BMJ Open

Nervous system drugs taken by future fathers and birth defects in offspring: a prospective registry-based cohort study

Maarten Wensink ,[1,2] Ying Lu,[3] Lu Tian,[3] Tina Kold Jensen,[4] Niels Erik Skakkebæk,[5] Rune Lindahl-Jacobsen,[1,2] Michael Eisenberg[6]

To cite: Wensink M, Lu Y, Tian L, *et al*. Nervous system drugs taken by future fathers and birth defects in offspring: a prospective registry-based cohort study. *BMJ Open* 2022;**12**:e053946. doi:10.1136/bmjopen-2021-053946

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-053946).

Received 28 May 2021
Accepted 04 February 2022

For numbered affiliations see end of article.

Correspondence to
Dr Maarten Wensink;
mwensink@health.sdu.dk

ABSTRACT

Objectives To evaluate the association of paternal intake of antipsychotics, anxiolytics, hypnotics and sedatives, antidepressants, selective serotonin reuptake inhibitors (SSRIs) and (benzo)diazepines during the development of fertilising sperm with birth defects in offspring.

Design Prospective registry-based cohort study.

Setting Total Danish birth cohort 1997–2016 using Danish national registries.

Participants All 1 201 119 Danish liveborn singletons born 1997–2016 were eligible, 39 803 (3.3%) of whom had at least one major birth defect.

Exposure Offspring were considered exposed if their father had filled at least one prescription in the relevant drug category during development of fertilising sperm (the 3 months prior to conception).

Primary and secondary outcome measures Primary outcome was the diagnosis, in the first year of life, of at least one major birth defect as categorised in the EUROCAT guidelines. Secondary outcome was the diagnosis, in the first year of life, of at least one major birth defect in any of the EUROCAT subcategories. Adjusted ORs (AORs) were calculated, along with their 95% CIs, adjusted for year, education, smoking status and age of the mother, and education, disposable income and age of the father.

Results This study found weak or null associations between birth defects and selected drugs. Specifically, antidepressants (17 827 exposed births) gave 3.5% birth defects (AOR 0.97 (0.89 to 1.05)). Diazepines, oxazepines, thiazepines and oxepines (as antipsychotics, 1633 offspring) gave 4.7% birth defects (AOR 1.22 (0.97 to 1.54)), attenuated to 1.13 when excluding by mothers' prescriptions. The study was well powered assuming 100% therapy adherence, while assuming 50% therapy adherence, the study remained well powered for the largest groups (SSRIs and antidepressants overall).

Conclusions Antipsychotics, anxiolytics, hypnotics and sedatives, antidepressants, SSRIs and benzodiazepine-derived anxiolytics, when taken by the father during development of fertilising sperm, are generally safe with regard to birth defects.

Strengths and limitations of this study

- ► High-quality registry data give full coverage of population.
- ► Highly powered study for most of the investigated drugs.
- ► Unable to assess actual drug intake.
- ► Unable to assess associations between drugs and fertility.

INTRODUCTION

Certain neurological drugs have been associated with adverse changes in semen quality. Beyond common reproductive outcomes like sperm motility, selective serotonin reuptake inhibitors (SSRIs) have been associated with increased frequencies of DNA fragmentation and abnormal sperm morphology.[1–4] Anxiolytics, in particular benzodiazepines, have been associated with chromosomal abnormalities in sperm[5 6]. Of concern, many of these drugs are commonly prescribed to prospective fathers with increasing use over time.[7] In Denmark, the proportion of births where the father had been prescribed neurological drugs in the 6 months preceding conception more than doubled between 1997 and 2017, from approximately 4% to almost 9%. Importantly, prescriptions of antidepressants, mostly SSRIs, increased threefold to 2.5%.[7]

It is known that paternal factors are associated with birth outcomes such as preterm birth, low birth weight and neonatal intensive care unit stays.[8 9] Given the association of sperm DNA damage in certain neurological drugs, the safety of neurological drugs regarding offspring health needs to be evaluated. In particular, it is unknown whether paternal use of these drugs during sperm development is associated with the risk of birth defects.

Hence, we performed a cohort study on all singleton live births in Denmark 1997–2016 (1 201 119 births), linking national registries: the birth registry, the prescription registry and the patient registry. We then assessed for any association between specific neurological drugs prescribed to the father to be in the 3 months just prior to conception (sperm

Table 1 Cohort characteristics by drug use for all liveborn singletons in Denmark 1997–2016

	None of the specified drugs (1 147 005)	Antipsychotics (4301)	Anxiolytics (4918)	Hypnotics and sedatives (5797)	Antidepressants (17 827)	(benzo) Diazepines(7057)
Age father, years (mean (Q1–Q3))	33.0 (29.2–36.4)	34.1 (29.0–38.7)	36.0 (30.9–40.4)	36.4 (31.5–40.6)	34.6 (30.2–38.4)	35.7 (30.6–40.2)
Age mother, years (mean (Q1–Q3))	30.4 (27.2–33.7)	30.0 (25.8–34.0)	31.0 (27.2–34.8)	31.3 (27.5–35.1)	31.0 (27.4–34.6)	30.8 (26.8–34.7)
Gestation age, days (mean (Q1–Q3))	279 (273–287)	277 (272–286)	277 (272–286)	277 (272–286)	278 (272–286)	277 (272–286)
Pre-term (% (N))	5.0% (57 395)	6.7% (288)	6.5% (319)	6.2% (357)	5.8% (1027)	6.6% (464)
Birth weight, kg (mean (Q1–Q3))	3.5 (3.2–3.9)	3.4 (3.1–3.8)	3.4 (3.1–3.8)	3.5 (3.1–3.8)	3.5 (3.2–3.9)	3.4 (3.1–3.8)
Birth length, cm (mean (Q1–Q3))	52 (50–54)	51 (50–53)	52 (50–53)	52 (50–53)	52 (50–53)	51 (50–53)
Apgar score <8 (% (N))	1.3% (15 001)	1.7% (73)	1.6% (78)	1.3% (74)	1.6% (288)	1.5% (108)
Low education father (% (N))	19.0% (213 009)	44.8% (1928)	38.8% (1908)	32.3% (1872)	28.6% (5101)	40.9% (2884)
High education father (% (N))	12.0% (136 825)	5.6% (241)	7.3% (358)	10.6% (614)	9.4% (1678)	6.9% (483)
Low education mother (% (N))	18.0% (204 933)	40.4% (1737)	35.2% (1730)	31.0% (1799)	24.8% (4423)	37.2% (2625)
High education mother (% (N))	11.0% (124 850)	5.2% (225)	6.2% (306)	8.8% (507)	9.2% (1637)	5.9% (413)
Mother quit smoking (% (N))	2.0% (26 752)	4.0% (171)	3.1% (151)	2.5% (146)	3.2% (569)	3.3% (229)
Mother smoked (% (N))	12.0% (138 007)	25.0% (1075)	26% (1281)	20.1% (1164)	17.8% (3176)	25.9% (1831)
Income father (mean (Q1–Q3))	205 (138–248)	147 (104–172)	158 (101–191)	183 (106–218)	181 (119–224)	153 (102–186)
Parity 0 (% (N))	45.6% (523 304)	45.7 (1967)	41.3% (2030)	42.6% (2467)	41.6% (7414)	42.9% (3042)
Parity 1 (% (N))	37.2% (426 557)	30.5% (1313)	33.4% (1643)	32.7% (1898)	36.2% (6450)	31.9% (2251)
Parity 2 (% (N))	13.2% (151 369)	15.4% (662)	15.8% (775)	15.5% (899)	15.3% (2727)	15.8% (1112)
Parity 3 or more (% (N))	4.0% (45 825)	8.3% (359)	9.6% (470)	9.2% (533)	6.9% (1236)	9.5% (670)
Boys (% (N))	51.4% (589 062)	50.8% (2185)	51.7% (2544)	51.4% (2979)	51.6% (9204)	51.8% (3657)
Major birth defect (% (N))	3.3% (38 194)	3.9% (167)	3.5% (173)	3.3% (190)	3.5% (617)	3.6% (254)

% are column percentages.
Group 'none of the specified drugs' refers to no drugs that occur in the other columns. Other columns may overlap. In particular, benzo(diazepines) are subgroups of antipsychotics (N05A), anxiolytics (N05B) and hypnotics and sedatives (N05C). Income father refers to disposable income in thousands of Danish crowns per year.
n, number; Q, quartile.

development) and birth defects diagnosed in the first year of life.

METHODS

Data and inclusion criteria

We obtained the Danish Medical Birth Registry (MFR[10]) 1997–2016, which contains all births in Denmark from 20 weeks of gestation onwards. In addition to characteristics of the newborn and pregnancy, such as gestational age and Apgar score, this registry contains the CPR (Centrale Personregister) number,[11] a unique identifier that all Danish citizens and residents have been given since 1968, for newborn, mother and father (if known). We used this CPR number to link registries, meaning that entries with unusable or missing CPR number of either parent or offspring were deleted. Stillbirths were also deleted due to dissimilar ascertainment of birth defects (see *Outcome*). Approximate conception date is contained in the MFR as birth date minus estimated gestational age.

We linked this registry to the Danish National Prescription Registry (LMDB[12]), which we obtained from 1995 to mid-2018. This registry gives complete coverage of all prescriptions filled in Denmark by persons with a CPR number. In Denmark, over-the-counter drug prescriptions are limited; common pain medication like paracetamol is not freely available in large packages. From this registry, we created indicator variables for exposure (see *Exposure*). We also used this registry to identify those births where the mother had taken any of the investigated drugs up to giving birth (see *Statistical Analyses*).

We further linked with the Danish National Patient Registry[13] from 1995 through mid 2018, which contains diagnoses for all inpatient and outpatient contacts, although not for diagnoses in the family doctor setting. This registry includes birth defects, which we classified according to the EUROCAT guidelines,[14] allowing 1 year of follow-up on birth. Birth defects that EUROCAT classified as minor were excluded.

We incorporated information from Statistics Denmark, the central authority on Danish statistics. These variables were paternal disposable income, the amount of money that a person or household has available for spending and saving after income taxes and interest expenses have been accounted for, and highest achieved education (both by year). We further linked with the Population Registry to obtain birth date and sex of the parents. Births with fathers of unknown or female sex were removed, as were births to mothers of male sex.

Outcome

The primary outcome was the diagnosis of at least one major birth defect in the first year of life (binary variable), categorised as per the EUROCAT guidelines,[14] which provide International Classification of Diseases (ICD) codes of birth defects that they classify as major. The secondary outcome was being diagnosed with at least one major birth defect (binary variable) in any of the EUROCAT subcategories (by organ or tract).

Exposure

As one spermatogenic cycle takes approximately 3 months,[15] we considered offspring whose father filled a prescription in the relevant category during the 3 months preconception as exposed. We examined the following medication categories: antipsychotics (Anatomical Therapeutic Chemical classification code N05A), among which diazepines, oxazepines, thiazepines and oxepines (N05AH); anxiolytics (N05B), among which benzodiazepine-derived anxiolytics (N05BA); hypnotics and sedatives (N05C), among which benzodiazepines (N05CD); and antidepressants (N06A), among which SSRIs (N06AB).

Missing data

Approximately 15% of the merged records had at least one entry missing, in particular maternal smoking status (online supplemental table 1). We imputed 10 datasets in a procedure described in detail in the Statistical appendix under the assumption of missingness at random. Reported results are estimates and SEs pooled under Rubin's rule. Imputation and pooling was handled with the R package *mice*[16] (V.3.8.0).

Statistical analyses

We employed flexible logistic regressions using generalised additive models with R package *mgcv*[17] V.1.8–33, which allow non-linear smooth associations between the exposure variable and the birth defect risk. Categorical variables were modelled by simple indicator variables for each level. From these models, we obtained ORs and their 95% CIs for being diagnosed with at least one major birth defect in the first year of life after adjusting for birth year, maternal factors (smoking status during pregnancy, highest achieved education, maternal age) and paternal factors (disposable income, highest achieved education and paternal age). These potential confounders were selected prior to the analysis for their potential relatedness to both the predictor and outcome[18–21] and were not selected based on their significance.

We compared exposed versus unexposed groups for each drug group separately, first for all liveborn singletons. As a sensitivity analysis, we then repeated this analysis excluding births where the mother had taken any of the investigated drugs at any time prior to delivery. We then compared, by conditional logistic regression, exposed versus unexposed offspring of the same father, adjusting for birth year, maternal age, and nulliparity. We then analysed the distribution across EUROCAT organ subgroups without excluding births based on maternal drug use.

All data analyses were carried out on the secure server of Statistics Denmark and run in R[22] V.3.6.3.

Minimum detectable risk and OR calculations

We calculated minimum detectable ORs at 80% and 90% power using the software *PS Power and Sample Size,* V.3.1.6[23] both for the actual exposure numbers and under the assumption that 50% of the fathers actually took their

Table 2 Specific neurological drugs associated with sperm damage and their adjusted ORs (AORs) for having at least one major birth defect

Drug class	Number of offspring	Number of fathers	Birth defects	AOR	95% CI
Antipsychotics (N05A)	4301	–	3.9% (167)	1.07	0.92 to 1.25
After exclusion	2590	–	3.3% (85)	0.95	0.77 to 1.18
Sibling analysis	5437	1971	3.4% versus 3.2%	1.00	0.74 to 1.37
Anxiolytics (N05B)	4918	–	3.5% (173)	1.07	0.92 to 1.24
After exclusion	3153	–	3.2% (102)	1.03	0.85 to 1.26
Sibling analysis	6196	2379	3.4% versus 3.1%	1.08	0.82 to 1.43
Hypnotics and sedatives (N05C)	5797	–	3.3% (190)	0.96	0.83 to 1.12
After exclusion	3706	–	3.2% (119)	0.99	0.83 to 1.19
Sibling analysis	8478	3220	3.1% versus 3.2%	0.97	0.76 to 1.25
Diazepines, oxazepines, thiazepines and oxepines (as antipsychotics, N05AH)	1633	–	4.7% (76)	1.22	0.97 to 1.54
After exclusion	902	–	4.1% (37)	1.13	0.81 to 1.57
Sibling analysis	2220	812	4.1% versus 3.2%	1.24	0.76 to 2.02
Benzodiazepine-derived anxiolytics (N05BA)	4742	–	3.5% (166)	1.06	0.91 to 1.24
After exclusion	3047	–	3.2% (97)	1.02	0.83 to 1.25
Sibling analysis	5885	2266	3.3% versus 3.1%	1.06	0.79 to 1.41
Benzodiazepines as hypnotics and sedatives (N05CD)	1153	–	3.1% (36)	0.96	0.69 to 1.34
After exclusion	736	–	2.9%[21]	0.93	0.60 to 1.43
Sibling analysis	1495	545	2.9% versus 3.4%	0.88	0.49 to 1.59
(Benzo)diazepines grouped (N05AH, N05BA or N05CD)	7057	–	3.6% (254)	1.06	0.93 to 1.20
After exclusion	4428	–	3.3% (147)	1.03	0.87 to 1.22
Sibling analysis	8777	3318	3.3% versus 3.3%	1.02	0.80 to 1.29
Antidepressants (N06A)	17827	–	3.5% (617)	0.97	0.89 to 1.05
After exclusion	11487	–	3.2% (372)	0.95	0.85 to 1.05
Sibling analysis	23400	9020	3.3% versus 3.6%	1.02	0.87 to 1.19
SSRIs (N06AB)	11902	–	3.3% (397)	0.94	0.85 to 1.04
After exclusion	7751	–	3.3% (254)	0.96	0.85 to 1.09
Sibling analysis	15971	6220	3.2% versus 3.6%	0.93	0.77 to 1.11

All liveborn singletons Denmark 1997–2016. Exclusion is by births where mothers used any of the investigated drug at any time prior to delivery. ORs adjusted for birth year, paternal age, income and education, and maternal age, smoking status and education, except for the sibling analysis. Separate models per drug. Exposure taken as binary: having at least one prescription in the 3-month preconception timeframe. Offspring numbers for the sibling analysis include exposed as well as unexposed offspring.
SSRIs, selective serotonin reuptake inhibitors.

prescriptions. Because some drugs induced highly selective groups (see Results), we conservatively assumed an exposed:unexposed ratio of 1:10 for these calculations (the larger groups tended to be less selective, see Results).

Patient and public involvement statement

Patients or the public were not involved in the planning, executing and communication of this study.

RESULTS

The cohort

The Birth Register had 1 276 229 records for 1997–2016. After exclusion of records with unusable CPR of the offspring (2888) or father (1150), 1 272 750 records could be linked to the patient register, the prescription register, (socioeconomic) variables held at Statistics Denmark and the population register. Excluding births to fathers with registered unknown or female sex (19 163), mothers of male sex (7), and stillbirths (1927) left 1 251 653 records for multiple imputation. After imputation, excluding records of non-singleton births (50 534) and records with missing gestational age (27 080) left 1 174 727 offspring. Exclusion of births with mothers who filled a prescription of any of the investigated drugs at any time up to delivery left 936 706 offspring.

Among the 1 174 727 births available for the main analysis, that is, liveborn singletons without missing gestational age, 17 827 offspring were exposed to antidepressants, including 11 902 to SSRIs; 4301 to antipsychotics, including 1633 to diazepines, oxazepines,

Table 3 EUROCAT subgroups (binary: ≥1) by drug class, all liveborn singletons, Denmark 1997–2016

Birth defect category	None of the specified drugs (1 147 055)	Antipsychotics (4301)	Anxiolytics (4918)	Hypnotics and sedatives (5797)	Antidepressants (17 827)	(Benzo)diazepines (7349)	N05AH (1633)
Digestive	0.22% (2471)	0.28% (12)	0.12% (6)	0.19% (11)	0.19% (33)	0.16% (11)	0.37% (6)
Urinary	0.26% (3020)	0.37% (16)	0.33% (16)	0.38% (22)	0.29% (51)	0.40% (28)	0.73% (12)
Heart	0.70% (8069)	0.77% (33)	0.79% (39)	0.78% (45)	0.70% (125)	0.79% (56)	0.73% (12)
Chromosomal	0.11% (1283)	≤0.12% (≤5)	≤0.10% (≤5)	≤0.09% (≤5)	0.11% (20)	≤0.07% (≤5)	≤0.30% (≤5)
Limb	0.93% (10 699)	0.93% (40)	0.96% (47)	0.67% (39)	0.94% (167)	0.91% (64)	1.16% (19)
Nervous	0.11% (1305)	0.19% (8)	0.14% (7)	0.16% (9)	0.08% (15)	0.13% (9)	≤0.30% (≤5)
Eye	0.12% (1384)	≤0.12% (≤5)	0.14% (7)	≤0.09% (≤5)	0.12% (22)	0.11% (8)	≤0.30% (≤5)
Genital	0.25% (2825)	0.21% (9)	0.22% (11)	0.24% (14)	0.30% (54)	0.24% (17)	≤0.30% (≤5)
Oro-facial clefts	0.15% (1684)	0.19% (8)	≤0.10% (≤5)	0.16% (9)	0.13% (23)	0.11% (8)	≤0.30% (≤5)
Other	0.64% (7287)	1.05% (45)	0.92% (45)	0.79% (46)	0.80% (142)	0.91% (64)	1.29% (21)

Notice that offspring may appear in more than one category. Publication of numbers smaller than 5 not permitted. Classified as recommended in EUROCAT guide 1.4, section 3.3, pages 92–96.

thiazepines and oxepines; 4918 to anxiolytics (primarily benzodiazepines, n=4742); and 5797 to hypnotics and sedatives, of which 1153 to benzodiazepines (tables 1 and 2). Grouping (benzo)diazepines resulted in 7057 exposed births. Exclusion of births where the mother had taken any of the investigated drugs prior to delivery reduced the exposure numbers (by approximately 1/3), representative of the correlation between parents for these drugs (table 2).

Fathers who were prescribed any neurological medications before conception were older, as were their partners (table 1). Differences in education, income, maternal smoking and parity were also noted. Preterm percentages were slightly higher in the drug exposed groups (>6%) versus the non-exposed group (5%). The sex ratio was similar for all exposure groups relative to the non-exposed group.

Multiple imputation results suggested that missing data were unlikely to have influenced the results from the complete case analysis. The regression results with or without multiple imputation showed only very modest associations for potential confounders, mostly maternal education with an adjusted OR (AOR) just below 1.1 for low education.

Birth defects analysis

Birth defects in children of fathers exposed to neurological drugs just before conception were generally similar to those in the unexposed population (3.3%–3.9% exposed vs 3.3% unexposed, table 1). After multivariable adjustment, all 95% CIs included one (table 2). Results were similar in the siblings analysis (table 2). For antidepressants and SSRIs, the ORs were 0.97 (0.89 to 1.05) and 0.94 (0.85 to 1.04), respectively (all liveborn singletons), and 0.95 (0.85 to 1.05) and 0.96 (0.85 to 1.09) after exclusion. There was a moderate but not statistically significant tendency towards higher birth defect risk among children whose fathers were prescribed diazepines, oxazepines, thiazepines and oxepines (N05AH), which showed an AOR of 1.22 (95% CI 0.97 to 1.54) for all liveborn singletons, and 1.13 (95% CI 0.81 to 1.57) after exclusion of births to mothers ever prescribed any drug in the groups investigated here. In this group, birth defects appeared especially elevated in the urinary tract (0.73% vs 0.26%, p<0.001 (p=0.04 after Šidàk correction for multiple testing), table 3).

Power and detectable odds

At 80% or 90% power, the minimum detectable OR was between 1.1 and 1.3 for the larger groups but approximately 1.5 for the smaller groups (N05AH and N05CD, table 4). Assuming a therapy adherence of 50%, minimum detectable ORs were approximately 1.3 for antidepressants or SSRIs, approximately 1.5 for antipsychotics, anxiolytics, hypnotics and sedatives, as well as for benzodiazepine-derived anxiolytics. For benzodiazepines as hypnotics and sedatives (N05CD), minimum detectable ORs could be as high as 2.1 (table 4).

Table 4 Minimum risks detectable as aberrant

Drug class	N	Minimum detectable OR			
		Assuming 100% therapy adherence		Assuming 50% therapy adherence	
		80% power	90% power	80% power	90% power
Antipsychotics (N05A)	4301	1.25	1.32	1.51	1.64
Anxiolytics (N05B)	4918	1.25	1.28	1.51	1.57
Hypnotics and sedatives (N05C)	5797	1.22	1.25	1.45	1.51
Diazepines, oxazepines, thiazepines and oxepines (as antipsychotics, N05AH)	1633	1.45	1.54	1.90	2.10
Benzodiazepine-derived anxiolytics (N05BA)	4742	1.25	1.28	1.51	1.57
Benzodiazepines as hypnotics and sedatives (N05CD)	1153	1.54	1.64	2.10	2.31
(Benzo)diazepines grouped (N05AH, N05BA or N05CD)	7057	1.19	1.22	1.38	1.45
Antidepressants (N06A)	17827	1.13	1.16	1.25	1.32
SSRIs (N06AB)	11902	1.16	1.19	1.32	1.38

Based on a univariate binomial model with population risk of 3.3% assuming a 1:10 exposed:unexposed ratio. The two rightmost columns assume a 50–50 mix between the population risk of 3.3% and the risk among the exposed.

DISCUSSION

Summary of findings

The current study found weak or null associations between offspring birth defects and prescriptions of common neurological drugs filled by the father during the 3 months preconception. The only medication group that suggested a possible association was diazepines, oxazepines, thiazepines and oxepines (as antipsychotics, N05AH), which showed a moderately elevated AOR of 1.22 (0.97 to 1.54)) for all liveborn singletons. The point estimate was similar in the sibling analysis, but reduced to 1.13 after excluding offspring whose mother had filled a prescription of any of the investigated drugs at any time prior to delivery. For SSRIs, a large group with the strongest prior evidence of associated sperm damage, the AOR was 0.94 (0.85 to 1.04) before exclusion and 0.96 (0.85 to 1.09) after exclusion. Results were similar when comparing exposed to unexposed siblings. The number of births with paternal exposure to each of the drugs was generally large enough to detect a what might be a clinically significant elevation in risk, for the larger groups even when assuming that only half the fathers took the medication that they had been prescribed.

Strengths and limitations

The design of a nationwide, registry-based cohort study allowed the inclusion of large numbers of fathers who were prescribed the investigated drugs just before conception and to ascertain whether their offspring had birth defects. The registries used are generally complete and of high quality, with (hospital) reimbursement generally depending on reporting and with cross-checks between registries in place. Further information can be found in references.[10–13] Although our measure of paternal exposure was indirect—filling a prescription does not equate with taking the drugs—the study had power to overcome exposure misclassification. Dosage and exact timing of exposure were not considered, which could have biased our results towards the null.

We did not have information on paternal lifestyle factors, such as exercise or smoking, and there may have been maternal factors (eg, genetic predisposition, lifestyle factors like exercise) for which we could not control. We saw significant differences in demographics between fathers prescribed drugs and those who were not. However, these factors are unlikely to have biased the results towards the null because that would require paternal drug prescriptions to correlate with protective maternal or paternal factors.

Even using registry data, there remains a possibility that offspring of fathers prescribed neurological drugs are less visible to the healthcare system because of the fathers' psychological or psychiatrical ailments. This could result in reduced birth defect ascertainment for these offspring and hence bias the results towards (or even below) the null. Nevertheless, Denmark has universal healthcare with scheduled check-ups for newborns, both at birth and in the first year of life, and we restricted to birth defects classified by EUROCAT as major. Thus, it seems reasonable to suspect that the majority of birth defects would be diagnosed. However, if there was an association between a paternal medication and an earlier reproductive outcome (e.g., failure to fertilize, miscarriage), the effect on birth defects could be interpreted as bias towards the null.

Interpretation, possible mechanism, comparison with the literature

Although sperm DNA damage suggests a risk to offspring, this risk may not materialise if sperm with damaged DNA fail to fertilise an egg cell, if the oocyte corrects any DNA damage, if the conceptus fails to develop into a viable fetus or if the fetus is aborted. Hence, sperm damage could

lead to subfertility or infertility, but not birth defects. As the Danish Medical Bi

rth Registry covers only pregnancies from week 20 onwards, further studies are necessary to explore this hypothesis.

Literature on paternal effects on offspring is limited. Certainly, it is reasonable to expect that the 9 months a fetus spends developing in *utero* gives more scope for teratogenic effects from maternal exposure than preconception spermatogenic paternal contribution. Yet there is increasing evidence that sperm contributes more than DNA alone,[24] and the early stages of pregnancy are also the most vulnerable stages with regard to birth defects.

The observation of a tendency towards increased risk in diazepines, oxazepines, thiazepines and oxepines (as antipsychotics, N05AH) may be due to the disease rather than drug, although antipsychotics as a whole only very mildly tended towards an increased OR, while neither birth defects of the nervous system nor chromosomal birth defects were elevated in this group. However, a prior study did suggest a possible association between paternal diazepam and perinatal mortality and growth retardation.[25] The attenuation of the point estimate seen when excluding births where mothers had been on any of these drugs may indicate confounding by maternal associations. However, if a significant share of the N05AH-exposed offspring were actually unexposed, because the father may not have taken the filled prescription, 1.22 may be an underestimate of the true association.

CONCLUSION

The current study found weak or null associations between prescriptions of neurological drugs (antipsychotics, anxiolytics, hypnotics and sedatives, antidepressants, and SSRIs, benzodiazepine-derived anxiolytics) filled by the father during the development of fertilising sperm (3 months before conception) and birth defects in the offspring. Paternal use of diazepines, oxazepines, thiazepines and oxepines (as antipsychotics, N05AH) during the development of fertilising sperm may be associated with mildly elevated birth defect frequencies, although a maternal pathway is not excluded here and although this observation could be due to chance. As such, men can be counselled that these medications likely do not increase the risk of birth defects. Further studies are necessary to investigate whether these drugs lead to higher rates of stillbirths, early abortions or failure to fertilise, as well as the group N05AH.

Author affiliations
[1]Department of Epidemiology, Biostatistics and Biodemography, University of Southern Denmark, Odense, Denmark
[2]Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Odense C, Denmark
[3]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, California, USA
[4]Department of Environmental Medicine, University of Southern Denmark, Odense, Denmark
[5]Department of Growth and Reproduction, Juliane Marie Centre, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[6]Male Reproductive Medicine and Surgery, Department of Urology, Stanford University School of Medicine, Stanford, California, USA

Contributors ME, RL-J, NES, YL and TKJ designed the study. MW, YL and LT handled data and statistical analysis. MW wrote the first draft. All authors interpreted the results, revised the manuscript and approved the final version. MW is the guarantor of the article.

Funding This study was funded by NIH grant HD096468 (to ME).

Disclaimer The funder had no role in the study design, collection, analysis and interpretation of the data and in the writing of the report.

Competing interests YL reports grants from Merck, and personal fees from United Health Care, Nektar, and Gilead, outside the submitted work. ME reports advisorships for Sandstone Diagnostics, Dadi, Hannah and Underdog, which are fertility-related companies.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants, but registry studies are exempted in Denmark from IRB review. They are subject to strict regulations. They are executed on a secure server at Statistics Denmark, and only aggregate data are allowed to leave the server (tables, regression coefficients) after being controlled by someone else than the primary analyst.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Data from Statistics Denmark cannot be made publicly available but can be applied for through the usual ways at DST.dk. The grant proposal is summarised here: https://reporter.nih.gov/search/C3FoCZUipkCJsZsijQP4LA/project-details/9585127#details

ORCID iD
Maarten Wensink http://orcid.org/0000-0001-6518-1015

REFERENCES

1 Safarinejad MR. Sperm DNA damage and semen quality impairment after treatment with selective serotonin reuptake inhibitors detected using semen analysis and sperm chromatin structure assay. *J Urol* 2008;180:2124–8.
2 Tanrikut C, Feldman AS, Altemus M, *et al*. Adverse effect of paroxetine on sperm. *Fertil Steril* 2010;94:1021–6.
3 Akasheh G, Sirati L, Noshad Kamran AR, *et al*. Comparison of the effect of sertraline with behavioral therapy on semen parameters in men with primary premature ejaculation. *Urology* 2014;83:800–4.
4 Koyuncu H, Serefoglu EC, Yencilek E, *et al*. Escitalopram treatment for premature ejaculation has a negative effect on semen parameters. *Int J Impot Res* 2011;23:257–61.
5 Adler I-D, Schmid TE, Baumgartner A. Induction of aneuploidy in male mouse germ cells detected by the sperm–FISH assay: a review of the present data base. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 2002;504:173–82.
6 Baumgartner A, Schmid TE, Schuetz CG, *et al*. Detection of aneuploidy in rodent and human sperm by multicolor fish after chronic exposure to diazepam. *Mutat Res* 2001;490:11–19.

7 Wensink MJ, Rizzi S, Jensen TK, *et al*. Paternal prescription medication before conception: a retrospective cohort study of all births in Denmark 1997-2017. *Scand J Public Health* 2021;49:884–90.
8 Kasman AM, Zhang CA, Li S, *et al*. Association of preconception paternal health on perinatal outcomes: analysis of U.S. claims data. *Fertil Steril* 2020;113:947–54.
9 Khandwala YS, Zhang CA, Lu Y, *et al*. The age of fathers in the USA is rising: an analysis of 168 867 480 births from 1972 to 2015. *Hum Reprod* 2017;32:2110–6.
10 Bliddal M, Broe A, Pottegård A, *et al*. The Danish medical birth register. *Eur J Epidemiol* 2018;33:27–36.
11 Schmidt M, Pedersen L, Sørensen HT. The Danish civil registration system as a tool in epidemiology. *Eur J Epidemiol* 2014;29:541–9.
12 Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, *et al*. Data resource profile: the Danish national prescription registry. *Int J Epidemiol* 2017;46:798-f.
13 Lynge E, Sandegaard JL, Rebolj M. The Danish national patient register. *Scand J Public Health* 2011;39:30–3.
14 EUROCAT. EUROCAT central registry uou. European surveillance of congenital Anormalities (EUROCAT 2013). In: *EUROCAT guide 1.4: instruction for the registration of congenital anomalies*. 11, 2019. https://eu-rd-platform.jrc.ec.europa.eu/sites/default/files/Full_Guide_1_4_version_28_DEC2018.pdf
15 Neto FTL, Bach PV, Najari BB, *et al*. Spermatogenesis in humans and its affecting factors. *Semin Cell Dev Biol* 2016;59:10–26.
16 van Buuren S, Groothuis-Oudshoorn K. Multivariate imputation by Chained equations in R. *Journal of Statistical Software* 2011;45:1–67.
17 Woods S. *Generalized additive models, an introduction with R*. 2 edn. New York: Chapman and Hall/CRC, 2017.
18 Harris BS, Bishop KC, Kemeny HR, *et al*. Risk factors for birth defects. *Obstet Gynecol Surv* 2017;72:123–35.
19 Rynn L, Cragan J, Correa A, *et al*. Update on overall prevalence of major birth defects--Atlanta, Georgia, 1978-2005. *MMWR Morb Mortal Wkly Rep* 2008;57:756–8.
20 Hackshaw A, Rodeck C, Boniface S. Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. *Hum Reprod Update* 2011;17:589–604.
21 Wasserman CR, Shaw GM, Selvin S, *et al*. Socioeconomic status, neighborhood social conditions, and neural tube defects. *Am J Public Health* 1998;88:1674–80.
22 R Development Core Team. *R: a language and environment for statistical computing: R foundation for statistical computing*, 2020.
23 Dupont WD, Plummer WD. Power and sample size calculations. A review and computer program. *Control Clin Trials* 1990;11:116–28.
24 Immler S. The sperm factor: paternal impact beyond genes. *Heredity* 2018;121:239–47.
25 Engeland A, Bjørge T, Daltveit AK, *et al*. Effects of preconceptional paternal drug exposure on birth outcomes: cohort study of 340 000 pregnancies using Norwegian population-based databases. *Br J Clin Pharmacol* 2013;75:1134–41.

Statistical appendix to: Nervous system drugs taken by future fathers and birth defects in offspring: a registry-based cohort study

Statistical appendix

M.J. Wensink, Y. Lu, L. Tian

Introduction and summary

Since some observations were partially missing, either because they were treated as such (see below), or, mostly, because they were missing in the original dataset, we planned to impute those observations before running generalized additive models on the imputed datasets, and combine the results through Rubin's rule. At the moment of imputation, we had variables such as gestational age, smoking status of the mother, and education of both parents, as well as a large number of binary variables for drug exposures. We had 973 drugs which at least one father was prescribed during the three months preconception, 945 drugs that at least one mother was prescribed during the first trimester of pregnancy, and 885 drugs that at least one mother was prescribed during the rest of pregnancy. We also had various indicator covariates for maternal and paternal conditions, as well as 14 subgroups of birth defects (for example birth defects of the heart, etc.). Because imputation on all these covariates was impractical giving computational limitations, we set up the following approach.

1. We ran a lasso algorithm on the complete cases, from which we selected an appropriate set of variables to impute from for each of the variables that had missing data.
2. We ran multiple imputation by chained equations, imputing each of the missing value of a variable based on variables selected by the lasso algorithm for the variable of interest.
3. We ran logistic regressions with the logit link function through generalized additive models based on entire datasets after imputing all missing values and pooled the results through Rubin's rule.

The lasso algorithm used linear relationships only and no interactions were considered. For multiple imputation and logistic regressions, we used approaches that allowed a maximum of flexibility with regard to departures from linearity. In particular, for imputation we used predictive mean matching for numeric variables and simple indicator variables for each group of categorical variables (without further assumptions such as the proportional odds assumption), while for the logistic regressions we used scatterplot smoothers (thin plate splines with four knots spaced uniformly, i.e. the default setting of the statistical software – changing to cubic splines did not change the results) for numeric variables and again simple indicator variables for each group of categorical variables. Details are given below.

Data preparation and missingness

This analysis concerns liveborn singletons only. Birth lengths below 21 cm (often 0 cm or 10 cm, 8412 births) or above 69 cm (538 births) were treated as missing. Birth weights below 366 g (often 0 g or 100 g, 2091 births) and above 6583 g (63 births) (+/- 5 standard deviations) were treated as missing. At least master-level education was relabeled as high education. Education ranging from upper secondary to bachelor level was relabeled as middle education. Primary and lower secondary education were relabeled as low education. This relabeling was based on divergent trends of preconception drug use in this data set published elsewhere (reference 3 of the main manuscript). Education not elsewhere classified was treated as missing (1883 and 3669 births for father and mother, respectively). Maternal smoking was relabeled as no smokers, current smokers, and quit during pregnancy (thus collapsing the various moments of quitting during pregnancy).

After applying exclusion criteria, i.e. for observations used in the final analysis, we had the following numbers of missing data:

Supplementary Table 1. Numbers (%) of data missing for each variable that had at least one missing entry (% missing = 0 for all other variables). Smoking status of the mother was missing most often (8.4%), but had a negligible effect on birth defects. Education of the mother (3.4%) and father (3.8%) was missing next most often, with education of the mother having a small but highly significant effect (adjusted odds ratio on the order of 1.07, depending on the model).

Variable	**Number missing net of exclusion criteria**
Apgar score	8,531 (0.8%)
Birth length	13,001 (1.2%)
Birth weight	5,594 (0.5%)
Hospital days upon birth	6,015 (0.5%)
Disposable income father	4,818 (0.4%)
Education father	41,895 (3.8%)
Education mother	37,314 (3.4%)
Smoking status mother	92,448 (8.4%)
At least one variable missing	168,871 (15.3%)

Notice that while Apgar score, birth weight and length, and number of hospital days were not involved in the main analysis, they had to be imputed because other imputations might rely on them. For the same reason we initially imputed the 27,080 missing gestation ages.

Lasso

For the lasso algorithm, all numeric variables were standardized by subtracting the mean and dividing by twice their standard deviation. Binary variables were not standardized. To allow for convergence in reasonable time, we first selected even years only, after which we sampled without replacement 1 in every 5 observation, giving approximately 10% of the original data or some 100 thousand (100K) observations. On this reduced dataset we then ran the lasso cross validation algorithm of R package *glmnet* (version 4.0-2), i.e. cv.glmnet() with alpha=1. From the cross validation results (representative example in Figure 1), we selected an appropriate number of variables from which to impute.

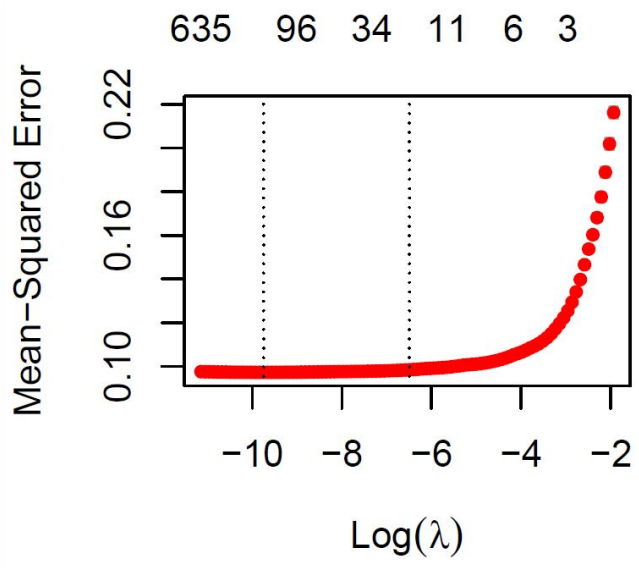

Figure 1. An example of a lasso result, in this case of gestation age on a sample of 100K observations. On the x-axes the number of non-zero covariates (top) along with the log of the penalty parameter λ (bottom). The loss (mean squared error) is given as the y-axis. Observe that the null model (0 covariates) corresponds to imputing the mean, giving a MSE (≈variance) of 0.25 as a result of our standardization procedure. After log(λ) reaches -6, diminishing returns lead the addition of further covariates to reduce the mean squared error only marginally, whereas the number of variables with non-zero coefficients grows swiftly. In this case we selected a model with 75 variables, striking the middle between sparsity and accuracy (see main text of this Appendix).

In selecting the set of variables to impute from, we considered that we had the following reasons to be generous, i.e., likely to choose a high number of variables:

1. A better, fuller model gives more accurate imputations and so improves the precision of estimates.
2. The relationships in the complete cases may differ somewhat from the relationships in the whole population (if we knew all missing values). Selecting a generous model gives a chance for these relationships to be picked up in the imputation procedure.
3. There would have been some sampling variability in selecting 100K observations from the dataset, in which the relationships may differ somewhat from the remaining (complete cases) dataset. Selecting a generous model gives a chance for this relationship to be picked up in the imputation procedure.
4. Overfitting was not our first concern, since we wished to maximally capture the variability in the data and our sample size is large relative to the number of variables.
5. We had particular reason to be generous for those variables with a high percentage of observations missing (smoking status in particular), so that *ceteris paribus* gains from accurate imputations would be larger (although variability in variables used in those imputations may also propagate).
6. Selecting a small model may lead imputations to depend on variables that are missing at the same time, which little opportunity to seize on other information in the data.

We also had the following reasons to be conservative, i.e. choose a low number of variables:

1. A small model runs faster, which was the aim of the lasso screening step. Multiple imputations based on the entire dataset is too slow and unnecessary.
2. Since there is variability in sampling the subset on which to run lasso, the variables near the minimum loss (the left vertical dashed line in Figure 1) may not be relevant for the rest of the data.

3. Some of the non-zero coefficients in the selected lasso model may not be at their full, unpenalized values. Removing the restriction of penalization may increase the predictive power of the model in the multiple imputation setting without need for additional covariates in the imputation model.

In practice we ended up choosing a model halfway between lambda for the minimum loss (lambda.min) and lambda at one standard error of the minimum loss (lambda.1se). Up to 50 covariates were added to the latter model depending on the location of lambda.min, the amount of missing data, a visual inspection of the plateau, and any pre-existing knowledge on the status of a variable as a confounder, always staying well away from lambda.min. The exceptions to this procedure were the education factor variables, which did not converge on 100K observations. For these variables we divided up the 100K variables in 4 datasets (of 25K variables each), ran the lasso algorithm on each of these, selected the lambda.min model for each, and took the intersection of the four sets of covariates. Hence, the first selection was generous (large number of variables selected), but taking the intersection of the four sets is conservative (small number of variables selected). For education, this procedure gave 59 variables for the father and 105 for the mother. Smoking status of the mother was imputed from 146 variables due to high missingness. The other variables were imputed from much smaller models (11-16 covariates). Thus, we used large models for variables with a relatively large amount of missing data, or that were difficult to predict.

Multiple imputation

Multiple imputation was done using chained equations (Gibbs sampler) under the assumption of missingness at random, implemented in R package *mice* (version 3.8.0) with a custom made predictor matrix set up from the models found through the lasso procedure outlined above. We created 10 imputed datasets using polytomous regression without further assumptions for categorical variables and predictive mean matching for numerical variables. The number of imputed datasets was lower than the 15% suggested by the "one dataset for each percent of missing data" rule, justified because most variables had only low percentages of missing data, and were used only indirectly. Because the default number of 5 iterations suggested that convergence was perhaps not fully achieved for some variables, we ran 10 iterations. Convergence was achieved after 7 iterations.

Logistics regressions (generalized additive model)

We ran logistic regression fully adjusted for birth year, paternal characteristics, and maternal characteristics, as implemented in R package *mgcv* (version 1.8-33). For example,

```
model<-gam(birth defect ~ drug name + s(birth year) +
                education father + s(income father) + s(age father) +
                education mother + smoking status mother + s(age mother),
                data = data,
                subset = liveborn singletons after exclusion criteria,
                family = binomial()
)
```

The scatterplot smoothers used most degrees of freedom for birth year and age mother, and least for income and age father.

Such models were run in functions that ran them for each of the datasets, pooled the estimates, and summarized the results, as implemented in R package *mice*:

estimates <- with(data = data, gam as above *omitting the data statement*)

pooled estimates <- pool(estimates)

model results <- summary(pooled estimates)