

BMJ Open The Copenhagen Primary Care Laboratory Pregnancy (CopPreg) database

Janet Janbek,^{1,2} Margit Kriegbaum,² Mia Klinten Grand,^{2,3} Ina Olmer Specht,⁴ Bent Struer Lind,⁵ Christen Lykkegaard Andersen,^{2,6} Berit Lilienthal Heitmann ^{2,4}

To cite: Janbek J, Kriegbaum M, Grand MK, *et al.* The Copenhagen Primary Care Laboratory Pregnancy (CopPreg) database. *BMJ Open* 2020;**10**:e034318. doi:10.1136/bmjopen-2019-034318

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-034318>).

CLA and BLH shared the last authorship.

CLA and BLH are joint senior authors.

Received 14 September 2019
Revised 27 November 2019
Accepted 02 March 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Berit Lilienthal Heitmann;
Berit.Lilienthal.Heitmann@regionh.dk

ABSTRACT

Purpose The Copenhagen Primary Care Laboratory Pregnancy (CopPreg) database was established based on data from The Danish Medical Birth Register and the Copenhagen Primary Care Laboratory (CopLab) database. The aim was to provide a biomedical and epidemiological data resource for research in early disease programming (eg, parental clinical biomarker levels and pregnancy/ birth outcomes or long-term health in the offspring).

Participants The cohort consisted in total of 203 608 women (with 340 891 pregnancies) who gave birth to 348 248 children and with 200 590 related fathers. In this paper, we focused on women and fathers who had clinical test requisitions prior to and during pregnancy, and on all children. Thus, the cohort in focus consisted of 203 054 pregnancies with requisitions on 147 045 pregnant women, 39 815 fathers with requisitions during periconception and 65 315 children with requisitions.

Findings to date In addition to information on pregnancy and birth health status and general socio-demographic data, over 2.2 million clinically relevant test results were available for pregnancies with requisitions, over 1.5 million for children and over 600 000 test results were available for the fathers with requisitions during periconception. These were ordered by general practitioners in the primary care setting only and included general blood tests, nutritional biomarkers (macronutrients and micronutrients) and hormone tests. Information on tests related to infections, allergies, heart and lung function and sperm analyses (fathers) were also available.

Future plans The CopPreg database provides ready to use and valid data from already collected, objectively measured and analysed clinical tests. With several research projects planned, we further invite national and international researchers to use this vast data resource. In a coming paper, we will explore and discuss the indication bias in our cohort.

INTRODUCTION

According to the *Developmental Origins of adult Health and Disease* theory,¹ maternal health during pregnancy contributes to the shaping of the offspring's mental and physical health.² Therefore, a better understanding of environmental influences in the prenatal period is crucial for potential early intervention and

Strengths and limitations of this study

- Unique data resource with millions of biomarker sample results available on pregnant women, fathers and children, collected over 15 years that are analysed, stored and cleaned.
- Provides valid and objective clinical test results as well as potential for individual level linkage to numerous Danish national health registers allowing for longitudinal research at low cost.
- Indication bias of the ordered clinical tests is a limitation that is currently being explored by biostatisticians.
- Clinical test results of pregnant women referred to hospitals, women not residing in the Copenhagen area during their pregnancy or tests performed directly at the general practitioner's practice are not included in our database.

for ensuring long-term physical and mental well-being.

Several studies have examined the influence of maternal health and lifestyle (eg, nutrition,^{3–5} body mass index,⁶ hormone and immune markers^{7 8} and infections^{9 10}) for offspring health. However, many of these studies—particularly those investigating influences of maternal nutrition—are with conflicting results.^{3–5} In addition, interpretation of such results from these previous studies has often been hampered due to small sample sizes, use of self-reported exposure and outcome measures and residual—or insufficient adjustment for—confounding.

A few studies suggest that paternal lifestyle factors and nutrition around conception may also play a role in shaping offspring health^{10–15} through incurring epigenetic changes in the offspring¹² that may influence general health,^{11 14} cognitive and neural functions¹⁵ or physical growth.¹⁶ However, results from the studies conducted so far, have also been inconsistent.

Thus, there remains a need for further research of the importance of parental

health and lifestyle in the months prior to and during pregnancy for offspring health and disease risk, and for data resources that are sufficiently extensive to offer the opportunity to examine prenatal risk factors for offspring health outcomes.¹⁷

Data from the Copenhagen Primary Care Laboratory Pregnancy (CopPreg) database is such a data resource. Although other laboratory data resources exist (eg, University of Utah, Elferink-Stinkens *et al* and THL),^{18–20} the CopPreg data resource is unique due to its sample sizes; availability of valid and already analysed clinical biomarker data; availability of such data on the pregnancies as well as the fathers and due to its potential for longitudinal follow-up of all of its population individuals for investigation of long-term health outcomes.

The Copenhagen Primary Care Laboratory Pregnancy data resource

Every individual residing in Denmark is assigned a unique personal identification number, the so-called CPR number, which is used in all national registers, enabling accurate linkage at the individual level between the registers. The CopPreg was established in 2017, by merging the clinical test results from the Copenhagen Primary Care Laboratory (CopLab) database to the Danish National Medical Birth Register (MBR)²¹ to form the CopPreg database. Recently, also using CPR linkage, data from other readily available Danish National Pregnancy, Birth, Disease and Death registries have been added to the CopPreg data resource. The CopPreg now provides a unique resource for research purposes as cost-effective analyses of the long-term influences from maternal and paternal clinical biomarker levels measured in the months around conception and during pregnancy on disease development in the offspring can be performed. The purpose of this article is to present the new CopPreg data resource, its history, population and content, in order to stimulate and encourage future research collaboration and use.

COHORT DESCRIPTION

History

The CopLab database

In the Copenhagen area (the Copenhagen Municipality and the former Copenhagen County) one laboratory only, the Copenhagen General Practitioners' Laboratory (CGPL), served all general practitioners (GPs) (n=750) and other private specialists (n=300) with clinical tests during the period from years 2000 through 2015. The CGPL was established in 1922 and was accredited by DANAK (Danish Accreditation Fund) according to International Organization for Standardization 17025 and 15189.²² Because of the accreditation, the quality of the tests performed at the laboratory, all changes in analytical methods and the history of the recommendations to the doctors from the laboratory are traceable and documented. The laboratory served the practitioners with a

broad range of blood, urine, semen, clinical physiological, cardiac-function and lung-function tests. When the CGPL was closed at the end of 2015, all clinical test results (n=176 million) were permanently saved in the CopLab database from 1.4 million different individuals. The total number of requisitions in the database is approximately 10.5 million and the mean number of requisitions per unique individual is 7.5. Part of the CopLab database, known as CopDiff, has been described in details elsewhere.²³

The Medical Birth Register

The MBR, established in 1973, contains continuously collected data on all live and stillbirths by women residing and giving birth in Denmark (until April 2004, the birth of a dead fetus before gestational week 28 was considered a miscarriage and therefore not registered in the MBR. Thereafter, delivery of a dead fetus after 22 completed weeks of pregnancy was considered a stillbirth).²¹ These data include pregnancy health information, maternal characteristics, lifestyle (such as smoking during pregnancy), health status (including pre-pregnancy BMI), delivery/birth outcomes and characteristics of the newborn.²¹ Several Danish national registers feed data into the MBR such as the National Patient Register (NPR) which adds specific variables on pregnancy and delivery²⁴ as well as the Danish Civil Registration System (CRS) which contributes with demographic information.²⁵ The MBR also registers fathers related to all births and provides information on the fathers' age (online supplementary table 1 presents an overview of the variables in MBR). Both parents of each registered birth in the MBR are assumed to be biological parents, but not necessarily also the registered legal parents (as registered in the CRS). The MBR has been described in details elsewhere.²¹

Specific biomarkers and examination for pregnant women in Denmark

In Denmark, according to the Danish Health Authorities in 1998 and 2013,^{26 27} each pregnant woman should have at least three antenatal consultation visits with her GP (during gestation weeks six to 10, and at week 25 and 32), two ultrasound scans (during gestation weeks 11+0 to 13+6 and at week 18) and four to seven antenatal consultations with the midwife. Thus, the GP is the pregnant woman's first contact. Here she is informed about the pregnancy course, introduced to the offered antenatal care and guided on the recommendations regarding vitamin supplementation, medicine use and lifestyle. The GP fills in the pregnancy journal and subsequently assigns a maternity care level score (one to four, where four is the highest-risk pregnancy) based on her medical history and lifestyle.²⁷ Additionally, the pregnant woman's general health is assessed by blood pressure measurement; urinary examination for asymptomatic bacteriuria, glucosuria and proteinuria; blood type and irregular blood type antibodies; screening for hepatitis B, HIV and syphilis; and body weighing and height measuring as a

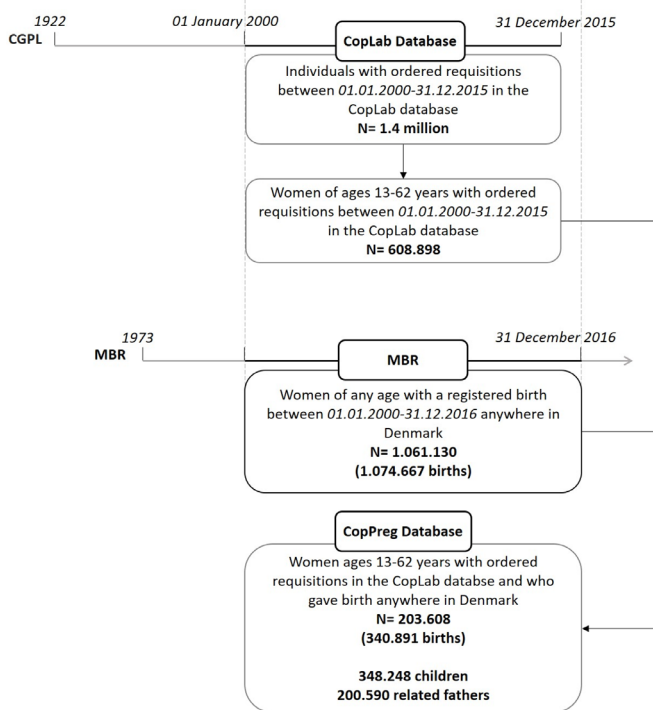


Figure 1 CopPreg database establishment process and data flow. CGPL, Copenhagen General Practitioners' Laboratory; CopLab, Copenhagen Primary Care Laboratory database; CopPreg, Copenhagen Primary Care Laboratory Pregnancy database; MBR, Medical Birth Register.

basis for advice on pregnancy weight gain, diet and physical activity. Women with high-risk pregnancies (maternity care levels scores two, three and four) are offered additional visits to the GP that include blood tests for chlamydia, gonorrhoea, haemoglobin, mean corpuscular volume, serum iron, vitamin D status, screening for gestational diabetes and a gynaecological examination.²⁷ They are also offered additional visits to the midwife or obstetrician. Other examples of further assessments offered to women with high-risk pregnancies include lifestyle interventions such as smoking cessation and weight management. Conversations regarding previous pregnancies where, for example, anxiety was present are also offered

to women of these pregnancies. All clinical tests ordered by the GP and the results of the tests analysed at CGPL were recorded in the CopPreg database.

Using the CPR number of all women aged 13 to 62 years (chosen a priori, based on information from Statistics Denmark on the ages of the youngest and eldest woman to give birth in Denmark during 2000 to 2016²⁸) in the CopLab database (n=608 898), a merge was performed with the MBR on the condition that the women had any registered birth between 1 January 2000 and 31 December 2016 (total registered births in the MBR during this period anywhere in Denmark was 1 074 667 births to 1 061 130 women of which the 608 898 women were identified in CopLab). As a result, there were 203 608 women in MBR with their registered births (n=340 891) which formed the CopPreg database. A total of 348 248 children were born to these women with 200 590 related fathers. **Figure 1** presents a graphical presentation of the establishment of the CopPreg database. As shown in the figure, although the CopLab database covers individuals with requisitions ordered between 2000 to 2015, we included births from the MBR from 2016 as well. This is to guarantee completeness of information on all women in the CopPreg database, because some women who were pregnant and had requisitions ordered in 2015 would have given birth in 2016.

Other Danish national registers

Detailed description of Danish national registers can be found in other papers.^{24 25 29-33} Examples include the CRS from which socio-demographic information can be obtained;²⁵ the NPR with information on hospitalisations and diagnoses of all somatic and psychiatric inpatients and outpatients;²⁴ the Danish National Prescription Register with information on all dispensed prescription medications;³⁰ the National Health Service Register for primary care with data on all GP consultations in primary care and referrals to other primary care treatments such as physiotherapy³¹ and the Student Register with grade-level information on schooling.³³ Data from these national registers are now linked to the CopPreg database

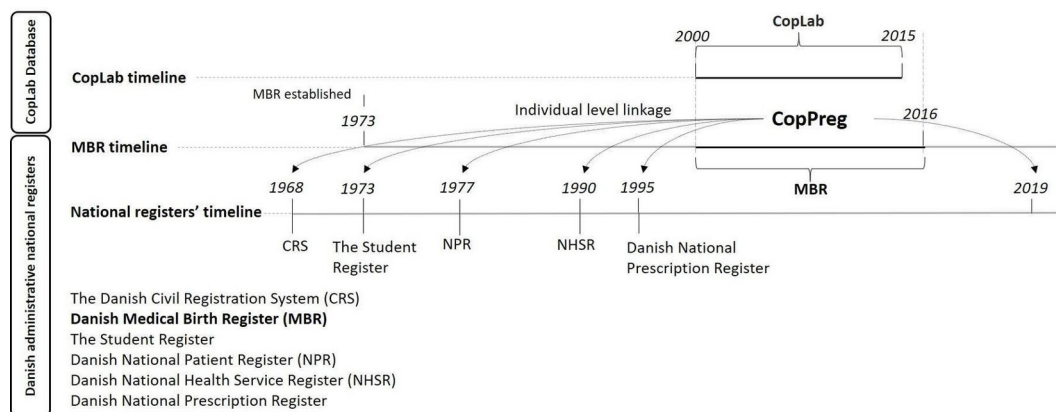


Figure 2 Timeline of national registers and CopPreg database. CopLab, Copenhagen Primary Care Laboratory database; CopPreg, Copenhagen Primary Care Laboratory Pregnancy database.

and provide information for each individual both when followed prospectively and retrospectively. [Figure 2](#) presents the linkage, timelines and data resources in relation to the CopPreg database.

Additionally, nationwide clinical databases are also available. Some of these include data concerned with specific diseases or conditions and contain supplementary diagnostic information, treatment types, rehabilitation measures and clinical outcomes. They also include data from interventions such as those for acute stroke and diabetes and with activities such as data on all transfusions and all anaesthetic procedures.³⁴ The Children's Health Database, is a non-clinical database, established by the Danish health nurse system from their visits to the pregnant women and to the families after birth. The database collects data from several municipalities in the greater Copenhagen area that chose to contribute to the database. From 2002 and onwards, the database included information on child's health status during the first years of life (birth to one year) based on four visits from nurses. The database contains information on several neurodevelopmental outcomes in the child, maternal and paternal mental well-being and child care-taking practices including breastfeeding.^{35 36} From 2007, the database further included physical and neurodevelopmental information on children at school entry and exit.

Ethical approval

Generally, legal and ethical considerations when conducting registry-based research in Denmark include requirements for data protection and privacy as a result of, amongst others, the Danish Act on Processing of Personal Data.³⁷ The CopPreg database is covered by a shared data protection agreement from The Danish Data Protection Agency,³⁴ therefore no ethical approval was required.

One important consideration when using this database (as well as all national registers) is that individuals cannot be identified personally, and will not be contacted at any point to obtain supplementary information by, for example, questionnaires. Thus, when conducting statistical analyses, results can only be reported at an aggregated level where no single person is identified.

Patient and public involvement

There was no patient or public involvement in the design or analysis of this study.

Findings to date

Population

Three population groups are in focus in this paper: The pregnant woman, the father and the offspring. Whether or not the pregnant woman and the father are included in the database depends on whether they had a requisition ordered in the six months prior to and three months after conception (father); or three months prior to conception and during the entire pregnancy (pregnant woman). These limits were chosen based on results from previous literature related to a potential for a programming influence of maternal and paternal health, biomarker levels and lifestyle on offspring future health.¹⁰ All live-born offspring to the included mothers were also included in the database. [Figure 3](#) presents the three population groups (A, B and C) that were defined as follows:

Pregnancies with requisitions (Group A in [figure 3](#)). Pregnancies of women who had at least one requisition ordered at any time during the period of three months preconception or the three pregnancy trimesters and until delivery. Gestational age was used to calculate the conception date and hence the pregnancy period for each woman. For women with a missing gestational age

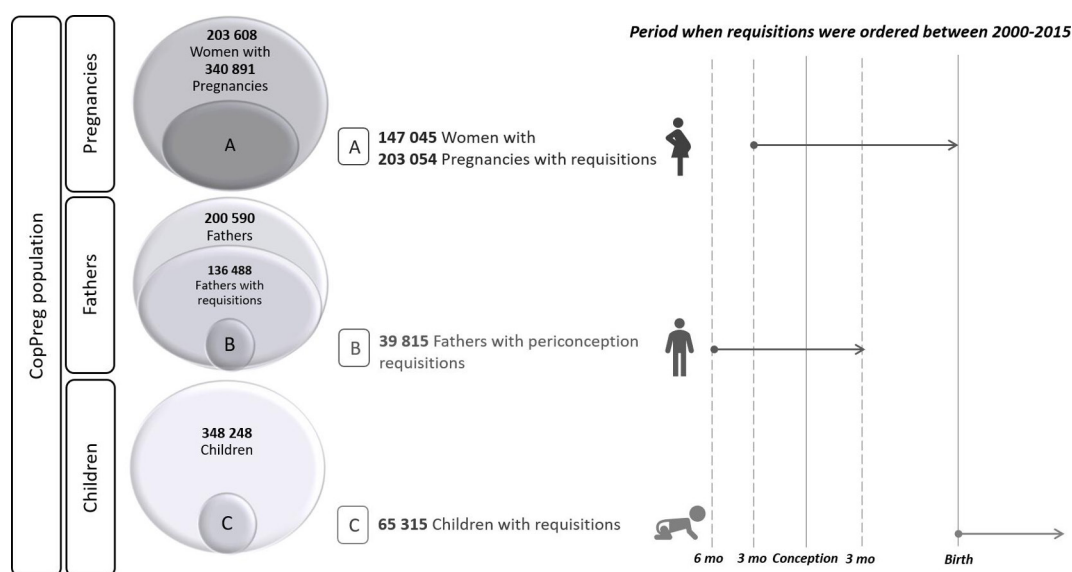


Figure 3 The CopPreg database population; presenting the three population groups in focus (A, B and C). *Gestational age was used to calculate the conception date and hence the pregnancy period for each woman. For women with a missing gestational age (n=6662 of all pregnancies), it was not possible to see whether requisitions were during the pregnancy or not and these pregnancies were excluded. CopPreg, Copenhagen Primary CareLaboratory Pregnancy database; mo, months.

(n=6662 of all pregnancies), it was not possible to see whether requisitions were during the pregnancy or not and these pregnancies were excluded. In total, 203 608 women had requisitions in the CopPreg database at any time, and 147 045 women (72%) had pregnancies with requisitions (in total 203 054 pregnancies, average number of requisitions per pregnancy=1.8). Of these, 99 141 women (67%) had one pregnancy with requisitions in the database, 28% had two pregnancies with requisitions and 5% of women had three or more (maximum eight) pregnancies with requisitions. Some of the children of these pregnancies and some of the related fathers also had requisitions ordered in the database. These children (n=42 492) and the related fathers (n=105 447, with requisitions at any time during a father's life with no focus here on the periconception period) are described in online supplementary tables 2-5 and will not be further described here.

The remaining women (56 563 women, 28% of all identified women in the database) did not have any requisitions ordered around the defined pregnancy period, but had requisitions ordered outside this period and between 2000 to 2015 (eg, women with requisitions ordered two, three or 10 years before or in the post-partum period, after the defined pregnancy period). These will not be further described here. Nevertheless, requisitions and test results taken outside of the defined pregnancy period are also of value in research, for example, clinical test results from two, four or six years prior to or post a pregnancy of interest are also of value as disease determinants. Data during such time periods are also available and traceable for the entire CopPreg population, although not focussed on in this paper.

Fathers with requisitions during periconception (Group B in figure 3). Fathers who had at least one requisition ordered at any time during the period of six months prior to conception and three months after conception of the related pregnancy. Of the total of 136 488 fathers who had requisitions in the CopPreg database at any time, 39 815 (29%) fathers had requisitions during periconception (average number of requisitions per father=1.6).

Children with requisitions (Group C in figure 3). Children/newborns who had at least one requisition from the time of birth until end of 2015. A total of 65 315 children had requisitions (average number of requisitions per child=1.8).

The pregnant women, fathers and children from the three population groups in focus are not necessarily all related, for example, *children with requisitions* were not all from *pregnancies with a requisition* or from *fathers with requisitions during periconception*. Similarly, *fathers with requisitions during periconception* were not all related to the same *pregnancies with requisitions*. This is largely dependent on the clinical test type, where some tests are frequently requested with tens-of-thousands of results available whereas others with a few thousands or less.

Table 1 Population characteristics; mothers/pregnancies

Mother/pregnancy characteristic		Pregnancies with requisitions (n=203 054)	
		Total	
<i>Births</i> N (% of total)	<i>Total</i>	207 509	–
	<i>Live-born</i>	206 787	99.7
	<i>Stillborn*</i>	722	0.3
<i>Age in years</i> N (%)	<i>Missing</i>	None	0.0
	<20	1694	0.8
	20–45	201 127	99.1
	>45	233	0.1
<i>Multiple pregnancies</i> N (%)	<i>Missing</i>	None	0.0
	<i>Singleton</i>	198 694	97.9
	<i>Twin</i>	4303	2.1
	<i>Triplet</i>	57	0.0
<i>Smoking status</i> N (%)	<i>Missing</i>	2937	1.5
	<i>Did not smoke</i>	178 154	87.7
	<i>Quit during pregnancy</i>	5200	2.6
	<i>Smoked</i>	16 763	8.3
<i>BMI (kg/m²) at start of pregnancy (calculated)†</i> N (%)	<i>Missing</i>	45 775	22.5
	<i>Underweight</i>	8176	4.0
	<i>Normal weight</i>	109 941	54.1
	<i>Overweight</i>	39 162	19.3
<i>Parity</i> N (%)	<i>Missing</i>	2860	1.4
	1	105 753	52.1
	2	66 178	32.6
	3 or more	28 263	13.9
<i>Midwife visits</i> N of visits (%)	<i>Missing</i>	10 659	5.2
	<i>Never</i>	38 251	18.8
	1–10	153 697	75.7
	>10	447	0.2
<i>Physician visits</i> N of visits (%)	<i>Missing</i>	12 196	6.0
	<i>Never</i>	38 456	18.9
	1–10	152 402	75.1
	>10	–	–
<i>Specialist visits</i> N of visits (%)	<i>Missing</i>	17 784	8.8
	<i>Never</i>	84 163	41.4
	1–10	101 107	49.8
	>10	–	–

The unit of count is pregnancies, thus one woman can be counted in different categories of each characteristic if she gave birth to more than one child during the period of 2000 to 2016.

Underweight defined as <18.5 kg/m²; overweight defined as ≥25 kg/m².

*Until April 2004, the birth of a dead fetus before gestational week 28+0 was considered a miscarriage and therefore not registered in the MBR. From 1st April 2004: delivery of a dead fetus after 22nd completed week of pregnancy was considered stillborn.²¹

†BMI was systematically registered in the MBR starting late 2003 to 2004.²¹ Here we have calculated BMI from maternal weight and height; we excluded women with extreme values, for example, BMI>60 kg/m², BMI<10 kg/m², body weight >200 kgs or height >200 cm, and these women were counted as missing BMI in the table. (n=793).

BMI, body mass index; MBR, Medical Birth Register.

Pregnancy and birth information

All pregnancy and birth information including maternal and newborn's anthropometric information was obtained

from the MBR.

Pregnancies with requisitions. Maternal characteristics and health information are summarised in table 1. The maternal mean age of *pregnancies with requisitions* at delivery was 30 years and 99% of pregnancies were of women who gave birth when they were aged 20 to 45 years. Most women reported that they did not smoke during pregnancy, were measured by the GP to be of normal weight, previously had a total of one to two children (in MBR, parity was counted at the index pregnancy in the cohort thus all women had at least one child), had singleton births and had had one to 10 antenatal visits to the midwife, a physician or a specialist during each of their pregnancies.

Children with requisitions. Child characteristics and birth information are summarised in table 2. Children were equally distributed in relation to sex with a slightly higher percentage of boys. Most of the children were full-term (>37 gestation weeks) and of normal birth weight (>2500 g). Using these two variables, together with infant sex, weight for gestational age can be calculated which is an important variable that can provide a better description of the infant and can be readily calculated using the available data in our database.

Fathers with requisitions during periconception. Table 3 shows information on fathers' age calculated at the time of childbirth. The mean age was 35 years and most fathers with requisitions during periconception (94.7%) were aged 20 to 45 years at the time of childbirth.

Clinical tests

We present here an overview of the clinical test results in the three populations. Table 4 shows the number of pregnancies, children and fathers with a test result and the total number of test results for each of these groups. As mentioned before, each ordered requisition might contain several tests. Some of the requisitions contained clinically irrelevant tests (technical and administrative help-tests) or clinically irrelevant test results (eg, results

Table 2 Population characteristics; children

Child characteristic		Children with requisitions (n=65 315)	
Sex N (%)	Missing	None	0.0
	Male	35 055	53.7
	Female	30 260	46.3
Birth weight (g) N (%)	Missing	1056	1.6
	Low birth weight*	3196	4.9
	Normal birth weight	61 063	93.5
Gestational age N (%)	Missing	916	1.4
	Preterm born†	4813	7.4
	Term born	59 586	91.2

*Low birth weight defined as <2500 g.

†Preterm born defined as <260 gestation days.

Table 3 Population characteristics; fathers

Father characteristic		Fathers with requisitions during periconception (n=39 815 related to 43 965 pregnancies)	
Age in years N (%)	Missing	None	0.0
	<20	50	0.1
	20–45	41 630	94.7
	>45	2285	5.2

The unit of count is pregnancies, and the age is the father's age at birth during each pregnancy where he was the registered father, thus one man can be counted in different age groups if he had more than one child during the period from 2000 to 2016.

reported as '+', '-', '!' or abbreviations for subcontractors that performed the analyses). We excluded all irrelevant tests and test results in CopLab (n=50 521 887 test results, 29% of all test results in the CopLab database) and defined a clinically relevant test result as one that is a numeric value or had a relevant alphanumeric format ('<' or '>'). As a result, around 12 000 *pregnancies with requisitions*, around 4000 *children with requisitions* and around 7000 *fathers with requisitions during periconception* were excluded, as they did not have any clinically relevant tests or test results, thus decreasing the numbers as shown in table 4. Over 2.2 million clinically relevant test results were available for *pregnancies with requisitions*, over 1.5 million test results were available for *children with requisitions* and over 600 000 test results were available for the *fathers with requisitions during periconception*.

The database contains information on reference limits, and all numeric and alphanumeric (but valid) tests results can be categorised according to the limits at the time of the analysis as either normal, below or above reference range. Furthermore, most of the numeric test results for each component are comparable over time (2000 to 2015) or can be corrected via a traceable documentable factorisation based on experimental results obtained as part of the CGPL's validation of new test assays.

The clinical tests included general blood tests, several nutritional biomarkers (macronutrients and micronutrients) and hormone tests in addition to numerous tests related to infections (antibodies), allergies and heart and lung function. Furthermore, the database contains requisitions on sperm analyses for the fathers and includes

Table 4 Clinical test results and numbers in population

Population group	N with test results	N of test results
<i>Pregnancies with requisitions</i>	191 413	2 244 397
<i>Children with requisitions</i>	61 733	1 511 597
<i>Fathers with requisitions during periconception</i>	33 074	639 394

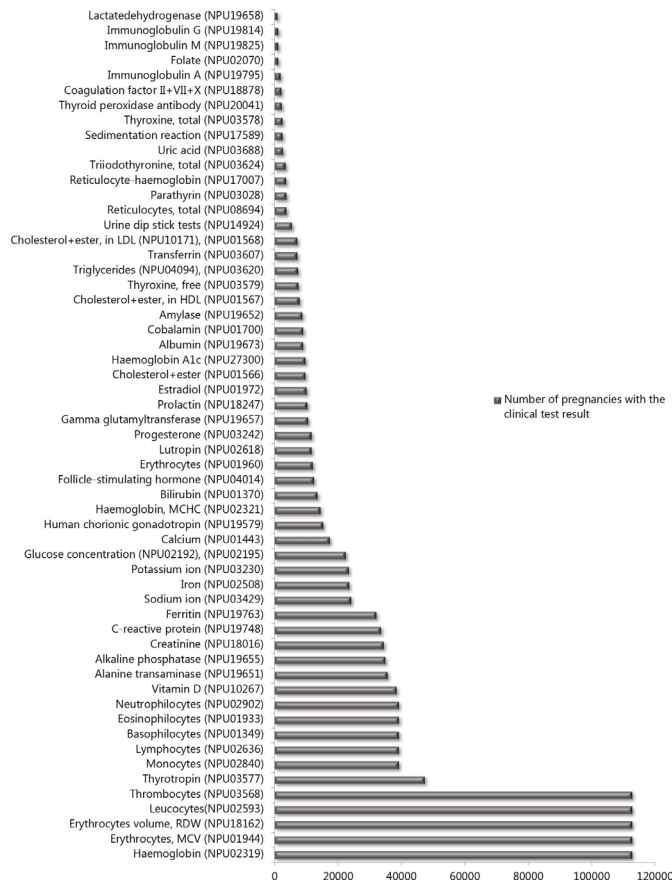


Figure 4 Overview of the type of clinical tests available in pregnancies with requisitions. For tests of glucose concentration, triglycerides and LDL-cholesterol, more than one test was available, depending on the fasting status of the patient. The percentages of test results from fasting patients in pregnancies was 39% (glucose), 73% (triglyceride) and 73% (LDL-cholesterol). In the figure, number of pregnancies presented was counted if a pregnancy had any of the tests for each of the three tests types. For urine dip stick tests, six different tests were available. In the figure, number of pregnancies and women presented were counted if a pregnancy/woman had test results for any of these six tests. Detailed numbers of each of these tests are presented in the online supplementary table 6. HDL, high density lipoprotein; IgA, immunoglobulin A; IgG, immunoglobulin G; LDL, low density lipoprotein; MCHC, mean corpuscular haemoglobin concentration; MCV, mean corpuscular volume; RDW, relative distribution width.

variables such as semen volume, clarity and motility. The results of heart and lung function tests as well as tests related to infections, allergies and sperm analyses are not presented in this paper because of the large variation in the type of these tests and the heterogeneity of the result format. We present in this paper only clinical tests where more than 1000 pregnancies, fathers or children had the test result available. Figures 4–6 show an overview of these tests in each of the population groups. Detailed numbers of these tests are presented in the online supplementary tables 6–8.

The five most common tests analysed in the groups of pregnancies, children and fathers (table 4) are those of



Figure 5 Overview of the type of clinical tests available in children with requisitions. For tests of glucose concentration, triglycerides and LDL-cholesterol, more than one test was available, depending on the fasting status of the patient. The percentages of test results from fasting patients were 29% (glucose), 62% (triglyceride) and 65% (LDL-cholesterol). In the figure, number of children presented was counted if a child had any of the tests for each of the three tests types. For urine dip stick tests, six different tests were available. In the figure, number of children presented was counted if a child had test results for any of these six tests. Detailed numbers of each of these tests are presented in the online supplementary table 7. HDL, high density lipoprotein; IgA, immunoglobulin A; IgG, immunoglobulin G; LDL, low density lipoprotein; MCHC, mean corpuscular haemoglobin concentration; MCV, mean corpuscular volume; RDW, relative distribution width.

haemoglobin, mean corpuscular volume, erythrocytes volume distribution width, leucocytes and thrombocytes. However, it is important to note that these tests were ordered on the same requisition where the individual test could not be requested alone. The next most common tests in pregnancies are those of thyrotropin, followed by equal numbers of pregnancies with tests of monocytes, lymphocytes, basophilocytes, eosinophilocytes, and neutrophilocytes. For each requisition of a clinical test, the exact date of requisition and clinical test result is available, and together with variables of gestation days for each pregnancy and child's birth date, gestational age when a

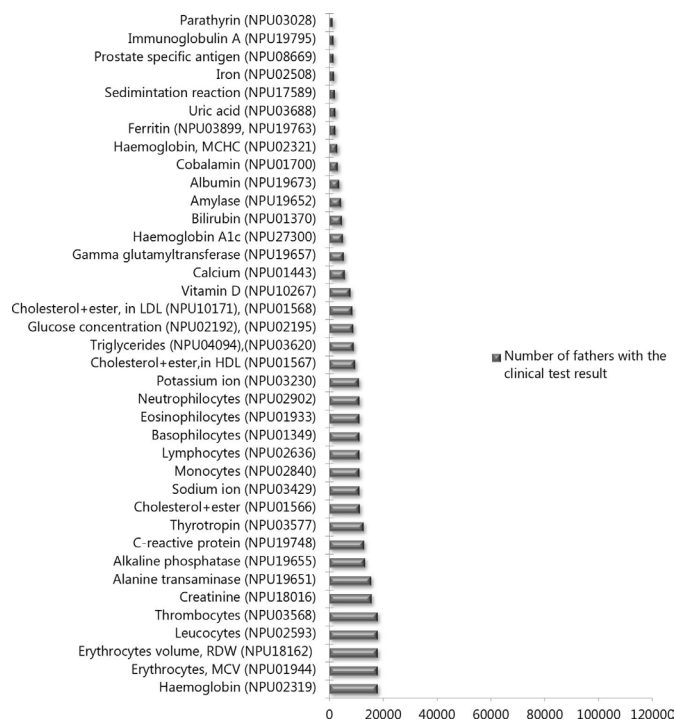


Figure 6 Overview of the type of clinical tests available in fathers with requisitions during periconception. For tests of glucose concentration, triglycerides and LDL-cholesterol, more than one test was available, depending on the fasting status of the patient. The percentages of test results from fasting patients were 54% (glucose), 75% (triglyceride) and 75% (LDL-cholesterol). In the figure, number of fathers presented was counted if a father had any of the tests for each of the three tests types. Detailed numbers of each of these tests are presented in the online supplementary table 8. HDL, high density lipoprotein; LDL, low density lipoprotein; MCHC, mean corpuscularhaemoglobin concentration; MCV, mean corpuscular volume; RDW, relative distribution width.

particular clinical test was analysed can be calculated. In children, the next most common tests are those of monocytes, lymphocytes, basophilocytes, eosinophilocytes, and neutrophilocytes followed by tests of C-reactive protein. Finally, in fathers, the next most common tests are those of creatinine and alanine transaminase, followed by tests of alkaline phosphatase and C-reactive protein.

Further, we looked at the distribution of children and the number of available test results per age group. When divided into two-year interval age groups, the biggest number of test results and children were taken/seen before the age of two years (figure 7).

FUTURE PLANS AND COLLABORATION

The CopPreg database provides an extensive clinical data resource for future research purposes. To our knowledge, CopPreg represents a unique and large data source of high-quality clinical data from pregnant women, fathers and the offspring—likely by far the largest in the world. The database can be used alone (where associations between parental clinical test results and newborn

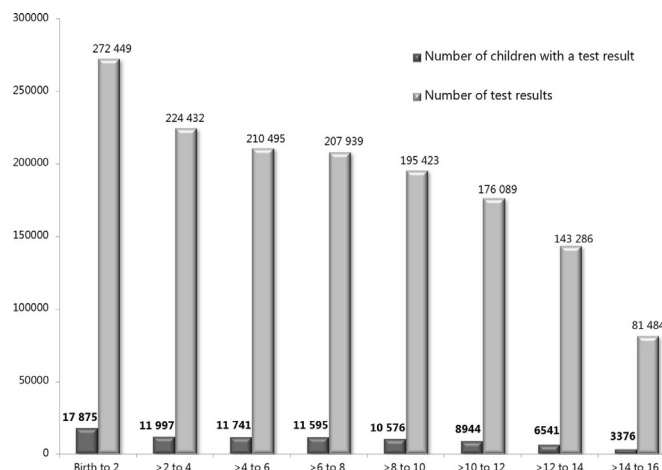


Figure 7 Distribution of children and clinical test results according to age groups. The figure is intended only for describing the distribution of test results in the different age groups of children. Children in the database are those born between 2000 and 2016 and thus, all children have had the chance to have clinical test results in the age group 0 to 2 years. This chance decreases with time and a child born in 2015, would only have the chance to have test results at birth or during the first year of life while a child born in 2000 would have the chance to have test results until age 16 years. Age groups were calculated based on 365 days in a year.

health/pregnancy health, are of interest) or through linkage to information from various other high quality Danish national registers like those described in the methods section.

The possibility to link data from the CopPreg database at the individual level to information from a variety of other national registers further allows for investigations of relations to numerous validated disease outcomes (eg, Green et al and Andersen *et al*),^{38 39} and for accounting for variation from socio-demographic and other needed covariables. Examples include: Research on the associations between both maternal and paternal biomarker levels around conception and offspring future health outcomes (asthma, neurodevelopment, obesity, psychiatric health, cancer and other diseases as well as death); research on the associations between maternal and/or paternal lifestyle and nutrition around conception and offspring future health and research on associations between maternal biomarker levels and pregnancy outcomes and complications.

All data from the CopPreg database are securely stored on safe servers at Statistics Denmark and the Research Unit for General Practice, Section for General Practice at the Department of Public Health of the University of Copenhagen. All data handling is done on a secure platform at Statistics Denmark, and no data is delivered to an external source. Researchers need to be from authorised institutions to get a remote access online where data sets can be constructed and remain stored at Statistics Denmark to ensure confidentiality and anonymity of all data subjects. Further information on ways to collaborate and on access to the data from the CopPreg database can

be obtained by contacting the corresponding author of this paper and by visiting the homepage for the CopLab database.⁴⁰

DISCUSSION, STRENGTHS AND LIMITATIONS

The CopPreg database is a new and large data resource for research purpose that provides a unique opportunity for performing cost-effective analyses, for instance those relating maternal and/or paternal clinical biomarker levels prior to and during pregnancy to pregnancy and birth outcomes or to long-term health and disease development in the offspring, from birth and at present until age around 19 years.

The database has several strengths. It provides a novel, extensive and valid clinical test data source with potential for testing numerous research questions at a low cost. The biomarker samples have already been collected, analysed, stored and cleaned and thus provide valid and objective clinical test results ready for research. Multiple biomarkers have been collected from the pregnant women and several of these have been taken more than once during pregnancy, as well as in the preconception and post-partum periods. The access to measured biomarker levels in fathers in the period around conception is another big strength that makes the CopPreg database unique, as few other data sources are available worldwide with paternal biomarker levels that can be analysed in relation to offspring biomarker and health status. Research using this database will thus contribute with new and important information for use by laboratories and authorities alike, to inform future national and global recommendations related to pregnancy as well as for supplementation policies and guidelines. This may provide grounds for implementing new or adjusting already existing antenatal care practices. Moreover, the very large number of clinical test results in the database allows for the examination of novel hypothesis-driven relationships using biomarkers or micronutrients of potential importance for early disease programming. Access to clinical test data from hundreds of thousands of individuals from over a period of 15 years, as well as the multiple measurements taken over time makes the data from the CopPreg database relevant in relation to both common and some rare disease outcomes and enables longitudinal research. Finally, as all the GPs in the greater Copenhagen area were served by one laboratory only, and the clinical test results were all from the same one accredited laboratory, the inter-laboratory differences in handling of samples, the laboratory recommendations and analytical performance are eliminated or strongly minimised.

The database also has some limitations. First, the clinical tests were ordered by the GPs for an unknown indication. Thus, we cannot assume that any of the clinical tests performed prior to or during pregnancy were routine ones. Our research team is currently working on exploring this matter in depth and finding needed solutions for how to specifically deal with indication bias.

Since the type of statistical approach will depend on each research question, the first step will be to consider a case study. The case study seeks to investigate the relation between a continuous exposure, which is not fully observed on the target population, and a time to event outcome. The exposure could be the result of a clinical test taken during pregnancy and the outcome could be the time when the child develops a certain disease. In this case the indication bias can be treated as a missing data problem for which there exists numerous methods to adjust for bias, such as inverse probability weighting and imputation. We anticipate that the structure of this case study will be common for many of the research questions relevant for CopPreg and so it will be a useful starting point.

A second limitation of the CopPreg database is that not all pregnancies had a requisition three months prior to conception or during pregnancy. Thus, the group of pregnant women included in CopPreg may not be representative of all pregnant women, and it is possible that the pregnant women without requisitions were high-risk pregnancies where women were referred to a hospital where clinical tests were performed. Results of such tests would thus not be included in the CopPreg database. It is technically possible however to link our data to hospital (inpatient and outpatient) laboratory data, but the challenge is often the lack of knowledge about the linked-to data. A Danish national laboratory database exists and contains clinical tests requested and analysed at hospitals. However, the laboratory only includes data from 2010 to 2015 and thus does not cover the whole CopPreg database period. Further, usage of data such as the laboratory database has several limitations. For example, several laboratories contribute with data, and the differences in test results compared with CopPreg data may be the result of different preanalytical, analytical and postanalytical methods and procedures used by the two laboratories rather than true biological differences. The limitations may not restrict the use of the results in patient treatment but may have undesirable consequences when used in research (comparability of test results over time, selection bias, etc). It is also possible that several women moved to or from the Copenhagen area during pregnancy, and thus may have clinical test results taken elsewhere, and therefore also not be in the CopPreg database. Some clinical test results may also be missing because some women had one or more tests performed directly at the GP's practice, and thus never requested by the laboratory. This procedure may be more prevalent in recent years where bench-top instruments that measure haemoglobin, blood glucose, lipids and inflammation markers have become more common in GP practices.⁴¹ These clinical test results would thus be missing in the CopPreg database.

Last, even though the large number of test results and the sample sizes of the database is an advantage, statistical significance may have limited clinical meaning when the sample size is large, and results from the database thus will have to be interpreted in a clinical relevance context.



Moreover, since multiple measurements are available for pregnancies, children and fathers, they are not necessarily independent, which is something that needs to be considered in the statistical analyses of research projects using the CopPreg data.

CONCLUSIONS

In this paper, we have introduced the new CopPreg research data resource that provides ready to use and valid data from already collected, objectively measured and analysed clinical test results from parents taken prior to and during pregnancy, and offspring health and disease development up to age 19 years. We have described and reviewed the information available in the CopPreg database and presented some perspectives for its utility in research. With its large population, millions of objective biomarker measures and extensive and valid information on several pregnancy, birth and delivery outcomes, together with health and disease outcomes from other valid Danish national registers, it represents a powerful biomedical and epidemiological database with extensive research perspectives. Despite the discussed limitations, the CopPreg database is a unique and highly valuable data resource for biomedical, primary care and public health research, and especially in relation to research focused on early disease programming and consequences. With several research projects already planned, we further invite and welcome other national and international researchers to use this vast data resource. Finally, in a coming paper, we will explore and discuss the generalisability of our population to the bigger Danish population of pregnancies in the Copenhagen area.

Author affiliations

¹Danish Dementia Research Centre, Department of Neurology, The Neuroscience Centre, Rigshospitalet, Copenhagen, Denmark

²Research Unit for General Practice and Section of General Practice, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

³Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

⁴Research Unit for Dietary Studies, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Frederiksberg, Denmark

⁵Department of Clinical Biochemistry, Copenhagen University Hospital, Hvidovre, Denmark

⁶Department of Hematology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

Acknowledgements We acknowledge Niels de Fine Olivarius for his major role in the establishment of the database (together with JJ, BLH and CLA). We also acknowledge Willy Karlslund for his data management role to start up the database. Finally, we acknowledge Volkert Siersma for his continuous input of statistical considerations related to the CopPreg database.

Contributors JJ, BLH and CLA were responsible for the establishment of the database by obtaining all necessary permissions from relevant data authorities and for conducting the merge of data. MK and JJ were responsible for the data management and providing descriptive statistics of the cohort. BSL was responsible for all clinical tests data content of the database. MKG contributed with statistical knowledge regarding biases in the database. JJ drafted the manuscript and MK, MKG, IOS, BSL, CLA and BLH together with JJ revised the manuscript with equal contributions to the intellectual content and manuscript format. All authors were responsible for the final version of the manuscript. CLA and BLH share last authorship and they have equally contributed to this work.

Funding The development of the CopLab database, from which the CopPreg database was derived, was supported by The Research Unit for General Practice and Section of General Practice, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. Furthermore, the establishment of the CopPreg database was supported by the Lundbeck and Lilly and Herbert Hansens Foundations. The foundations had no role in the design, conduction or reporting of the database.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data from the CopPreg database are stored securely and in an anonymised form at the Research Unit for General Practice, Section of General Practice at the Department of Public Health, University of Copenhagen. Data access and utility is detailed in the manuscript. Researchers are invited to contact the corresponding author upon interest to use data from the CopPreg database.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Berit Lilienthal Heitmann <http://orcid.org/0000-0002-6809-4504>

REFERENCES

- 1 Wadhwa PD, Buss C, Entringer S, *et al*. Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. *Semin Reprod Med* 2009;27:358–68.
- 2 Urrutia-Pereira M, Solé D. [Vitamin D deficiency in pregnancy and its impact on the fetus, the newborn and in childhood]. *Rev Paul Pediatr* 2015;33:104–13.
- 3 Veena SR, Gale CR, Krishnaveni GV, *et al*. Association between maternal nutritional status in pregnancy and offspring cognitive function during childhood and adolescence; a systematic review. *BMC Pregnancy Childbirth* 2016;16:220.
- 4 Iglesias L, Canals J, Arija V, *et al*. Effects of prenatal iron status on child neurodevelopment and behavior: a systematic review. *Crit Rev Food Sci Nutr* 2018;58:1–11.
- 5 Bleker LS, de Rooij SR, Roseboom TJ. Malnutrition and depression in pregnancy and associations with child behaviour and cognitive function: a review of recent evidence on unique and joint effects. *Can J Physiol Pharmacol* 2019;97:158–73.
- 6 Papachatz E, Dimitriou G, Dimitropoulos K, *et al*. Pre-pregnancy obesity: maternal, neonatal and childhood outcomes. *J Neonatal Perinatal Med* 2013;6:203–16.
- 7 Opazo MC, Haensgen H, Bohmwald K, *et al*. Imprinting of maternal thyroid hormones in the offspring. *Int Rev Immunol* 2017;36:240–55.
- 8 Schepanski S, Buss C, Hanganu-Opatz IL, *et al*. Prenatal Immune and Endocrine Modulators of Offspring's Brain Development and Cognitive Functions Later in Life. *Front Immunol* 2018;9:2186.
- 9 Simanek AM, Meier HCS. Association between prenatal exposure to maternal infection and offspring mood disorders: a review of the literature. *Curr Probl Pediatr Adolesc Health Care* 2015;45:325–64.
- 10 Murphy SK, Fineberg AM, Maxwell SD, *et al*. Maternal infection and stress during pregnancy and depressive symptoms in adolescent offspring. *Psychiatry Res* 2017;257:102–10.
- 11 Lambrot R, Xu C, Saint-Phar S, *et al*. Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. *Nat Commun* 2013;4:ncmms3889.
- 12 McPherson NO, Fullston T, Kang WX, *et al*. Paternal under-nutrition programs metabolic syndrome in offspring which can be reversed by antioxidant/vitamin food fortification in fathers. *Sci Rep* 2016;6:27010.
- 13 Safi J, Joyeux L, Chalouhi GE. Periconceptional folate deficiency and implications in neural tube defects. *J Pregnancy* 2012;2012:1–9.
- 14 McPherson NO, Fullston T, Aitken RJ, *et al*. Paternal obesity, interventions, and mechanistic pathways to impaired health in offspring. *Ann Nutr Metab* 2014;64:231–8.

- 15 Ryan DP, Henzel KS, Pearson BL, *et al.* A paternal methyl donor-rich diet altered cognitive and neural functions in offspring mice. *Mol Psychiatry* 2018;23:1345-1355.
- 16 Science News. Father's pre-conception vitamin D intake linked to child height and weight at 5 years old - ScienceDaily, 2017. Available: <https://www.sciencedaily.com/releases/2017/05/170518221002.htm> [Accessed 29 Aug 2019].
- 17 Li J, Tsuprykov O, Yang X, *et al.* Paternal programming of offspring cardiometabolic diseases in later life. *J Hypertens* 2016;34:2111-26.
- 18 | University of Utah. Utah population database. Available: <https://uofuhealth.utah.edu/huntsman/utah-population-database/data/> [Accessed 25 Nov 2019].
- 19 Elferink-Stinkens PM, Van Hemel OJ, Brand R, *et al.* The perinatal database of the Netherlands. *Eur J Obstet Gynecol Reprod Biol* 2001;94:125-38.
- 20 THL. Drugs and pregnancy. Available: <https://thl.fi/en/web/thlfi-en/research-and-expertwork/projects-and-programmes/drugs-and-pregnancy> [Accessed 25 Nov 2019].
- 21 Bliddal M, Broe A, Pottegård A, *et al.* The Danish medical birth register. *Eur J Epidemiol* 2018;33:27-36.
- 22 DANAK - The Danish Accreditation Fund. DANAK - Den danske akkrediteringsfond. Available: <http://english.danak.dk/> [Accessed 26 Apr 2019].
- 23 Andersen CL, Siersma VD, Karlslund W, *et al.* The Copenhagen primary care differential count (CopDiff) database. *Clin Epidemiol* 2014;6:199.
- 24 Lyng E, Sandegaard JL, Rebolj M. The Danish national patient register. *Scand J Public Health* 2011;39:30-3.
- 25 Pedersen CB. The Danish civil registration system. *Scand J Public Health* 2011;39:22-5.
- 26 Sundhedsstyrelsen. *Svangreomsorg*. 2nd ed. Sundhedsstyrelsen, 1998.
- 27 Sundhedsstyrelsen. *Anbefalinger for svangreomsorgen 2013 (Kap. 2, 5,6)*. Sundhedsstyrelsen, 2013.
- 28 Statistics Denmark. *FODP: live births by age of mother and live-birth order*, 2019.
- 29 Schmidt M, Schmidt SAJ, Adelborg K, *et al.* The Danish health care system and epidemiological research: from health care contacts to database records. *Clin Epidemiol* 2019;11:563-91.
- 30 Kildemoes HW, Sørensen HT, Hallas J. The Danish national prescription registry. *Scand J Public Health* 2011;39:38-41.
- 31 Andersen JS, Olivarius NDF, Krasnik A. The Danish National health service register. *Scand J Public Health* 2011;39:34-7.
- 32 Helweg-Larsen K. The Danish register of causes of death. *Scand J Public Health* 2011;39:26-9.
- 33 Statistics Denmark. Documentation of statistics: The Student Register - Statistics Denmark. Available: <https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/the-student-registre> [Accessed 27 Aug 2018].
- 34 Green A. Danish clinical databases: an overview. *Scand J Public Health* 2011;39:68-71.
- 35 Databasen Børns Sundhed - SIF. Available: https://www.sdu.dk/da/sif/forskning/projekter/databasen_boerns_sundhed/om_databasen [Accessed 26 Apr 2019].
- 36 Universitet S. *Årsrapport Amning I 14 kommuner*, 2009: 1-97.
- 37 Thygesen LC, Daasnes C, Thaulow I, *et al.* Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scand J Public Health* 2011;39:12-16.
- 38 Green A, Sortsø C, Jensen PB, *et al.* Validation of the Danish national diabetes register. *Clin Epidemiol* 2015;7:5.
- 39 Andersen TF, Madsen M, Jørgensen J, *et al.* The Danish national Hospital register. A valuable source of data for modern health sciences. *Dan Med Bull* 1999;46:263-8.
- 40 Webmaster. CopLab - The Copenhagen Primary Care Laboratory Database, 2018. Available: <https://publichealth.ku.dk/research/databases-for-collaboration/coplab/>
- 41 Siersma V, Kousgaard MB, Reventlow S, *et al.* The effectiveness of computer reminders versus postal reminders for improving quality assessment for point-of-care testing in primary care: a randomized controlled trial. *J Eval Clin Pract* 2015;21:13-20.