DEPARTMENT OF PUBLIC HEALTH

Jingyuan Xiao

How I Wrote My Prospectus

I began my prospectus with a review of existing literature to establish a good understanding of the topic. I then outlined the entire document to set clear directions for each section with guiding topic sentences for each paragraph. After setting up the structure, I focused on writing the critical need and significance, which took the most time for me in terms of writing. Following that, I tackled the specifics by identifying necessary steps for the analysis and performed some preliminary analysis on data I had access to beforehand. The final stage involved refining the document for better clarity and logical flow, making sure that all information was relevant and well-connected.

Advice for Prospectus Writers

1. When writing a prospectus, it's important to highlight the critical need and the 'why' behind your proposal. This clarifies the importance of your work and sets a strong foundation for smoothly transitioning to your proposed aims and methodology.

2. Don't worry too much about having a "perfect" draft right from the beginning! Break down your prospectus into manageable sections and tackle one section at a time. If you hit a roadblock, move on to the next sentence, paragraph, or section, and revisit it later.

Multigenerational cohort studies of grandmaternal factors and neurodevelopment in grandchildren

Jingyuan Xiao

Dissertation prospectus for the degree of
Doctoral in Philosophy in Environmental Health Sciences,
Yale School of Public Health

Committee
[List of committee members]

Yale SCHOOL OF PUBLIC HEALTH

Specific Aims

Worldwide, neurodevelopmental disorders (NDDs) are rising in incidence, with evidence supporting environmental risk factors contributing to the increase. Prior etiological research of NDDs has focused on the roles of pre- and early postnatal exposures. However, emerging evidence suggests that the effects of grandmaternal factors affecting early life development of the parent generation can subsequently influence NDD risk in the grandchildren, possibly through inheritable epigenetic alterations of the germ cells and/or disparities in important social determinants of health. Current multigenerational NDD risk research has been limited by the lack of prospectively collected data that spans over several generations, and few grandmaternal exposures have been studied (e.g., diethylstilbestrol and smoking). The objective of this dissertation is to build on my previous work – birth characteristics of the parents (i.e., grandmaternal delivery characteristics) and autism in their offspring – and assess the multigenerational effects of two uninvestigated grandmaternal factors: educational attainment and breastfeeding. Finally, from my previous research, I have discovered a specific type of selection bias that may affect multigenerational analyses, which I propose to formally evaluate in my thesis.

Low grandmaternal educational attainment is of particular interest given its well-documented correlations with environmental exposures and experiences during the early life of the immediate offspring and possible impacts on disease risks across the lifespan. Despite the long tradition of education research, the understanding of whether educational attainment may influence the third generation is limited, with the only evidence showing associations with grandchildren's preterm birth and lower birthweight (risk factors for NDDs). Another factor that merits investigation is grandmaternal breastfeeding, a possible indicator of socioeconomic inequality, which may influence grandchild health through epigenetic modulators. For instance, nutrition, enhanced infant attachment, and gut microbiota have been recently suggested to play an important function in epigenome remodeling, raising the possibility of multigenerational transmission of their effects. Further, although multigenerational analyses require data from three generations, the consequences of selection bias that would arise from excluding those who have not reproduced during the study period due to the exposure under study are not well understood.

Therefore, I propose to leverage two highly unique and high-quality data sources, secured through Dr. Liew's laboratory and address these important gaps in research. The first is country-wide population data from Denmark (1978-2017), a prospective register-based cohort with extensive follow-up. The second is survey data from the Nurses' Health Study II (1989-2009), one of the largest and longest cohorts in the U.S. that uniquely holds detailed information on grandmothers' breastfeeding. Analyses will be guided by directed acyclic graphs to adjust for important covariates and assess possible mediating pathways. These studies will greatly expand the current multigenerational literature in neurodevelopment and inform future intervention strategies that may have impacts lasting over generations. To achieve this, my dissertation will carry out three specific aims as follows:

Aim 1: To evaluate the associations of grandmaternal educational attainment with grandchildren's neurodevelopment. I will evaluate the associations of educational attainment (before delivery and highest achieved) among the grandmothers, with the diagnoses of three common NDDs – autism spectrum disorder, attention deficit hyperactivity disorder, and intellectual disability – in ~500,000 grandchildren born in Denmark (1994-2013). The joint effects of grandmaternal and parental education will be assessed; NDD risks in offspring of discordant siblings of parents associated with grandmaternal education before delivery will be compared.

Aim 2: To evaluate the associations of grandmaternal breastfeeding with grandchildren's neurodevelopment. Many factors associated with breastfeeding have been recently suggested to have epigenetic potentials (e.g., nutrition, enhanced attachment, and gut microbiota). However, no studies have investigated the relation between grandmaternal breastfeeding and grandchildren's NDD risks. This aim will particularly examine whether breastfeeding in the grandmothers is associated with risk for autism spectrum disorder and attention deficit hyperactivity disorder in their grandchildren, using data of ~47,000 nurses (and their mothers) collected in the Nurses' Health Study II.

Aim 3: To quantitatively assess and adjust for selection bias in multigenerational cohort studies. I will conduct a case study to investigate the influence of selection bias in multigenerational studies by examining whether such bias plays a role in the association between parental preterm birth and offspring ASD risk that was found in my previous three-generational linkage study from Denmark. Specifically, I will investigate whether individuals born preterm are less likely to have a child therefore not selected into the multigenerational cohort. Results from Aims 1 & 2 will also be re-evaluated under this aim.

Significance

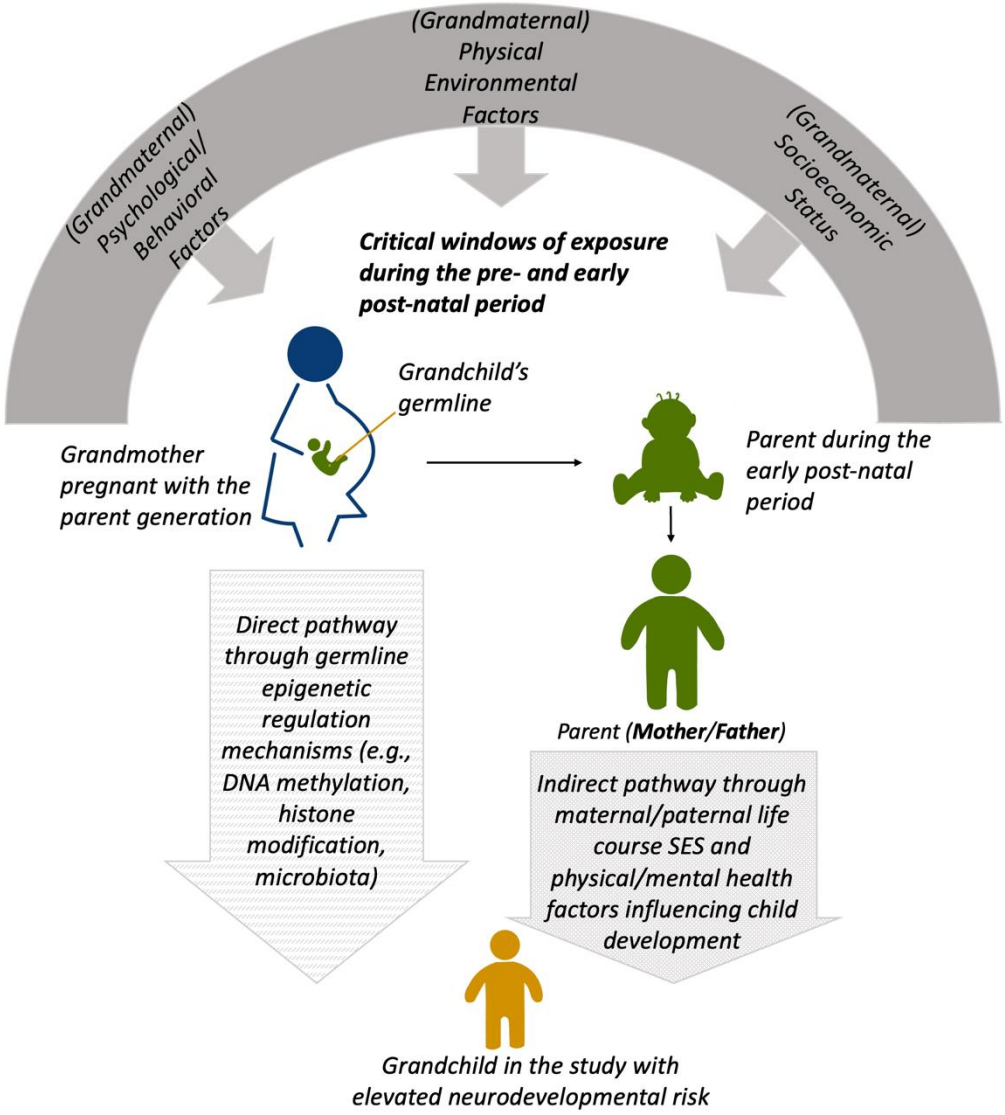
Neurodevelopmental disorders (NDDs) are a class of neurologic conditions characterized by affected brain development and a delay or disturbance in the attainment of cognitive, social, language, and psychomotor developmental milestones.[1] **NDDs are considered to be a leading cause of morbidity in children (affecting >15% children worldwide),**[2] **and because the impairments will oftentimes persist into adulthood, these disorders bring enormous psychological and economic burden to the affected individuals, their families, and society.**[3,4] Some of the most commonly diagnosed neurodevelopmental disorders are autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), and intellectual disability,[5,6] which will be the focus of the proposed research.

Specifically, ASD is characterized by impaired social communication and the presence of restricted and repetitive behaviors and interests. The prevalence of ASD has increased dramatically in recent decades: in the US, the prevalence increased by 240.9% between 2000 (0.7%) and 2018 (2.3%).[7] ADHD is another neurodevelopmental condition, which is marked by difficulties in sustaining attention, hyperactivity, and impulsivity, and associated functional impairment. As with ASD, ADHD has also seen a substantial increased prevalence, which has increased by 20.5% between 2003 (7.8%) and 2016 (9.4%),[8,9] and these increases cannot be fully explained by changes in diagnostic criteria and parental awareness.[10] Finally, intellectual disability is characterized by significant impairments in cognitive, language, motor, and social abilities that "contribute to the overall level of intelligence".[11] The prevalence of intellectual disability, 2%, is lower than for ASD and ADHD. All three of these disorders have an early age of onset in childhood, are more frequent in boys, and have complex etiologies that remain obscure.[12]

Although genetic factors play a role in many NDDs including ASD, ADHD, and intellectual disability,[12] **a number of pre- and early postnatal factors have been suggested to contribute to the complex etiology of these neurodevelopmental disorders**. This is supported by the David Barker's developmental origins of health and disease (DOHAD) hypothesis that recognizes that the environment in which early life develops may have a long-term impact on disease risks later ln life.[13] For instance, according to previous literature reviews, exposures to a range of environmental toxins such as metals,[14] endocrine disrupting chemicals,[15] and air pollutants[16] during the prenatal period have been linked to many neurodevelopmental conditions including ASD, ADHD, and intellectual disability. The effects of psychological and behavioral determinants of health surrounding pregnancy, such as stress,[17] inadequate nutrition,[18] obesity,[19] as well as smoking and alcohol drinking also affect neurodevelopment while inutero.[20] **Among many factors that may influence neurodevelopment, the importance of maternal socioeconomic status has been particularly highlighted because it may affect children's exposure to many of the aforementioned (physical) environmental and psychological/ behavioral risk factors for neurodevelopment.**[21]

While earlier studies have primarily focused on neurodevelopmental risks among those directly exposed, a recent extension of the DOHAD hypothesis has attracted attention.[22] This hypothesis suggests that **the effects of certain pre- and early postnatal factors may be transmitted across generations and affect future offspring.**[23,24] One plausible mechanism underlying such transmission is **disparities in important socioenvironmental determinants of health, which may track across generations and result in elevated disease risk.**[23,24] Recent evidence has suggested another possible mechanism, which is through gene-environment interaction modulated via **epigenetic modifications** that leads to **heritable changes in germline gene expression (Figure 1).**[22,22] For example, prenatal exposure to environmental toxins (e.g., bisphenol A, smoke component benzopyrene) and maternal stress have been shown to impose multigenerational effects on brain mRNA and DNA methylation patterns, and increase mutation frequency in the germ cell.[25–27] A few epidemiological studies have examined the possibility of multigenerational effects in human populations and linked several grandmaternal factors to increased neurodevelopmental risks in their grandchildren. The Nurses' Health Study II (NHSII) in the U.S. has associated elevated risks of attention-deficit/hyperactivity Disorder (ADHD) in grandchildren with grandmaternal exposure to diethylstilbestrol during pregnancy[28] ,periconceptional grandmaternal weight,[29] and smoking.[30] The Avon Longitudinal Study of Parents and Children has found associations between grandmaternal smoking during pregnancy with autistic traits among grandchildren in south-west England.[31] Two registry-based linkage studies from Sweden[32] and Denmark[33] have reported that advanced grandmaternal age may influence ASD risk in their grandchildren.

Figure 1. A schematic of possible multigenerational effects of germline and early developmental exposures on offspring neurodevelopment



Within this multigenerational context, my master's thesis based on population data from Denmark, provided preliminary evidence that adverse grandmaternal delivery characteristics, i.e., birth characteristics of the parents, specifically preterm birth (<37 weeks of completed gestational weeks) in the parent generation, are associated with increased ASD risk in the third-generation children.[34] **However, owing to the limited availability of multigenerational data, many other important factors that have been suggested to influence the pre- and perinatal experiences of the parent generation, including important socioeconomic status indicators of grandmothers, have not been explored yet,** warranting the proposed investigation. Multigenerational effects transmitted via the paternal line have also been suggested; however, due to potentially different mechanisms underlying such transmission of effects, the current study will focus on factors related to the grandmothers.

Project 1:

One factor that has not yet been explored is **educational attainment**, **which has been suggested to be the strongest and the most consistent predictor for health among the three most important indicators of socioeconomic status** (i.e., education, income, and occupation).[35] Compared to income and occupation, education may contribute more greatly to shaping critical personal (e.g., health literacy) and psychosocial (e.g., social support and self-efficacy) resources, and determining the utilization of preventative and therapeutic healthcare services, the choice of more healthful environments in which to work and live, and the adoption and continuation of healthy behaviors.[36] Educational attainment among the grandmothers may also act as a proxy for the effects of the overall socioenvironmental and environmental conditions that influence the pre- and early postnatal experiences of the parent generation and neurodevelopment in the child generation through different mechanisms.[37]

I postulate that one of these mechanisms is that grandmaternal educational attainment may **directly affect the health of the grandchild via heritable epigenetic modifications, independent of later health outcomes of the parents,** given epigenetic reprogramming potentials of several aforementioned exposures.[38] While assessing the contribution of this pathway is beyond the scope of my dissertation, I will examine the second plausible mechanism which is **through the health and well-being of the parent generation**. For example, one factor most strongly associated with educational attainment is tobacco smoking, which is also an important risk factor for adverse health outcomes in the immediate offspring.[39] In addition, educational attainment is related to (grand)parenting skills, rearing environment,[40,41] psychological stress, as well as access to medical care and economic security, which are all important determinants of healthy development of the parents and are potentially risk factors for neurodevelopmental disorders.[21] Furthermore, (grand)maternal education has been associated with exposures to environmental toxins and stressors in the parent generation, such as heavy metals,[42] air pollutants,[43] and endocrine disrupting chemicals that have been suggested to influence neurodevelopment.[44,45] Thus, understanding the relationship between grandmaternal educational attainment and grandchildren's development is crucial for advancing knowledge on the potential longer-term consequences of low levels of education. Nevertheless, studies on whether educational attainment may influence the third generation remain limited. Currently, three studies have shown that higher risk for lower birthweight and preterm birth in grandchildren whose grandmothers had lower attained education,[46–48] yet **no studies to date have specifically focused on neurodevelopment.** Additionally, the understanding of the total contribution of grandmaternal education to offspring neurodevelopment is important because it guides future work to investigate the unique effect of specific socioenvironmental and environmental mediators.

Policy Implications: Besides its scientific value, this research will provide direct evidence to further support policies that promote higher education particularly in women, which are likely to be important not only for themselves but also for their potential future grandchildren.

Project 2:

Another factor requiring investigation is **grandmaternal breastfeeding** during the parent's early life. Breastfeeding, as **a modifiable health behavior and a possible indicator of socioeconomic inequality**,[49] has long been of great interest to individuals, clinicians, policymakers, as well as researchers.[50] There is ample epidemiologic evidence for **the potential benefits of breastfeeding to the mother** [51,52] **and the immediate offspring.**[51,53] In detail, longer duration of breastfeeding has been associated with lower rates of breast and ovarian cancer, metabolic syndrome, obesity and diabetes, hypertension, and adverse cardiovascular outcomes[52,54,55] Breastfeeding has also been associated with better child health outcomes in the short and long term, including decreased mortality and reduced risk for childhood obesity, asthma and other allergic diseases, as well as neurodevelopmental disorders.[51,53,56,57] Given its well-documented benefits, The World Health Organization (WHO) has recommended exclusive breastfeeding for 6 months and continued breastfeeding for at least 2 years, along with nutritionally adequate and safe complementary foods.[58] Although there have been recent concerns over possible accumulation of persistent organic pollutants, pesticides, heavy metals, and other contaminants in human breast milk, breastfeeding continues to be regarded as a health priority for women and children because certain components (e.g., whey protein) in human milk are found to buffer children from the potential effects of environmental toxicants and the net benefits of human milk can override the potential adverse effects.[59] **Unfortunately, overall rates for breastfeeding initiation and continuation have remained far below the WHO target across the world,** including in the U.S. (only 58.3% of infants were breastfeeding at 6 months).[60] This may be due to various social (e.g., lack of support), economic (e.g., lower economic and political status[61]), medical and health-care related factors (e.g., antenatal visits), and personal constraints (e.g., the belief that breast milk cannot provide sufficient nutrition to infants).[58,62–64]

Recent literature has suggested **potential epigenetic effects** of breastfeeding based on observed associations between breastfeeding and differential DNA methylation patterns in their children.[65,66] Many benefits of breastfeeding, such as nutrition,[67] increased mother-child interaction and enhanced infant attachment,[68] hormones, and microbiota,[69] could also have **an important function in epigenome remodeling**[70–74] and thus underlie a potential multigeneration transmission of breastfeeding effects. For instance, infant gut microbiota, heavily influenced by breastfeeding,[75] has been shown to modify the germ cell epigenome and regulates metabolism, immune responses, and behavioral outcomes across generations.[70-72] In addition, the effects of grandmaternal breastfeeding could influence their grandchildren **through the parental generation**. Breastfeeding has been shown to protect against infectious diseases, reduces the risk of chronic diseases (e.g., obesity and high blood pressure) and mental illnesses, and increase the likelihood of breastfeeding in the immediate offspring, which have all been associated with offspring neurodevelopment.[76]

No studies, however, have assessed the multigenerational association between grandmaternal breastfeeding during the parents' infancy and the neurodevelopment of the third-generation offspring, warranting the proposed investigation.

Policy Implications: In addition to addressing a key knowledge gap in multigenerational research, the present study may help form another compelling rationale for 1) increased advocacy and action to increase rates of breastfeeding worldwide to optimize the health of current and future generations; and 2) concomitant public investment to remove roadblocks to breastfeeding with timely policies (e.g., paid maternity leave and breaks during the workday for breastfeeding or breast milk extraction[77]) and interventions that support women who are inclined to breastfeed.

Project 3:

Finally, **I propose to investigate a unique methodological issue in multigenerational research that may affect all cohort studies of multiple generations**. As described above, there is growing epidemiological interest in understanding whether the effects of pregnancy exposures and experiences among grandmothers, namely the prenatal exposure of the parent generation, could persist to influence the child generation.[38] **Current research of multiple generations, however, faces various methodological challenges that may undermine study validity.**[78] Several methodological issues have been previously described and studied, including 1) exposure misclassification in the grandmaternal generation because exposure ascertainment requires long-term recall or sometimes is reported by the parent generation,[29,78–83] 2) inadequate adjustment for potential confounding variables[33,34,84,85] and 3) inappropriate adjustment of potential mediators (e.g., parental income, smoking status, or pre-pregnancy weight)[80,84] that could result in collider bias.[86] **However, insufficient attention has been given to sources of selection bias in a multigenerational context.**

Selection bias, which can distort the association between exposure and health outcome in the target population due to differences in individuals who completed the study, is one of the primary methodological challenges faced by all epidemiological studies, and those with multiple generations are no exception.[87] The term selection bias entails various issues surrounding study participation: volunteer bias, healthy worker bias, non-response bias, and in case-control studies the incorrect selection of controls.[88] Notably, there is one potential source of bias that is unique to studies on the effects of grandmaternal factors, which is informatively empty clusters in the following generations, discussed in two recent simulation studies.[78,89] In other words, while multigenerational analyses require information from three generations, selection bias could occur if the parent generation not have a child within the study period due to adverse pre-and postnatal exposures and therefore be excluded from the multigenerational cohort.[82] **One potential factor that may lead to this specific form of selection bias is preterm birth status in the parent generation.** Preterm birth in the parents was associated with elevated child ASD risk in my previous work;[34] **however, it may be of concern to all multigenerational research alike if it also leads to an empty child-generation cluster.**[78] This potential issue has not been formally assessed, partially due to the lack of prospectively collected reproductive data that spans generations, necessitating the proposed project. Upon completion, **the results from this investigation can be broadly applied to multigenerational research designs and inform strategies to evaluate and mitigate the influence of such bias in future multigenerational cohort studies.**

Data sources for the proposed research:

Through prior research and collaborations, **I have secured access to two of the largest prospective data sources, namely the Danish Medical Birth Registry (DMBR) and the Nurses' Health Study (NHS) II, which will provide the basis for the proposed dissertation.** Specifically, with whole population data from Denmark (1978-2017), I will study the association of grandmaternal educational attainment with risk for ASD, ADHD, and intellectual disability in their grandchildren. Using the Danish data, I will also explore one possible source of selection bias unique to studies of multiple generations, which may arise from excluding the second-generation individuals who were born preterm. In addition, the NHS II, one of the largest and longest cohorts in the United States that uniquely holds information on grandmaternal self-reported breastfeeding, will allow me to investigate the potential multigenerational effects of breastfeeding on neurodevelopment. Based on directed acyclic graphs (under "*Approach*"), important demographic covariates will be identified and adjusted, and the possible mediating role of second-generation health and third-generation perinatal conditions will be evaluated for the assessment of grandmaternal educational attainment and breastfeeding practices. The specific variable(s) included in each individual aim will be slightly different depending on the information available in the corresponding database (more details can be found under "*Approach*").

In summary, while the possibility of multigenerational transmission of disease risk has received increasing research interest, hampered by inadequate data spanning over multiple generations, epidemiological evidence in this area remains limited. The projects outlined in my prospectus will examine the relationship between two important grandmaternal factors (i.e., educational attainment and breastfeeding) and their grandchildren's neurodevelopmental risk and evaluate one specific source of selection bias that could be relevant to all multigenerational cohort analyses. **The proposal is innovative in several ways: (1)** it will use large longitudinal cohorts spanning over three generations in two different countries; **(2)** it is interdisciplinary in nature as it considers important socioenvironmental exposure factors and pre- and postnatal influences on neurodevelopment in the third-generation; **(3)** it addresses three important research questions in multigenerational research of neurodevelopment that have not been investigated; and **(4)** it employs both advanced causal mediation and bias analysis that have not been applied in previous research in this field.

Upon completion, these studies will add to the limited literature concerning a recent hypothesis of multigenerational neurodevelopmental effects of early-life exposures and experiences, point to important topics for future mechanistic research, and inform future preventative public health policies and interventions to promote optimal neurodevelopment for future generations.

Approach

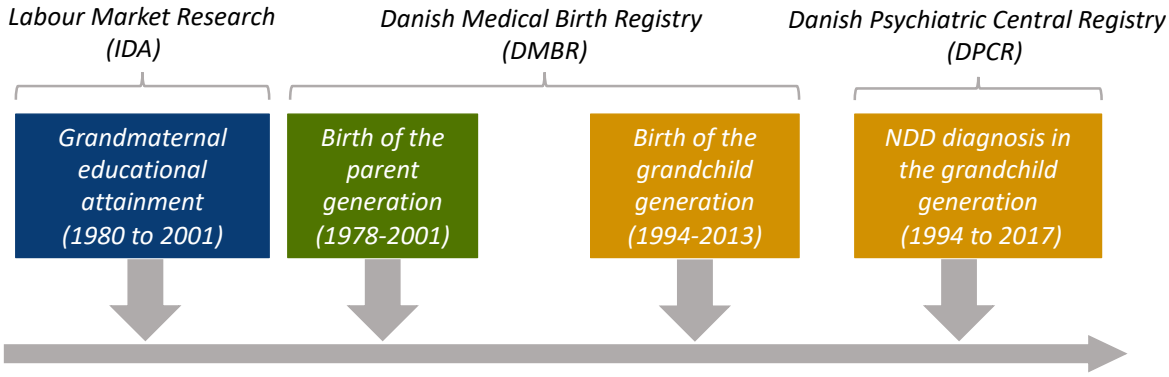
Aim 1: To evaluate the associations of grandmaternal educational attainment on grandchildren's neurodevelopment.

Rationale Education level is possibly the strongest socioeconomic predictor for health[35] and a potential composite indicator of lifestyle, socioenvironmental, and environmental risk factors that may affect disease risks across generations. However, the only evidence in the field is associations between grandmaternal educational attainment and fetal growth and gestational age in the grandchildren.[46,47,90] In order to investigate the relationship between grandmaternal educational attainment and their grandchildren's neurodevelopmental outcomes, I propose to carry out a population-based registry study with high-quality socioeconomic and health data from multiple Danish national registries. Specifically, I will investigate whether grandmaternal educational attainment is associated with elevated risk for three childhood neurodevelopmental disorders – ASD, ADHD, and intellectual disability – in their grandchildren. Joint effects of grandmaternal and parental education (maternal and paternal education considered separately) on grandchildren's neurodevelopment will also be examined (e.g., the effects of changes in educational attainment from the grandmaternal generation to the parental generation on neurodevelopment).

Study Setting and Design The proposed study will be a nationwide cohort study based on data from multiple Danish nationwide registries.[91] The source population will be identified from the Danish Medical Birth Registry (DMBR) which has registered pregnancy and birth information for nearly all Danish residents since 1968 (computerized in 1973).[92] Information on educational attainment will be obtained from Labour Market Research (IDA) in Denmark.[93,94]. All neurodevelopmental diagnostic records reported by specialists in psychiatry will be retrieved from the Danish Psychiatric Central Register (DPCR), which are established based on a thorough evaluation with final conclusion by a child and adolescent psychiatrist.[95] Covariate data will be obtained from multiple registries detailed below. A unique civil registration number (CPR) assigned to each Danish individual is used in all registries, which allows for accurate individual linkage of information across registries.[91] The CPR number of the newborn is also linked to the CPR number of the parents, which enables accurate linkage across generations.[91]

Study Population Danish individuals with information on three generations (grandmother, mother/father, and grandchild) in the DMBR will be included, specifically, women giving birth to a child between 1978-2001 will be referred to as the grandmother generation; the children (of either sex) of these grandmothers will be referred to as the parent generation (mother or father); children of the parent generation born during the period of 1994-2013 will be referred to as the grandchildren in this study. Grandmothers and mothers/fathers whose information on educational attainment is missing will be excluded (<2%, without significant biasing patterns). The final analytical cohort will be further restricted to individuals from singleton live births, considering possible differences in intrauterine experiences and development between singletons and multiplets.[96] The expected yield of grandmother-mother-child triads for analysis is 230,174, and the expected yield of grandmother-father-child triads is 157,926.

Figure 2. An illustration of the multigenerational study design for Specific Aim 1, evaluating associations of grandmaternal educational attainment on grandchildren's neurodevelopment using registry data from Denmark



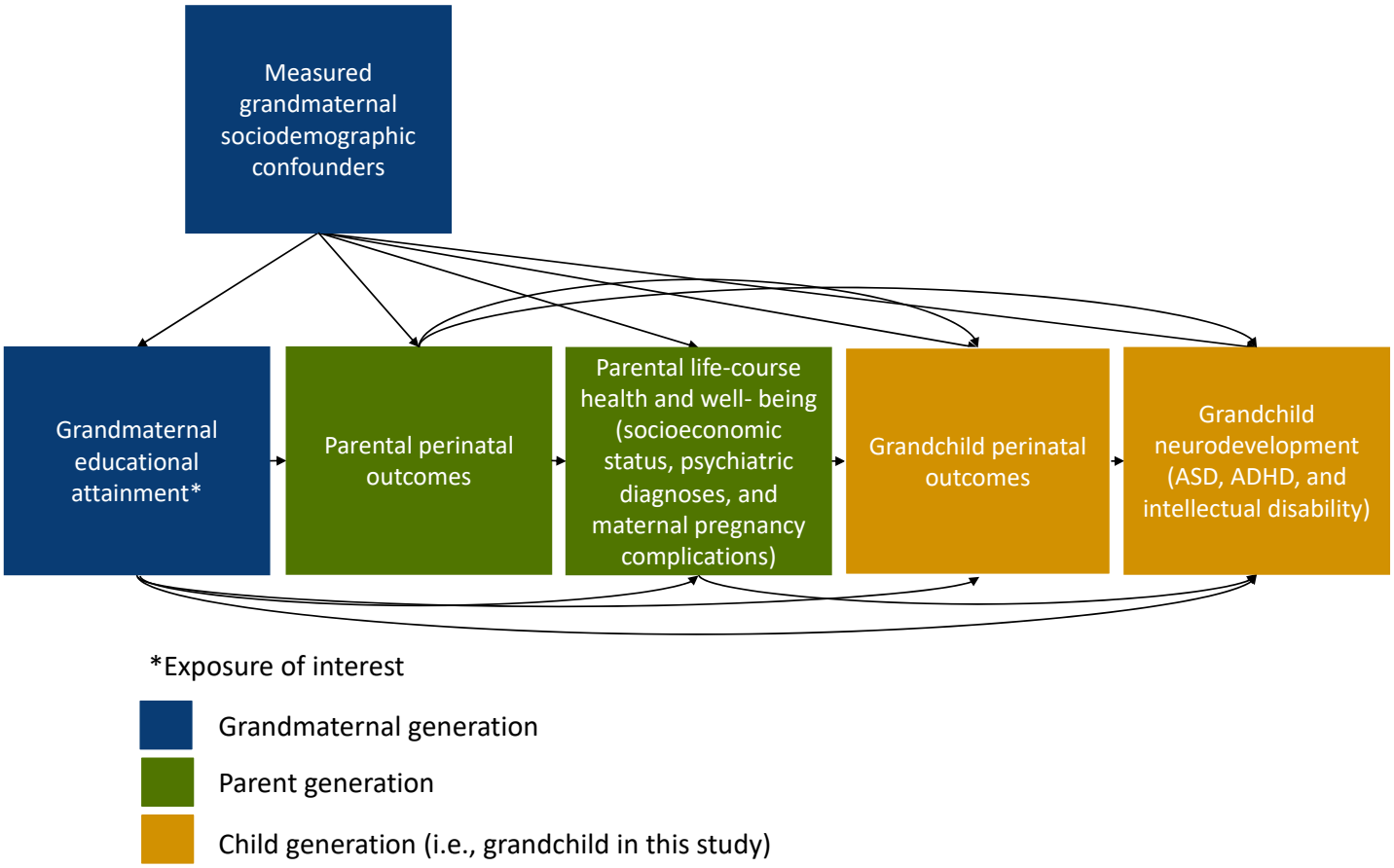
Grandmaternal Educational Attainment Information on educational attainment among the grandmothers will be obtained from the Integrated Database for Labour Market Research (IDA) in Denmark.[93,94] This database, with high validity and coverage, has registered information on education (length and type), employment status, workplace, as well as wages and income of (nearly) the entire Danish population yearly since 1980.[97,98] In this study, grandmaternal education level at the time of delivery and highest education will be classified into basic (primary and lower secondary education, ≤9 years), intermediate (upper secondary including vocational upper secondary education, 10-12 years) and high (>12 years) education, based on the framework of the International Standard Classification of Education (ISCED), 2011 version.[99]

Grandchildren's Neurodevelopmental Outcomes Grandchildren born between 1994-2017 will be assessed for neurodevelopmental diagnoses using the International Classification of Diseases (ICD) 10th Edition codes recorded in the Danish Psychiatric Central Registry (DPCR), which has collected electronic information on all psychiatric hospital admissions in Denmark since 1969 and all outpatient visits since 1995.[95] The primary neurodevelopmental outcomes are 1) pervasive developmental disorder, which has been used in Denmark to denote a broad class of autistic disorders (ICD-10 code F84) based on the ICD-10 diagnostic classification system;[100] 2) ADHD (F90.0, F90.1,F90.2, F90.8, F90.9);[5] and 3) intellectual disability (ICD-10 codes F70-F79). The estimated prevalence of these specific NDDs in Denmark are: pervasive developmental disorders ~1.4%, ADHD ~2.0%, and intellectual disability ~1% (lower than in the U.S.).[101,102]

Covariates The following potential factors will be retrieved for the entire cohort from multiple Danish national registries and treated as covariates based on Direct Acyclic Graphs (DAG, shown in **Figure 3**). The following factors will be treated as confounders: grandmaternal age at delivery (12-19, 20-24, 25-29, 30-34, 35-39, ≥ 40), grandmaternal place of residence (Capital city [Copenhagen], other major cities [Aarhus, Odense, Aalborg], and other), and grandmaternal country of origin (Denmark or others) registered in the Danish Civil Registration Service;[91], and calendar period of grandmaternal delivery (i.e., birth year of the parent generation, 1973-1977, 1978-1982, 1983-1987, 1988-2001) to control for possible cohort effects).

A range of variables that have previously been suggested to be consequences of low grandmaternal educational attainment and risk factors for neurodevelopmental disorders will be evaluated as mediators. **These mediators will be grouped into several broad categories that reflect the health and well-being of the parent and the grandchild generation**: **1) parental perinatal outcomes,** i.e., grandmaternal delivery outcomes (preterm birth and low birthweight) obtained from the Danish Medical Birth Registry, **2) parental socioeconomic status including** educational attainment (primary and lower secondary, intermediate, high) and income (five groups based on the sex-and calendar year specific income distribution (quintiles) of the entire Danish population),[103] obtained from the IDA **3) parental diagnosis of psychiatric illness** (any diagnosis of ICD-10 F00-F99), obtained from the DPCR **4) maternal pregnancy diagnoses that may influence the intrauterine environment for the third generation**, including hypertension/cardiovascular disease (any diagnosis of ICD-8 400-401, ICD-10 I10, ICD-8: 390-459, or ICD-10 I00-I99) and obesity/diabetes (any diagnosis of ICD-8 277, ICD-10 E66, ICD8 249, or ICD-10 E10-E14), ascertained based on information in the Danish National Patient Registry; and **5) perinatal characteristics among the grandchildren** (preterm and low birthweight) recorded in the DMBR.

Figure 3. The directed acyclic graph (DAG) for the study variables for Specific Aim 1



Statistical Analysis The distributions of grandmaternal education and included sociodemographic covariates have been compared according to grandchildren's ASD status in the initial phase of the project (**Table 1a. and Table 1b)**. The preliminary data suggests a higher proportion of grandchildren diagnosed with ASD were descendants of grandmothers with lower educational attainment or grandmothers with younger (<25) or older (>35) age at the time of the delivery of the parent generation. These results are limited to ASD and have not adjusted for potential confounding factors. Logistic regression models will be used to estimate odds ratios (OR) and corresponding confidence intervals (CIs) for neurodevelopmental disorders of interest, respectively (i.e., ASD, ADHD, and intellectual disability), with generalized estimating equations to account for inclusion of more than one grandchild from each grandmother (detailed below).[104] Each disorder will be analyzed as an independent outcome. Both crude and adjusted odds ratios will be estimated (the adjusted model is detailed below). As described previously, grandmaternal age at delivery, grandmaternal place of residence, grandmaternal country of origin, and calendar period of grandmaternal delivery (i.e., birth year of the parent generation, 1973-1977, 1978-1982, 1983-1987, 1988-2001) will be included as covariates in the adjusted estimates.

$$\mu_{ij} = E(y_{ij}), \qquad \mathrm{g}(\mu_{ij}) = \begin{bmatrix} 1 \\ X_{ij} \\ C_{ij} \end{bmatrix}' \beta$$

Where μ_{ij} is the marginal response of y_{ij}, whether grandchild i having ASD at time point j or not, X'_{ij} is the exposure, C_{ij} is a $p * 1$ vector of covariates and β is a $(p + 2) * 1$ vector of regression coefficients.

In addition, the **joint effects** of grandmaternal and maternal or paternal level of education will be assessed. Because I am interested in understanding the effects of intergenerational changes in educational attainment from the grandmaternal generation to the parental generation, I will additionally compare the effect estimates

(i.e., odds ratios) for neurodevelopmental diagnoses between children of mothers/fathers who have attained different and equivalent education levels as their grandmothers.

The analyses will then be stratified according to the sex of the grandchildren to examine potential sex differences, and further be restricted to first-born grandchildren to eliminate potential effects of parity on fetal development.[105] To further adjust for unmeasured confounding by shared genetic, environment, and familial risk factors, additional **sibling analysis** will be carried out to compare NDD risks in children of discordant siblings of parents for grandmaternal educational attainment achieved before delivery. Concerning the possibility that the pre-selected confounders do not suffice to control for potential mediator-outcome relationship,[106] **mediation** of the dichotomous form of aforementioned mediators will be assessed using a stabilized inverse-probability-weighted marginal structural model under the counterfactual framework described by VanderWeel and Robins et al [107,108] Specifically, I will estimate the controlled direct effect, which indicates the effect of grandmaternal education on offspring neurodevelopment, conditional on the covariates, when the pre-identified mediators is fixed to the reference. I will also estimate the natural indirect effect, which indicates the effect of the exposure through the pre-specified mediator. Each of the mediators will be evaluated separately, with adjustment of the same covariates as before by adding exposure-covariate product terms into the models. Sensitivity analyses for the causal mediation effects will be additionally performed, taking into consideration of possible presence of exposure-outcome, exposure-mediator, and mediator-outcome confounders.[109,110]

Additionally, the interrelationships, and the independent and joint effects of grandmaternal income and occupation will be assessed to investigate the multigenerational effects of other financial and social resources on offspring health that may not be captured by education alone,[35,111] given that different socioeconomic indicators (i.e., education, income, and occupation) could measure related yet different aspects of socioeconomic status and may affect health in unique ways.[36] Gross maternal income (wages, pensions, unemployment benefits, social security benefits and bank interest) during the calendar year of the birth of the parents (as a crude measure of income during F1's early life) will be divided into three levels: low (<150,000 DKK), medium (150,000–<300,000 DKK), high (>300,000 DKK) (US $1 or 1 Euro corresponds to approximately 7.5 DKK).[112] Information on occupation (unemployed, student, receiving public assistance, employed, receiving a pension (i.e., disability benefits), and self-employed (including working at spouse's company) will be classified based on Statistics Denmark classification of occupation and labour market status.[113]

Assuming two-sided alpha = 0.05, this aim will have sufficient power (>0.99) to detect an effect size of 1.2, 1.4. 1.6, or 1.8 when estimating odds ratios for all neurodevelopmental outcomes of interest therefore a detailed power calculation table is not presented here.

Complete-case analyses will be performed given the very low level (<1.7%) of missing covariate data (**Table 1a. and 1b.**) All analyses will be performed using SAS statistical software, version 9.4 (SAS Institute, Cary, NC).

Limitations The proposed study has some limitations. First, information on several factors related to the exposures in the parent generation that may impact neurodevelopment in the subsequent generation, such as prenatal smoking and maternal body mass index is not available in Denmark until more recent years to be evaluated in this study.[92] Moreover, the Danish population during the study period is predominantly of White ethnic/racial origin, which may limit the generalizability of this study to more ethnically diverse populations. Further, this study uses educational attainment in the grandmothers as a socioeconomic indicator for the totality of effects of (physical) environmental, social, and behavioral/psychological factors that may influence their offspring. If associations found, future studies are encouraged to disentangle the role of individual exposure factors on third-generation neurodevelopmental outcomes.

Strengths Despite the limitations, this proposed nationwide registry study will be the first assessment of the potential multigenerational effects of educational attainment among grandmothers on the risk of developing three common childhood neurodevelopmental disorders among their grandchildren. Both grandmaternal education at the time of delivery and the highest-completed education that may better reflect the final socioeconomic level reached will be studied. Information on educational attainment and third-generation neurodevelopmental diagnoses used for study analysis will be from nationwide data collected in routine registers in Denmark that have high quality and completeness.[92,93,103,114] The estimations of effect sizes will be

based on both hospitalization for neurodevelopmental disorders and less severe cases diagnosed and treated in outpatient care.[27] The analysis in the proposed study will be adjusted for several important sociodemographic confounders, as well as age at delivery and calendar period that addresses influences of cohort effects, potential mediating pathways underlying the potential association will be evaluated, and the effects of two other important socioeconomic indicators (income and occupation) will be assessed. Upon completion, this study will expand the small body of current literature in the field of multigenerational research on neurodevelopment, and the study results will highlight the importance of women's educational attainment, which should be further uplifted through national and regional policies.

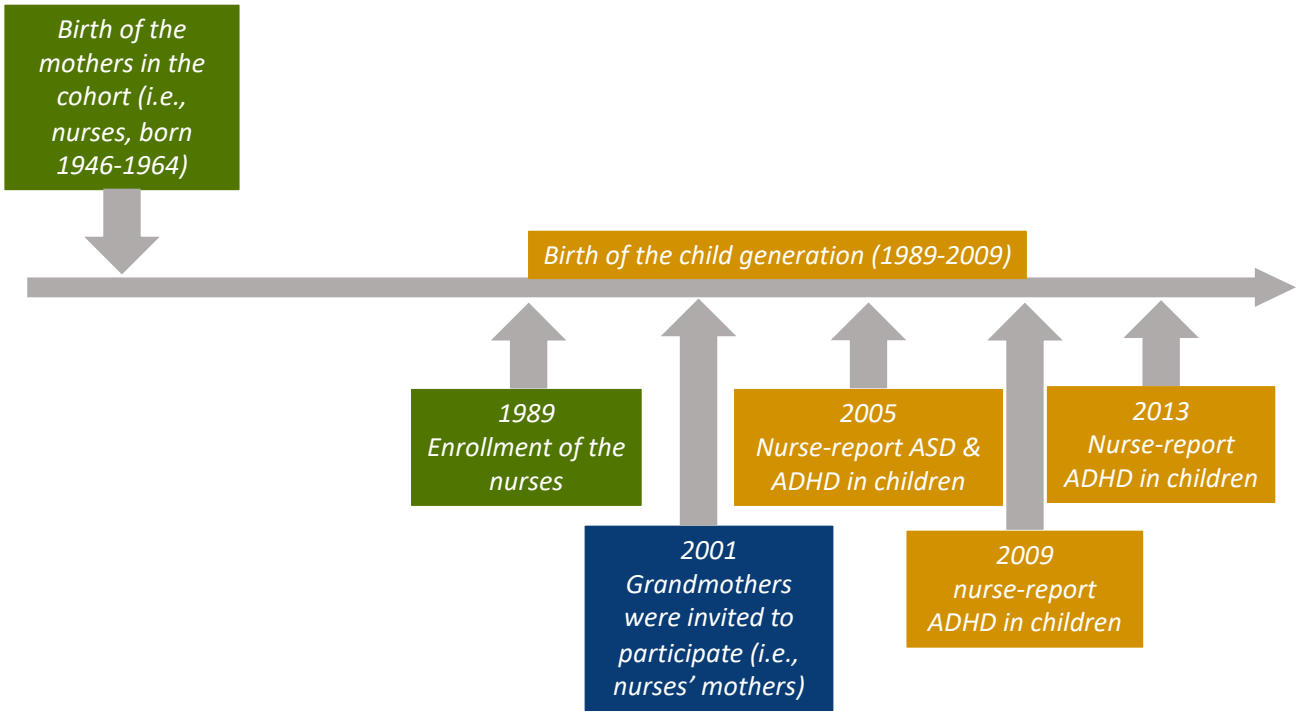
C2. Aim 2. To investigate whether grandmaternal breastfeeding during mother's infancy is associated with lower grandchildren's ASD and ADHD risk.

Rationale Breastfeeding has many well-documented health benefits, to mothers and their children, both in the short term and the longer term.[55,115,116] Recent evidence has suggested potential multigenerational effects of breastfeeding, [70,74] through mechanism such as improved nutrition,[70,72] enhanced infant attachment,[73] and stabilized infant gut microbiota[70,71] that may have the potential to induce epigenetic remodeling of germ cells. However, no study to date has evaluated whether the effects of breastfeeding among grandmothers can influence the neurodevelopment of their grandchildren.

Therefore, I propose to investigate whether grandmaternal breastfeeding is associated with decreased risk for two of the most common childhood neurodevelopmental disorders, ASD and ADHD, in their grandchildren. I will utilize data on grandmother-reported breastfeeding practices and on ASD and ADHD diagnostic status in children of the nurses enrolled in the Nurses' Health Study II (NHS II). The NHS II is one of the largest and longest cohorts in the U.S. that has uniquely collected detailed information on grandmaternal breastfeeding practices of the nurses' mothers as part of the Mother's Questionnaire and diagnosis of ASD and ADHD in their children reported by the nurses.

Study Setting and Design The Nurses' Health Study (NHS) II is a nationwide prospective cohort consisted of 116,430 American female registered nurses who were at 25-42 years of age and living in the 11 most populous statues in the United States in 1989. Since 1989, the nurses were followed-up biennially every odd calendar year to report demographic and lifestyle factors and as well as their child's health status including diagnosis of two neurodevelopmental disorders of interest (i.e., ADHD and ASD).[43–45] In March 2001, mothers of nurses in NHS II (the grandmothers of the index grandchild) who were alive and free of cancer were invited to participate in the **Nurses' Mothers' Cohort Study** to answer mailed questionnaires about their pregnancy and early-life experiences of their nurses daughters.[43] NHS II and the Mothers' Cohort can be linked, providing data on grandmaternal and grandchildren's neurodevelopment, as well as rich information on sociodemographic covariates and potential mediators.[117,118]

Figure 4. An illustration of the study setting and design of the NHS II cohort for Specific Aim 2



Study Population A total of 35,830 (76.5%) nurses whose mothers participated in the Nurses' Mothers' Cohort Study and reported data on the entire lifespan of the nurse from conception to adult life will serve as the source population.[43] Nurses who have had a child from 1989-2009 (~80%) will be identified.[119] Nurses and their children who were adopted, and families with missing data of grandmaternal breastfeeding practices or the grandchildren's diagnostic status of ASD and ADHD will be excluded from the final analytical cohort. Only singleton-born individuals will be included in the study.

Grandmaternal Breastfeeding Grandmothers participated in the Nurses' Mothers' Cohort Study were asked whether they ever breastfed their nurse daughter [yes, no] and the duration of breastfeeding of their nurse daughters [less than one week, 1 week-3 months, 3-6 months, 6-9 months, 9-12 months, and one year or more]. Grandmothers were also asked whether they ever fed the nurse daughter infant formula, evaporated milk, commercial infant formula, soy-based infant formula, regular cow's milk, and solid food, and if so, the time of introduction and duration. The effects of the status of formula feeding, mixed feeding, and exclusive breastfeeding will be compared (not supplementing with formula, evaporated milk, or solid food). Duration of exclusive and partial breastfeeding will be additionally calculated using the information on breastfeeding and supplemental feeding and the possible effects of the duration of breastfeeding will be assessed only among those who have ever breast-fed their daughters, considering potential sociodemographic and health differences among those who have never breastfed. My preliminary results show that the ever-breastfed rate among the grandmothers in the study cohort is 47.12%.

Grandchildren's ADHD and ASD The outcomes of interest are mother (nurse) reported ASD diagnosis (Autism, Asperger's, and other ASD not listed) in their children born 1989-2009 (the third generation in this cohort), assessed in the 2005 and 2009 questionnaires and mother (nurse)-reported ADHD diagnosis in children, assessed in the 2005 and 2013 questionnaires. Both mother-report ASD and ADHD in the NHS II have found to be reliable. Diagnosis of ASD has been validated by telephone administered Autism Diagnostic Interview–Revised (ADI-R) to 50 randomly selected nurses, with 86% of the nurses who completed the ADI met full criteria for a clinical diagnosis of ASD.[120] A prior validation study with 92 participating nurses who reported ADHD diagnosis in their children has found that their children scored high on the ADHD Rating Scale-IV (all girls scored above 90%, 81.1% of boys scored above 80%, and 63.8% of boys scored above 90%).[81,121]

Covariates Mothers of the nurses (i.e., grandmothers in the cohort) provided information on their sociodemographic and pregnancy-related factors that will be examined as covariates. These factors will include **1) sociodemographic characteristics:** grandmaternal education level at the birth of the nurse daughter [high school or less, some college, college or above] and grandmaternal occupation during

pregnancy with the nurse daughter (professional, executive/manager, sales/clerical work, mechanic/skilled, machine operator/driver, service worker, laborer/unskilled work, farming, or house maker); **2) lifestyle and health-related factors:** grandmaternal pre-pregnancy body mass index [<18.5, underweight; 18.5–24.9, healthy; 25.0-29.9, overweight; ≥ 30.0, obese] and grandmaternal smoking [yes, no], 3**) grandmaternal conditions** diagnosed during pregnancy (e.g., diabetes and high blood pressure), and grandmaternal medication (for nausea, sleep, or DES) used during pregnancy [yes, no]; and 4) **grandmaternal delivery characteristics**: grandmaternal age at delivery [<=20, 21-24, 25-29, 30-34, 35-39, ≥ 40], delivery method [vaginal delivery, cesarean section), preterm and low birthweight status of the delivery (i.e. the birth of the nurse), as well as grandmaternal year of delivery, namely the birth year of the nurses, to account for secular trends that may influence breastfeeding practices,[118]

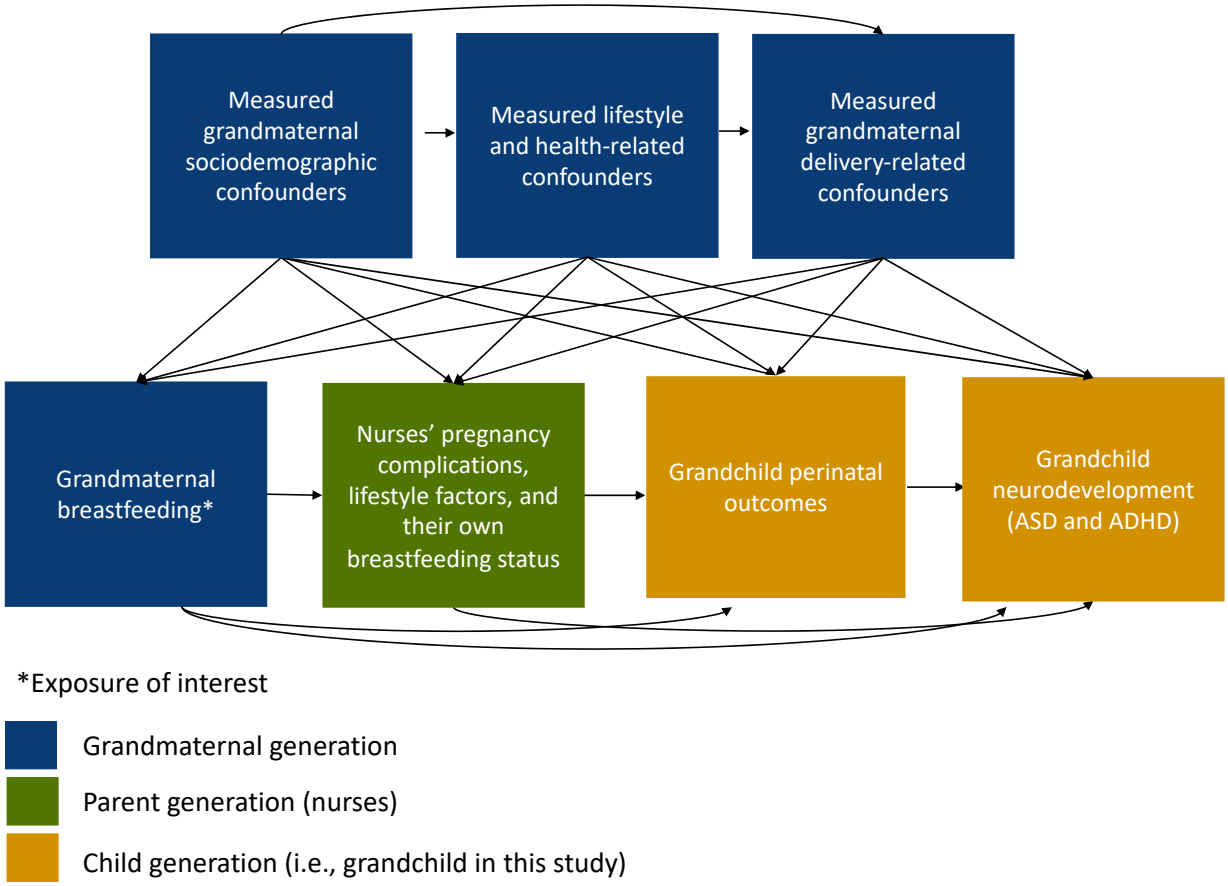
Potential Mediators To understand the pathways through which grandmaternal breastfeeding practices influence grandchildren's neurodevelopment, a range of variables will be considered as mediators based on current evidence. These variables include: **1) nurses' pregnancy related complications** (i.e., preeclampsia or toxemia, pregnancy-related high blood pressure, or gestational diabetes; yes or no); **2) nurses' lifestyle factors**, specifically smoking and drinking during pregnancy; **3) nurses' own breastfeeding status** during the child's infancy (ever/never and duration); as well as **4) the child's perinatal characteristics**: preterm birth (less than 37 completed weeks of gestation) and low birthweight (less than 5 pounds, 8 ounces [2,500 grams]) status

Statistical Analysis The distributions of some sociodemographic covariates by grandmaternal breastfeeding status have been generated in the initial phase of the project and presented in **Table 2**. All analyses will be guided by a DAG that has been developed *a priori* (**Figure 5**). Logistic regression models will be used to estimate the associations between grandmaternal breastfeeding and odds with 95% confidence intervals (CI) for ADHD or ASD diagnoses in grandchildren adjusting for potential confounders. Three exposure classifications will be compared: ever breastfeeding [yes vs. no], duration of breastfeeding among those who have ever breastfed [less than one week, 1 week-3 months, 3-6 months, >6months], and methods of breastfeeding [formula, mixed vs. exclusively breastfeeding]. Both unadjusted and adjusted odds for ASD and ADHD diagnosis in grandchildren according to grandmaternal breastfeeding of the nurse mothers will be estimated, with previously described covariates included in the adjusted models.

To assess the possibility that breastfeeding effect is an artifact of residual confounding,[122] a third set of logistic regression models only including grandmaternal education and occupation will be constructed and the effect estimates will be compared to those obtained from the main analyses and stratified analyses by grandmaternal education and occupation will be performed, respectively. To further understand the possible relationships between the grandmaternal and grandchild variables of interest, the potential effect modification by grandmaternal mode of delivery [vaginal delivery, caesarean section] and grandchildren's sex will be additionally evaluated by stratified analysis. Test of heterogeneity will be performed by assessing the P value and corresponding CI for the interaction term. To what extent the potential association between grandmaternal breastfeeding and grandchildren's ASD and ADHD can be explained by the previously described mediators will be examined using a regression-based mediation analysis approach (detailed description of the approach can be found under Specific Aim 1). Complete case analysis will be conducted given the low level of missingness in covariate data in the NHS II (≤ 8.1%).

The estimated power to detect an odds ratio of 0.90 is 0.89 and >0.99 for odds ratios $\leq$ 0.85 for ADHD; for ASD, the estimated power to detect an odds ratio of 0.90, 0.85, 0.80 and 0.75 is 0.27, 0.54, 0.79, and 0.94, respectively, assuming two-sided alpha = 0.05. All statistical analyses will be performed using SAS version 9.4 (SAS Institute Inc.).

Figure 5. **The directed acyclic graph (DAG) for the study variables for Specific Aim 2**



Limitations The proposed study has several limitations to be noted. First, the cohort of the mother of the nurses is only a subset of NHS-II, making selection bias possible, although a recent study on ADHD has shown that such bias may be minimal as they have found little difference in grandchildren's ADHD prevalence by grandmother's participation status in the study.[29] Second, misclassification of grandmaternal breastfeeding practices is also possible as a result of self-report and long-term recall. Nevertheless, such misclassification is most likely to be non-differential by the grandchildren health, and it has been previously shown that events surrounding pregnancy and birth, including breastfeeding, are special in a mother's life and likely to be accurately recalled even after many decades.[123] As previously described, the extent to which inaccurate recall may influence the study results will also be evaluated in this study by comparing the effect size estimates among grandmothers who only had one child to those who had several children. Further, although a variety of sociodemographic confounders will be adjusted, residual confounding from unknown and unmeasured factors is still likely. Finally, the generalizability of the study results will need to be determined with future research as the NHS II participants are predominantly Caucasians.[118]

Strengths This nationwide, prospective cohort of nurses provides a unique opportunity to investigate the relation between grandmaternal breastfeeding and grandchildren's ASD and ADHD risk, which is of important public health relevance yet has not been studied. The ascertainment of ASD and ADHD in this study has been found to be reliable in previous validation studies.[121,124] Further, a rich set of covariates directly obtained from the grandmothers allows for more reliable effect estimates, and data on the pregnancy and life-style factors of the nurses enables exploration of the possible mediating paths that may underlying the hypothesized associations. If associations are found, this will be the first study to suggest that grandmaternal breastfeeding may influence ASD and/or ADHD in their grandchildren. This study will help identify and prioritize future research directions in the field (e.g., replication and reevaluation from cohort studies with prospectively collected breastfeeding and offspring neurodevelopment data), and potentially call for additional social, economic, and political efforts to ensure breastfeeding support to all women.

C3. Aim 3. To assess selection bias in multigenerational studies that arise from excluding individuals who have not reproduced during the study follow-up

Rationale Research spanning multiple generations is subject to many methodological challenges that are shared in all epidemiological studies, e.g., exposure misclassification, residual confounding, and inappropriate adjustment for potential mediators.[25,28,29, 71–78] Notably, there is one potential source of bias that is unique to studies on the effects of grandmaternal factors, which is informatively empty clusters in the following generations, as pointed to in two recent simulation studies and considered in an epidemiological study on periconceptional weight in grandmothers.[25,71,82] This issue has also been discussed in my previous work, which has recognized preterm birth of the parents (i.e., delivery characteristics of the grandmothers) as a possible contributor to empty clusters in the third generation that may result in selection bias. To formally assess this potential issue, I propose to carry out a case study using my previous three-generation linkage study on parental preterm birth and offspring ASD. Specifically, I will first investigate whether individuals born preterm are less likely to have a child and thus not selected into the multigenerational analysis, compared to their non-preterm counterparts. Then, I will use inverse probability weighting to assess if the association observed between preterm birth status in the parent generation and elevated risk for ASD is affected by such bias.

Study Setting and Design From the Danish Medical Birth Register (DMBR), I will identify all singleton births (parent generation) in Denmark during the period of 1978-2001 for follow-up **until August 10, 2017**. The DMBR has collected detailed records of pregnancy and delivery characteristics (e.g., gestational age and birthweight) that are registered by the midwife immediately upon delivery[125] for more than 98% births in Denmark since 1973.[126] All individuals registered in the DMBR have been assigned to an unique 10-digit civil registration identifier, and will be mother's or father's identifier if they have ever had a birth recorded in the DMBR, allowing for accurate construction of familial relationships.[127] Age at which the index individuals have had their first child was also recorded in the child's civil registration file that feeds into the DMBR.[91,127] The 10-digit personal identifier also allows for linkage across registries,[127] such as the Danish Psychiatric Central Registry (DPCR) to obtain ASD diagnostic records of the child.

Study Population Women and men needed to be at reproductive age (15-49, in accordance with the definition by Statistics Denmark)[128] at the end of study follow-up. A total of 662,966 women and 699,697 men with complete and realistic information on gestational age (20-45 completed weeks) will be included in the final cohort for analysis.

Exposure Variables The primary exposure variable of interest is preterm birth. Gestational age estimated based on reported last menstrual period and clinical evaluation at delivery until 1999 in the DMBR[129] will be used to classify birth characteristics of the index individuals into preterm (born before 37 completed weeks of gestation[130]), very preterm (born before 32 completed weeks) and the non-preterm reference (37-45 completed weeks).

Subsequent Reproductive Outcomes Amongst Index Women/Men Women and men in the study cohort will be classified into ever versus never giving a live birth to or fathered a child based on whether they have had a birth registered in the DMBR. According to the number of existing records, individuals will be further divided into three groups: having one child, two children, and more than three children.

Covariates Sociodemographic characteristics in index women and men and their mothers (i.e., the grandmothers in the study) are selected as covariates *a priori*. Covariate data are retrieved from multiple national registers in Denmark. These variables include: birth year of the index individuals (1978-1982, 1983-1987, 1988-2001), maternal parity status (1,2, ≥ 3), maternal age at delivery (12-19, 20-24, 25-29, 30-34, 35-39, ≥ 40) and maternal place of residence (Capital city [Copenhagen], other major cities [Aarhus, Odense, Aalborg], and other) registered in the Danish Civil Registration Service;[91] and maternal education level (primary and lower secondary, upper secondary education and academy profession degree, bachelor and above) collected in the Integrated Database for Labor Market Research (IDA).[94]

Statistical Analysis First, sociodemographic characteristics among individuals in the baseline cohort and the selected cohorts are compared for women and men, respectively. Some pilot study results comparing sociodemographic characteristics among individuals in the baseline cohort and the selected cohorts have been obtained and are presented in Tables 3a and 3b. Overall, 463,415 women and 544,374 men had never had a

live birth registered during the follow up, while 84,457 women and 74,226 men had one birth, 89,650 women and 64,791 men had two births, and 25,444 women and 16,306 men had more than three. Individuals who had not had any child (therefore not selected into the multigenerational analysis) at the end of study follow-up were more likely to be born to mothers with lower education and age at delivery younger than 25.

Then, I will construct logistic regression models to evaluate whether the exposure, namely the preterm birth status in the parent generation (women and men, respectively), influence reproduction later in life and ultimately their selection into multigenerational analyses. I will estimate the odds ratios (ORs) and 95% confidence intervals (CIs) for the binary reproductive status (ever/never) and total number of children (compared to those with no recorded live birth) according to the exposure status (yes/no being born preterm birth) in the index individuals.

To account for potential selection bias that rises when only population with available child data is included, **stabilized inverse probability of selection weights will be estimated using a two-step approach**:[88] **1)** calculating the probability of being selected into the analytical cohort, conditional to the aforementioned covariates; and **2)** taking the inverse of the conditional probability in logistic regression analyses among the selected, i.e., index individuals who have had children within the specified study period. This weighting created a pseudo population in which the distributions of measured factors represent the unselected study source population.[88] **Further**, using my previous multigenerational work that has reported associations between F1 preterm birth and elevated third-generation ASD risk as a case study,[101] I will assess the influences of potential selection bias due to never having a live-born child among individuals born preterm by comparing the results with or without accounting for such selection. Finally, under guidance of Dr. Liew and Dr. Rogne, I will develop DAGs of possible scenarios showing hypothesized relationships among study variables after inverse probability of weighting adjustment.

Several sensitivity analyses are planned. To evaluate the effects of potential exposure misclassification, I will estimate ORs for the very preterm birth group comparing to the non-preterm birth reference, given the suggested higher validity of the classification of gestational ages less than 32 weeks.[131] We also limited our study population to those without congenital malformations through linkage with the National Patient Register[132] and re-evaluated our results. Congenital malformations were defined as having any diagnoses made based on the International Statistical Classification of Diseases and Related Health Problems, eighth revision (ICD-8, prior to 1994) and tenth Revision (ICD-10, since1994), codes 740–759 in ICD-8 and Q00–Q99 in ICD-10. All effect size estimates were adjusted for birth year of the F1 individuals and maternal (F0) pregnancy-related factors, including parity status, age at delivery, place of residence, and education level. Only individuals with complete covariate data were analyzed, given the small number of missing values (<1.5%).

Under Aim 3, I will also use the inverse probability of selection weighting method to re-evaluate the results obtained from Aims 1 & 2 by comparing the effect estimates for the association between grandmaternal educational attainment/breastfeeding and grandchildren's neurodevelopment before and after adjustment for selection bias associated with the exposures and the preterm status of the parent generation.

All statistical analyses will be performed using SAS statistical software, version 9.4 (SAS Institute, Cary, NC).

Limitations Several possible limitations should be pointed out. First, this study focuses on selection bias related to preterm birth. Selection bias structures due to other pre-and postnatal exposures as well as later health outcomes and events (e.g., emigration and death) should be evaluated in future research. In common with other studies involving age of gestation, there might be some misclassification in preterm birth status based on gestational age.[133] Nevertheless, the misclassification is probably non-differential and the analyses of very preterm birth, which are less prone to misclassification,[131] will likely provide evidence on the robustness of study results. Further, the estimations of effect size may be under the influence of unmeasured sociodemographic or environmental factors such as maternal smoking, which is only available until late 1997 in Denmark therefore cannot be controlled for in this study.[127]

Strengths The proposed study involves nationwide data coverage (~98% of Danish individuals), long-term prospective data collection and follow-up (~40 years), and comparison between women and men. Moreover, data collected in the Danish data registries is overall high in quality and validity.[134] Because of the low emigration rate in Denmark,[135] bias introduced by differential loss to follow-up was also expected to be very

minimal. Furthermore, linkage across generations and registries was highly accurate with the use of the unique personal identifier, and the linkage of registry data additionally prevented recall and reporting biases. Taken together, this will be the first study to formally study the presence and effects of potential selection bias that are unique in multigenerational research. Knowledge of this sources of bias will inform future studies to appropriately address it when estimating effects of grandmother-related factors.

D. Estimated Timeline

The approximate timeline for my dissertation process is presented in the table below. Access to the Danish Medical Birth Registry (DMBR) and other national Danish registries has been secured through existing collaborations with Dr. Jiong Li, who is a reproductive epidemiologist and an associate professor at Department of Epidemiology, Aarhus university, Denmark. Access to the Nurses' Health Study (NHS) II has been obtained through collaborations with Dr. Marc G. Weisskopf, who is a professor of Environmental Epidemiology and Physiology, Environmental Health at Harvard University. To deepen my understanding of the Danish registry data that is an essential part of my proposed dissertation research, a three-month research stay at Aarhus University, Denmark will take place in early 2023, during which the statical analysis and a part of my writing for Specific Aim 1 will also be completed.

References

1. Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental Disorders: From Genetics to Functional Pathways. *Trends in Neurosciences*. 2020;43(8):608-621. doi:10.1016/j.tins.2020.05.004

2. Romero-Ayuso D. Future Challenges in Research in Children with Neurodevelopmental Disorders. *Children (Basel)*. 2021;8(5):328. doi:10.3390/children8050328

3. Hansen BH, Oerbeck B, Skirbekk B, Petrovski BÉ, Kristensen H. Neurodevelopmental disorders: prevalence and comorbidity in children referred to mental health services. *Nordic Journal of Psychiatry*. 2018;72(4):285-291. doi:10.1080/08039488.2018.1444087

4. Atladóttir HÓ, Parner ET, Schendel D, Dalsgaard S, Thomsen PH, Thorsen P. Time Trends in Reported Diagnoses of Childhood Neuropsychiatric Disorders: A Danish Cohort Study. *Archives of Pediatrics & Adolescent Medicine*. 2007;161(2):193-198. doi:10.1001/archpedi.161.2.193

5. Doernberg E, Hollander E. Neurodevelopmental Disorders (ASD and ADHD): DSM-5, ICD-10, and ICD-11. *CNS Spectrums*. 2016;21(4):295-299. doi:10.1017/S1092852916000262

6. Jeste SS. Neurodevelopmental Behavioral and Cognitive Disorders. *CONTINUUM: Lifelong Learning in Neurology*. 2015;21(3):690. doi:10.1212/01.CON.0000466661.89908.3c

7. CDC. Data and Statistics on Autism Spectrum Disorder | CDC. Centers for Disease Control and Prevention. Published March 2, 2022. Accessed April 30, 2022. https://www.cdc.gov/ncbddd/autism/data.html

8. CDC. Data and Statistics About ADHD | CDC. Centers for Disease Control and Prevention. Published November 16, 2020. Accessed April 30, 2022. https://www.cdc.gov/ncbddd/adhd/data.html

9. Attention-Deficit/Hyperactivity Disorder (ADHD). National Institute of Mental Health (NIMH). Accessed April 30, 2022. https://www.nimh.nih.gov/health/statistics/attention-deficit-hyperactivity-disorder-adhd

10. Dall'Aglio L, Muka T, Cecil CAM, et al. The role of epigenetic modifications in neurodevelopmental disorders: A systematic review. *Neuroscience & Biobehavioral Reviews*. 2018;94:17-30. doi:10.1016/j.neubiorev.2018.07.011

11. Toward Understanding Intellectual Disability Stigma: Introduction | SpringerLink. Accessed April 30, 2022. https://link.springer.com/chapter/10.1057/978-1-137-52499-7_1

12. Cardoso AR, Lopes-Marques M, Silva RM, et al. Essential genetic findings in neurodevelopmental disorders. *Human Genomics*. 2019;13(1):31. doi:10.1186/s40246-019-0216-4

13. Heindel JJ, Skalla LA, Joubert BR, Dilworth CH, Gray KA. Review of developmental origins of health and disease publications in environmental epidemiology. *Reproductive Toxicology*. 2017;68:34-48. doi:10.1016/j.reprotox.2016.11.011

14. Merced-Nieves FM, Arora M, Wright RO, Curtin P. Metal mixtures and neurodevelopment: Recent findings and emerging principles. *Current Opinion in Toxicology*. 2021;26:28-32. doi:10.1016/j.cotox.2021.03.005

15. O'Shaughnessy KL, Fischer F, Zenclussen AC. Perinatal exposure to endocrine disrupting chemicals and neurodevelopment: How articles of daily use influence the development of our children. *Best Practice & Research Clinical Endocrinology & Metabolism*. 2021;35(5):101568. doi:10.1016/j.beem.2021.101568

16. Volk HE, Perera F, Braun JM, et al. Prenatal air pollution exposure and neurodevelopment: A review and blueprint for a harmonized approach within ECHO. *Environmental Research*. 2021;196:110320. doi:10.1016/j.envres.2020.110320

17. Manzari N, Matvienko-Sikar K, Baldoni F, O'Keeffe GW, Khashan AS. Prenatal maternal stress and risk of neurodevelopmental disorders in the offspring: a systematic review and meta-analysis. *Soc Psychiatry Psychiatr Epidemiol*. 2019;54(11):1299-1309. doi:10.1007/s00127-019-01745-3

18. Cortés-Albornoz MC, García-Guáqueta DP, Velez-van-Meerbeke A, Talero-Gutiérrez C. Maternal Nutrition and Neurodevelopment: A Scoping Review. *Nutrients*. 2021;13(10):3530. doi:10.3390/nu13103530

19. Álvarez-Bueno C, Cavero-Redondo I, Lucas-de la Cruz L, Notario-Pacheco B, Martínez-Vizcaíno V. Association between pre-pregnancy overweight and obesity and children's neurocognitive development: a systematic review and meta-analysis of observational studies. *Int J Epidemiol*. 2017;46(5):1653-1666. doi:10.1093/ije/dyx122

20. Polańska K, Jurewicz J, Hanke W. Smoking and alcohol drinking during pregnancy as the risk factors for poor child neurodevelopment – A review of epidemiological studies. *Int J Occup Med Environ Health*. 2015;28(3):419-443. doi:10.13075/ijomeh.1896.00424

21. Yu J, Patel RA, Gilman SE. Childhood disadvantage, neurocognitive development and neuropsychiatric disorders: Evidence of mechanisms. *Current Opinion in Psychiatry*. 2021;34(3):306-323. doi:10.1097/YCO.0000000000000701

22. Wu W, Wu P, Tang Q, Lu C. Transgenerational Epigenetic Inheritance of Developmental Origins of Health and Disease. In: Xia Y, ed. *Early-Life Environmental Exposure and Disease: Facts and Perspectives*. Springer; 2020:229-239. doi:10.1007/978-981-15-3797-4_14

23. Calkins K, Devaskar SU. Fetal Origins of Adult Disease. *Current Problems in Pediatric and Adolescent Health Care*. 2011;41(6):158-176. doi:10.1016/j.cppeds.2011.01.001

24. Hagemann E, Silva DT, Davis JA, Gibson LY, Prescott SL. Developmental Origins of Health and Disease (DOHaD): The importance of life-course and transgenerational approaches. *Paediatric Respiratory Reviews*. 2021;40:3-9. doi:10.1016/j.prrv.2021.05.005

25. Wolstenholme JT, Edwards M, Shetty SRJ, et al. Gestational Exposure to Bisphenol A Produces Transgenerational Changes in Behaviors and Gene Expression. *Endocrinology*. 2012;153(8):3828-3838. doi:10.1210/en.2012-1195

26. Franklin TB, Russig H, Weiss IC, et al. Epigenetic transmission of the impact of early stress across generations. *Biol Psychiatry*. 2010;68(5):408-415. doi:10.1016/j.biopsych.2010.05.036

27. Meier MJ, O 'Brien Jason M., Beal MA, Allan B, Yauk CL, Marchetti F. In Utero Exposure to Benzo[a]Pyrene Increases Mutation Burden in the Soma and Sperm of Adult Mice. *Environmental Health Perspectives*. 2017;125(1):82-88. doi:10.1289/EHP211

28. Kioumourtzoglou MA, Coull BA, O'Reilly ÉJ, Ascherio A, Weisskopf MG. Association of Exposure to Diethylstilbestrol During Pregnancy With Multigenerational Neurodevelopmental Deficits. *JAMA Pediatr*. 2018;172(7):670-677. doi:10.1001/jamapediatrics.2018.0727

29. Yim G, Roberts A, Ascherio A, Wypij D, Kioumourtzoglou MA, Weisskopf MG. Association Between Periconceptional Weight of Maternal Grandmothers and Attention-Deficit/Hyperactivity Disorder in Grandchildren. *JAMA Network Open*. 2021;4(7):e2118824. doi:10.1001/jamanetworkopen.2021.18824

30. Yim G, Roberts A, Ascherio A, Wypij D, Kioumourtzoglou MA, Weisskopf MG. Smoking During Pregnancy and Risk of Attention-Deficit/Hyperactivity Disorder in the Third Generation. *Epidemiology*. Published online February 28, 2022. doi:10.1097/EDE.0000000000001467

31. Golding J, Ellis G, Gregory S, et al. Grand-maternal smoking in pregnancy and grandchild's autistic traits and diagnosed autism. *Sci Rep*. 2017;7. doi:10.1038/srep46179

32. Frans EM, Sandin S, Reichenberg A, et al. Autism Risk Across Generations: A Population Based Study of Advancing Grandpaternal and Paternal Age. *JAMA Psychiatry*. 2013;70(5):516-521. doi:10.1001/jamapsychiatry.2013.1180

33. Gao Y, Yu Y, Xiao J, et al. Association of Grandparental and Parental Age at Childbirth With Autism Spectrum Disorder in Children. *JAMA Netw Open*. 2020;3(4):e202868-e202868. doi:10.1001/jamanetworkopen.2020.2868

34. Xiao J, Gao Y, Yu Y, et al. Associations of parental birth characteristics with autism spectrum disorder (ASD) risk in their offspring: a population-based multigenerational cohort study in Denmark. *International Journal of Epidemiology*. 2021;00(00):11.

35. Kachmar AG, Connolly CA, Wolf S, Curley MAQ. Socioeconomic Status in Pediatric Health Research: A Scoping Review. *The Journal of Pediatrics*. 2019;213:163-170. doi:10.1016/j.jpeds.2019.06.005

36. Herd P, Goesling B, House JS. Socioeconomic Position and Health: The Differential Effects of Education versus Income on the Onset versus Progression of Health Problems. *J Health Soc Behav*. 2007;48(3):223-238. doi:10.1177/002214650704800302

37. Sauerbrun-Cutler MT, Segars JH. Do In Utero Events Contribute to Current Health Disparities in Reproductive Medicine? *Semin Reprod Med*. 2013;31(5):325-332. doi:10.1055/s-0033-1348890

38. Hamm NC, Hamad AF, Wall-Wieler E, Roos LL, Plana-Ripoll O, Lix LM. Multigenerational health research using population-based linked databases: an international review. *Int J Popul Data Sci*. 6(1):1686. doi:10.23889/ijpds.v6i1.1686

39. Maternal Active Smoking During Pregnancy and Low Birth Weight in the Americas: A Systematic Review and Meta-analysis | Nicotine & Tobacco Research | Oxford Academic. Accessed February 7, 2022. https://academic.oup.com/ntr/article-abstract/19/5/497/3589578

40. Davis-Kean PE, Tighe LA, Waters NE. The Role of Parent Educational Attainment in Parenting and Children's Development. *Curr Dir Psychol Sci*. 2021;30(2):186-192. doi:10.1177/0963721421993116

41. Domina T, Roksa J. Should Mom go back to school? Post-natal educational attainment and parenting practices. *Social Science Research*. 2012;41(3):695-708. doi:10.1016/j.ssresearch.2011.12.002

42. Taylor CM, Golding J, Hibbeln J, Emond AM. Environmental Factors Predicting Blood Lead Levels in Pregnant Women in the UK: The ALSPAC Study. *PLOS ONE*. 2013;8(9):e72371. doi:10.1371/journal.pone.0072371

43. Ou CQ, Hedley AJ, Chung RY, et al. Socioeconomic disparities in air pollution-associated mortality. *Environmental Research*. 2008;107(2):237-244. doi:10.1016/j.envres.2008.02.002

44. Sagiv SK, Rifas-Shiman SL, Webster TF, et al. Sociodemographic and Perinatal Predictors of Early Pregnancy Per- and Polyfluoroalkyl Substance (PFAS) Concentrations. *Environ Sci Technol*. 2015;49(19):11849-11858. doi:10.1021/acs.est.5b02489

45. De Felice A, Ricceri L, Venerosi A, Chiarotti F, Calamandrei G. Multifactorial Origin of Neurodevelopmental Disorders: Approaches to Understanding Complex Etiologies. *Toxics*. 2015;3(1):89-129. doi:10.3390/toxics3010089

46. Jain NJ, Faiz AS, Ohman-Strickland PA, Smulian JC, Rhoads GG. Educational Attainment of Grandmothers and Preterm Birth in Grandchildren. *Matern Child Health J*. 2021;25(2):293-301. doi:10.1007/s10995-020-03021-2

47. Are Early-Life Socioeconomic Conditions Directly Related to Birth Outcomes? Grandmaternal Education, Grandchild Birth Weight, and Associated Bias Analyses | American Journal of Epidemiology | Oxford Academic. Accessed October 5, 2021. https://academic.oup.com/aje/article/182/7/568/107922?login=true

48. Astone NM, Misra D, Lynch C. The effect of maternal socio-economic status throughout the lifespan on infant birthweight. *Paediatric and Perinatal Epidemiology*. 2007;21(4):310-318. doi:10.1111/j.1365-3016.2007.00821.x

49. Abegunde D, Hutchinson P, Anaba U, et al. Socioeconomic inequality in exclusive breastfeeding behavior and ideation factors for social behavioral change in three north-western Nigerian states: a cross-sectional study. *International Journal for Equity in Health*. 2021;20(1):172. doi:10.1186/s12939-021-01504-4

50. Lawrence RA. Breastfeeding Triumphs. *Birth*. 2012;39(4):311-314. doi:10.1111/birt.12007

51. Binns C, Lee M, Low WY. The Long-Term Public Health Benefits of Breastfeeding. *Asia Pac J Public Health*. 2016;28(1):7-14. doi:10.1177/1010539515624964

52. Alimi R, Azmoude E, Moradi M, Zamani M. The Association of Breastfeeding with a Reduced Risk of Postpartum Depression: A Systematic Review and Meta-Analysis. *Breastfeeding Medicine*. Published online December 29, 2021. doi:10.1089/bfm.2021.0183

53. Sankar MJ, Sinha B, Chowdhury R, et al. Optimal breastfeeding practices and infant and child mortality: a systematic review and meta-analysis. *Acta Paediatrica*. 2015;104(S467):3-13. doi:10.1111/apa.13147

54. Schwarz EB, Ray RM, Stuebe AM, et al. Duration of lactation and risk factors for maternal cardiovascular disease. *Obstet Gynecol*. 2009;113(5):974-982. doi:10.1097/01.AOG.0000346884.67796.ca

55. Schwarz EB, Nothnagle M. The Maternal Health Benefits of Breastfeeding. *AFP*. 2015;91(9):602-604.

56. Ghozy S, Tran L, Naveed S, et al. Association of breastfeeding status with risk of autism spectrum disorder: A systematic review, dose-response analysis and meta-analysis. *Asian Journal of Psychiatry*. 2020;48:101916. doi:10.1016/j.ajp.2019.101916

57. Oddy WH. Breastfeeding, Childhood Asthma, and Allergic Disease. *ANM*. 2017;70(Suppl. 2):26-36. doi:10.1159/000457920

58. Infant and young child feeding. Accessed March 17, 2022. https://www.who.int/news-room/fact-sheets/detail/infant-and-young-child-feeding

59. Mead MN. Contaminants in Human Milk: Weighing the Risks against the Benefits of Breastfeeding. *Environmental Health Perspectives*. 2008;116(10):A426-A434. doi:10.1289/ehp.116-a426

60. Breastfeeding Report Card United States, 2020. *New York*.:6.

61. Smith PH, Hausman B, Labbok M. *Beyond Health, Beyond Choice: Breastfeeding Constraints and Realities*. Rutgers University Press; 2012.

62. Louis-Jacques A, Deubel TF, Taylor M, Stuebe AM. Racial and ethnic disparities in U.S. breastfeeding and implications for maternal and child health outcomes. *Seminars in Perinatology*. 2017;41(5):299-307. doi:10.1053/j.semperi.2017.04.007

63. Li R, Darling N, Maurice E, Barker L, Grummer-Strawn LM. Breastfeeding Rates in the United States by Characteristics of the Child, Mother, or Family: The 2002 National Immunization Survey. *Pediatrics*. 2005;115(1):e31-e37. doi:10.1542/peds.2004-0481

64. Balogun OO, Dagvadorj A, Anigo KM, Ota E, Sasaki S. Factors influencing breastfeeding exclusivity during the first 6 months of life in developing countries: a quantitative and qualitative systematic review. *Maternal & Child Nutrition*. 2015;11(4):433-451. doi:10.1111/mcn.12180

65. Hartwig FP, Mola CL de, Davies NM, Victora CG, Relton CL. Breastfeeding effects on DNA methylation in the offspring: A systematic literature review. *PLOS ONE*. 2017;12(3):e0173070. doi:10.1371/journal.pone.0173070

66. Hartwig FP, Davey Smith G, Simpkin AJ, Victora CG, Relton CL, Caramaschi D. Association between Breastfeeding and DNA Methylation over the Life Course: Findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Nutrients*. 2020;12(11):3309. doi:10.3390/nu12113309

67. Breastfeeding, Polyunsaturated Fatty Acid Levels in Colostrum and Child Intelligence Quotient at Age 5-6 Years - PubMed. Accessed April 27, 2021. https://pubmed.ncbi.nlm.nih.gov/28081886/

68. Jackson DB. The Association Between Breastfeeding Duration and Attachment: A Genetically Informed Analysis. *Breastfeed Med*. 2016;11(6):297-304. doi:10.1089/bfm.2016.0036

69. Kim H, Sitarik AR, Woodcroft K, Johnson CC, Zoratti E. Birth Mode, Breastfeeding, Pet Exposure, and Antibiotic Use: Associations With the Gut Microbiome and Sensitization in Children. *Curr Allergy Asthma Rep*. 2019;19(4):22. doi:10.1007/s11882-019-0851-9

70. Gabbianelli R, Bordoni L, Morano S, Calleja-Agius J, Lalor JG. Nutri-Epigenetics and Gut Microbiota: How Birth Care, Bonding and Breastfeeding Can Influence and Be Influenced? *International Journal of Molecular Sciences*. 2020;21(14):5032. doi:10.3390/ijms21145032

71. Indrio F, Martini S, Francavilla R, et al. Epigenetic Matters: The Link between Early Nutrition, Microbiome, and Long-term Health Development. *Front Pediatr*. 2017;5. doi:10.3389/fped.2017.00178

72. Franzago M, Fraticelli F, Stuppia L, Vitacolonna E. Nutrigenetics, epigenetics and gestational diabetes: consequences in mother and child. *Epigenetics*. 2019;14(3):215-235. doi:10.1080/15592294.2019.1582277

73. Jones-Mason K, Allen IE, Bush N, Hamilton S. Epigenetic marks as the link between environment and development: examination of the associations between attachment, socioeconomic status, and methylation of the SLC6A4 gene. *Brain and Behavior*. 2016;6(7):e00480. doi:10.1002/brb3.480

74. Darling Rasmussen P, Storebø OJ. Attachment and Epigenetics: A Scoping Review of Recent Research and Current Knowledge. *Psychol Rep*. 2021;124(2):479-501. doi:10.1177/0033294120901846

75. Derrien M, Alvarez AS, de Vos WM. The Gut Microbiota in the First Decade of Life. *Trends in Microbiology*. 2019;27(12):997-1010. doi:10.1016/j.tim.2019.08.001

76. Di Manno L, Macdonald JA, Knight T. The Intergenerational Continuity of Breastfeeding Intention, Initiation, and Duration: A Systematic Review. *Birth*. 2015;42(1):5-15. doi:10.1111/birt.12148

77. Pérez-Escamilla R. Breastfeeding in the 21st century: How we can make it work. *Social Science & Medicine*. 2020;244:112331. doi:10.1016/j.socscimed.2019.05.036

78. McGee G, Perkins NJ, Mumford SL, et al. Methodological Issues in Population-Based Studies of Multigenerational Associations. *Am J Epidemiol*. doi:10.1093/aje/kwaa125

79. Golding J, Ellis G, Gregory S, et al. Grand-maternal smoking in pregnancy and grandchild's autistic traits and diagnosed autism. *Sci Rep*. 2017;7(1):46179. doi:10.1038/srep46179

80. Yim G, Roberts A, Wypij D, Kioumourtzoglou MA, Weisskopf MG. Grandmothers' endocrine disruption during pregnancy, low birth weight, and preterm birth in third generation. *International Journal of Epidemiology*. 2021;(dyab065). doi:10.1093/ije/dyab065

81. Kioumourtzoglou MA, Coull BA, O'Reilly ÉJ, Ascherio A, Weisskopf MG. Association of Exposure to Diethylstilbestrol During Pregnancy With Multigenerational Neurodevelopmental Deficits. *JAMA Pediatr*. 2018;172(7):670-677. doi:10.1001/jamapediatrics.2018.0727

82. Golding J, van den Berg G, Northstone K, et al. Grandchild's IQ is associated with grandparental environments prior to the birth of the parents. *Wellcome Open Res*. 2021;5:198. doi:10.12688/wellcomeopenres.16205.2

83. Yim G, Roberts A, Ascherio A, Wypij D, Kioumourtzoglou MA, Weisskopf MG. Smoking During Pregnancy and Risk of Attention-Deficit/Hyperactivity Disorder in the Third Generation. *Epidemiology*. Published online February 28, 2022. doi:10.1097/EDE.0000000000001467

84. Ahrén-Moonga J, Silverwood R, Klinteberg B af, Koupil I. Association of Higher Parental and Grandparental Education and Higher School Grades With Risk of Hospitalization for Eating Disorders in Females: The Uppsala Birth Cohort Multigenerational Study. *American Journal of Epidemiology*. 2009;170(5):566-575. doi:10.1093/aje/kwp166

85. Golding J, Steer C, Pembrey M. Parental and Grandparental Ages in the Autistic Spectrum Disorders: A Birth Cohort Study. *PLOS ONE*. 2010;5(4):e9939. doi:10.1371/journal.pone.0009939

86. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *International Journal of Epidemiology*. 2013;42(5):1511-1519. doi:10.1093/ije/dyt127

87. Smith LH. Selection Mechanisms and Their Consequences: Understanding and Addressing Selection Bias. *Curr Epidemiol Rep*. 2020;7(4):179-189. doi:10.1007/s40471-020-00241-6

88. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615-625. doi:10.1097/01.ede.0000135174.63482.43

89. McGee G, Weisskopf MG, Kioumourtzoglou MA, Coull BA, Haneuse S. Informatively empty clusters with application to multigenerational studies. *Biostatistics*. 2020;21(4):775-789. doi:10.1093/biostatistics/kxz005

90. Kwok MK, Leung GM, Lam TH, Leung SSL, Schooling CM. Grandparental education, parental education and child height: evidence from Hong Kong's "Children of 1997" birth cohort. *Annals of Epidemiology*. 2013;23(8):475-484. doi:10.1016/j.annepidem.2013.05.016

91. Pedersen CB. The Danish Civil Registration System. *Scand J Public Health*. 2011;39(7_suppl):22-25. doi:10.1177/1403494810387965

92. Bliddal M, Broe A, Pottegård A, Olsen J, Langhoff-Roos J. The Danish Medical Birth Register. *Eur J Epidemiol*. 2018;33(1):27-36. doi:10.1007/s10654-018-0356-1

93. Jensen VM, Rasmussen AW. Danish education registers. *Scand J Public Health*. 2011;39(7_suppl):91-94. doi:10.1177/1403494810394715

94. Li J, Vestergaard M, Obel C, Cnattingus S, Gissler M, Olsen J. Cohort Profile: The Nordic Perinatal Bereavement Cohort. *Int J Epidemiol*. 2011;40(5):1161-1167. doi:10.1093/ije/dyq127

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

96. Avnon T, Ovental A, Many A. Twin versus singleton pregnancy in women ≥ 45 years of age: comparison of maternal and neonatal outcomes. *The Journal of Maternal-Fetal & Neonatal Medicine*. 2021;34(2):201-206. doi:10.1080/14767058.2019.1602115

97. Timmermans B. The Danish Integrated Database for Labor Market Research: Towards Demystification for the English Speaking Audience. *DRUID, Copenhagen Business School, Department of Industrial Economics and Strategy/Aalborg University, Department of Business Studies, DRUID Working Papers*. Published online January 1, 2010.

98. Petersson F, Baadsgaard M, Thygesen LC. Danish registers on personal labour market affiliation. *Scand J Public Health*. 2011;39(7_suppl):95-98. doi:10.1177/1403494811408483

99. UNESCO Institute for Statistics. *International Standard Classification of Education: ISCED 2011*. UNESCO Institute for Statistics; 2012.

100. Fombonne E, Quirke S, Hagen A. Prevalence and Interpretation of Recent Trends in Rates of Pervasive Developmental Disorders. *Mcgill J Med*. 2009;12(2):73.

101. Xia Y, Xiao J, Yu Y, et al. Rates of Neuropsychiatric Disorders and Gestational Age at Birth in a Danish Population. *JAMA Network Open*. 2021;4(6):e2114913-e2114913. doi:10.1001/jamanetworkopen.2021.14913

102. Schendel DE, Thorsteinsson E. Cumulative Incidence of Autism Into Adulthood for Birth Cohorts in Denmark, 1980-2012. *JAMA*. 2018;320(17):1811-1813. doi:10.1001/jama.2018.11328

103. Danish registers on personal income and transfer payments - Mikkel Baadsgaard, Jarl Quitzau, 2011. Accessed January 30, 2022. https://journals.sagepub.com/doi/abs/10.1177/1403494811405098

104. Wilson JR, Lorenz KA. Generalized Estimating Equations Logistic Regression. In: Wilson JR, Lorenz KA, eds. *Modeling Binary Correlated Responses Using SAS, SPSS and R*. ICSA Book Series in Statistics. Springer International Publishing; 2015:103-130. doi:10.1007/978-3-319-23805-0_6

105. Ong KKL, Preece MA, Emmett PM, Ahmed ML, Dunger DB. Size at Birth and Early Childhood Growth in Relation to Maternal Smoking, Parity and Infant Breast-Feeding: Longitudinal Birth Cohort Study and Analysis. *Pediatr Res*. 2002;52(6):863-867. doi:10.1203/00006450-200212000-00009

106. Vanderweele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*. 2014;25(2):300-306. doi:10.1097/EDE.0000000000000034

107. Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*. 2000;11(5):550-560.

108. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*. 2009;20(1):18-26. doi:10.1097/EDE.0b013e31818f69ce

109. VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. *Annual Review of Public Health*. 2016;37(1):17-32. doi:10.1146/annurev-publhealth-032315-021402

110. Smith LH, VanderWeele TJ. Mediational E-values: Approximate sensitivity analysis for unmeasured mediator–outcome confounding. *Epidemiology*. 2019;30(6):835-837. doi:10.1097/EDE.0000000000001064

111. Geyer S, Hemström Ö, Peter R, Vågerö D. Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. *Journal of Epidemiology & Community Health*. 2006;60(9):804-810. doi:10.1136/jech.2005.041319

112. Nielsen MW, Hansen EH, Rasmussen NK. Prescription and non-prescription medicine use in Denmark: association with socio-economic position. *Eur J Clin Pharmacol*. 2003;59(8):677-684. doi:10.1007/s00228-003-0678-z

113. Andrés AR, Collings S, Qin P. Sex-specific impact of socio-economic factors on suicide risk: a population-based case–control study in Denmark. *European Journal of Public Health*. 2010;20(3):265-270. doi:10.1093/eurpub/ckp183

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

115. Sankar MJ, Sinha B, Chowdhury R, et al. Optimal breastfeeding practices and infant and child mortality: a systematic review and meta-analysis. *Acta Paediatrica*. 2015;104(S467):3-13. doi:10.1111/apa.13147

116. Tseng PT, Chen YW, Stubbs B, et al. Maternal breastfeeding and autism spectrum disorder in children: A systematic review and meta-analysis. *Nutritional Neuroscience*. 2019;22(5):354-362. doi:10.1080/1028415X.2017.1388598

117. Xue F, Colditz GA, Willett WC, Rosner BA, Michels KB. Parental age at delivery and incidence of breast cancer: a prospective cohort study. *Breast Cancer Res Treat*. 2007;104(3):331-340. doi:10.1007/s10549-006-9424-4

118. Michels KB, Willett WC, Graubard BI, et al. A longitudinal study of infant feeding and obesity throughout life course. *International Journal of Obesity*. 2007;31(7):1078-1085. doi:10.1038/sj.ijo.0803622

119. James-Todd TM, Karumanchi SA, Hibert EL, et al. Gestational age, infant birth weight, and subsequent risk of type 2 diabetes in mothers: Nurses' Health Study II. *Prev Chronic Dis*. 2013;10:E156. doi:10.5888/pcd10.120336

120. Roberts AL, Lyall K, Hart JE, et al. Perinatal Air Pollutant Exposures and Autism Spectrum Disorder in the Children of Nurses' Health Study II Participants. *Environmental Health Perspectives*. 2013;121(8):978-984. doi:10.1289/ehp.1206187

121. Gao X, Lyall K, Palacios N, Walters AS, Ascherio A. RLS in middle aged women and attention deficit/hyperactivity disorder in their offspring. *Sleep Medicine*. 2011;12(1):89-91. doi:10.1016/j.sleep.2010.05.006

122. McCrory C, Murray A. The Effect of Breastfeeding on Neuro-Development in Infancy. *Matern Child Health J*. 2013;17(9):1680-1688. doi:10.1007/s10995-012-1182-9

123. Tomeo CA, Rich-Edwards JW, Michels KB, et al. Reproducibility and validity of maternal recall of pregnancy-related events. *Epidemiology*. 1999;10(6):774-777.

124. Faraone SV, Biederman J, Milberger S. How reliable are maternal reports of their children's psychopathology? One-year recall of psychiatric diagnoses of ADHD children. *J Am Acad Child Adolesc Psychiatry*. 1995;34(8):1001-1008. doi:10.1097/00004583-199508000-00009

125. Iversen DS, Kesmodel US, Ovesen PG. Associations between parity and maternal BMI in a population-based cohort study. *Acta Obstetricia et Gynecologica Scandinavica*. 2018;97(6):694-700. doi:10.1111/aogs.13321

126. The Danish Medical Birth Registry. - Abstract - Europe PMC. Accessed June 8, 2020. https://europepmc.org/article/med/9675544

127. Bliddal M, Broe A, Pottegård A, Olsen J, Langhoff-Roos J. The Danish Medical Birth Register. *Eur J Epidemiol*. 2018;33(1):27-36. doi:10.1007/s10654-018-0356-1

128. Fertility. Accessed February 20, 2021. https://www.dst.dk/en/Statistik/emner/befolkning-og-valg/foedsler/fertilitet

129. Olesen AW, Westergaard JG, Thomsen SG, Olsen Jø. Correlation between self-reported gestational age and ultrasound measurements. *Acta Obstetricia et Gynecologica Scandinavica*. 2004;83(11):1039-1043. doi:10.1111/j.0001-6349.2004.00193.x

130. Who: Recommended Definitions, Terminology and Format for Statistical Tables Related to The Perinatal Period And Use of A New Certificate For Cause of Perinatal Deaths. *Acta Obstetricia et Gynecologica Scandinavica*. 1977;56(3):247-253. doi:10.3109/00016347709162009

131. Validation of maternal reported pregnancy and birth characteristics against the Medical Birth Registry of Norway. Accessed June 15, 2020. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0181794

132. Vallgårda S. Why did the stillbirth rate decline in Denmark after 1940? *Population Studies*. 2010;64(2):117-130.

133. Green A, Sortsø C, Jensen PB, Emneus M. Validation of the Danish National Diabetes Register. *Clin Epidemiol*. 2014;7:5-15. doi:10.2147/CLEP.S72768

134. Nørgård B, Fonager K, Sørensen HT, Olsen J. Birth outcomes of women with ulcerative colitis: a nationwide Danish cohort study. *The American Journal of Gastroenterology*. 2000;95(11):3165-3170. doi:10.1016/S0002-9270(00)02083-9

135. Maimburg RD, Bech BH, Væth M, Møller-Madsen B, Olsen J. Neonatal Jaundice, Autism, and Other Disorders of Psychological Development. *Pediatrics*. 2010;126(5):872-878. doi:10.1542/peds.2010-0052