Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Validation of Obstetric Diagnosis and Procedure Codes in the Danish National Patient Registry in 2017

Kamille Herskind[1], Peter Bjødstrup Jensen[1], Christina Anne Vinter[2–4], Lone Krebs[5,6], Lene Friis Eskildsen[5], Anne Broe[1,7], Anton Pottegård[1], Mette Bliddal[1,8]

[1]Clinical Pharmacology, Pharmacy, and Environmental Medicine, Department of Public Health, University of Southern Denmark, Odense, Denmark; [2]Department of Gynaecology and Obstetrics, Odense University Hospital, Odense, Denmark; [3]Department of Clinical Research, University of Southern Denmark, Odense, Denmark; [4]Steno Diabetes Center Odense, Odense University Hospital, Odense, Denmark; [5]Department of Obstetrics and Gynecology, Copenhagen University Hospital-Amager and Hvidovre, Copenhagen, Denmark; [6]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; [7]IQVIA, London, UK; [8]Research Unit OPEN, Department of Clinical Research, University of Southern Denmark, Odense, Denmark

Correspondence: Mette Bliddal, OPEN, Heden 16, St.tv, Odense, 5000, Denmark, Tel +45 2943 7912, Email mette.bliddal@rsyd.dk

Purpose: This study aimed to systematically evaluate the validity of variables related to pregnancy, delivery, and key characteristics of the infant in the Danish National Patient Register using maternal medical records as the reference standard.

Patients and Methods: We reviewed medical records of 1264 women giving birth in the Region of Southern Denmark during 2017. We calculated positive (PPV) and negative (NPV) predictive values, sensitivity, and specificity to estimate the validity of 49 selected variables.

Results: The PPV was ≥0.90 on most pregnancy-related variables including parity, pre-gestational BMI, diabetes disorders, and previous cesarean section, while it was lower for hypertensive disorders, especially mild to moderate preeclampsia (0.49, 95% CI 0.32–0.66). Sensitivity ranged from 0.80 to 1.00 on all pregnancy-related variables, except hypertensive disorders (sensitivity 0.38–0.71, lowest for severe preeclampsia). On most delivery-related variables including obstetric surgical procedures (eg cesarean section and induction of labor), pharmacological pain-relief, and gestational age at delivery, PPV's ranged from 0.98 to 1.00 and the corresponding sensitivities from 0.87 to 1.00. Regarding infant-related variables, both the APGAR score registered five minutes after delivery and birthweight yielded a PPV of 1.00.

Conclusion: Obstetric coding in the Danish National Patient Register shows very high validity and completeness making it a valuable source for epidemiologic research.

Plain Language Summary: Danish register data are often used for epidemiological research in reproduction. The registers are based on coded information to the registers based on information from medical records. The quality of the register data is highly dependent of the validity of the codes. Yet there is a lack in our knowledge of the validity of data related to pregnancy, childbirth, and the characteristics of the newborn baby. We therefore aimed to validate the Danish National Patient Registry data related to pregnancy and childbirth by comparing the registered code with information from the medical records.

We scrutinized medical records from 1264 women giving birth in the Region of Southern Denmark during 2017. We compared the registration in the medical record with the registered code in the Danish National Patient Registry by calculating how accurate the register data are according to 49 different variables.

Results showed that registered codes in the Patient Registry for pregnancy- and childbirth-related conditions and key infant characteristics were to a high degree in agreement with the data from the medical report with few exceptions.

In conclusion, the study revealed that the Danish National Patient Register provides highly accurate and comprehensive data for most pregnancy, delivery, and infant-related variables. This underscores the register's value as a reliable source for epidemiologic research in reproductive health.

Keywords: registries, sensitivity and specificity, validity, epidemiology, pregnancy, delivery

Received: 13 October 2023
Accepted: 6 January 2024
Published: 21 February 2024

Introduction

The Danish National Patient Registry (Patient Registry) was established in 1977 and is a nationwide register, containing routinely collected data on all hospital contacts, including contacts related to pregnancies and childbirths.[1] It holds detailed information on maternal characteristics, pregnancy- and delivery-related variables, and characteristics of the infant, and it feeds directly into the Danish Medical Birth Register.[2] Both registries are valuable and much used resources in reproductive and obstetric epidemiologic research.[3,4] The internal validity of such research is closely connected to the validity of the variables registered in the Patient Registry.

The latest systematic validation of obstetric variables in the Patient Registry is a report published in 2003 in Danish based on data from 2001.[5] Since then, only few validation studies on specific variables from the Patient Registry have been published, validating specific diagnoses including gestational diabetes mellitus,[6] uterine rupture,[7] spontaneous abortion,[8] second trimester miscarriages and deliveries,[9] and preeclampsia-related diagnoses.[10,11] Considering the changes in both clinical practice and coding practices over the last 20 years, an updated and comprehensive validation of obstetric variables in the Patient Registry for obstetric epidemiological research is long overdue.

The aim of this study was to comprehensively evaluate the validity of obstetric variables in the Patient Registry, using medical records as the reference standard. To do this, our objective was to estimate the positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity, on pregnancy-, delivery-, and infant-related variables.

Materials and Methods

We conducted a validation study using a sample of all women giving birth in the Region of Southern Denmark during 2017. To obtain the PPV, NPV, sensitivity, and specificity of the Patient Registry we compared diagnosis and procedure codes from the Patient Registry to information extracted from maternal medical records.

Data Source

Denmark is divided into five regions each responsible for public tax-based healthcare services including hospitals and primary care. The regions are comparable regarding sociodemographic and health-related characteristics.[12] In 2017, 61,250 births were registered in Denmark.[13] Of these, 11,700 were in the Region of Southern Denmark. The Region of Southern Denmark has five obstetric departments of different size and functions, and pregnant women are followed by midwives, general practitioners, and obstetricians during antenatal care programs regardless of planning a hospital birth or home birth. Uncomplicated pregnancies are exclusively monitored by general practitioners and midwives. The departments consists of one highly specialized obstetric department at Odense University Hospital (OUH) with one small satellite department in Svendborg (4725 births in 2017), and three less specialized obstetric departments affiliated with neonatal departments, located at Hospital of Southern Jutland (N=1815), Lillebaelt Hospital, Kolding (N=3340), and Hospital South-West Jutland (N=1915). We retrieved maternal medical records and personal pregnancy forms from each of these five obstetric departments. The medical records included all records from relevant healthcare professionals including obstetricians, midwives, and nurses. The personal pregnancy form is a paper form held by the pregnant woman and filled out at the first antenatal interview around week 10 (week 6–10) with her general practitioner and continuously updated throughout pregnancy by clinicians responsible for the antenatal care. It includes information on pre-pregnancy height, weight, and body mass index (BMI), ultrasound-based due date, maternal smoking status, relevant comorbidities, and obstetric history. The personal pregnancy forms are transferred to the medical records at the delivery. Medical records and pregnancy forms were used as the reference standard.

The Patient Registry is a population-based registry with complete nationwide coverage on all hospital admissions since 1977. Since 1995 it also includes outpatient and psychiatric contacts. For every hospital contact, one primary and several optional secondary diagnoses are registered according to the Danish version of the International Classification of Diseases, 10th revision (ICD-10) codes.[1,14] Treatments, including surgical procedures, medical treatments, and anesthesia, are registered according to the Danish version of the Nordic Medico-Statistical Committee (NOMESCO) classification of surgical procedures and Danish classification systems provided in the Health Care Classification System.[1] All data are registered at the hospitals and automatically transferred to the Patient Registry. Since 1997, data on all births including detailed information on diagnosis, procedures, and treatments related to pregnancy and delivery for both the mother and the child have been fed directly into the Medical Birth

Register from the Patient Registry.[2] All Danish residents hold a unique 10-digit identification number from the Danish Civil Registration System.[15] This enables unambiguous linkage between registries and further linkage to medical records.

Study Population

To estimate the validity of obstetric variables in the Patient Registry, we planned to identify 1500 women giving birth in the Region of Southern Denmark between January 1, 2017, and December 31, 2017.

We sampled women with live births according to birthplace by annual birth count for each hospital location to ensure a sample that was representative of live births within the region. Our sample was planned for two validation studies. The first on pregnancy- and delivery-related variables conducted on the women (this study), and the second study on the validity of malformation registrations conducted using the children. Therefore, the women were sampled according to the congenital malformation status of their newborn child in the following groups: 300 with hip dysplasia, 300 with a major congenital malformation, 300 with a minor congenital malformation, and 600 without any registered congenital malformation. Congenital malformations were defined according to the EUROCAT definitions[16,17] and were sampled if registered in the child's record in the Patient Registry within the first year after birth. The same child could be present in more than one group in case of either multiple malformation diagnoses or if the malformation(s) could be classified in more than one malformation group (N=141). In the hip dysplasia group only 213 children with hip dysplasia were registered in 2017. We used the unique mother–child link from the Medical Birth Registry to link each child to their biological mother. For women with more than one pregnancy in the study period (n<5), only information from their first pregnancy was used. Finally, 1272 children of 1264 pregnancies were identified for the study population. We disregarded information from 34 pregnancies due to missing key information, eg birthdate, delivery details or child identifier (Figure 1).

Variables

The obstetric variables considered in this study are listed in Table 1 and are all defined in line with national guidelines from the Danish Society of Obstetrics and Gynecology (DSOG) and codes from the Patient Registry used in 2017.[18] We divided the variables into pregnancy-, delivery-, and infant-related variables (for definitions and codes see Table S1). Pregnancy-related variables included: parity, pre-gestational BMI, maternal smoking status registered at the end of pregnancy, hypertensive and diabetes disorders, and previous cesarean section. Delivery-related variables included: induction and stimulation of labor, pharmacological and non-pharmacological pain-relief, vacuum extraction, cesarean section, episiotomy, perineal laceration, uterine rupture, postpartum hemorrhage, the primary diagnosis of delivery, and gestational age at delivery. We subdivided several of the variables and performed additional analyses to specify the validity of specific diagnoses and procedures (see Table 1 for subgroups). When more than one method was used to induce labor, only the first method mentioned in the medical record was recorded as induction, whereas the subsequent methods were recorded as stimulation. Regarding pharmacological pain-relief, we included both procedure codes and supplementary codes from the Patient Registry. Infant-related variables included: APGAR score five minutes after delivery and birthweight. The limited number of infant-related variables were due to the limited availability of this information in the maternal medical records. We did not have access to the medical record of the child.

We included pregnancy-related variables from the Patient Registry registered from first week of pregnancy, defined as date of delivery minus gestational age (in weeks) at delivery, until seven days after the date of delivery to minimize the risk of including irrelevant codes. For delivery- and infant-related variables we included data registered seven days before or after the date of delivery.

Data Handling

Data from medical records and pregnancy forms were obtained from each hospital, reviewed by a medical student (KH), and entered into RedCAP, an electronic data capture tool hosted at the research support unit OPEN, Odense University Hospital, Region of Southern Denmark.[19] When in doubt, the record was conferred with experts within obstetrics (CV), obstetric coding (LFE), or register-based reproductive epidemiology (MB). Approximately 100 records were entered with help from other health science students. Information from the medical records was transferred to a secure server at the

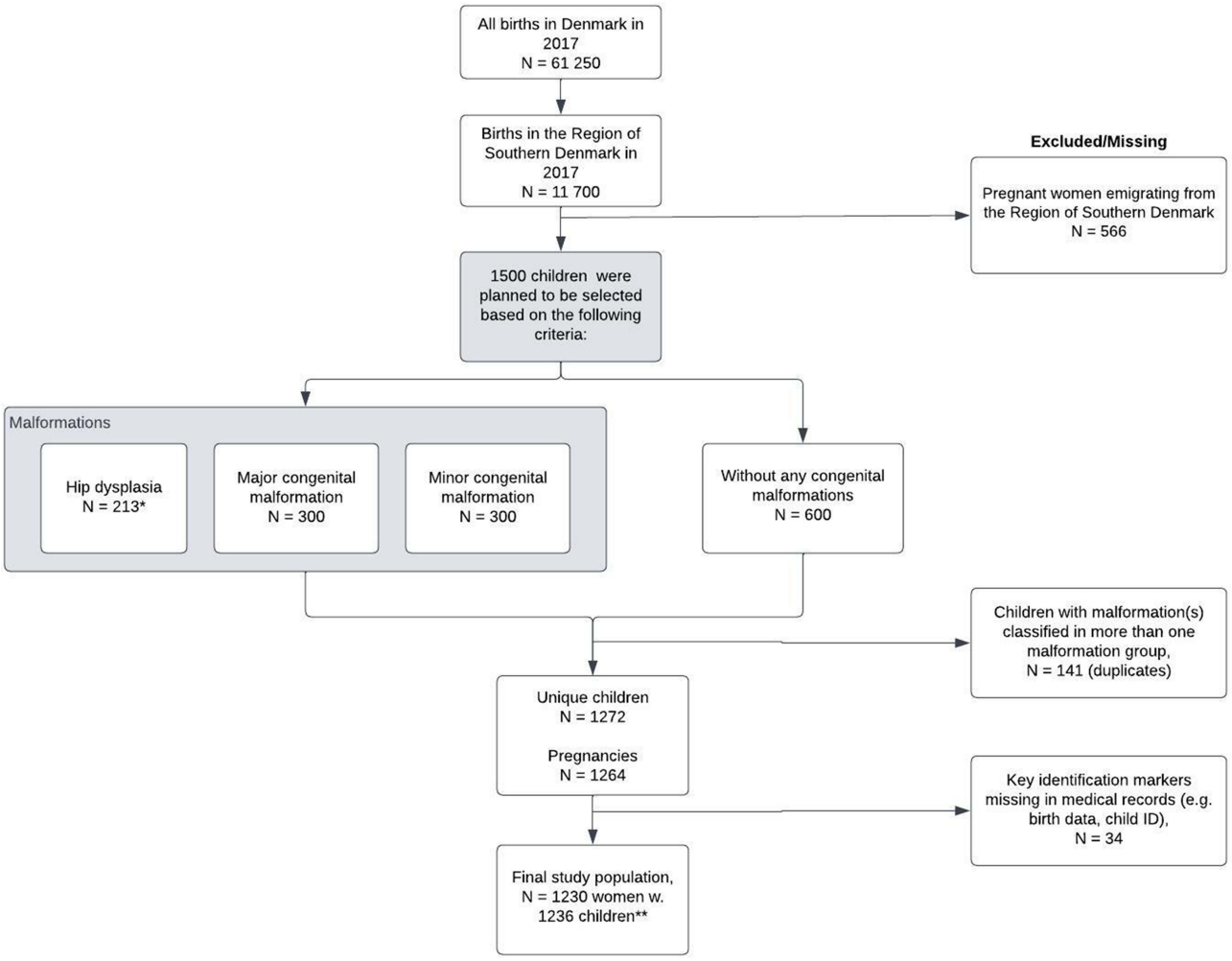

Figure 1 Flowchart of sampling for the study population.
Notes: *Only 213 children with hip dysplasia were registered in 2017. **Twelve children were from a multifold pregnancy where both children were included in the study population (1 pregnancy=2 children). Each pregnancy could only be included once.

Danish Health Data Authorities for linkage to the Patient Registry data and analysis. All data were pseudonymized before handling.

Statistical Analysis

We calculated the PPV, NPV, sensitivity, and specificity comparing the registration in the Patient Registry with the medical records as our reference standard. The main outcomes of interest were the PPV and sensitivity. The PPV was defined as the proportion of women registered with the variable of interest in the Patient Registry confirmed by the medical records divided by the total number of women registered in the Patient Registry, and thus denotes the trustworthiness of records in the registry. The sensitivity was defined as the number of confirmed women with the variable of interest divided by the total number of women with the variable according to the medical records, and thus denotes the completeness of recording in the registry.

For the non-binary variables (eg BMI), the validity was analyzed in two steps. First, we assessed the completeness of the registration by examining whether any registration was present in both the Patient Registry and the medical record. Thus, the calculated PPV, NPV, sensitivity, and specificity, for these variables refer to whether any registration is noted. Secondly, we divided the numeric variables into a priori defined categories (see Table 1 for definitions) and analyzed whether the given variable was correctly registered according to category to validate the values. In this step, only women

Table 1 Key Obstetric Variables Including Subdivisions, Selected for Validation and Grouped into "Pregnancy-Related Variables", "Delivery-Related Variables", and "Infant-Related Variables"

Pregnancy-Related Variables	Delivery-Related Variables	Infant-Related Variables
Parity	Induction of labor, overall	APGAR score five minutes after delivery
1	With prostaglandin	1–3
≥2	With oxytocin	4–6
Pregestational BMI (kg/m^2)	Stimulation of labor, overall	≥7
Underweight (<18.5)	With oxytocin	Birthweight
Normal weight (18.5–24.9)	Pharmacological pain-relief, overall	Low (<2500 g)
Pre-obesity (25.0–29.9)	Spinal block	Normal (2500–3999 g)
Obesity (>30.0)	General anesthesia	High (≥4000 g)
Maternal smoking status registered at the end of pregnancy	Epidural block	
Hypertensive disorders, overall	Infiltration analgesia	
Preeclampsia	Pudendal nerve block	
Mild to moderate	Strong analgesia	
Severe	Non-pharmacological pain-relief	
Eclampsia	Vacuum extraction	
Diabetes disorders, overall	Cesarean section	
Gestational diabetes	Emergency cesarean section*	
Previous cesarean section	1st degree	
	2nd degree	
	3rd degree	
	Episiotomy	
	Perineal laceration (3rd–4th degree)	
	Uterine rupture	
	Postpartum hemorrhage (>500 mL)	
	Primary diagnosis of delivery	
	Spontaneous single birth	
	Single birth with instrumental delivery	
	Single birth with cesarean section	
	Single birth after induction	
	Single birth after abortion procedure	
	Spontaneous multiple birth	
	Multiple birth with instrumental delivery	
	Multiple birth with cesarean section	
	Multiple birth after induction	
	Multiple birth with vaginal birth A and acute cesarean section B	
	Multiple birth after abortion procedure	
	Gestational age at delivery	
	<28+0	
	28+0–31+6	
	32+0–36+6	
	>37+0	

Notes: *Degree of emergency cesarean section: 1, 2, or 3 are based on urgency. Maximum tolerated time from decision to delivery is 15, 30, or 60 minutes, respectively.
Abbreviation: BMI, Body mass index.

who had any value available from both medical records and the Patient Registry were used (=true positives), and therefore only the PPV was calculated.

When substantial data in a medical record were missing, such as no information on the entire delivery or missing child ID, the pregnancy was excluded in the final analysis. Otherwise, missing information in medical records were registered as a "negative", meaning "not registered in DNPR".

We performed a sensitivity analysis restricted to data from women with children not registered with any malformation, to test whether our sampling strategy including a large proportion of children with malformations influenced the results. We calculated 95% confidence intervals for proportions. All statistical analyses were performed using STATA 17 (StataCorp, College Station, TX, USA).

Results

The final study samples included 1230 women (1230 pregnancies) giving birth to 1236 children. Few women were misclassified in the pregnancy-related variables (Table 2) leading to PPVs and sensitivities ≥0.89 for all variables (parity, pregestational BMI, diabetes disorders (overall), gestational diabetes, previous cesarean section) except for two overall categories being maternal smoking at time of delivery with a sensitivity of 0.80 (95% confidence interval (CI) 0.71–0.87) and hypertensive disorders (overall) with a PPV of 0.86 (95% CI 0.76–0.93) and a sensitivity of 0.71 (95% CI 0.60–0.80). When analyzing hypertensive disorders in subcategories, the PPV and sensitivity decreased with PPV for preeclampsia being 0.70 (95% CI 0.54–0.83) and sensitivity being 0.71 (95% CI 0.55–0.84). Further, when subdivided on severity of disease (mild to moderate vs severe) the PPV was lowest for mild to moderate preeclampsia (0.49, 95% CI 0.32–0.66) compared to the more severe cases which yielded a PPV of 0.71 (95% CI 0.29–0.96). The sensitivity was 0.62 (95% CI 0.42–0.79) for mild to moderate preeclampsia and 0.38 (95% CI 0.14–0.68) for severe preeclampsia. To be noted, only few cases were registered with severe preeclampsia leading to wide confidence intervals.

The vast majority of delivery-related variables yield PPVs ≥0.89 (Table 3). Induction of labor overall showed a PPV of 0.98 (95% CI 0.96–0.99), but when subdivided into specific drugs, induction with oxytocin only yielded a PPV of 0.34 (95% CI 0.24–0.45). The PPV on stimulation of labor with oxytocin was 0.91 (95% CI 0.86–0.95). Notably the overall use of oxytocin (ie disregarding indication) yielded a PPV of 0.98 (95% CI 0.95–0.99).

Regarding pain-relief during delivery, PPV ranged from 0.82–0.99 and was lowest for less "severe" intervention such as infiltration analgesia with PPV of 0.82 (95% CI 0.75–0.87) and non-pharmacological pain-relief of 0.87 (95% CI 0.83–0.90).

The sensitivity estimates for delivery-related variables ranged from 0.50 to 1.00, where the majority yielded sensitivities ≥0.81. The lower values included overall stimulation of labor with a sensitivity of 0.56 (95% CI 0.51–0.61), stimulation with oxytocin where the sensitivity was 0.51 (95% CI 0.45–0.57), and infiltration analgesia with a sensitivity of 0.55 (95% CI 0.48–0.61).

Table 2 Counts of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), Positive Predictive Value (PPV), Sensitivity, and 95% Confidence Interval (CI) for Pregnancy-Related Variables in the Danish National Patient Register (DNPR) Using Medical Records as the Reference Standard

Pregnancy-related variables	n=1230							
	Counts				Positive Predictive value (PPV)		Sensitivity	
	TP	TN	FP	FN	Estimate	95% CI	Estimate	95% CI
Available parity information	1230	0	0	0	1.00	1.00–1.00	1.00	1.00–1.00
Same parity groups[a] (n=1230[b])	1209	0	21	0	0.98	0.97–0.99	–	–
Available pre-gestational BMI information	~1205	<5	6	17	1.00	0.99–1.00	0.99	0.98–0.99
Same pre-gestational BMI group[c] (n~1205[b])	~1165	0	~40	0	0.97	0.95–0.97	–	–
Maternal smoking status at the end of pregnancy (smoker)	83	1116	10	21	0.89	0.81–0.95	0.80	0.71–0.87
Hypertensive disorders, overall	61	1134	10	25	0.86	0.76–0.93	0.71	0.60–0.80
Preeclampsia	30	1175	13	12	0.70	0.54–0.83	0.71	0.55–0.84
Mild to moderate	18	1182	19	11	0.49	0.32–0.66	0.62	0.42–0.79
Severe	5	~1215	<5	8	0.71	0.29–0.96	0.38	0.14–0.68
Eclampsia	0	~1230	0	<5	–	–	0.00	0.00–0.97
Diabetes disorders, overall	~95	~1125	7	<5	0.93	0.86–0.97	0.98	0.93–1.00
Gestational diabetes	~90	~1130	9	<5	0.91	0.83–0.96	0.99	0.94–1.00
Previous cesarean section	135	1077	9	9	0.94	0.88–0.97	0.94	0.88–0.97

Notes: Cell counts between 1 and 4 are represented as "<5". To avoid identification of actual low counts from the total, some numbers have been rounded (~). [a]Groups: parity=1, parity≥2. [b]Only women who had any value available from both medical records and DNPR (TP) were included in the analysis. [c]Group division: underweight (<18.5), normal weight (18.5–24.9), pre-obesity (25.0–29.9), obesity (>30.0).
Abbreviation: BMI, Body mass index.

Table 3 Counts of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), Positive Predictive Value (PPV), Sensitivity, and 95% Confidence Interval (CI) for Delivery-Related Variables in the Danish National Patient Register (DNPR) Using Medical Records as the Reference Standard

Delivery-related variables	n=1230							
	Counts				Positive predictive value (PPV)		Sensitivity	
	TP	TN	FP	FN	Estimate	95% CI	Estimate	95% CI
Induction, overall	312	880	7	31	0.98	0.96–0.99	0.91	0.87–0.94
With oxytocin	30	1131	59	10	0.34	0.24–0.45	0.75	0.59–0.87
With prostaglandin	189	1026	7	8	0.96	0.93–0.99	0.96	0.92–0.98
Stimulation, overall	256	730	42	202	0.86	0.81–0.90	0.56	0.51–0.61
With oxytocin	161	899	15	155	0.91	0.86–0.95	0.51	0.45–0.57
Oxytocin exposure during delivery	251	870	6	103	0.98	0.95–0.99	0.71	0.66–0.76
Pharmacological pain–relief, overall	765	349	17	99	0.98	0.97–0.99	0.89	0.86–0.91
Spinal block	207	961	15	47	0.93	0.89–0.96	0.81	0.76–0.86
General anesthesia	~35	~1175	<5	17	0.93	0.80–0.98	0.69	0.54–0.80
Epidural block	~275	~930	<5	24	0.99	0.97–1.00	0.92	0.88–0.95
Infiltration analgesia	141	941	31	117	0.82	0.75–0.87	0.55	0.48–0.61
Pudendal nerve block	~85	~1095	<5	~45	0.98	0.92–1.00	0.64	0.55–0.72
Strong analgesic (eg morphine)	58	1131	7	34	0.89	0.79–0.96	0.63	0.52–0.73
Non–pharmacological pain–relief	308	773	46	103	0.87	0.83–0.90	0.75	0.70–0.79
Vacuum extraction	~40	~1175	<5	12	0.95	0.84–0.99	0.77	0.64–0.88
Cesarean section	~265	~965	0	<5	1.00	0.99–1.00	0.99	0.97–1.00
Emergency	~175	~1045	<5	8	0.98	0.94–0.99	0.96	0.92–0.98
1st degree	11	~1215	0	<5	1.00	0.72–1.00	0.73	0.45–0.92
2nd degree	~60	~1155	<5	10	0.94	0.85–0.98	0.86	0.76–0.93
3rd degree	72	1125	8	25	0.90	0.81–0.96	0.74	0.64–0.83
Episiotomy	~45	~1170	<5	13	0.98	0.88–1.00	0.78	0.65–0.87
Perineal laceration (≥3rd degree)	16	~1210	<5	<5	0.89	0.65–0.99	0.89	0.65–0.99
Uterine rupture	<5	~1230	<5	0	0.50	0.01–0.99	1.00	0.03–1.00
Postpartum hemorrhage (>500 mL)	309	872	7	42	0.98	0.95–0.99	0.88	0.84–0.91
Birth primary diagnosis								
Spontaneous single birth	646	530	49	5	0.93	0.91–0.95	0.99	0.98–1.00
Single birth with instrumental delivery	47	1176	0	7	1.00	0.92–1.00	0.87	0.75–0.95
Single birth with cesarean section	~255	~975	<5	<5	0.99	0.97–1.00	0.99	0.97–1.00
Single birth after induction	204	974	8	44	0.96	0.93–0.98	0.82	0.77–0.87
Single birth after abortion procedure	<5	~1230	0	0	1.00	0.03–1.00	1.00	0.03–1.00
Spontaneous multiple birth	<5	~1225	0	0	1.00	0.40–1.00	1.00	0.40–1.00
Multiple birth with instrumental delivery	<5	~1230	0	<5	1.00	0.03–1.00	0.50	0.01–0.99
Multiple birth with cesarean section	10	~1215	<5	<5	0.91	0.59–1.00	0.83	0.52–0.98
Multiple birth after induction	<5	~1225	<5	0	0.75	0.19–0.99	1.00	0.29–1.00
Multiple birth with vaginal birth A and acute cesarean section B[a]	0	1230	0	0	–	–	–	–
Multiple birth after abortion procedure[a]	0	1230	0	0	–	–	–	–
Available "gestational age at delivery" information	~700	0	<5	~525	1.00	0.99–1.00	0.57	0.54–0.60
Correct "gestational age at delivery"–group[b] (n~700[c])	~700	0	0	0	1.00	0.99–1.00		

Notes: Cell counts between 1 and 4 are represented with "<5". To avoid identification of actual low counts from the total, some numbers have been rounded (~). [a]When no data were registered in either the Patient Registry or the medical records, no results were presented (–). [b]Groups: gestational age <28+0, 28+0–31+6, 32+0–36+6, ≥37 +0. [c]Only women who had any value available from both medical records and DNPR (TP) were included in the analysis.

Regarding gestational age at delivery, we found a PPV of 1.00 and a sensitivity of 0.57. When restricting the analysis to include only women who had any value available from both medical records and DNPR (N~700) and checking whether the numeric value on gestational age was correctly placed in one of four categories (gestational age <28+0, 28 +0–31+6, 32+0–36+6, >37+0) the PPV remained 1.00.

On infant-related variables, the validity estimates show that most children had a registered value of an APGAR score five minutes after delivery (missing=90) (Table 4). In children with a registration (N=1146), five individuals had

Table 4 Counts of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), Positive Predictive Value (PPV), Sensitivity, and 95% Confidence Interval (CI) for "Infant-Related Variables" in the Danish National Patient Register (DNPR) Using the Mothers' Medical Records as the Reference Standard

Infant-related variables	n=1236							
	Counts				Positive Predictive Value (PPV)		Sensitivity	
	TP	TN	FP	FN	Estimate	95% CI	Estimate	95% CI
Available APGAR 5 score information	1146	15	75	0	0.94	0.92–0.95	1.00	1.00–1.00
Correct APGAR 5 score group[a] (n=1146[b])	1141	0	5	0	1.00	0.99–1.00	–	–
Available birthweight information	324	0	912	0	0.26	0.24–0.29	1.00	0.99–1.00
Correct birthweight group[c] (n=324[b])	324	0	0	0	1.00	0.99–1.00	–	–

Notes: [a]Groups: APGAR 5 score=1–3, 4–6, ≥7. [b]Only children who had any value available from both medical records and DNPR (TP) were included in the analysis. [c]Groups: birthweight=low (<2500 g), normal/intermediate (2500–3999 g), high (≥4000 g).

diverging value categories leading to a PPV of 1.00 (95% CI 0.99–1.00). Maternal medical records lacked information on child birthweight to a large degree (missing 912 of 1236). In children with registration of birthweight, all 324 individual's birthweights were categorized correctly in one of three categories (low (<2500 g), normal (2500–3999 g), and high (≥4000 g)) in the Patient Registry, leading to a sensitivity of 1.00 (95% CI 0.99–1.00).

The results on specificity and NPV for all variables are available in Table S2. Overall, the specificity was ≥0.92 on all included variables, except for some of the non-binary variables including pre-gestational BMI information, gestational age at time of delivery, APGAR score five minutes after birth, and birthweight information. All NPV values were within the range 0.78 to 1.00, when disregarding BMI information and gestational age at time of delivery.

When restricting the analyses to pregnancies without malformations (N=535), the estimates were virtually identical on the vast majority of variables and we concluded that the sampling strategy did not influence our results (Table S3), justifying our results to be representative for all births in the register. To be noted, the 95% confidence intervals were generally wider due to the smaller sample size.

Discussion

In this comprehensive validation study, we found that the validity of obstetric variables in the Patient Registry was very high for the vast majority of variables using medical records as the reference standard. There was a high completeness of data in the maternal medical records and only information on child birthweight was restricted due to lack of access to the infant's medical records. Our findings support that the obstetric information from the Patient Registry are reliable for epidemiological research.

The main strength of the study was our systematic inclusion of numerous variables related to both pregnancy, delivery, and the newborn child, which allows a comprehensive overview of the validity of obstetric information from the Patient Registry. Due to our large sample and selection strategy, inclusion in the study was independent of values of variables, which enabled us to estimate sensitivity and specificity along with the predictive values. The study also had several limitations. Firstly, each medical record was accessed by one reviewer only; however, when in doubt, each case was discussed among the authors and agreement was reached based on clinical knowledge and extensive specific data source experience. Secondly, the personal pregnancy forms were not available for all women. Most information from the personal pregnancy forms which were of our interest was also available in the maternal medical record, so we do not expect this to influence the results. Thirdly, even though we included a large sample we did not have many cases on rare diagnoses such as uterine rupture (n<5), which hinders meaningful assessment of these outcomes. Lastly, when reviewing the women's medical records, data on the child were often missing. We did not have access to the children's medical records where data on birthweight and APGAR score would be registered. This explains the low PPV (0.26) for birthweight (high amount of false positive). Reassuringly, when including only children with information on birthweight, there was no misclassification in birthweight categories, and we do not expect the validity to be affected due to registration in the child's medical record.

Although the PPV was high on almost all variables, our results showed that only 34% of women registered with induced labor with oxytocin were confirmed by medical records. In 2017, coding guidelines state that only the first used method should be registered as induction and subsequent methods as stimulation, but clinical experience indicates some coding inconsistency which might explain the low PPV on induction with oxytocin. Importantly, if the focus is on exposure to oxytocin at any point during labor disregarding indication, we found an excellent validity; 98% of women registered with either induction or stimulation with oxytocin were confirmed by medical records.

Preeclampsia is a challenging diagnosis to establish, which is confirmed by our findings. The clinical picture may vary considerably, change quickly, and advance rapidly, and notably, treatment may be initiated to avoid progression yet the indication for treatment may lead the registration. We experienced several challenges in the review of this diagnosis and in classifying the severity of the cases. First, in some cases there was conflicting information in the medical record regarding the clinical assessment from the doctor and the objective measures on blood pressure, urine protein, and blood samples. All cases with conflicting information were reviewed and classified by a senior obstetrician (CV). Secondly, we did not have access to all biochemistry on blood samples which is important in the diagnosis of severe preeclampsia. Lastly, the preeclampsia diagnostic criteria were under review in 2017, which can have caused some confusion and inconsistent coding practice. Therefore, a dedicated validation study regarding preeclampsia evaluating the diagnosis based on objective measures (blood pressure, urine protein, and blood samples) should be considered.

Some of the included variables have previously been validated. Luef et al[11] validated the diagnosis of preeclampsia and found a similar overall PPV on 80.5%, but, contradictory to us, reported a low sensitivity of 56% indicating that nearly half of the preeclampsia patients were unregistered in the registry during 2010–2012. When restricting the analysis to only include severe cases they found a low sensitivity of 18.6%, which is comparable to our findings. This indicates that especially severe cases of preeclampsia remain hard to classify.

Langhoff-Roos and Rasmussen[5] validated a number of obstetric variables from the Patient Registry by reviewing medical records from women giving birth during one week in January 2001. They concluded that the Patient Registry generally had high validity but estimates on sensitivity for variables on pain-relief were low especially on epidural block (sensitivity≈60%). Our finding indicates that the sensitivity has increased for these variables, especially regarding epidural block where 92% of the women having epidural block were registered correctly in the Patient Registry.

Conclusion

Our findings provide strong evidence that the Danish National Patient Registry has very high validity on pregnancy-related, delivery-related and two infant-related key variables, supporting that the Patient Registry is a valid and valuable source to identify obstetric variables for epidemiologic research.

Abbreviations

PPV, Positive predictive value; NPV, Negative predictive value; BMI, Body mass index.

Ethical Approval

Permission to access medical journals was obtained from the Region of Southern Denmark (21/49588, October 6, 2021). The use of data from the national registers was approved by the Danish Data Protection Agency (FSEID-00004511). Due to Danish legislation, approval from Ethics Committee was not needed.

Acknowledgments

We would like to thank Emma Simonsen, Maria H. Stoltz-Andersen, Sophie Lund Elkrog, and Debbie Rosenlyst Ollerup for their work on entering data into REDCap.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically

reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Disclosure

The authors report no financial or non-financial conflicts of interest in this work.

References

1. 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. doi:10.2147/CLEP.S91125
2. Bliddal M, Broe A, Pottegård A, Olsen J, Langhoff-Roos J. The Danish medical birth register. *Eur J Epidemiol*. 2018;33(1):27–36. doi:10.1007/s10654-018-0356-1
3. Otkjaer AM, Jørgensen HL, Clausen TD, Krebs L. Maternal short-term complications after planned cesarean delivery without medical indication: a registry-based study. *Acta Obstet Gynecol Scand*. 2019;98(7):905–912. doi:10.1111/aogs.13549
4. Hedegaard M, Lidegaard Ø, Skovlund CW, Mørch LS, Hedegaard M. Perinatal outcomes following an earlier post-term labour induction policy: a historical cohort study. *BJOG*. 2015;122(10):1377–1385. doi:10.1111/1471-0528.13299
5. Langhoff-Roos J, Rasmussen S. [Validering af Landspatientregistret (LPR) mhp. obstetrisk forskning og kvalitetssikring – et kvalitetsudvikling-sprojekt]. Danish Health and Medicines Authority; 2003. Danish.
6. Olsen SF, Houshmand-Oeregaard A, Granström C, et al. Diagnosing gestational diabetes mellitus in the Danish National Birth Cohort. *Acta Obstetricia et Gynecologica Scandinavica*. 2017;96(5):563–569. doi:10.1111/aogs.13083
7. Thisted DLA, Mortensen LH, Hvidman L, Rasmussen SC, Larsen T, Krebs L. Use of ICD-10 codes to monitor uterine rupture: validation of a national birth registry. *Eur J Obstetrics Gynecol Reprod Biol*. 2014;173:23–28. doi:10.1016/j.ejogrb.2013.10.033
8. Lohse SR, Farkas DK, Lohse N, et al. Validation of spontaneous abortion diagnoses in the Danish national registry of patients. *Clin Epidemiol*. 2010;2:247–250. doi:10.2147/CLEP.S13815
9. Sneider K, Langhoff-Roos J, Sundtoft IB, Christiansen OB. Validation of second trimester miscarriages and spontaneous deliveries. *Clin Epidemiol*. 2015;7:517–527. doi:10.2147/CLEP.S85107
10. Klemmensen AK, Olsen SF, Osterdal ML, Tabor A. Validity of preeclampsia-related diagnoses recorded in a national hospital registry and in a postpartum interview of the women. *Am J Epidemiol*. 2007;166(2):117–124. doi:10.1093/aje/kwm139
11. Luef BM, Andersen LB, Renäult KM, Nohr EA, Jørgensen JS, Christesen HT. Validation of hospital discharge diagnoses for hypertensive disorders of pregnancy. *Acta Obstet Gynecol Scand*. 2016;95(11):1288–1294. doi:10.1111/aogs.13006
12. Henriksen DP, Rasmussen L, Hansen MR, Hallas J, Pottegård A. Comparison of the five Danish regions regarding demographic characteristics, healthcare utilization, and medication use--a descriptive cross-sectional study. *PLoS One*. 2015;10(10):e0140197. doi:10.1371/journal.pone.0140197
13. eSundhed. Nyfødte og fødsler; 1997. Available from: https://www.esundhed.dk/Emner/Graviditet-foedsler-og-boern/Nyfoedte-og-foedsler-1997-#tabpanel6B95F8298EB444F48C3403E7B75B7202. Accessed August 7, 2023.
14. ICD-10 Version:2019. Available from: https://icd.who.int/browse10/2019/en. Accessed September 10, 2023.
15. Schmidt M, Pedersen L, Sørensen HT. The Danish civil registration system as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541–549. doi:10.1007/s10654-014-9930-3
16. Broe A, Damkier P, Pottegård A, Hallas J, Bliddal M. Congenital Malformations in Denmark: considerations for the use of Danish health care registries. *Clin Epidemiol*. 2020;12:1371–1380. doi:10.2147/CLEP.S278596
17. European platform on rare disease registration. Available from: https://eu-rd-platform.jrc.ec.europa.eu. Accessed September 10, 2023.
18. DSOG. Obstetrik. Available from: https://www.dsog.dk/obstetrik. Accessed September 10, 2023.
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J biomed informat*. 2009;42(2):377–381. doi:10.1016/j.jbi.2008.08.010