kesmodel US, Landro NI et al. 2012. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on selective and sustained attention in 5-year-old children. BJOG 119: 1211-1221 available from: PM:22712829

Figure 1: timeline showing the period of which information from each interview was used

		Cumulated alcohol intake						
Alcohol group	Full sample	0	>0-10	>10-30	>30-90	>90	P	Linear trend
N	63,464	7204	19111	15054	16619	5476		
Socio-demographic factors								
Age (M)[1,2]	30.5	29.5	29.8	30.5	31.1	32.5	****	<0.0001
Age (P)[1,3]	32.3	31.5	31.6	32.2	32.9	34.6	****	<0.0001
Married (no)[4]	2.3%	3.5%	2.5%	2.1%	1.9%	2.4%	****	0.75, NS
Tenant[5]	26.7%	30.6%	28.0%	26.2%	25.5%	22.8%	****	<0.0001
Education								
Mandatory (M)[2,6]	8.4%	17.8%	9.3%	6.6%	5.4%	6.5%	****	<0.0001
University (M)[3,6]	14.6%	5.8%	11.3%	15.4%	19.2%	21.2%	****	<0.0001
Mandatory (P)[2,6]	13.3%	22.9%	14.8%	11.5%	9.8%	10.6%	****	<0.0001
University (P)[3,6]	15.3%	7.1%	12.4%	15.8%	19.6%	21.7%	****	<0.0001
Psychological problems/ Psychiatric diagnoses								
self-rep. pre-preg. prob.[7]	7.5%	9.4%	7.2%	6.9%	7.2%	8.6%	****	0.0004***
Pre-preg. (M)[2,8]	2.9%	4.8%	3.0%	2.4%	2.4%	3.0%	****	0.40, NS
Pre-preg. (P)[3,8]	1.9%	2.8%	1.9%	1.7%	1.5%	1.9%	****	0.37, NS
self-rep. psych. prob. in preg.[9]	1.2%	1.6%	1.1%	1.0%	1.1%	1.6%	****	0.05*
After birth (M)[2,10]	3.6%	5.4%	3.9%	3.1%	2.9%	3.7%	****	0.05*
After birth (P)[3,10]	2.4%	3.8%	2.5%	2.1%	1.9%	2.3%	****	0.09, NS
Lifestyle factors								
Binge drinking[11]	0.5	0.1	0.3	0.5	0.7	1.4	****	

Binge drinking (yes)[12]	30.3%	10.2%	22.4%	31.7%	39.8%	52.7%	****	<0.0001
Cum. alc. intake[13]	32 (0/84)	0 (0/0)	4.6 (1/9)	19 (12/28)	53 (34/79)	143 (95-209)	****	<0.0001
Pre-preg. alc. intake [14]	3.0	0	1.4	2.8	4.6	8.3	****	<0.0001
Alc. intake 1st intake (yes)[15]	44.8%	0%	12.2%	50.8%	80.4%	93.9%	****	<0.0001
Alc. intake 2nd intake (yes)[15]	49.5%	0%	8.5%	60.6%	92.2%	98.0%	****	<0.0001
Alc. intake 3rd intake (yes)[15]	46.9%	0%	8.7%	54.5%	87.7%	97.0%	****	<0.0001
Alc. intake pre-preg. (yes)[15]	87.2%	0%	97.1%	98.2%	99.6%	99.9%	****	<0.0001
Cumulated smoking (cig.)[16]	63	100	62	54	50	78	****	<0.0001
Smoking (yes)[17]	25.2%	31.5%	24.1%	23.3%	23.5%	30.9%	****	<0.0001
nicotine substitutes[18]	2.2%	2.6%	1.8%	2.2%	2.3%	3.0%	****	<0.0001
Partner smoking (yes)[19]	29.6%	37.0%	29.6%	27.9%	27.1%	31.2%	****	0.05, NS
Coffee (yes)[20]	41.9%	33.4%	34.1%	41.0%	49.0%	65.0%	****	<0.0001
Vitamin (yes)[21]	84.1%	81.8%	85.2%	84.8%	84.1%	81.5%	****	<0.0001
Iron (yes)[21]	71.2%	69.4%	71.8%	72.0%	72.1%	68.2%	****	<0.0001

Fish oil (yes)[21]	5.1%	5.2%	5.1%	5.2%	5.0%	5.4%	P=0.69, NS	0.36, NS
Fish eating (never)[22]	3.4%	6.9%	4.3%	2.7%	2.1%	1.8%	****	<0.0001
BMI[23]	67.6%	56.8%	64.0%	68.4%	72.8%	76.0%	****	<0.0001
Cola[24]	16.5%	24.0%	17.2%	15.0%	14.1%	15.4%	****	<0.0001
TV[25]	21.1%	29.6%	23.4%	20.1%	17.4%	15.7%	****	<0.0001
Exercise (yes)[26]	24.3%	17.5%	22.3%	25.2%	27.8%	27.2%	****	<0.0001
Painkillers (yes)[21]	25.0%	26.5%	23.3%	24.4%	25.8%	27.7%	****	<0.0001
Sleep medication (yes)[21]	0.7%	0.8%	0.7%	0.6%	0.7%	0.7%	P=0.51, NS	0.09, NS
Diabetes (yes)[27]	2.1%	3.0%	2.4%	2.1%	1.6%	1.6%	****	<0.0001
Asthma (yes)[27]	3.3%	4.3%	3.7%	3.0%	2.9%	2.8%	****	<0.0001
Anaemia (yes)[27]	3.7%	3.7%	3.5%	3.6%	3.7%	4.4%	****	0.00**

Table 1: background characteristics across levels of cumulated alcohol intake in pregnancy including information from the early unrecognised part of pregnancy. P-value: one way ANOVA (for means) and chi square (for percentages) tests, two-tailed probability. NS: non-significant; * Significance at <0.05; ** Significance at <0.01; *** Significance at <0.001; **** Significance at <0.0001. [1] Age at birth. [2] M = Maternal. [3] P = Paternal. [4] Married or cohabiting with the child's biological father six months post-partum. [5] Tenant, homeless or living with parents. [6] Register-based information on educational level in year 2010. [7] Self-reported psychological problems prior to pregnancy. [8] Register-based information on contact with the psychiatric system prior to pregnancy. [9] Self-reported psychological problems in pregnancy. [10] Contact with the psychiatric system in the first seven years of the child's life. [11] Average number of binge episodes in pregnancy. [12] Binge episodes in pregnancy (yes/no). [13] Maternal cumulated alcohol intake In pregnancy including information from the unrecognized part of pregnancy. [14] Reported weekly average alcohol intake prior to pregnancy. [15] reporting an average alcohol intake in pregnancy (yes/ no) in the interviews conducted approximately in weeks 16 and 30 in pregnancy and 6 months post-partum. [16] Cumulated smoking in pregnancy calculated in the same manner as the cumulated alcohol exposure variable. [17] Reported smoking in pregnancy (yes/no). [18] Use of nicotine substitution in the early part of pregnancy. [19] Partner smoking in early part of pregnancy (yes/no). [20] Maternal intake of coffee in early part of pregnancy (yes/no). [21] Maternal intake in last part of pregnancy (yes/no). [22] information from

early part of pregnancy. [23] Pre-pregnant body mass index (BMI) outside normal range. BMI normal range: 18.5-24.99. [24] Intake of ≥ 1 liter of Cola per week in early part of pregnancy. [25] Television watching ≥ 2 hours/ day in last part of pregnancy. [26] Maternal exercise in last part of pregnancy. [27] Information from last part of pregnancy.

	Cumulated alcohol intake			
Alcohol group	Full sample	All time alcohol abstainers	Pregnancy abstainers	P
N	63,464	7,204	16,563	
Socio-demographic factors				
Age (M)[1,2]	30.5	29.5	29.6	P=0.11, NS
Age (P)[1,3]	32.3	31.5	31.4	P=0.71, NS
Married (no)[4]	2.3%	3.5%	3.0%	*
Tenant[5]	26.7%	30.6%	30.0%	P=0.35, NS
Education				
Mandatory (M)[2,6]	8.4%	17.8%	9.4%	****
University (M)[3,6]	14.6%	5.8%	12.6%	****
Mandatory (P)[2,6]	13.3%	22.9%	14.8%	****
University (P)[3,6]	15.3%	7.1%	13.5%	****
Psychological problems/ Psychiatric diagnoses				
self-rep. pre-preg. prob. (M)[7]	7.5%	9.4%	7.1%	****
Pre-preg. (M)[2,8]	2.9%	4.8%	3.1%	****
Pre-preg. (P)[3,8]	1.9%	2.8%	1.9%	****

self-rep. psych. prob. in preg. (M)[9]	1.2%	1.6%	1.1%	**
After birth (M)[2,10]	3.6%	5.4%	4.0%	****
After birth (P)[3,10]	2.4%	3.8%	2.5%	****
Lifestyle factors				
Binge drinking[11]	0.5	0.1	0.4	****
Binge drinking (yes)[12]	30.3%	10.2%	26.0%	****
Cum. alc. Intake[13]	32 (0/84)	0 (0/0)	6.7 (1/13)	****
Pre-preg. alc. Intake[14]	3.0	0	2.3	****
Cumulated smoking[16]	62	99	66	****
Smoking (yes)[17]	25.2%	31.5%	25.9%	****
nicotine substitutes[18]	2.2%	2.6%	2.0%	**
Partner smoking[19]	29.6%	37.0%	30.2%	****
Coffee (yes)[20]	41.9%	33.4%	33.5%	P=0.98, NS
Vitamin (yes)[21]	84.1%	81.8%	85.1%	****
Iron (yes)[21]	71.2%	69.4%	72.3%	****
Fish oil (yes)[21]	5.1%	5.2%	5.3%	P=0.74, NS
Fish eating (never)[22]	3.4%	6.9%	4.4%	****
BMI[23]	67.6%	56.8%	64.6%	****
Cola[24]	16.5%	24.0%	17.8%	****
TV[25]	21.1%	29.6%	23.9%	****
Exercise (yes)[26]	24.3%	17.5%	23.0%	****
Painkillers[21]	25.0%	26.5%	21.8%	****
Sleep medication (yes)[21]	0.7%	0.8%	0.6%	P=0.1277, NS

Diabetes[27]	2.1%	3.0%	2.3%	***
Asthma[27]	3.3%	4.3%	3.7%	*
Anaemia[27]	3.7%	3.7%	3.4%	P=0.17, NS

Table 2: Background characteristics of the subdivided abstaining group. Divided into an all-time abstaining group and a pregnancy-only abstaining group. P: ANOVA (means) and Chi square (percentages) significance tests, two-tailed probability. NS: non-significant; * Significant at <0.05; ** Significant at <0.01; Significant at <0.001; Significant at <0.0001.[1] Age at birth. [2] M = Maternal. [3] P = Paternal. [4] Married or cohabiting with the child's biological father six months post-partum. [5] Tenant, homeless or living with parents. [6] Register-based information on educational level in year 2010. [7] Self-reported psychological problems prior to pregnancy. [8] Register-based information on contact with the psychiatric system prior to pregnancy. [9] Self-reported psychological problems in pregnancy. [10] Contact with the psychiatric system in the first seven years of the child's life. [11] Average number of binge episodes in pregnancy. [12] Binge episodes in pregnancy (yes). [13] Maternal cumulated alcohol intake In pregnancy including information from the unrecognized part of pregnancy. [14] Reported weekly average alcohol intake prior to pregnancy. [15] Reporting on an average alcohol intake in pregnancy (yes). [16] Cumulated smoking in pregnancy calculated in the same manner as the cumulated alcohol exposure variable. [17] Reported smoking in pregnancy (yes). [18] Use of nicotine substitution in the early part of pregnancy (yes). [19] Partner smoking in early part of pregnancy (yes). [20] Maternal intake of coffee in early part of pregnancy (yes). [21] Maternal intake in last part of pregnancy (yes). [22] Information from early part of pregnancy. [23] Pre-pregnant body mass index (BMI) within normal range. BMI normal range: 18.5-24.99. [24] Intake of ≥ 1 liter of Cola per week in early part of pregnancy (yes). [25] Television watching ≥ 2 hours/ day in last part of pregnancy (yes). [26] Maternal exercise in last part of pregnancy (yes). [27] Information from last part of pregnancy.

	Cumulated alcohol intake			
Alcohol group	Full sample	>90-180	>180	P
N	63,464	4,605	871	
Socio-demographic factors				
Age (M)[1,2]	30.5	32.3	33.4	P=0.10, NS
Age (P)[1,3]	32.3	34.3	36.0	****
Married (no)[4]	2.3%	2.1%	4.3%	****
Tenant[5]	26.7%	22.9%	22.4%	P=0.77, NS
Education				
Mandatory (M)[2,6]	8.4%	6.0%	9.1%	***
University (M)[3,6]	14.6%	21.7%	18.2%	*
Mandatory (P)[2,6]	13.3%	10.1%	13.2%	*
University (P)[3,6]	15.3%	22.2%	18.7%	*
Psychological problems/ Psychiatric diagnoses				
self-rep. pre-preg. Prob. (M)[7]	7.5%	7.7%	13.1%	****
Pre-preg. (M)[2,8]	2.9%	2.6%	5.2%	****
Pre-preg. (P)[3,8]	1.9%	1.6%	3.2%	**
self-rep. psych. prob. in preg. (M)[9]	1.2%	1.5%	2.2%	P=0.14, NS
After birth (M)[2,10]	3.6%	3.5%	4.7%	P=0.09, NS
After birth (P)[3,10]	2.4%	2.0%	3.7%	**
Lifestyle factors				
Binge drinking[11]	0.5	1.2	2.3	****
Binge drinking (yes)[12]	30.3%	50.7%	63.4%	****

Cum. alc. intake[13]	32 (0/84)	121 (94/157)	263 (187/355)	****
Pre-preg. alc. intake[14]	3.0	7.7	11.4	****
Alc. intake 1st intake (yes)[15]	55.2%	6.0%	6.3%	P=0.75, NS
Alc. intake 2nd intake (yes)[15]	50.5%	1.6%	4.1%	****
Alc. intake 3rd intake (yes)[15]	53.1%	2.7%	4.4%	**
Alc. intake pre-preg. (yes)[15]	87.2%	99.9%	97.8%	P=0.38, NS
Cumulated smoking[16]	62	66	138	****
Smoking (yes)[17]	25.2%	28.6%	43.3%	****
nicotine substitutes[18]	2.2%	2.9%	3.8%	P=0.14, NS
Partner smoking[19]	29.6%	29.8%	42.5%	****
Coffee (yes)[20]	41.9%	64.0%	70.6%	**
Vitamin (yes)[21]	84.1%	77.1%	84.1%	***
Iron (yes)[21]	71.2%	68.8%	65.4%	P=0.05, NS
Fish oil (yes)[21]	5.1%	5.3%	6.0%	P=0.46, NS
Fish eating (never)[22]	3.4%	1.8%	1.7%	P=0.87, NS
BMI[23]	67.6%	76.1%	75.2%	P=0.95, NS
Cola[24]	16.5%	14.9%	17.7%	*
TV[25]	21.1%	15.1%	19.1%	**
Exercise (yes)[26]	24.3%	28.1%	22.7%	**
Painkillers (yes)[21]	25.0%	27.4%	29.4%	P=0.22, NS
Sleep medication (yes)[21]	0.7%	0.7%	1.2%	P=0.12, NS
Diabetes (yes)[27]	2.1%	1.5%	2.0%	P=0.32, NS
Asthma (yes)[27]	3.3%	2.7%	3.3%	P=0.32, NS
Anaemia (yes)[27]	3.7%	4.1%	6.1%	**

Table 3: Background characteristics of the subdivided high exposure group. Divided into a >90-180 cumulated alcohol intake group and a >180 cumulated alcohol intake group. P: ANOVA (means) and Chi square (percentages) significance tests, two-tailed probability. NS: non-significant; * Significant at <0.05; ** Significant at <0.01; Significant at <0.001; Significant at <0.0001. [1] Age at birth. [2] M = Maternal. [3] P = Paternal. [4] Married or cohabiting with the child's biological father six months post-partum. [5] Tenant, homeless or living with parents. [6] Register-based information on educational level in year 2010. [7] Self-reported psychological problems prior to pregnancy. [8] Register-based information on contact with the psychiatric system prior to pregnancy. [9] Self-reported psychological problems in pregnancy. [10] Contact with the psychiatric system in the first seven years of the child's life. [11] Average number of binge episodes in pregnancy. [12] Binge episodes in pregnancy (yes). [13] Maternal cumulated alcohol intake In pregnancy including information from the unrecognized part of pregnancy. [14] Reported weekly average alcohol intake prior to pregnancy. [15] Reporting an average alcohol intake in pregnancy (yes/ no) in the interviews conducted approximately in weeks 16 and 30 in pregnancy and 6 months post-partum. [16] Cumulated smoking in pregnancy calculated in the same manner as the cumulated alcohol exposure variable. [17] Smoking in pregnancy (yes). [18] Use of nicotine substitution in the early part of pregnancy (yes). [19] Partner smoking in early part of pregnancy (yes). [20] Maternal intake of coffee in early part of pregnancy (yes). [21] Maternal intake in last part of pregnancy (yes). [22] Maternal intake in early part of pregnancy. [23] Pre-pregnancy body mass index (BMI) within normal range. BMI normal range: 18.5-24.99. [24] Intake of ≥ 1 liter of Cola per week in early part of pregnancy (yes). [25] Television watching ≥ 2 hours/ day in last part of pregnancy (yes). [26] Maternal exercise in last part of pregnancy (yes). [27] Information from last part of pregnancy.

Principal component	1	2	3	4	5	6
Initial Eigenvalue	4.4	2.8	1.9	1.7	1.4	1.3
Initial explained variance	11.5	7.4	4.9	4.4	3.7	3.4
Extracted Factors	**Average alcohol consumption**	**Stimulants**	**Parental age**	**Maternal mental health**	**Binge drinking**	**Educational-related lifestyle**
Cumulated alc. Intake	0.88 (0.88)					
Alc. intake 2nd int.	0.84 (0.83)					
Alc. intake 3rd int.	0.82 (0.82)					
Alc. intake 1st int.	0.78 (0.76)					
Pre.-preg alc. Intake	0.72 (0.77)					
Alc. intake pre-preg.	0.71 (0.77)					
Cumulated smoking (cig.)		0.83 (0.88)				
Smoking (yes)		0.82 (0.87)				
Coffee (yes)		0.53 0.53)				
Partner smoking (yes)		0.45 0.50)				
Nicotine		-0.41 (-0.37)				
Age (M)			0.85 (0.89)			
Age (P)			0.82 (0.88)			
Tenant			0.37 (0.40)			

Self-rep. pre-preg. prob.				0.75 (0.79)		
Pre-preg. (M)				0.63 (0.65)		
Self-rep. psych. prob. in preg.				0.63 (0.66)		
After birth (M)				0.45 (0.46)		
Binge drinking (yes)					0.88 (0.92)	
Binge drinking					0.86 (0.92)	
Education (M)						0.59 (0.74)
Education (P)						0.58 (0.73)
BMI						-0.49 (-0.40)
Exercise						0.35 (0.38)
TV						-0.31 (-0.45)

Table 4: Principal component analysis (PCA) with Varimax rotation for the full sample (N=63,464). Data from the 1-factor Principal Component Analyses are shown in brackets. Factor loading between +/- 0.30 has been omitted.