Phenome-Wide Analysis of Short- and Long-Run Disease Incidence Following Recurrent Pregnancy Loss Using Data From a 39-Year Period

Westergaard, David; Nielsen, Anna Pors; Mortensen, Laust Hvas; Nielsen, Henriette Svarre; Brunak, Søren

Published in:
Journal of the American Heart Association

DOI:
10.1161/JAHA.119.015069

Publication date:
2020

Document version
Publisher's PDF, also known as Version of record

Journal of the American Heart Association

ORIGINAL RESEARCH

Phenome-Wide Analysis of Short- and Long-Run Disease Incidence Following Recurrent Pregnancy Loss Using Data From a 39-Year Period

David Westergaard, PhD; Anna Pors Nielsen, MD; Laust Hvas Mortensen, PhD; Henriette Svarre Nielsen, MD, DMSc; Søren Brunak, PhD

BACKGROUND: It is unclear how recurrent pregnancy loss (RPL) impacts disease risk and whether there is a difference in risk between women with or without a live birth before RPL (primary versus secondary RPL). We investigated the disease risk following RPL, and whether there was a difference between primary and secondary RPL.

METHODS AND RESULTS: Using population-wide healthcare registries from Denmark, we identified a cohort of 1 370 896 ever-pregnant women aged 12 to 40 years between 1977 and 2016. Of this cohort, 10 691 (0.77%) fulfilled the criteria for RPL (50.0% primary RPL). Average follow-up was 15.8 years. Incidence rate ratios were calculated in a phenome-wide manner. Diagnoses related to assessment and diagnosis of RPL and those appearing later in life were separated using a mixture model. Primary RPL increased the risk of subsequent cardiovascular disorders, including atherosclerosis, cerebral infarction, heart failure, and pulmonary embolism, as well as systemic lupus erythematosus, chronic obstructive pulmonary disease, anxiety, and obsessive-compulsive disorder. Women with secondary RPL had no increased risk of cardiovascular disorders. However, we observed an increased risk of gastrointestinal disorders such as irritable bowel syndrome and intestinal malabsorption, as well as mental disorders and obstetric complications.

CONCLUSIONS: RPL is a risk factor for a spectrum of disorders, which is different for primary and secondary RPL. Screening following RPL explains some associations, but the remaining findings suggest that RPL influences or shares cause with cardiovascular disorders, autoimmune disorders, and mental disorders. Research into the pathophysiology of RPL and later diseases merits further investigation.

Key Words: Bayesian statistics ■ epidemiology ■ pregnancy ■ recurrent pregnancy loss ■ women's health

Recurrent pregnancy loss (RPL) is a condition that affects 1% to 3% of couples, depending on its precise definition.[1,2] While the definition of RPL is not fully consistent across the world, it has historically been defined as 3 or more consecutive pregnancy losses (PL), and this is the definition used in Denmark and this study.[3] With increasing number of losses, the frequency of euploid losses increases while the chance of a live birth decreases.[4–6] RPL is divided into 2 categories: primary (no live birth before losses) and secondary RPL (at least 1 live birth before losses). Prior Nordic studies report that 40% to 48% of cases are secondary RPL, but in some countries this may be as high as 61%.[1,7,8]

RPL is a multifactorial condition that, in addition to fetal causes, have multiple known female risk factors

Correspondence to: Søren Brunak, PhD, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3A, DK-2200 Copenhagen, Denmark. E-mail: soren.brunak@cpr.ku.dk or Henriette Svarre Nielsen, MD, DMSc, Recurrent Pregnancy Loss Unit, Department of Obstetrics and Gynecology, Hvidovre Hospital, DK-2650 Hvidovre, Denmark. E-mail: henriette.svarre.nielsen@regionh.dk

Supplementary Materials for this article are available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.119.015069

For Sources of Funding and Disclosures, see page 8.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Recurrent pregnancy loss (RPL) is associated with disease later in life; understanding the link between RPL and subsequent disease patterns is crucial to identify prevention and treatment.
- Women experiencing primary and secondary RPL have different subsequent disease patterns, notably within the cardiovascular, autoimmune, mental, and endocrine disease spectrum.
- Only primary RPL is associated with later cardiovascular disease.

What Are the Clinical Implications?

- Primary and secondary RPL should be considered as 2 separate risk factors for later disease.
- Clinicians should include pregnancy history (including information about miscarriages) when assessing disease risk in a clinical setting and in study designs.
- Primary RPL is especially important for the assessment of cardiovascular disease risk.

Nonstandard Abbreviations and Acronyms

PL	pregnancy loss
RPL	recurrent pregnancy loss
ICD-10	*International Classification of Diseases, Tenth Revision*

such as autoimmune diseases, endocrine disorders, and uterine malformations.[2,9] Risk factors differ between women with primary and secondary RPL. Male risk factors have been investigated less, but include chromosomal abnormalities, sperm DNA fragmentation, and age.[10]

In recent years, PLs and RPL have been suggested or identified as a risk factor for cardiovascular disease (CVD) and malignancies, although the evidence is ambiguous for the latter.[11–16] The association with CVD has been further explored in a study that found first-degree relatives of women with RPL had an increased risk of CVD.[17,18] It has yet to be investigated if the risk is different between women with primary and secondary RPL. Some factors associated with a delayed time to pregnancy and infertility, such as endometriosis, polycystic ovary syndrome, and diabetes mellitus, have a clear association to disease later in life, including malignancies, autoimmune diseases, and CVD.[19–21] Moreover, obstetric complications have also been associated with an increased risk of CVD.[22]

Despite RPL having a major impact on the affected couples, both mentally and physically, the cause and lifelong consequences are poorly understood.[23,24] Our hypothesis is that RPL and PLs not attributed to aneuploidy are associated with later disease, and that women with primary and secondary RPL have quite different risk profiles in terms of later disease. Identifying novel associations could lead to a pathophysiological understanding that can improve the take-home baby rate, long-term maternal health, and reduce the health gap between men and women.

In this study, we compare primary and secondary RPL in a phenome-wide analysis using a nationwide cohort based on 1 370 896 women.

METHODS

The study was approved by the Danish Data Protection Agency (ref: 2015-54-0939 and SUND-2017-57) and Danish Health Authority (ref: FSEID-00001627 and FSEID-00003092). Informed consent and assessment of the proposal in scientific ethical committees are not required for registry-based research in Denmark. Permission to access and analyze data can be obtained following approval from the Danish Data Protection Agency and the Danish Health Authority.

Study Design and Setting

This observational population was identified using the nationwide Danish Patient Registry and the Danish Medical Birth Registry. We included women aged 12 to 40 years with at least 1 live birth or PL in the period between January 1, 1977 and October 5, 2016. Women with multiple births (twins or higher order) were excluded. The population totaled 1 370 896 women.

Women with RPL were followed from the date of meeting the exposure until the end of follow-up (criteria until death or October 5, 2016).

Each woman with RPL was matched with 20 women from the population without RPL. The women were matched based on year of birth and number of previous live births. The matched comparison group of women without RPL was followed for the outcomes from the same date of the woman with RPL they were matched to (ie, only the hospital admissions occurring on or after the date of RPL) and matching was used in the analysis. Women with RPL were divided into primary and secondary RPL based on the parity history before the date of RPL.

Outcomes were identified using several nationwide registries that cover all hospital admissions in Denmark (see Data S1 for a description).[25]

We identified 10 691 women who fulfilled the criteria for RPL (50.0% primary) (defined in the next section). The number of years the women were

followed varied by outcome, but the average number of person-years was 1 782 238 years (range 1 231 217–1 799 076 years) for women with primary RPL and the matched group, and 1 744 736 years (range 1 155 427–1 763 733 years) for women with secondary RPL and the matched group. Across all outcomes, the median follow-up time for women was 15.8 years, ranging from 11.6 to 16.4 years for each specific outcome.

Pregnancy Loss and Recurrent Pregnancy Loss

PLs were identified from hospital admissions in the Danish National Patient Registry (Table). PLs occurring 8 weeks before or after a molar pregnancy, induced abortion, or extrauterine pregnancy were disregarded (Table). In Denmark, every child showing signs of life at delivery is categorized as a live birth irrespective of gestational age. If there is no sign of life at delivery it is considered a PL before the 28th completed weeks of gestation and a stillbirth thereafter. This definition was changed starting from 2004 when stillbirth was counted from 22nd completed weeks of gestation. We filtered out diagnoses that were repeated within a medically unreasonable period of time: (1) <22 weeks between 2 live or still births, (2) <90 days between 2 PLs, termination of pregnancy, molar or ectopic pregnancies, (3) <22 weeks between a PL, induced abortion, or an extrauterine or molar pregnancy and a stillbirth or livebirth, and (4) <30 days between a live or still birth and a PL, termination of pregnancy, or an extrauterine or molar pregnancy.

Cases of RPL were identified using hospital discharge codes (Table) or by 3 consecutive PLs. The date of RPL diagnosis was defined as the date of the third PL or RPL diagnosis, whichever came first.

In the matching process, we used 1 370 339 women from the total population. The median number of unique matched women per outcome was 186 752, ranging from 133 875 to 187 355.

Outcomes

In the phenome-wide analysis, we investigated *International Classification of Diseases, Tenth Revision* (*ICD-10*) codes at the third level that had a prevalence of ≥0.1% in women with RPL (626 diagnoses). *ICD-10* codes related to birth, PL, and termination of pregnancy were excluded (O00–O08 and O80–O84), as well as codes from *ICD-10* chapters 19, 20, 21, and 22. Only the first instance of a diagnosis code was included.

Differentiating Early From Late Disease Occurrences

Some of the diagnoses with an increased risk are because of investigations at an RPL clinic. To differentiate the diagnoses found during investigations at RPL clinics and later in life, we fitted a log-normal mixture model. The optimal number of components was determined using the Bayesian Information Criteria. For each diagnosis code with an increased risk, the time from RPL diagnosis until the outcome occurred was summarized as the median. Time distributions were visualized as histograms that binned multiple diagnoses into 1 bin.

Statistical Analysis

Incidence rate ratios (IRR) were estimated using a hierarchical Bayesian Poisson log-linear model with time in 3-year bands. As covariates we included the age, binned into 11 groups (Table S1), and previous number of live births. A model was fit separately for women with primary and secondary RPL, respectively. Bayesian posterior distributions were summarized as 95% uncertainty intervals (UI). To account for multiple

Table. *ICD-8* and *ICD-10* Codes Used to Identify Recurrent Pregnancy Loss, Pregnancy Loss, and Molar and Extrauterine Pregnancies

Diagnosis	*ICD-8/ICD-10*	Surgery	Procedures
Recurrent pregnancy loss	6430x, Y6439, N96.x, O26.2	...	...
Missed abortion	6346x, 6451x, O02.1, O02.1A	...	...
Miscarriage	6438x, 6439x, O03.x	...	...
Extrauterine pregnancy	63109, 6311x, 63129, 63139, 63149, 6315x, 63169, 63199, O00.x	66100, KLBCx, KMCBx,	BKHE0, BKHE8x
Molar pregnancy	63190, 63429, 63460, 6450x, D39.2, O01.x, O02.0,		
Abortion (induced)	6400x, 6401x, 6402x, 6409x, 6410x, 6411x, 6412x, 6413x, 6414x, 6415x, 6416x, 6417x, 6419x, 64209, 64219, 64229, 64239, 64299, 6455x, 6456x, 6458x, O04.x, O05.x, O06.x, O07.x	63680, 94520, KLCHx, KLWW00,	BKKG1, BKXG1

Lowercase x denotes codes including all subcodes. *ICD-8* and *ICD-10* indicate *International Classification of Diseases, Eighth and Tenth Revisions.*

testing, we calculated the false discovery rate using the Benjamini and Hochberg method. An association was deemed significant if the false discovery rate was <5%. We defined a prior structure that takes into account the chapter structure in the *ICD-10* terminology (see Data S1 and Figure S1 for a thorough evaluation of the priors and the model). Distributions of time from RPL diagnoses until outcome were modeled using a log-normal mixture model, and the number of components was selected using the Bayesian Information Criteria (Table S2). Additional details on the statistical procedures are provided in Data S1. All data were analyzed using stan v2.18, python v2.7, and R v3.1.3.

RESULTS

Out of a pool of 1 310 unique *ICD-10* codes assigned to women after RPL, we investigated 615 diagnoses with a cumulative incidence proportion of at least 0.1% in women with RPL. We found 151 and 127 diagnoses with an elevated IRR in primary and secondary RPL, respectively. These covered very heterogeneous types of disease (Figure 1, Data S1), and included CVDs, obstetric diagnoses, autoimmune diseases, mental disorders, gastrointestinal disorders, and respiratory diseases. We also observed 7 diagnoses with a RR lower than 1 (all in women with primary RPL), which were all related to complications of labor and delivery (Data S1).

We identified an "early" and "late" component for both primary and secondary RPL (Figure 2, Data S1). The median time to each diagnosis for women with RPL is available in Data S1. In women with secondary RPL, the "late" component had a mean value of 11.06 (SD=3.16) years after the diagnosis. Diagnoses that only had an increased risk in women with secondary RPL included irritable bowel syndrome (IRR=1.32, 1.10–1.60 95% UI) and intestinal malabsorption (IRR=1.67, 1.11–2.45 95% UI). Mental disorders included recurrent depressive disorders (IRR=1.35, 1.17–1.55 95% UI) and mixed personality disorders (IRR=1.58, 1.11–2.24 95% UI).

The late component was found to have a mean of 12.6 (SD=4.53) after the primary PRL diagnosis and contained 141 diagnoses. Strikingly, for many CVDs we did not observe an increased risk in women with secondary RPL (Figure 3). This included atherosclerosis (IRR=1.90, 1.30–2.75 95% UI), cerebral infarction (IRR=1.83, 1.36–2.38 95% UI), heart failure (IRR=1.93, 1.40–2.63 95% UI), and pulmonary embolism (IRR=1.64, 1.16–2.29 95% UI). Other diseases included lupus erythematosus (IRR=3.79, 2.45–5.78 95% UI), systemic lupus erythematosus (IRR=2.52, 1.34–4.31 95% UI), chronic obstructive pulmonary disease (IRR=1.56, 1.24–1.96 95% UI), and type 2 diabetes mellitus (IRR=1.92, 1.62–2.27 95% UI). Several mental disorders were also found, including anxiety (IRR=1.29, 1.08–1.53 95% UI) and substance abuse because of smoking (IRR=2.0, 1.65–2.42 95% UI).

Several disorders were found to have an increased risk in both subtypes, such as thyroid disorders (hypothyroidism and thyroiditis), asthma, reaction to severe stress, depressive episodes, risk of mental disorders because of substance abuse of alcohol, and gastrointestinal disorders (gastroesophageal reflux disease, and gastritis and duodenitis). The disorders all belonged to the late component.

The early component was found to have a mean of 2.26 (SD=0.78) years for primary RPL and 1.68 (SD=0.56) years for secondary RPL. The component contained diagnoses routinely investigated as part of the RPL evaluation (eg, uterine malformations, balanced chromosomal rearrangements, and abnormal immunological findings in serum). Cervical malformations, which is not a part of the standard evaluation, was also found to have an increased risk. Additionally, there were many diagnoses related to obstetric complications and pregnancy for both primary and secondary RPL. Certain obstetric complications were only found to have an increased risk in secondary RPL, such as placenta previa (IRR=3.67, 95% UI 2.80–4.77), premature rupture of membrane (IRR=2.67, 95% UI 2.30–3.06), intrapartum hemorrhage (IRR=2.53, 95% UI 1.55–3.98), gestational hypertension (IRR=1.96, 95% UI 1.45–2.63), preeclampsia (IRR=2.29, 95% UI 1.81–2.87), puerperal sepsis (IRR=2.55, 95% UI 1.80–3.54), and placental abruption (IRR=2.91, 95% UI 2.15–3.86).The latter also had an increased risk in primary RPL, albeit lower (IRR=1.32, 95% UI 1.05–1.67).

DISCUSSION

In this study, we performed the largest registry-based phenome-wide study to date of longitudinal disease incidence in RPL. The population comprised 1 370 896 women, of which 10 691 had 3 or more consecutive PLs. By investigating diagnoses occurring after RPL, we identified distinct spectrums of complications for primary and secondary RPL. The diagnoses were divided into "early" and "late" complications: The "late" component included multiple complications in heterogeneous disease domains, such as CVDs, autoimmune diseases, mental disorders, digestive system disorders, and respiratory diseases. The disorders occurred, on average, >10 years after the RPL diagnosis, and there was a distinct set of diseases associated with primary and secondary RPL. The "early" complications contained many previously established risk factors for RPL that are part of the routine screening at RPL clinics.[2]

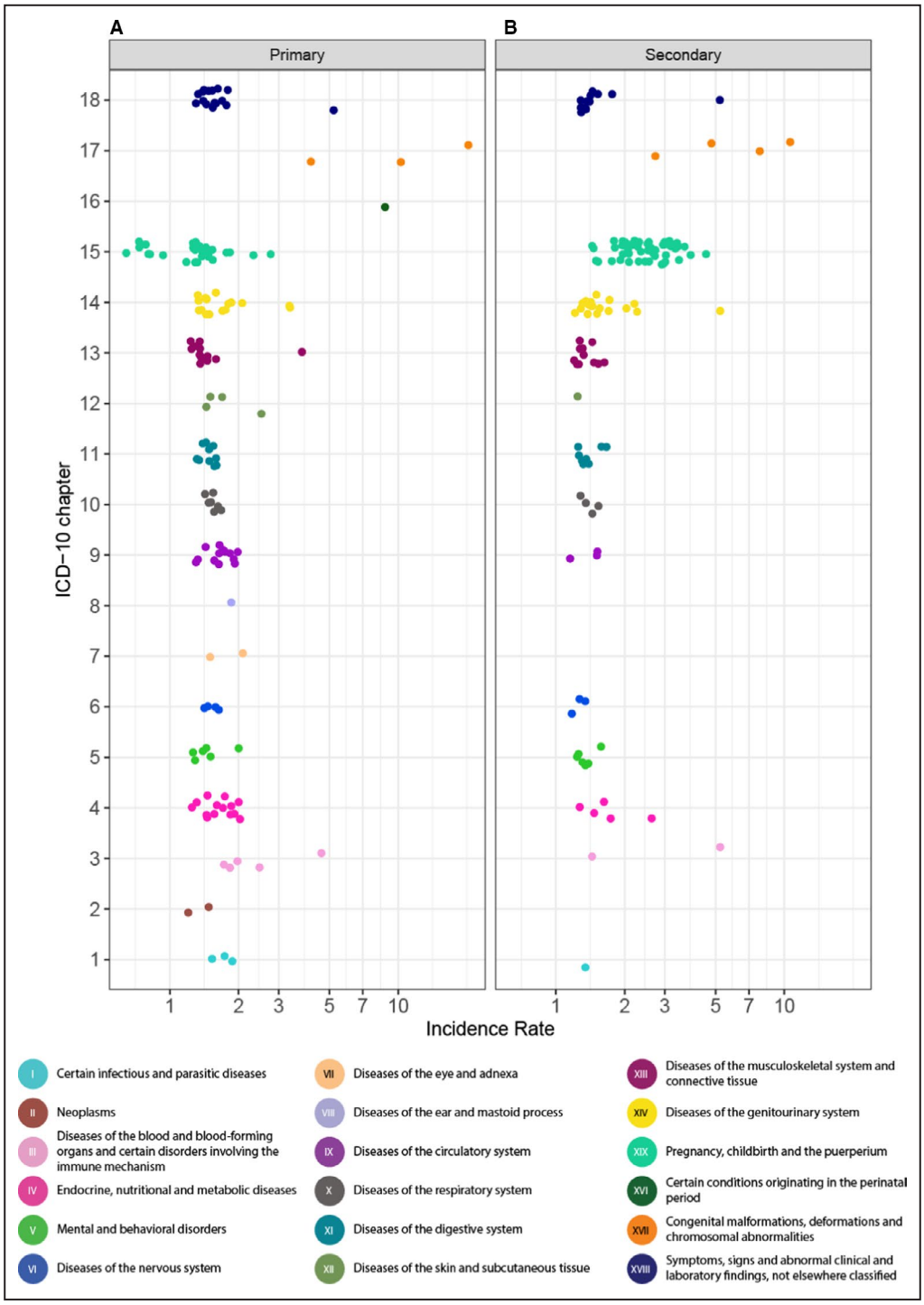


Figure 1. Diagnoses occurring more frequently following recurrent pregnancy loss (RPL) across 18 *ICD-10* chapters.
The points have been scattered in the vertical direction to improve readability. Each point represents the median value from the posterior distribution. **A**, Diagnoses with an elevated risk for women with primary RPL (180 diagnoses). **B**, Diagnoses with an elevated risk for women with secondary RPL (172 diagnoses). *ICD-10* indicates *International Classification of Diseases, Tenth Revision.*

These included coagulation disorders, congenital malformations of the female genital organs, and autoimmune diseases. The early component also contained obstetric complications, with the majority seen only after secondary RPL. We found no evidence that women with RPL had an increased risk of malignancies, irrespective of the subtype. The large population of women, both with and without RPL, allowed us to

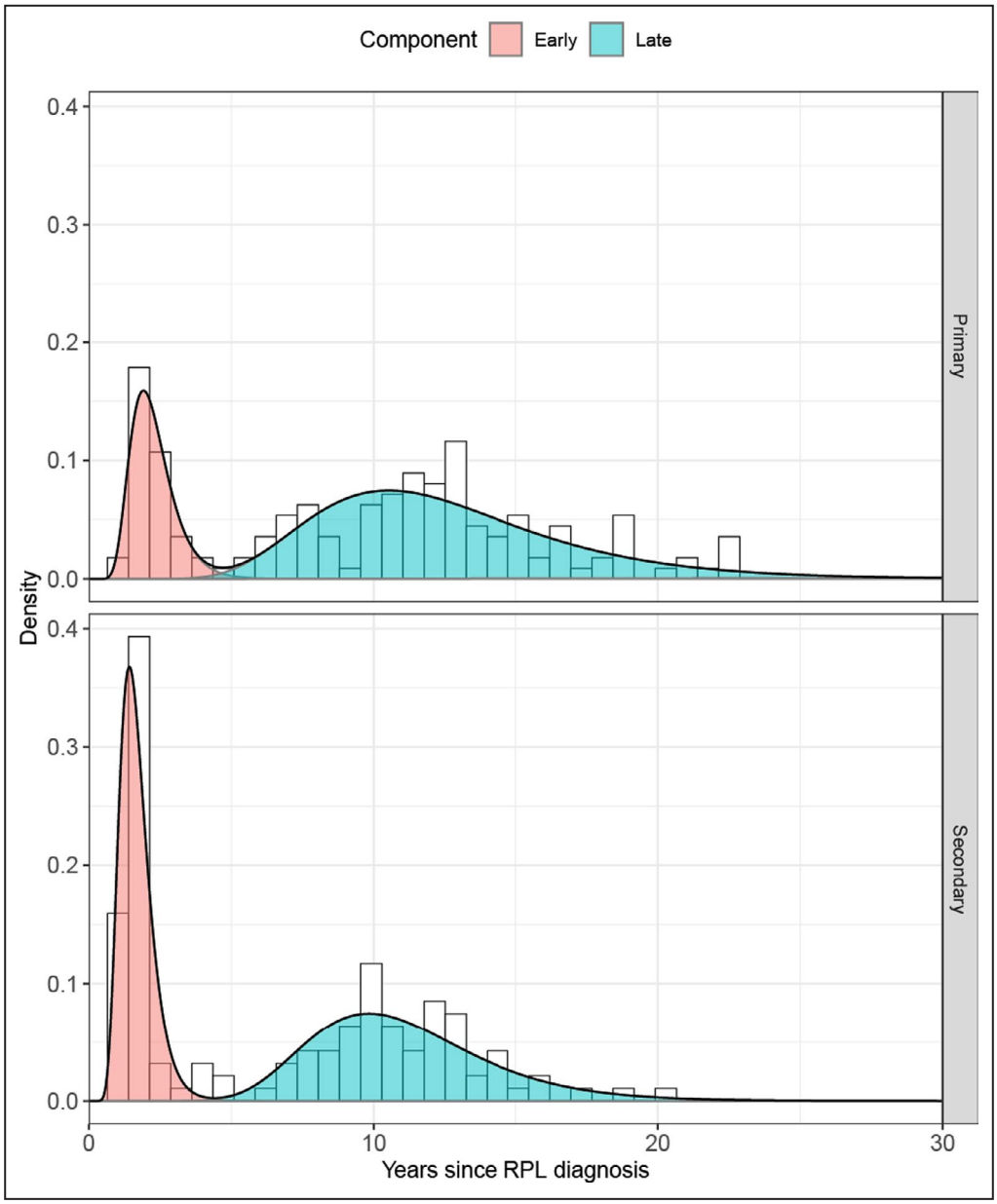


Figure 2. Distribution of the median time between a recurrent pregnancy loss (RPL) diagnosis and 1 of the 180 and 172 diagnoses that are more frequent following primary RPL and secondary RPL, respectively.
The histogram indicates the observed data points, whereas the 2 density plots indicate the 2 components of the mixture model.

investigate late-state complications in a phenome-wide manner. This has historically not been possible, because of the small cohort sizes often used. The findings we present here are of wide interest as we identify primary and secondary RPL as a risk factor and an early indicator of later disease.

The basis for this study was nationwide data collected over a 39-year period. Reporting to the Danish registries used in this study is mandatory. An important source of bias is unregistered PLs. This includes PLs handled at home, in general or gynecological practice. This leads to some women being erroneously classified as not having RPL, leading to an underestimation of the IRR. Furthermore, the RPL definition depends on the reliability of the identification of PLs in the Danish National Patient Registry. However, this condition has previously been estimated to have a positive predictive value of 97%.[26] Lack of information from the Danish Medical Birth Register could result in women being classified wrongly as primary and secondary RPL; nonetheless, the Danish Medical Birth Register is considered

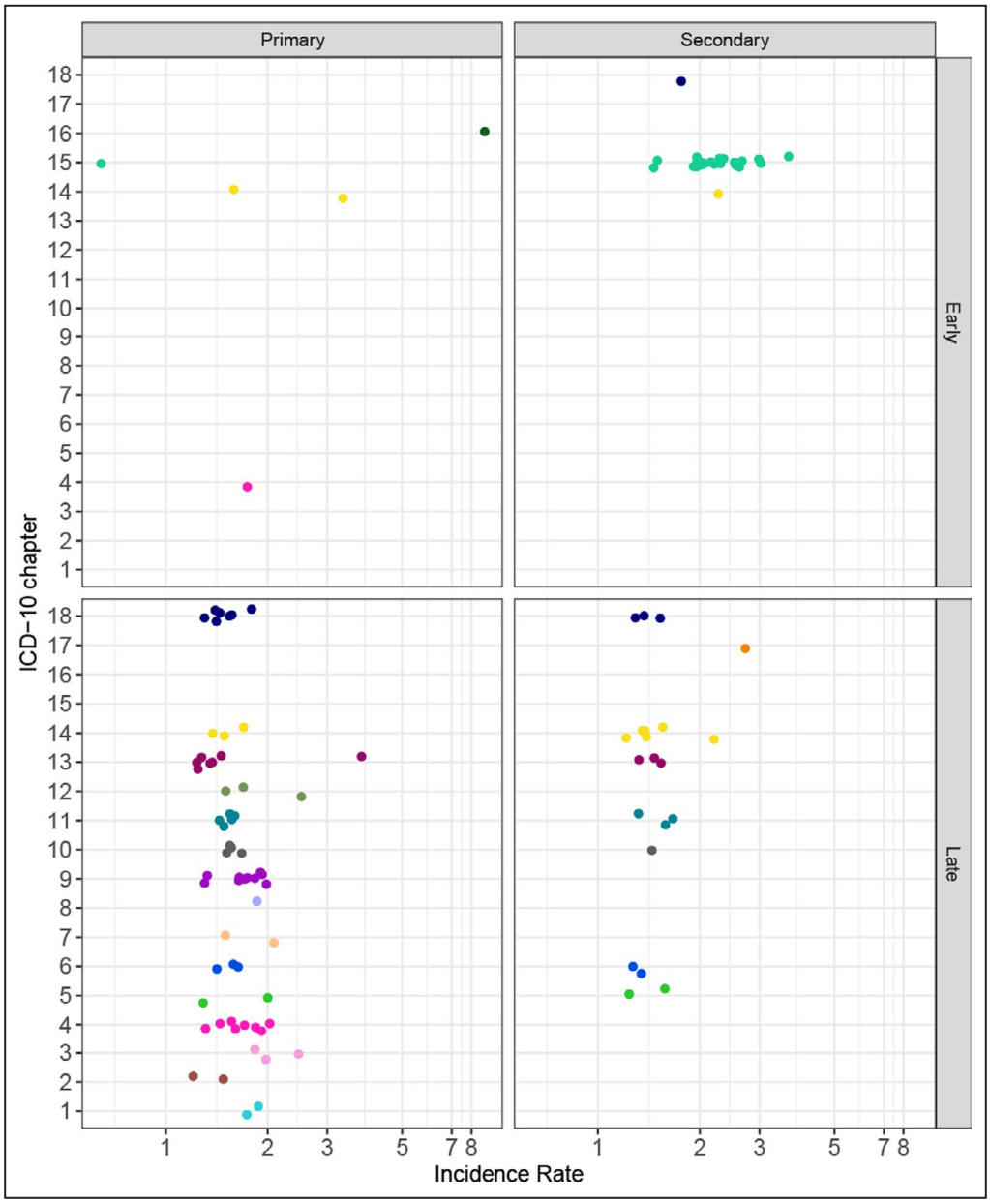


Figure 3. Diagnoses unique to either primary or secondary recurrent pregnancy loss, divided into the 2 "early" and "late" groups determined from the mixture model analysis.
Each point represents the median value from the posterior distribution. The coloring scheme is as in Figure 1. *ICD-10* indicates *International Classification of Diseases, Tenth Revision.*

to be complete.[27] The registries do not contain complete information on smoking status or body mass index. Additionally, we did not have any information on socioeconomic status or whether the partner was the same for all pregnancies, which could confound our findings. Furthermore, we did not include the disease history before RPL, which could also confound the results. Still, many women experiencing RPL are not properly investigated before referral to a specialist clinic, and it is therefore not possible to take this into account. We also note that the infant and maternal mortality in Denmark is extremely low.[28,29] This is partially because of universal health care and referral to specialist clinics involved in the monitoring and treatment of, for example pregnant women with diabetes mellitus and heart disease. Therefore, we would not expect to see many PLs in the cohort because of, for example, uncontrolled diabetes mellitus. Additionally, there are no known causal factors for RPL with the exception of embryonal malformations, and it is thus difficult to control for pre-existing conditions related to RPL. The probability of an early PL is highly correlated with age. Since we did not have information on the karyotype of the product

of conception, we have tried to minimize the confounding effect from aneuploidy by only including women younger than 40 years of age. Moreover, because of the lack of paternal information, we could not account for those cases where there is a paternal chromosomal translocation. A mixture of cases because of aneuploidy or paternal translocations would weaken the signal, and we could thereby have missed some associations, or underestimated the relative risks. This study was based on population data from a 39-year period. While this is helpful for a long follow-up period, changes and developments in medical practice may influence the results. However, we did attempt to mitigate this effect by matching women born in the same year. Lastly, because of the large number of outcomes studied in this phenome-wide analysis, we cannot rule out chance findings. Nevertheless, we have attempted to mitigate this by advanced statistical modeling and we also note that the large number of diagnoses associated with RPL remains striking and is consistent within certain domains of disease, which merits further investigation.

We found that primary and secondary RPL, when considered separate phenotypes, are followed by a unique spectrum of complications later in life. Our observations in the domain of CVDs and obstetric complications are in accordance with prior studies.[12,13,30,31] Yet, previous studies of CVD and PL have not been stratified by subtype, and here we demonstrate that primary RPL drives the observed associations. This is an important distinction, because this could point towards shared pathophysiology only present in women with primary RPL and could serve as the basis for future screening and risk-assessment profiles. Furthermore, we found that systemic lupus erythematosus (SLE), a recognized risk factor for RPL, also was significantly more frequent in women after an RPL diagnosis and belonged to the "late" component. This can be explained in at least 2 different ways. First, if SLE and RPL share a common pathophysiological cause, our findings indicate that this cause has not always manifested fully at the time of RPL diagnosis and evaluation or it could be that RPL itself increases the risk of SLE, possibly through increased microchimerism.[32,33] Lastly, the increased risk of mental disorders and substance abuse could possibly be mitigated by a closer follow-up and referral to therapy.

There are large discrepancies across health in men and women.[34] Here, we have tried to elucidate how the pregnancy history contributes to disease. The identification of complications across the full spectrum of disease present a new set of challenges that must be examined in-depth clinically to uncover the cause and close the gap in health care between men and women. Identification of clinically relevant subtypes is an important aspect of precision medicine, both to tailor screening and therapy to reduce the disease burden, but also to prevent overdiagnosis and unnecessary invasive procedures in an already vulnerable patient group. We have demonstrated this aspect in our distinction between primary and secondary RPL, in which CVD only had an increased risk in 1 subtype. Indeed, RPL may very well be an early initiator of disease, disease maintenance, and disease progression. However, the exact relationship between RPL and chronic diseases remains to be elucidated. Most importantly, our results indicate that RPL is an early marker of a wide range of diseases. We therefore speculate that, in the future, fertility history will be a core part of most risk prediction models across diverse clinics in health care. Considering the full pregnancy history including prior PLs is thus relevant when evaluating and predicting the future risk of disease in women.

ARTICLE INFORMATION

Received November 15, 2019; accepted March 4, 2020.

Affiliations

From the Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences (D.W., A.P.N., S.B.), Department of Public Health, Faculty of Health and Medical Sciences (L.H.M.), and Department of Clinical Medicine (H.S.N.), University of Copenhagen, Denmark; Methods and Analysis, Statistics Denmark, Copenhagen, Denmark (D.W., L.H.M.); Recurrent Pregnancy Loss Unit, Fertility Clinic, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark (D.W., H.S.N.); The Recurrent Pregnancy Loss Unit, Department of Obstetrics and Gynaecology, Copenhagen University Hospital, Hvidovre Hospital, Copenhagen, Denmark (D.W., H.S.N.); Department of Gynecology and Obstetrics, Rigshospitalet, Copenhagen University Hospital, DK-2200 Copenhagen, Denmark (A.P.N.).

Sources of Funding

The work is carried out as a part of the BRIDGE – Translational Excellence Programme (bridge.ku.dk) at the Faculty of Health and Medical Sciences, University of Copenhagen, funded by the Novo Nordisk Foundation (grant agreement NNF18SA0034956). Funding from other Novo Nordisk Foundation grants (NNF14CC0001 and NNF17OC0027594) and the Ole Kirk Foundation and Rigshospitalet's Research Fund is also acknowledged.

Disclosures

None.

Supplementary Materials

Data S1
Tables S1–S2
Figure S1
References 35–47

REFERENCES

1. Roepke ER, Matthiesen L, Rylance R, Christiansen OB. Is the incidence of recurrent pregnancy loss increasing? A retrospective register-based study in Sweden. *Acta Obstet Gynecol Scand*. 2017;96:1365–1372.
2. Bashiri A, Harlev A, Agarwal A, eds. *Recurrent Pregnancy Loss*. Cham, Switzerland: Springer International Publishing; 2016.
3. Jauniaux E, Farquharson RG, Christiansen OB, Exalto N. Evidence-based guidelines for the investigation and medical treatment of recurrent miscarriage. *Hum Reprod*. 2006;21:2216–2222.
4. Ogasawara M, Aoki K, Okada S, Suzumori K. Embryonic karyotype of abortuses in relation to the number of previous miscarriages. *Fertil Steril*. 2000;73:300–304.
5. Lund M, Kamper-Jørgensen M, Nielsen HS, Lidegaard Ø, Andersen A-MN, Christiansen OB. Prognosis for live birth in women with recurrent miscarriage. *Obstet Gynecol*. 2012;119:37–43.

6. Egerup P, Kolte AM, Larsen EC, Krog M, Nielsen HS, Christiansen OB. Recurrent pregnancy loss: what is the impact of consecutive versus non-consecutive losses? *Hum Reprod*. 2016;31:2428–2434.
7. Christiansen O, Kolte A, Nielsen H. Secondary recurrent miscarriage—a unique entity with respect to etiology and treatment. *Curr Womens Health Rev*. 2006;2:119–124.
8. Shapira E, Ratzon R, Shoham-Vardi I, Serjienko R, Mazor M, Bashiri A. Primary vs. secondary recurrent pregnancy loss—epidemiological characteristics, etiology, and next pregnancy outcome. *J Perinat Med*. 2012;40:389–396.
9. El Hachem H, Crepaux V, May-Panloup P, Descamps P, Legendre G, Bouet P-E. Recurrent pregnancy loss: current perspectives. *Int J Womens Health*. 2017;9:331–345.
10. Puscheck EE, Jeyendran RS. The impact of male factor on recurrent pregnancy loss. *Curr Opin Obstet Gynecol*. 2007;19:222–228.
11. Peters SAE, Yang L, Guo Y, Chen Y, Bian Z, Tian X, Chang L, Zhang S, Liu J, Wang T, et al. Pregnancy, pregnancy loss, and the risk of cardiovascular disease in Chinese women: findings from the China Kadoorie Biobank. *BMC Med*. 2017;15:148.
12. Oliver-Williams CT, Heydon EE, Smith GCS, Wood AM. Miscarriage and future maternal cardiovascular disease: a systematic review and meta-analysis. *Heart*. 2013;99:1636–1644.
13. Kharazmi E, Dossus L, Rohrmann S, Kaaks R. Pregnancy loss and risk of cardiovascular disease: a prospective population-based cohort study (EPIC-Heidelberg). *Heart*. 2011;97:49–54.
14. Charach R, Sheiner E, Beharier O, Sergienko R, Kessous R. Recurrent pregnancy loss and future risk of female malignancies. *Arch Gynecol Obstet*. 2018;298:781–787.
15. Heida KY, Bots ML, de Groot CJ, van Dunné FM, Hammoud NM, Hoek A, Laven JS, Maas AH, Roeters van Lennep JE, Velthuis BK, et al. Cardiovascular risk management after reproductive and pregnancy-related disorders: a Dutch multidisciplinary evidence-based guideline. *Eur J Prev Cardiol*. 2016;23:1863–1879.
16. Mikkelsen AP, Egerup P, Ebert JFM, Kolte AM, Nielsen HS, Lidegaard Ø. Pregnancy loss and cancer risk: a nationwide observational study. *EClinicalMedicine*. 2019;15:80–88.
17. Smith G, Wood A, Pell J, Hattie J. Recurrent miscarriage is associated with a family history of ischaemic heart disease: a retrospective cohort study. *BJOG*. 2011;118:557–563.
18. Ranthe MF, Diaz LJ, Behrens I, Bundgaard H, Simonsen J, Melbye M, Boyd HA. Association between pregnancy losses in women and risk of atherosclerotic disease in their relatives: a nationwide cohort study. *Eur Heart J*. 2016;37:900–907.
19. Parazzini F, Esposito G, Tozzi L, Noli S, Bianchi S. Epidemiology of endometriosis and its comorbidities. *Eur J Obstet Gynecol Reprod Biol*. 2017;209:3–7.
20. Ziller V, Heilmaier C, Kostev K. Time to pregnancy in subfertile women in German gynecological practices: analysis of a representative cohort of more than 60,000 patients. *Arch Gynecol Obstet*. 2015;291:657–662.
21. Hanson B, Johnstone E, Dorais J, Silver B, Peterson CM, Hotaling J. Female infertility, infertility-associated diagnoses, and comorbidities: a review. *J Assist Reprod Genet*. 2017;34:167–177.
22. Grandi SM, Filion KB, Yoon S, Ayele HT, Doyle CM, Hutcheon JA, Smith GN, Gore GC, Ray JG, Nerenberg K, et al. Cardiovascular disease-related morbidity and mortality in women with a history of pregnancy complications. *Circulation*. 2019;139:1069–1079.
23. Kolte AM, Olsen LR, Mikkelsen EM, Christiansen OB, Nielsen HS. Depression and emotional stress is highly prevalent among women with recurrent pregnancy loss. *Hum Reprod*. 2015;30:777–782.
24. Rai R, Regan L. Recurrent miscarriage. *Lancet*. 2006;368:601–611.
25. Pedersen CB. The Danish civil registration system. *Scand J Public Health*. 2011;39:22–25.
26. Lohse SR, Farkas DK, Lohse N, Skouby SO, Nielsen FE, Lash TL, Ehrenstein V. Validation of spontaneous abortion diagnoses in the Danish National Registry of Patients. *Clin Epidemiol*. 2010;2:247.
27. Bliddal M, Broe A, Pottegård A, Olsen J, Langhoff-Roos J. The Danish Medical Birth Register. *Eur J Epidemiol*. 2018;33:27–36.
28. Kassebaum NJ, Barber RM, Bhutta ZA, Dandona L, Gething PW, Hay SI, Kinfu Y, Larson HJ, Liang X, Lim SS, et al. Global, regional, and national levels of maternal mortality, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388:1775–1812.
29. Wang H, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, Abraha HN, Abu-Raddad LJ, Abu-Rmeileh NME, et al. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390:1084–1150.
30. Ranthe MF, Andersen EAW, Wohlfahrt J, Bundgaard H, Melbye M, Boyd HA. Pregnancy loss and later risk of atherosclerotic disease. *Circulation*. 2013;127:1775–1782.
31. Nielsen HS, Steffensen R, Lund M, Egestad L, Mortensen LH, Andersen AMN, Lidegaard O, Christiansen OB. Frequency and impact of obstetric complications prior and subsequent to unexplained secondary recurrent miscarriage. *Hum Reprod*. 2010;25:1543–1552.
32. Adams KM, Nelson JL. Microchimerism. *JAMA*. 2004;291:1127.
33. Hovinga ICLK, Koopmans M, Baelde HJ, van der Wal AM, Sijpkens YWJ, de Heer E, Bruijn JA, Bajema IM. Chimerism occurs twice as often in lupus nephritis as in normal kidneys. *Arthritis Rheum*. 2006;54:2944–2950.
34. Westergaard D, Moseley P, Sørup FKH, Baldi P, Brunak S. Population-wide analysis of differences in disease progression patterns in men and women. *Nat Commun*. 2019;10:666.
35. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449–490.
36. Gjerstorff ML. The Danish Cancer Registry. *Scand J Public Health*. 2011;39:42–45.
37. Knudsen LB, Olsen J. The Danish Medical Birth Registry. *Dan Med Bull*. 1998;45:320–323.
38. Pedersen MK, Pedersen HK, Simon C, Eriksson R, Sørup FKH, Damgaard KA, Birch AM, Larsen M, Nielsen AP, Brunak S. Mapping of all international and Danish ICD-8 disease codes to ICD-10 codes. Technical report.
39. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. Stan: a probabilistic programming language. *J Stat Softw*. 2017;76. Available at: https://www.jstatsoft.org/v076/i01.
40. Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res*. 2014;15:30.
41. Betancourt M. A conceptual introduction to Hamiltonian Monte Carlo. *arxiv*. 2017. Available at: https://arxiv.org/abs/1701.02434.
42. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7:457–472.
43. Stan Development Team. Stan modeling language: user's guide and reference manual. Available at: http://mc-Stan.Org/. Accessed January 2, 2020.
44. Betancourt M. Diagnosing biased inference with divergences. 2017.
45. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3rd ed. New York: Chapman and Hall/CRC; 2013.
46. Kruschke JK. *Doing Bayesian Data Analysis: A Tutorial With R, JAGS, and Stan*. 2nd ed. Saint Louis: Elsevier Science & Technology; 2014.
47. Gelman A, Hill J, Yajima M. Why we (usually) don't have to worry about multiple comparisons. *J Res Educ Eff*. 2012;5:189–211.

SUPPLEMENTAL MATERIAL

Data S1.

Supplemental Methods

Registry Data

The basis for the study was the Danish National Patient Registry (NPR), the Danish Psychiatric Registry (PSYK), the Danish Cancer Registry (TUMOR), and the Danish Medical Birth Registry (DMBR).[35–37] Hospital admissions from NPR and PSYK included all inpatient, outpatient, and emergency contacts. Discharge diagnoses were coded according to ICD version 8 (ICD-8) in the period 1977-1993, and ICD-10 starting from 1994. ICD-8 codes were mapped to ICD-10 using the translation provided by Pedersen et al.[38] ICD-10 has a hierarchical structure and consists of four levels (chapters, blocks, level 3 codes, and level 4 codes). Level 4 codes are the most specific codes. Here, we used the level 3 codes. Not using the most specific code is a trade-off between power and specificity, as the most specific codes will be assigned to a lower number of people. The NPR and PSYK also contains operations and procedures, coding according to a Danish classification until 1995 (available at https://sundhedsdatastyrelsen.dk/da/rammer-og-retningslinjer/om-klassifikationer/sks-klassifikationer/download-sks, last visited 16th September 2018). Since 1995, operations have been coded according to the Nordic NOMESCO Classification of Surgical Procedures. Procedures are coded according to a Danish classification, (available at http://www.medinfo.dk/sks/brows.php?s_nod=1616, last visited 16th September 2018) starting from 1995. The Danish Cancer Registry is a population-wide registry that was established in 1943. Reporting to the registry has been mandatory since 1987. The diagnoses from the Danish Cancer Registry has been coded using ICD-10 since 1978 (in 2003 codes from 1978-2003 was translated into ICD-10). The MFR has collected information on all births in Denmark since 1973. This includes date of birth, gestational age, outcome of birth (liveborn / stillborn), and parental information. Women living in Greenland or the Faroe Islands were excluded from the analysis since the Danish National Patient Registry (DNPR) does not cover these countries.

Bayesian Estimation and Inference

All models were fitted using Stan v. 2.18.0 (http://mc-stan.org/).[39] Stan is an open-source implementation of the Hamiltonian Monte Carlo No-U-Turn-Sampler (HMC-NUTS).[40,41] HMC-NUTS is a Markov Chain Monte Carlo (MCMC) algorithm. For each model, four chains were run in parallel with different starting seeds. Each chain was run for 20,000 steps, of which 10,000 was used for warm-up and discarded prior to inference, with default parameters. HMC-NUTS uses the warm-up steps to tune the hyper-parameters of the algorithm. The number of steps is much lower, compared to e.g. Gibbs sampling. This is due to the HMC-NUTS being much more efficient at exploring the posterior distribution and thus reaching the stationary distribution faster.[40] We assessed converge of each model by computing the R-hat statistics, inspecting the tree-depth, and counting the number of divergences. The R-hat number is a quantitative description of the within and between chain variation for the different MCMC chains for each parameter, and is sometimes also referred to as the potential scale reduction factor.[42] A large R-hat value indicates non-convergence. If the chains have converged, the R-hat value will be close to one. The tree-depth is a HMC-NUTS specific characteristic that puts an upper limit on the number of computations in each iteration. If the max tree-depth is reached post-warmup the HMC-NUTS will be similar to a random-walk MCMC, and can bias estimates.[43] Lastly, divergences are a count of how many times HMC-NUTS had numerical problems (e.g. underflowing / overflowing, division by zero). This can be due to a bad model, problems in data, or ill-suited priors. Divergences are serious, as HMC-NUTS cannot explore the region of the posterior where they happen.[44]

Here we define convergence if all R-hat values < 1.1, the tree-depth is not maxed at any point post warm-up, and there are no divergences. Inference was made by summarizing the posterior distribution into the Bayesian Uncertainty Interval (UI) and the probability that the interval is different from zero.[45] We define the UI to be the 95% Highest Density Interval (HDI), which indicates the narrowest interval that spans 95% of the posterior distribution.[46] We calculated the P-value as the probability that the effect was equal to or less than zero when the estimated mean value was greater than zero, and vice versa when the estimated mean value was less than zero,

$$P = \begin{cases} \Pr(x \leq 0), & \bar{x} > 0 \\ \Pr(x \geq 0), & \bar{x} < 0 \end{cases}$$

The false discovery rate (FDR) was controlled using the Benjamini-Hochberg approach. We concluded that RPL was associated with an increased or decreased risk of later diagnosis if the FDR did not exceed 5%.

Incidence Rate Ratio Estimation

Estimation of the incidence rate ratio (IRR), was done using a Poisson log-linear model. We estimated the IRR by selecting a matched comparison group. Age was controlled by grouping the year of birth into twelve five-year intervals (Table S1). For every woman with RPL, twenty women without RPL were selected matched by the year of birth and number of live births priors to the matched time. If the outcome had occurred prior to the RPL diagnosis date, the woman was excluded from the analysis. Likewise, if the outcome had occurred prior to the exposure date in a matched woman, the woman was not eligible for the matched comparison group.

The IRR was estimated using a Poisson log-linear model specified as shown in **Equation 1** and **Equation 2**. The Poisson model is a piece-wise exponential survival model that approximates a Cox model under certain assumptions, one of them being a constant hazard over time. To relax this assumption, we divided the total time into three year intervals and included a covariate for each segment.

$$\boldsymbol{y_i \sim Poisson}(\mathbf{exp}(\boldsymbol{\eta_i}))$$

Equation 1

$$\boldsymbol{\eta_i = \beta_0 + \beta_{RPL} * x_{RPL} + \sum_i \beta_i * x_i + \sum_j \beta_j * x_j + \sum_p \beta_p * x_p + \log(c)}$$

Equation 2

In the equation, y_i is the number of women with the outcome, x_{RPL} is an indicator variable for the RPL status, i is one of the fourteen hazard intervals (setting the first interval as reference), j is one of the eleven age bins defined in **Table S1 (**using the seventh bin as the reference level), and x_j is an indicator variable for the age group, p is the parity group (1, 2, >=3, 1 being the reference level) and x_p is an indicator for the parity group, c is the total number of person-years within that stratum. The parity coefficient was only included in the model for secondary RPL. The first age bin comprised years prior to 1945, as there were only very few women in the cohort born prior to 1945. The last age bin comprised everyone born in 1991 and onwards, as very few women with RPL was born after 1991.

From the Poisson model, the parity and age-adjusted IRR is equal to the exponentiated value of the coefficient, i.e. exp (β_{RPL}).

Table S1. Age bins used for adjusting the RR for age.

Age bin	Years
1	<=1945
2	1946-1950
3	1951-1955
4	1956-1960
5	1961-1965
6	1966-1970
7	1971-1975

8	1976-1980
9	1981-1985
10	1986-1990
11	>=1991

All outcomes, subscripted *A*, are jointly modelled in a Bayesian hierarchical model, in which each outcome has its own set of coefficients for the model. To complete the Bayesian model specification, we assign a prior to each coefficient in the model, Equation3-Equation 10.

$$\beta_{0,A} \sim N(0,1)$$

Equation 3

$$\beta_{i,A} \sim N(0,1)$$

Equation 4

$$\beta_{j,A} \sim N(0,1)$$

Equation 5

$$\beta_{p,A} \sim N(0,1)$$

Equation 6

$$\beta_{RPL,A} \sim t(7,0,\sigma_\tau)$$

Equation 7

$$\sigma_\tau \sim N_+(0,\sigma_\omega)$$

Equation 8

$$\sigma_\omega \sim N_+(0,1)$$

Equation 9

The priors for the intercept (β_0), time segment (β_i), age-bin (β_j), and parity (β_{parity}) were assigned weakly informative normal distributions centered around zero. The prior for the coefficient of interest for each outcome, β_{RPL}, is assigned a student-t distribution with seven degrees of freedom, centered around zero, with a standard deviation determined from the data. The centering around zero a priori expects that the IRR is equal to one, given that $\exp(0) = 1$. The student-t distribution also has a large body of its probability mass close to zero, further augmenting the choice. The long tails may accommodate values that may fall outside a normal distribution.

Rather than using an unpooled prior, that is, an independent prior on each outcome, we instead encoded the ICD-10 hierarchy into the prior. E.g., diagnoses regarding the circulatory system all have the same prior. In total, we included eighteen ICD-10 chapters in the present study, each indexed by τ in Equation 8. This is a natural grouping of diagnosis based on anatomical or functional group, and may improve how well the model fits the data, thereby leading to better inference. The hierarchical prior induces regularization of estimates, a process known as Bayesian shrinkage.[47]. The shrinkage process can safeguard against spurious findings, which can sometimes be driven by small sample sizes where it is difficult to get a sufficiently certain estimate of the coefficients. To ascertain the effect of the shrinkage procedure, we compared the beta values for the RPL coefficient estimated from our Bayesian procedure, with the ones estimated from a Maximum Likehood Estimate (Figure S1). We observed that in the vast majority of cases, the values estimated using the full Bayes procedure are below the diagonal, indicating that they are being shrunk towards the null value. We therefore conclude our shrinkage procedure works as expected.

Lastly, we randomly shuffled the exposure status (RPL) to ascertain that there were no errors in data. In the randomly shuffled data set, we did not observe any increased incidence rate ratios.

Mixture Models

A log-normal mixture model was employed to model the multimodal distributions observed in the time between an RPL diagnosis and complications. A mixture model is a probabilistic model that assumes that the observed data has been generated by a discrete number of distributions, *k*, **Equation 10**.

$$\mathbf{f(x) = \sum_{k} \theta_k * lnorm(\mu_k, \sigma_k)}$$

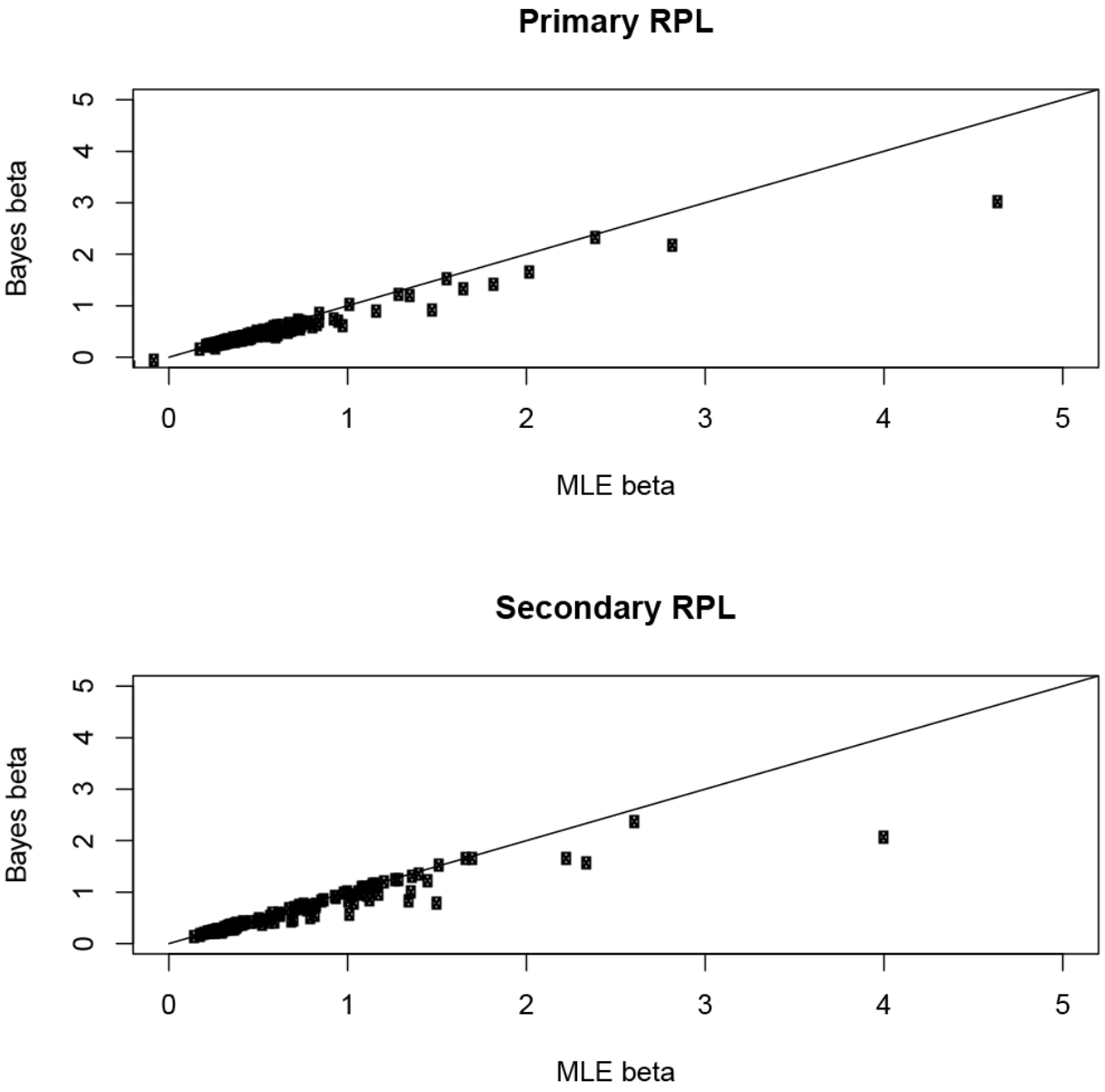
Equation 10

A log-normal distribution was chosen since only the time after an RPL diagnosis was investigated, and hence the distribution was bounded to positive numbers. Each diagnosis with an elevated RR was considered a data point, and the time between the RPL and diagnosis was summarized as the median across all patients due to the heavy tail arising from the long follow-up period. The model assigns a probability to each diagnosis of belonging to a specific component in the model. A diagnosis was assigned to the component that had the highest probability. The model was fit using the R library mixR v2.3.14. The number of components was selected using the Bayesian Information Criteria (BIC, Table S2). Inspection of the model post-fit was performed to visually determine if the fit looked reasonable.

Table S2. Bayesian Information Criteria for mixture model.

	K=1	K=2	K=3
Primary	991.75	**910.76**	918.19
Secondary	745.17	**631.00**	635.40

Figure S1. Shrinkage plot comparing the Bayesian median estimate of the beta value versus the maximum likelihood estimate.



The values for the beta coefficient are typically below the diagonal, indicating that the values are being shrunk towards the null value of the prior.