Optimizing an algorithm for the identification and classification of pregnancy outcomes in German claims data

Nadine Wentzell, Tania Schink, Ulrike Haug, Sandra Ulrich, Marieke Niemeyer, Rafael Mikolajczyk

DOI
10.1002/pds.4588

Published in
Pharmacoepidemiology and Drug Safety

Document version
Accepted manuscript

Online publication date
18 July 2018

Corresponding author
Nadine Wentzell

Citation
Wentzell N, Schink T, Haug U, Ulrich S, Niemeyer M, Mikolajczyk R. Optimizing an algorithm for the identification and classification of pregnancy outcomes in German claims data. Pharmacoepidemiol Drug Saf. 2018;27(9):1005-10.

Optimizing an Algorithm for the Identification and Classification of Pregnancy Outcomes in German Claims Data

Running head: Pregnancy Outcomes in German Claims Data

Nadine Wentzell[1]; Tania Schink[1]; Ulrike Haug[1,2], Sandra Ulrich[3]; Marieke Niemeyer[3]; Rafael Mikolajczyk[4]

[1] Department of Clinical Epidemiology, Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany

[2] Faculty of Human and Health Sciences, University of Bremen, Bremen, Germany

[3] Department of Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany

[4] Institute for Medical Epidemiology, Biometry and Informatics, Martin-Luther-University Halle-Wittenberg, Halle, Germany

Corresponding author

Nadine Wentzell

Leibniz Institute for Prevention Research and Epidemiology - BIPS

Achterstraße 30

28359 Bremen

Germany

E-Mail: wentzell@leibniz-bips.de

Phone: +49 421 218-56879

Fax: +49 421 218-56941

Keywords: Pharmacoepidemiology**,** Pregnancy outcomes, German claims data

Key points

- We optimized an existing algorithm to specifically identify and classify births in German claims data, which is the first essential methodological prerequisite for using these data for the subsequent study of drug utilization and safety during pregnancy.
- Overall, we identified 1,235,261 pregnancy outcomes, of those 94% live births (preterm (10%), term (78%) and (12%) births after the expected delivery date), 0.3% still births, 3.6% induced abortions, 1.3% ectopic pregnancies and 0.5% spontaneous abortions.
- Most results were plausible regarding the age distribution (median age 32 years) and sequence of outcomes (>99%) and in agreement (95%) with case profile review by clinical experts.

Word count

3,235

Name(s) of any sponsor(s) of the research contained in the paper, along with grant number(s)

No external funding source.

Statement about prior postings and presentations

None

Abstract

Purpose: For studying drug utilization and safety in pregnancy based on administrative health care data, the reliable identification and classification of pregnancy outcomes in the data is essential. We aimed to optimize an existing algorithm for the identification and classification of pregnancy outcomes in the German Pharmacoepidemiological Research Database (GePaRD) with a particular focus on births.

Methods: We reconsidered all codes used by the original algorithm and applied it to data of GePaRD from 2006 to 2014. Longitudinal records of pregnancies were used to identify targets for enhancing the algorithm's specificity. We checked the plausibility of the results, e.g. regarding the age distribution of persons with pregnancy outcomes. Based on 20 longitudinal records of pregnancies, we compared the outcome classification by clinical experts with the results of the modified algorithm.

Results: Our algorithm identified 1,235,261 pregnancy outcomes in the database, with the majority (94%) being live births, classified as preterm (10%), term (78%) and (12%) births after the expected delivery date. The median age of pregnant women was 32 years (Q1 28; Q3 35). Implausible sequence of outcomes (for example an induced abortion within a pregnancy categorized as ending in a live birth) were rare (0.03%). The case profile review by clinical experts resulted in the same outcome type and date as the algorithm in 95%.

Conclusion: Our algorithm led to plausible results regarding the identification and classification of pregnancy outcomes. It will be an important foundation for studies on drug utilization and drug safety during pregnancy based on GePaRD.

Introduction

A considerable proportion of women use drugs during pregnancy, but at the time of market approval the safety of drugs during pregnancy is mostly unknown[1-3]. For an appropriate benefit-risk assessment of drug treatment during pregnancy, post-marketing safety evaluation is of major public health importance. Administrative health care (i.e. claims) databases are an important data source for studying drugs during pregnancy as they allow monitoring drug utilization and investigating the association between individual exposures and pregnancy outcomes in a timely manner and avoid recall bias[4,5]. Before a claims data base can be used for drug utilization and safety studies, the following three methodological procedures need to be established: (i) the reliable identification of pregnancy outcomes such as live births including preterm, term births, and births after the expected delivery date and classification of still births, induced or spontaneous abortions, and ectopic pregnancies[6] and their outcome date, (ii) the estimation of the beginning of pregnancy to assess gestational age at exposure during pregnancy and (iii) a linkage of the mothers' and the children's records to conduct drug safety studies focusing on the health of the child, e.g. malformations or developmental disorders.

In our study, we focused on the first of these methodological procedures. For the identification of pregnancy outcomes in claims or medical record data, particular caution is required because on person-level several codes like diagnoses and procedures may represent the same pregnancy outcome or diagnoses may be coded for administrative purposes. For example, a diagnosis could be recorded upon suspicion or to justify treatment (e.g. "preterm birth" as the reason for a hospital stay to avoid preterm birth). Furthermore, coding errors, implausible codes and implausible sequence of codes like an induced abortion shortly before a live birth may occur. Algorithms are therefore needed to identify and classify pregnancy outcomes and assign their date, taking into account the reliability and plausibility of the respective codes and their combinations. Mikolajczyk et al. (2013)[6] have developed such an algorithm for the German Pharmacoepidemiological Research Database (GePaRD). We aimed to optimize this algorithm in terms of increasing its specificity, with a particular focus on live births for the subsequent analysis of drug-mediated outcomes in the child.

Methods

Data source and study population

We used data from 2006 to 2014 from GePaRD which contains information from four statutory health insurance providers (SHI) covering more than 20 million people from all regions in Germany. Data available in GePaRD include sex and year of birth, inpatient delivery dates, inpatient diagnoses (i.e. admission, discharge, and potentially secondary diagnoses) and outpatient diagnoses according to the German Modification of the International Classification of Diseases (ICD-10-GM), in- and outpatient procedures according to the Operations and Procedures Coding System (OPS), outpatient services, procedures according to the Doctors' Fee Scale within the Statutory Health Insurance Scheme (Einheitlicher Bewertungsmaßstab, EBM), and outpatient drug dispensations[1,7]. The sources differ in terms of the information provided, e.g. for outpatient procedures (OPS) only the quarter and the year are available.

Our study population included all persons with at least one pregnancy outcome between 2006 and 2014. The rationale for not restricting our population to women of reproductive age was to investigate how often pregnancy outcomes were coded in women outside of the childbearing age or in men.

Algorithm to identify pregnancy outcomes and their corresponding date in GePaRD[6]

For each pregnancy, a number of ICD, OPS, and EBM codes (and respective dates) are recorded over time in claims data, providing information that may be consistent or redundant but that could also be inconsistent (e.g. a live birth shortly after an induced abortion). To select the most plausible and valid information and to assign the correct date to the outcome, the previously developed algorithm[6] uses two nested hierarchical approaches. The first hierarchy determines an order for the different pregnancy outcome types, namely births, induced abortions, ectopic pregnancies, and spontaneous abortions (see Table 1, left column). This order is based on the assumption that indicators for births actually reflect births whereas ectopic pregnancies can also be coded in case of suspected (and later ruled out) diagnoses. Also, spontaneous abortions may be suspected diagnoses or could be coded as reasons for treatment to avoid the outcome. Therefore, the outcome “birth” with the respective date is considered most unambiguous and identified first. Furthermore, it is assumed that codes indicating induced abortions do not represent suspected diagnoses but that the induced abortion actually occurred. Hence, these outcomes are considered in the second place. As indicators for ectopic pregnancies may reflect suspected diagnoses but are more specific and less common than spontaneous abortions, they are identified in the third place. Finally, spontaneous abortions are assumed to be the most unspecific outcome and are therefore assessed last. In such way, if codes indicating ectopic pregnancies or spontaneous abortions were followed by live births within the same pregnancy, we could discard these earlier codes as not being true outcomes.

The second hierarchy is applied within each pregnancy outcome type. It takes into account that in the claims records there are typically various codes for a pregnancy outcome from different sources of information. This information may be redundant, but could also be conflicting regarding the date of the outcome. The algorithm therefore ranks the different sources of information according to their presumed level of validity when defining the date of the pregnancy outcome: inpatient delivery dates, inpatient OPS and ICD codes and outpatient EBM and OPS codes and their respective dates. Inpatient delivery dates, for example, are assumed to have the highest validity for births and specifically indicate the date of birth. Inpatient OPS codes are considered next because they also have a specific date. Inpatient diagnoses do not have an exact date, but generally, indicators from the inpatient setting are considered as more valid than from the outpatient sector. Outpatient OPS codes are considered last because they do not have a specific date.

Given that a woman can have multiple pregnancy outcomes during the study period - either of the same or a different type - it is essential to define which codes (and their dates) reflect the same pregnancy outcome and which sequences of codes are separate pregnancy outcomes. Hence, outcome-specific time frames, more specifically, a minimum “pregnancy duration” are defined as 259 days before a term birth (or birth after the expected delivery date), 154 days before a preterm or still birth, 49 days for an induced abortion and 42 days for ectopic pregnancies and spontaneous abortions. If more than one code for the same outcome occurs within these time frames, the most plausible outcome date is chosen specifically for each outcome type, e.g. for two delivery dates the date with an OPS code is assumed to reflect the true date of birth.

These time frames were also used to disregard codes according to the hierarchy of pregnancy outcomes, e.g. an ectopic pregnancy before a live birth.

Modification of algorithm

Our modification of the algorithm was performed in an iterative process: First, we assessed all codes used in the original algorithm[6] in cooperation with gynecologists and updated them (codes available from the authors). Based on longitudinal records of all pregnancy related outcome codes during manually selected pregnancies, we investigated the plausibility of the algorithm's outcome classification. This also allowed identifying targets to enhance the algorithm's specificity. We adapted the selection of codes and the structure of the algorithm accordingly. In detail, we limited the selection of ICD, OPS, and EBM codes to specific codes for each pregnancy outcome type. For example, our updated algorithm does not consider codes for providing care to a woman recently having given birth or wound infection after an obstetrical procedure as indicators for a birth because these codes often were used with a long delay after the pregnancy and thus might not be a good indicator of a birth. Regarding the identification of births, we restricted the inpatient diagnoses to main discharge and other main diagnoses. For induced and spontaneous abortions and ectopic pregnancies, we only included main discharge and other main diagnoses given that the review of longitudinal records during pregnancy showed inconsistencies when considering other types of inpatient diagnoses. In addition to the original algorithm, we also considered outpatient OPS codes for the identification of ectopic pregnancies. Regarding the classification of births, we introduced the category "birth after the expected delivery date" in order to be able to account for potentially longer pregnancy durations when estimating the beginning of pregnancy. To define births after the expected delivery date, we considered all inpatient diagnoses as it is very unlikely that these represent suspected diagnoses. To classify births as preterm and still births, we included all types of inpatient diagnoses except for admission diagnoses as those could be suspected diagnoses that are either ruled out or later confirmed in a main discharge diagnosis.

After several adaptations, we re-implemented the modified algorithm.

Plausibility check of the modified algorithm

We assessed the number of pregnancy outcomes and the frequency of the various types of pregnancy outcomes based on the modified algorithm. We evaluated the contribution of the various sources of information (i.e. delivery dates, inpatient OPS codes, etc.) in the determination of the outcome. To check the plausibility of the results of the modified algorithm, we analyzed the sex and age distribution of persons with a pregnancy outcome and the frequency of implausible sequences of outcomes within a pregnancy such as an induced abortion within a pregnancy ending in a live birth.

In addition, 20 randomly selected longitudinal records of codes during pregnancy covering all types of outcomes were reviewed by two clinical experts. For this review, we only selected pregnancies that have not been used to refine the algorithm. The aim of this review was to assess whether medical experts would use the same or other information for the identification and classification of pregnancy outcomes and process it in a similar way as our algorithm.

Results

Pregnancy outcomes: Identification, date assignment, and classification

Overall, the algorithm identified 1,235,261 pregnancy outcomes in the study period, of those 1,164,743 were live births (94.3%), 3,190 still births (0.3%), 44,013 induced abortions (3.6%), 16,659 ectopic pregnancies (1.3%) and 6,656 spontaneous abortions (0.5%).

Most of the live births (86.3%) were identified based on the source of information with the highest validity, i.e. an inpatient delivery date, and this information was used to assign the date of birth. For 12.5% of live births, an inpatient OPS code was used to determine the date of the outcome and for a minority (1.0%), inpatient ICD codes were used (Table 2). Live births in the outpatient setting (i.e. for which only outpatient codes were available) were rare (0.2%) and most of them were assigned an exact date based on an EBM code.

Overall, 10% of live births were classified as preterm, 78% as term, and 12% as birth after the expected delivery date.

The contribution of the various sources of information to determine the date of induced abortions, ectopic pregnancies and spontaneous abortions is shown in Table 2

Plausibility of the results

The median age at the end of the year of the pregnancy outcome for women was 32 years (interquartile range IQR 28-35, see Figure 1 for further information).

We observed a considerable number of codes for a pregnancy outcome recorded for newborns like an OPS code “postnatal care for a newborn”. This explained the vast majority (>99%) of pregnancy outcomes identified in male persons (i.e. codes were assigned to male newborns). The remaining number of pregnancy outcomes identified in men was 604 out of more than 1,150,000 entries. In these men, we identified a variety of different codes for pregnancy outcomes that may be - either mistakenly or for administrative reasons - recorded in the health record of the father. Some combinations, e.g. a single code for induction of labor in an 85-year old are obvious coding errors. Further coding errors may occur regarding the sex in core data. Another explanation for pregnancy outcomes in men may be actual pregnancies before changing sex.

The results of the algorithm showed few implausible sequences of outcomes, such as a code for an induced abortion during a pregnancy (=within 259 days before and up to 42 days after an outcome classified as live birth) which were initially identified, but discarded by the algorithm in 393 pregnancies.

For all 20 longitudinally analyzed records reviewed by two clinical experts (blinded to the results of the algorithm), the outcomes identified and classified by the experts were the same as the outcomes from the algorithm (100%). For 19 profiles (95%), also the same outcome date was selected by the algorithm and the clinical experts and for one profile (spontaneous abortion) a discrepancy regarding the outcome date was observed. This discrepancy was caused by a code ('missed abortion') that we didn't select for our algorithm as it is not necessarily a final outcome.

Discussion

We optimized and implemented an algorithm that more specifically identifies and classifies pregnancy outcomes, particularly preterm, term and births after the expected delivery date based on German claims data, which is the first of three procedures for using these data to investigate drug utilization and drug safety during pregnancy.

In our study, 1,235,261 pregnancy outcomes were identified, with the majority being live births. More than 98% of live births were identified through an inpatient delivery date or an inpatient OPS code which means that the date of birth could be assigned based on these most unambiguous sources of information in almost all live births.

The study implementing the original algorithm on GePaRD[6] showed that the percentages of births and ectopic pregnancies agreed well with representative data for Germany, whereas induced and spontaneous abortions were identified to a lesser extent[6]. The percentages in our study support this observation: The proportion of stillbirths in our study of 0.3% was well in line with national statistics[8], whereas the proportion of spontaneous abortion of 0.5% was much smaller than expected based on literature, e.g. in a multicenter study in three different gynecological centers in Germany (average 3.3%)[9]. This can occur due to several reasons. For example, pregnancy outcomes for which no medical care is sought like unrecognized early losses - although relevant - can generally not be identified in studies based on administrative claims data. Further, we only used inpatient data for the identification of spontaneous abortions as outpatient diagnoses are only available on a quarterly basis and are generally less reliable than hospital diagnoses (e.g. they may present suspected diagnoses or history). We assume that by this we most likely only identified relatively late spontaneous abortions, which may also explain the low proportion. Induced abortions without a medical or criminological indication are usually not reimbursed by health insurance providers and can thus not be identified based on claims data either. We focused, however, on live births as an essential prerequisite for the investigation of potentially drug-mediated outcomes in the child such as developmental disorders. Other potentially drug-mediated outcomes such as spontaneous abortions, induced abortions, and perinatal death shortly after birth are also important outcomes to be studied based on GePaRD, but were not - although relevant - in the focus of our analysis.

Our data do not include claims of midwives, so non-identification of live births would be relevant for deliveries solely accompanied by midwives. However, in 2014 only 1.3% of all deliveries in Germany were out-of-hospital deliveries[10] and births taking place solely in midwife-led-units in hospitals are assumed to be few due to only 15 units of this kind in Germany[11]. To identify those children in the data it is possible to use proxies such as claims for routine examinations for small children, but this was not in the focus of our study. Overall, we think that an underestimation of the total number of live births played only a minor role.

We classified birth outcomes into still births and preterm or births after expected delivery date according to inpatient diagnoses. The proportion of preterm births was well comparable with national statistics (10% as compared to 8.9%[8]). The coding of preterm, term and births after the expected delivery date in the database is based on the reimbursement rules. In claims data, birth after the expected delivery date is any birth after the expected day of delivery, as the reimbursable care (e.g. midwife visits) increases after the expected day of delivery. As expected, the proportion of deliveries after the expected date is considerably higher than the proportion of postterm births in national statistics (12% as compared to 0.6%[8]). The relevance of this sub-classification will be evaluated in future studies that aim to determine the beginning of pregnancy in GePaRD (e.g. based on the date of specific examinations).

In our study, an implausible sequence of codes occurred only very rarely. For example, in 393 out of 1.16 million pregnancies (0.03%) classified as ending in a live birth by the algorithm, we observed a code indicating an induced abortion 154-259 days before or shortly after the birth outcome. The codes for induced abortion in these cases may reflect coding errors which justifies discarding them from the analysis. For 65 births, the algorithm identified a code indicating an induced abortion on the same day as the birth outcome, which may reflect late termination of pregnancy. Late abortions were not in the focus of this modified outcome algorithm, but will be assessed in more detail in the context of drug safety studies.

The age distribution of persons with a pregnancy outcome in our study was plausible and - in terms of live births - well in agreement with national statistics. The mean age of women giving birth in Germany overall was 31 years in 2015[12] as compared to 31.5 years during our study period. For the vast majority of persons with a pregnancy outcome at an implausible age, we could identify the reason. This was predominantly caused by codes referring to newborns (e.g. postnatal care for a newborn) that may be recorded either for the mother or her child. When these codes were recorded for the child, the algorithm identified newborns aged 0 years as having a birth outcome. In some cases, the corresponding mother may be identified by linkage of mother-baby pairs[1]. This is an important prerequisite to investigate the association between the mothers' drug utilization during pregnancy and the fetal outcome. This linkage is the third procedure for studying drug safety in pregnancy and an existing algorithm[1] will be updated based on experiences with our study.

When using German claims data for studies on drug utilization and drug safety during pregnancy, the second of three essential procedures is to determine the beginning of pregnancy (and thus the gestational age) as there are critical time windows of exposure for many drugs. Since the beginning of pregnancy is not directly available in administrative claims data, several algorithms were developed to estimate the beginning of pregnancy[5]. German administrative claims data have not been explored in this regard yet, but our algorithm identifying pregnancy outcomes and assigning the respective date is an important prerequisite to estimate, in a next step, the beginning of pregnancy based on GePaRD.

Conclusion

We optimized an algorithm to identify and classify pregnancy outcomes, particularly births, in GePaRD and to assign the respective dates. The algorithm showed plausible results and a high level of agreement based on clinical expert review of pregnancy profiles. This algorithm is the first procedure for using GePaRD for studies on drug utilization and drug safety during pregnancy.

Acknowledgements

We thank the statutory health insurance providers AOK Bremen/Bremerhaven, DAK-Gesundheit, hkk Krankenkasse, and Techniker Krankenkasse (TK) for contributing the underlying data for this analysis. We would like to acknowledge the gynecologists, Cornelia von Herder and Dr. Elisabeth Holthaus-Hesse, for the medical counseling regarding the selection of diagnostic and procedure codes. We also like to thank Dr. Marlies Onken and PD Dr. Katarina Dathe for the review of longitudinal records of pregnancy related codes and Dr. Heike Gerds for editing the manuscript.

Conflict of interest

NW, TS, MN, SU and UH are working at the Leibniz Institute for Prevention Research and Epidemiology – BIPS. Unrelated to this study, BIPS occasionally conducts studies financed by the pharmaceutical industry. Almost exclusively, these are post-authorization safety studies (PASS) requested by health authorities. The studies and the resulting publications are not influenced by the pharmaceutical industry.

RM declares no conflict of interest.

References

1. Garbe E, Suling M, Kloss S, Lindemann C, Schmid U. Linkage of mother-baby pairs in the German Pharmacoepidemiological Research Database. *Pharmacoepidemiol Drug Saf* 2011;20(3):258-64.
2. Cea-Soriano L, Garcia Rodriguez LA, Fernandez Cantero O, Hernandez-Diaz S. Challenges of using primary care electronic medical records in the UK to study medications in pregnancy. *Pharmacoepidemiol Drug Saf* 2013;22(9):977-85.
3. Charlton RA, Neville AJ, Jordan S, Pierini A, Damase-Michel C, Klungsøyr K, Andersen AMN, Hansen AV, Gini R, Bos JH. Healthcare databases in Europe for studying medicine use and safety during pregnancy. *Pharmacoepidemiol Drug Saf* 2014;23(6):586-594.
4. Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernandez-Diaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf* 2013;22(1):16-24.
5. Margulis AV, Palmsten K, Andrade SE, Charlton RA, Hardy JR, Cooper WO, Hernandez-Diaz S. Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results. *Pharmacoepidemiol Drug Saf* 2015;24(4):335-42.
6. Mikolajczyk RT, Kraut AA, Garbe E. Evaluation of pregnancy outcome records in the German Pharmacoepidemiological Research Database (GePaRD). *Pharmacoepidemiol Drug Saf* 2013;22(8):873-80.
7. Pigeot I, Ahrens W. Establishment of a pharmacoepidemiological database in Germany: methodological potential, scientific value and practical limitations. *Pharmacoepidemiol Drug Saf* 2008;17(3):215-23.
8. Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH (AQUA). Qualitätsreport 2013. 2014. <https://www.sqg.de/sqg/upload/CONTENT/Qualitaetsberichte/2013/AQUA-Qualitaetsreport-2013.pdf>. (accessed February 9, 2018).
9. Berle P, Weiss E. [Rate of spontaneous abortion in relation to the time of fetal viability assessment]. *Geburtshilfe Frauenheilkd* 1990;50(12):959-63.
10. Loytved C. Qualitätsbericht - Außerklinische Geburtshilfe in Deutschland 2014. 2016. <http://www.quag.de/downloads/QUAG_bericht2014.pdf>. (accessed June 23, 2017).
11. Deutscher Hebammen Verband e.V. .Hebammenkreißsäle in Deutschland. 2015. <https://www.hebammenverband.de/familie/hebammen-kreisssaele/>. (accessed February 9, 2018).
12. Statistisches Bundesamt (Destatis) Alter der Mutter - Durchschnittliches Alter der Mutter bei der Geburt des Kindes 2015 (biologische Geburtenfolge) nach Bundesländern. 2017. <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Geburten/Tabellen/GeburtenMutterAlterBundeslaender.html>. (accessed June 23, 2017).

Tables and Figures

Table 1: Hierarchical structure of the algorithm i) regarding the pregnancy outcomes and ii) the sources of information

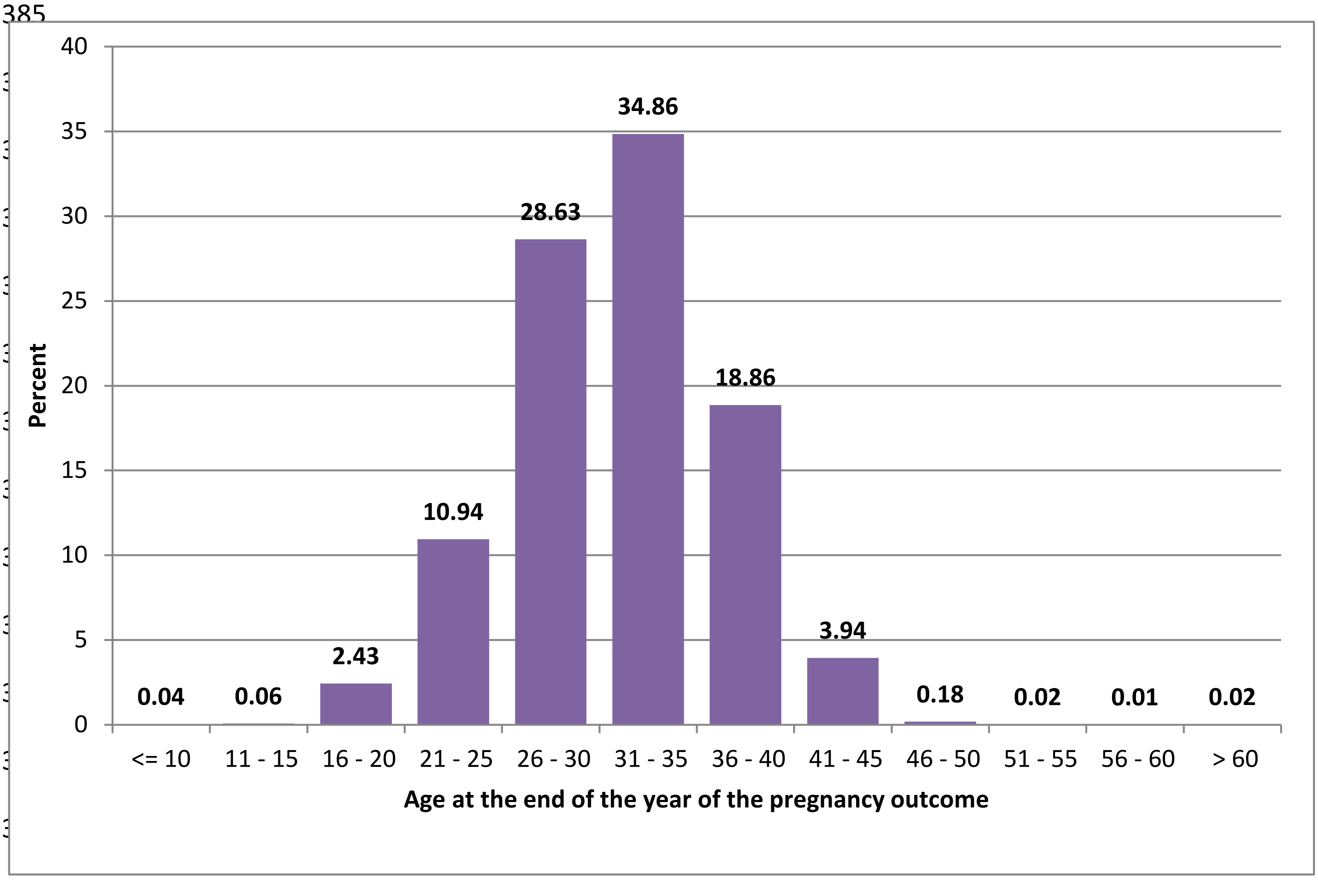
Hierarchy of the outcomes	Hierarchy of the sources of information	Availability of an exact outcome date
I. Births		
	I.a. Inpatient delivery date	Yes
	I.b. Inpatient procedure (OPS)	Yes
	I.c. Inpatient diagnosis (ICD)	Yes†
	I.d. Outpatient service (EBM)	Yes
	I.e. Outpatient procedure (OPS)	No
II. Induced abortions		
	II.a. Inpatient procedure (OPS)	Yes
	II.b. Outpatient service (EBM)	Yes
	II.c. Inpatient diagnosis (ICD)	Yes†
	II.d. Outpatient procedure (OPS)	No
III. Ectopic pregnancies		
	III.a. Inpatient procedure (OPS)	Yes
	III.b. Inpatient diagnosis (ICD)	Yes†
	III.c. Outpatient procedure (OPS)	No
IV. Spontaneous abortions		
	IV.a. Inpatient diagnosis (ICD)	Yes†

† Here, the date of hospital admission was used as it is likely to be closer to the date of the outcome than the date of discharge from hospital.

Table 2: Contribution of sources for pregnancy outcomes

Outcome/ Source of information	Live births N Column %	Still births N Column %	Induced abortions N Column %	Ectopic pregnancies N Column %	Spontaneous abortions N Column %	∑ Row %
Inpatient delivery date	1,004,965 86.28	2,917 91.44	NA	NA	NA	1,007,882 81.59
Inpatient procedure (OPS)	146,029 12.54	244 7.65	7,542 17.14	10,780 64.71	NA	164,595 13.32
Inpatient diagnosis (ICD)	11,203 0.96	26 0.82	2,099 4.77	5,642 33.87	6,656 100	25,626 2.07
Outpatient service (EBM)	2,221 0.19	3 0.09	34,207 77.72	NA	NA	36,431 2.95
Outpatient procedure (OPS)	325 0.03	0 0.0	165 0.37	237 1.42	NA	727 0.06
∑ Column %	1,164,743 94.29	3,190 0.26	44,013 3.56	16,659 1.35	6,656 0.54	1,235,261 100

Figure 1: Distribution of age among female persons with at least one pregnancy outcome during the study period



This includes female persons with information on the year of birth. Newborns were excluded.